2e

Principles of
Business Forecasting

Keith Ord | Robert Fildes | Nikolaos Kourentzes

PREFACE
to the Second Edition

Business forecasting is art woven into science and principle teamed with pragmatism. Virtually every manager has to make plans or decisions that depend on forecasts. Research over the past 50 years or more has shown that taking an analytical approach rather than just relying on informal intuition leads to more accurate forecasts and more effective plans and decisions. However, forecasting is often the poor relation of more theoretical material, available through courses in regression, and time series analysis. Despite our backgrounds as statisticians and management scientists we believe this to be misguided; a student leaving a business school should know how to produce a reasonable forecast and how to evaluate the forecasts of others. Forecasting is important — every day we are confronted with forecasts ranging from the weather to stock prices and the state of the economy. Our motivation in writing this book is provide students of business and management with the tools and the insight to make effective forecasts. But we are also aiming to assist the many practitioners who are working in industry or government and need guidance on how to improve their performance.

This book not only provides an introduction to both standard and advanced approaches to forecasting, but also presents general principles that underlie forecasting practice. In turn, we show how good practice can be established within a systematic framework for the forecasting enterprise. The book builds unashamedly on Scott Armstrong's reference book, *Principles of Forecasting* (2001), although our focus here is on putting these principles to work to produce forecasts. What makes this book unique is its emphasis on incorporating the latest research findings to help practicing forecasters carry out their job and to enable students to prepare for a managerial or analytical career. Ambitiously, it also aims to act as a reference book that will guide those needing to study a particular forecasting topic in depth.

As a starting point, we recognize that there can never be just one approach to forecasting that meets all needs; rather we must invest in "horses for courses". To achieve that goal we consider a series of steps summarized by the mnemonic PIVASE (pronounced 'pi-vase'). The letters stand for the six elements that must be incorporated into the forecasting task: **P**urpose, **I**nformation, **V**alue, **A**nalysis, **S**ystem, and **E**valuation. This revises and extends our ideas from the first edition to emphasize both the analysis required to make successful forecasts and the organizational system in which they are developed. Chapter 1 elaborates on these concepts.

Good forecasting is a challenge. This observation means that the reader must engage in working on real data with all its aberrations. Wherever possible, we have used series from governmental or business sources, so that it is possible to go back to the source and update the database. Not only does this activity make forecasting more realistic but we also get to see how well our forecasts did with the new "out-of-sample" data. This second edition has offered us the opportunity, as we've updated the data used in our many examples, to emphasize the importance of stability (or change) in the system being modeled: and of course to comment on how to cope with the problems that arise.

The so-called “Great Recession” which started in 2007/08 and the subsequent slow recovery have provided an unusual basis from which to accomplish these objectives. Many of the economic models analyzed in the book initially use only the data up to the end of 2008 to estimate model parameters. In this second edition, most of the data sets we provide online run through the end of 2015 so that further, detailed assessments of model stability and performance can be carried out. The end-of-chapter exercises have also been used to examine further the effects of expanding the data bases.[1]

As technology improves, databases expand exponentially and most forecasting activities become unthinkable without a computer and well-designed software. Good programs are continuously evolving and tend to focus on different aspects of the forecasting enterprise (e.g. sales, macroeconomics). Accordingly, we have made use of a variety of software in the book, so that the user can link the methodology to his or her available resources. Our only caveat is that you, the reader, should have effective forecasting software available! To make the requirement more practical we have added supplementary material from the open-source R software that makes all the models we consider freely available to the student and the practitioner.

At the end of each chapter we list a set of principles that the forecaster should keep in mind. Some of these may seem obvious, such as “Check the data for outliers” yet failure to do so has led to forecasting disasters. Just as driving an automobile requires simultaneous attention to multiple indicators, so the effective forecaster needs to internalize these principles in everyday practice.

No book can include all aspects of forecasting and answer every question. Nevertheless, we have tried to meet that challenge in several ways. We include many exercises which lead the reader from the classroom towards the role of practicing forecaster. We also include *discussion questions* that aim to stimulate thinking beyond the narrow confines of the technical issues. In addition, more references are included than is usual in a textbook so that users can quickly access the more advanced research literature. In the end, the successful forecaster has to engage with all the complexities of real data in an organizational setting. We hope that we’ve met the challenge of providing a set of operational principles that help in this quest.

Structure of the Book

Introduction (Chapters 1–2) Chapter 1 provides an overview of forecasting including a variety of situations where both time series and cross-sectional data can help the forecaster’s understanding; then Chapter 2 introduces the basic statistical tools that are needed in later chapters. The remainder of the book falls into three component parts.

Extrapolative methods (Chapters 3–6) The focus here is upon a single time series and forecasting in the short to medium term. In Chapter 3 we consider regular (non-seasonal) time series and introduce exponential smoothing methods for series with and without a trend. Exponential smoothing was one of the first forecasting methods used extensively in industry and remains so today. We extend the discussion in Chapter 4 to include seasonal series

1 Many government agencies periodically update their databases and this operation may include the revision of historical data. More recent downloads may produce slightly different values for a series than those we have provided.

particularly the so-called Holt-Winters methods, but also give attention to seasonal adjustment procedures that are important for macroeconomic series.

The discussion in these two chapters is limited to forecasting *methods*, which can provide useful point forecasts but do not produce measures of uncertainty. Thus in Chapter 5 we consider the class of state-space models, which provides a natural framework for exponential smoothing and allows the creation of prediction intervals. State-space models are closely linked to the ARIMA (AutoRegressive Integrated Moving Average) models developed by George Box and Gwilym Jenkins, two of the famous names in forecasting, so Chapter 6 explores this connection and develops prediction intervals using these models. Also in Chapter 6 we consider models for changing uncertainty (ARCH/GARCH models) which are widely used in financial analysis. Just as point forecasts change, so do the related measures of uncertainty. These too need to be forecast.

Statistical model building (Chapters 7–9) Business forecasting involves both the analysis of time series and the use of cross-sectional databases (for example decisions by banks and companies on extending credit to new customers). We begin these developments by considering simple linear regression in Chapter 7, where we examine the use of a single predictor variable to assist in explaining the variations in the dependent variable to be forecast. This discussion leads naturally to multiple regression, the use of two or more predictor variables, in Chapter 8.

Although it sometimes appears that regression modeling is just a matter of downloading the database and running a suitable statistical program, genuine applications involve careful variable selection and building the database. Chapter 9 introduces various extensions to the basic linear regression model including indicator (dummy) variables, lag variables which are fundamental to time series forecasting and non-linearities. Even when the initial model is specified and estimated, the result must be checked to ensure that the final form satisfies the underlying statistical assumptions, at least approximately. These models also need to produce forecasts that are more accurate than those provided by simpler alternatives. In particular, the model should be structurally stable — that is it remains unchanging over time. The extension of the data bases in this second edition allows an extensive discussion of model stability, model development and checking: the themes of Chapter 9.

Advanced methods and forecasting practice (Chapters 10–13) The material in later parts of the book might be beyond a first course in forecasting, but it forms an essential knowledge base for the modern forecaster. As the forecasting literature expands, new methods emerge and solutions to new problems are developed. We attempt to capture these novel developments in these later chapters. Thus, Chapter 10 describes more advanced techniques including classification and regression trees, logistic regression, neural networks and vector autoregressive models. These topics have seen considerable theoretical development in the last few years and also have clear implications for improved practice.

Forecasting practice often relies heavily on subjective judgments by those who are expert (and not so expert) in the field. In Chapter 11 we consider different approaches to judgmental forecasting and discuss when judgmental inputs can add value. Like many such choices, the one between judgmental and quantitative forecasting methods is a false dichotomy and the correct answer is often to use both in combination, capitalizing on their respective strengths.

In the two final chapters we focus attention on forecasting in practice. Chapter 12 first considers the core question of how forecasts and forecasting methods should be evaluated. Procedures for comparing forecasting methods, often a "hot topic" for an organization, are fraught with difficulties. This is a core topic for any forecasting course. Section two examines how forecasts are prepared in organizations through a software-based forecasting support system. We consider what characterizes an effective support system and how such systems and the methods they contain should be evaluated. Three important application areas are then considered in more detail: operations and marketing as well as models focused on individual customer behavior. Finally, Chapter 13 examines the construction of a forecasting system within an organization with particular attention being paid to the interaction between the forecaster and the user of the forecasts. Ultimately, the purpose of forecasting is to aid planning and the reason for planning is to ensure that "things don't just happen". If the forecaster and the user are not communicating properly, the best models in the world will not help.

Use of the Book

Most readers will have had the benefit of a first course in applied statistics, or an equivalent background, although a brief refresher of key statistical methods is included in Chapter 2 (and online Appendix A). Our aim in the book is to show users how to forecast (and that forecasting is fun). In our experience forecasting is a topic that many students are interested in beyond the formulae and routines that can sometimes form the core of statistical or econometric courses. Every day there are examples in the media of important, interesting or even bizarre forecasts, which form a backdrop to the more technical material at the book's core. However, occasionally we have been forced to include material which is particularly demanding mathematically and/or statistically — we have designated such sections with '*'; similarly, more advanced exercises are labeled with '*'. Thus, we anticipate that the book might appeal to four broad groups of readers.

MBA students and advanced undergraduates in business
Chapters 1–4 and 7–9 would form the core of the course, supported by Chapters 11 and 13 to provide a more managerial focus.

Undergraduates in the management sciences and statistics
Chapters 1–9 would be the principal components of a course in this area, with the expectation that much of Chapters 2 and 7 could be omitted. Chapter 11 would offer such students (and their lecturers) light relief, but also make an important point, often omitted from more technical courses, that judgment has a key role to play. Chapter 9, on model building, would be given particular emphasis as many more technical statistical courses neglect this aspect in favor of a more mathematical approach.

Undergraduates in business analytics
Business analytics is becoming more popular as an undergraduate specialism. As yet the course content has not become established. We firmly believe that *Forecasting* is a critical component for any such course as it forms the basis of *Predictive Analytics* without which a degree in business analytics would be incomplete. Chapters 1–4, 7–9, and in particular the data mining elements of chapters 10 and 12 would provide the core material. Standardized R code for these chapters offers a major benefit.

Masters programs in applied statistics, management science and business analytics
Chapters 1–10 and 12 would be the principal components of a course in this area, with a stronger emphasis on Chapters 5 and 6 and with the expectation that Chapters 2 and 7 could be omitted.

Forecasting practitioners

The modern forecaster who wishes to be on top of the latest ideas should ultimately become familiar with all the material in the book, although we would expect him or her to chart a course like one of those just outlined, and then build on that knowledge with experience. However, some forecasters have a more limited developmental agenda focused on their particular organizational responsibilities. For an operations forecaster, chapters 1-4 should be supplemented by Chapter 7, Chapter 11 and the sections of Chapter 12 focused on evaluation and operations. Marketing forecasters also need to develop their skills in regression (Chapters 8 and 9) as well as those sections of Chapter 12 that consider marketing models. Finally, Chapter 13 provides a necessary framework for effective forecasting practice.

Additional Materials

A critical element in every good forecasting text is the provision of data sets for readers to test out their developing skills. The associated website includes all the data sets used in the book, both in the examples and the exercises. In addition we have provided basic software (The Exponential Smoothing Macro, or ESM) for carrying out the exponential smoothing methods of Chapters 3–4. This is for two good reasons: exponential smoothing software is not available (or if it is, it is not well designed) in standard statistical packages. Secondly, despite its apparent simplicity, there are a number of hazards if the student is asked to carry out the calculations relying only on Excel. In addition, a macro for analyzing non-linear trend curves is also provided and used in market modeling. But we are the first to admit these programs are not the equivalent of a professional forecasting package such as ForecastPro or an econometrics package like EViews. We discuss the important issue of computer programs to support forecasting in online Appendix B. Our basic principle is that the forecaster should use well-validated commercial packages wherever possible; many are available for regression analysis and its extensions (Chapters 7–10). Since the first edition, R has become widely available and online materials are provided in Appendix C (Forecasting in R: Tutorial and Examples) to support all the analyses we discuss.

The website resources for instructors include PowerPoint slides that may be used to develop course materials and outline solutions to many of the exercises.

Key Innovations in the Second Edition

This second edition embodies some key changes:

- R programs and tutorial material are provided so that all analyses can be carried out (for free!) wherever the user is working.
- Expanded coverage of model building in regression.
- New material on judgment to address some of the political shocks in the last few years.
- Greater coverage of data analytics, in particular neural nets together with software.
- Expanded material on applications that uniquely includes new research findings relevant and immediately applicable to operations, such as hierarchical modelling and temporal aggregation.

- A new colleague, Nikolaos Kourentzes, has joined the authoring team with complementary expertise, particularly in R.
- Updated data sets.

And of course the various minor corrections that we and others have found.

Other Resources

These days the web provides many resources to support every activity humans contemplate. Forecasting is no exception. We list here a few of the resources a practicing forecaster would wish to consider:

- Two key academic publications that an ambitious reader would need to consult (in writing a dissertation for example) are the *Journal of Forecasting* and the *International Journal of Forecasting* (*IJF*). Occasionally survey articles are written accessible to all but most of the articles published are technical in nature.
- International Institute of Forecasters (IIF) (*https://forecasters.org/*): The IIF is a non-profit organization and the publisher of the *IJF.* The Institute organizes an annual symposium that attracts a world-wide group of participants and also sponsors occasional professional workshops in various topic areas.
- *Foresight: The International Journal of Applied Forecasting:* This journal, published by the IIF aims to provide practicing forecasters with easy-to-read material on important topics that nevertheless summarize the latest research ideas.
- Principles website (*www.forecastingprinciples.com*): The Forecasting Principles site presents a personal view of developments in forecasting (some of which we agree with and some we don't). It aims to offer advice to both researchers and practitioners and is designed to make available the material in Armstrong (2001), suitably updated. It contains much that is useful from a method selection 'tree' to a dictionary.
- Online bibliography: This can be found by accessing: Fildes, R. and Allen, P.G. (2015). *Forecasting*. In Oxford Bibliographies in Management. Ed. Griffin, R.W., New York: Oxford University Press. DOI 10.1093/obo/9780199846740-0064
- Software resources (see online Appendix B): Good software for forecasting is essential but there is no perfect package. We discuss some of the best known alternatives in this appendix.
- Institute of Business Forecasting (IBF) (*https://ibf.org/*): The IBF organizes professional conferences and publishes the *Journal of Business Forecasting*. Its primary focus is on forecasting for supply chain organizations. The web site offers a job search facility.
- Applied Forecasting (*www.appliedforecasting.com*): This site provides a compendium of current news and the latest research in forecasting.
- Principal data resources are listed in Chapter 1.

Acknowledgments

In the preparation of this book we have benefitted from comments from forecasting colleagues from around the world. Particular thanks are due to Geoff Allen (University of Massachusetts, Amherst), Paul Goodwin (University of Bath), Herman Stekler (George Washington University, Washington) and the late Peg Young (from the U.S. Department of Transportation). Thanks are also due to colleagues associated with the Lancaster University Centre for Marketing Analytics and Forecasting, John Boylan, Sven Crone, Steven Finlay, Marwan Izzeldin, Fotios Petropoulos, Patrick Saoud, Oliver Schäer, and Ivan Svetunkov. However, the book would not have been conceived and developed without the research that has been stimulated by the founding of the International Institute of Forecasters and the publication of the *Journal of Forecasting* and the *International Journal of Forecasting*.

We also wish to acknowledge the support we have drawn from our families. For the first edition, both KO and RF were given support including tea, coffee, and encouragement. For this second edition, these sources have become more limited. For RF, support has consisted of putting up with the odd complaint and only occasional sustenance. For KO, both tea and coffee remained available but his clutter was moved to a separate part of the house. As for NK, he has finally started appreciating tea and coffee. Like most projects, there was considerable optimism from all parties as to its completion date, a failure to learn from the first edition: hopefully the revenue projections won't suffer the same fate. We appreciate the responsiveness of our new publishers, Wessex Press, Inc.: Paul Capon (CEO), Alisa Matlovsky (Project Manager), and Anna Botelho (Graphic Design/Production).

Reference

Armstrong, J.S. (ed.), (2001), *Principles of Forecasting*. Boston and Dordrecht: Kluwer.

CHAPTER 1

Forecasting, the Why and the How

Table of Contents

One cannot divine nor ***forecast*** *the conditions that will make happiness; one only stumbles upon them by chance, in a lucky hour, at the world's end somewhere, and hold fast to the days, as to fortune or fame.*

— Willa Sibert Cather (American author and winner of the Pulitzer Prize)
(*http://thinkexist.com/quotes/*)

Introduction

Many of us follow familiar rituals after climbing out of bed in the morning. One of these activities is to check the weather forecast. Between sips of coffee, we may scan the newspaper, listen to the TV or radio, or check an online weather service. At the simplest level, we obtain a short summary, such as, "Cloudy early in downtown Washington, some sun in the afternoon, with a high of 76°F (24°C) expected." There might also be a travelers' warning for those proposing to visit Mexico in the next couple of days, with a hurricane force 6 expected to hit land south of Cancún. More detail is often forthcoming and may include such information as:

- A summary of expected temperatures at intervals during the day.
- The probability of precipitation or of the storm.
- Summary forecasts for the next few days.

Meteorologists provide a reasonable degree of detail in their forecasts for up to 7–10 days ahead, after which broader assessments, such as "above or below average temperatures," are about as much as can be usefully provided. If we are concerned about the weather several months ahead, we rely on long-term average values for the time of year.

What does the weather report have to do with business forecasting? In fact, weather is an important factor affecting the sales of many products, such as swimsuits, fresh fruit, and barbecue food. In addition, it affects holiday and business travel.

The example also raises the most basic question of what we mean by a forecast:

FORECAST

A prediction or estimate of an actual outcome expected at a future time or for another situation.

Here we have two types of outcome: the expected temperatures over the course of the day and a predicted event, a hurricane. In this book, we will be concerned with predicting both types of outcome. When we are told the forecast, we understand it to mean the most likely outcome (e.g., that the hurricane occurs or that the temperature will reach 76°F, 24°C). Sometimes, forecasts include estimated probabilities that the outcome will happen, such as "There is a 70 percent chance of rain this evening." In the United States such probability weather forecasts are common, whereas they are absent in Great Britain. The other key feature of this weather forecast is that it is precise: It tells us where and when these events are expected to happen. So this simple example serves to illustrate a number of basic features of any forecasting problem.

We now explore these features in more detail, asking why, what, and how, as well as providing a number of examples of different forecasting issues.

DISCUSSION QUESTION: *What other forecasts are regularly included in the daily newspapers, such as the* Wall Street Journal*, the* New York Times*, or the* Times of London*?*

1.1 Why Forecast?

Why do we check the weather forecast? Partly, we may do so out of idle curiosity, but the information also serves to guide our planning for the days ahead. The planning activity may include such decisions as what to wear, whether to carry (and maybe lose) an umbrella, or whether to go ahead with a scheduled outdoor activity or travel plan. The forecast helps us plan, and planning, in turn, may improve our quality of life, compared with just rushing out the door. If we examine the elements of this simple vignette, we arrive at the following motivation for forecasting:

FORECASTING AND PLANNING

The purpose of forecasting is to inform the process of planning future actions.

The purpose of planning is to develop a course of action so that current activities don't "just continue" based on a no-change forecast.

Before embarking on a forecasting exercise, we should always consider the "Why?". That is, we need to specify the reasons for generating the forecast, how it fits with possible plans over the planning horizon, and the kind of forecast we need. Once the specifications are settled we must resolve the "How?" by careful analysis of the available information, its potential value in improving the planning decisions and the development of a forecasting system to implement the resulting forecasts. Indeed, "How?" is the question that we seek to answer throughout most of this book. Finally, last but by no means least, we must check whether the forecast system is doing the job for which it was designed – a question that is all too often ignored. Using the mnemonic PIVASE, we may identify these components as follows:

- **P**urpose
- **I**nformation
- **V**alue
- **A**nalysis
- **S**ystem
- **E**valuation.

1.1.1 Purpose

What do we hope to achieve by generating the forecast? What plans depend on the results of the forecasting exercise? The meteorologist provides information to a broad range of individuals and organizations, and they are all interested for their own reasons. For the weather service, it suffices to know that there is a demand for forecasting services. We tailor the level of detail in the forecast to satisfy our planning needs. If the weather forecast was for downtown Washington, but we were planning an outing to the nearby coast, the city forecast would give us only limited help. Similarly, the overall forecast demand for a company's cars is not sufficiently detailed to plan the production line. An integral part of this question is how far ahead do we wish to forecast? We refer to this period as the *forecasting horizon*. In turn, the horizon depends on our purpose in forecasting and will drive the choice of method. The methods we employ for short-term forecasting will consider only the factors that change rapidly, whereas longer term forecasts need to take into account a larger number of factors that may change during the time frame of interest. We also refer to the *forecast origin*, the point in time from which the forecasts start. The horizon often affects the accuracy and usefulness of a forecast. As we know, weather forecasts are quite accurate up to 7–10 days ahead, but they tend to be more like long-term seasonal averages beyond that time frame.

1.1.2 Information

What do we know that may help us in forecasting, and when will we know it? Detailed information is useful only if it is available in a timely fashion. Tomorrow's financial reports may provide an excellent explanation of today's events, but they are of no use in forecasting stock prices today. Likewise, information is of value only if it has an impact on the forecasting procedure we seek to implement. Population changes may have a major effect upon sales ten years from now, but such changes take place at a relatively slow rate and would be irrelevant to forecasting sales in the next three months.

What do we know and when will we know it? A large-scale forecasting model may take into account a broad range of key factors, but if we have to wait several months for the data to become available, the forecasts may be too dated to be useful. Thus, a forecaster often relies upon leading indicators, such as closing levels of a major stock index or a survey of consumer sentiment, to signal changes in short-term economic conditions. These factors do not provide a cause-and-effect description of what is happening, but they may well yield a timely assessment of potential changes.

An often critical factor in producing an accurate forecast is knowledge of plans made in other parts of the organization that will affect the variable to be forecast. Producing a forecast of incoming calls to a call center without information on the corporate marketing plans over the horizon will usually lead to poor forecasts. Thus, identifying potentially important drivers that will affect future outcomes is an important step in developing a forecasting system.

The distinction between forecasting and explanation is often somewhat blurred. On the one hand, we may consider so-called pure forecasts that use only currently available information. On the other hand, a detailed description of a process (e.g., a macroeconomic model of the economy or a model of the world's climate) may enable us to answer what-if questions based upon the understanding that the detailed description provides. Such models typically require more information than pure forecasting exercises. When seeking to answer what-if questions, it is entirely reasonable to use data that are available only after the fact, because they provide an indication of the likely effects of certain policy decisions. They may also provide an explanation of the forecast errors we have made. In this book, we focus primarily upon pure forecasts, although the methods discussed in later chapters are useful in both contexts.

1.1.3 Value

How valuable is the forecast? What would you pay to have perfect knowledge of a future event? For example, a weather forecast is useful to an individual in that she is better able to decide on appropriate clothing. However, most of us would not pay very much for such information, and we do not need to know exactly how much rain will fall.

In contrast, the agricultural sector is very interested in using weather forecasts to plan irrigation and the planting of crops, and it is willing to pay for more accurate location-specific forecasts. In a different context, consider a company with thousands of product lines. Although the value of forecasts for any single line may be modest, an effective forecasting system for the complete range of products is very valuable in making production and inventory decisions.

1.1.4 Analysis

Once the purpose is clear and the information has been assembled, we turn to the analysis of the data. This process includes the development of a forecasting model, which will involve the consideration of several different approaches and the final selection from within these alternatives. Once the model has been selected, we estimate any unknown model parameters and then test the model's performance using a *hold-out sample* (that is, data that are held in reserve and are separate from the information used to fit the model). For example, we might have six years of monthly sales figures and use the first four years to fit the model and keep the last two years as the hold-out sample to test performance.

1.1.5 System

Most of the examples in this book refer to a single time series or forecasting issue. However, practical applications may involve the simultaneous forecasting of thousands of items (products, stock prices, etc.). Further, an organization may generate forecasts at several different levels (e.g., individual products, distinct sales regions, weekly or monthly sales) with many people involved and these forecasts must be integrated into a consistent description of total sales as the basis for budgeting, marketing, and production planning. Thus, we must develop a forecasting system and process that is capable of meeting these needs. Such systems are usually computer based. Chapters 12 and 13 focus on these issues.

1.1.6 Evaluation

How do we know whether a particular forecasting exercise was effective? The obvious answer is that we should compare past results with the forecasts that were made, possibly comparing several different forecasting methods as we proceed. We examine several such criteria in Chapter 2, all of which are based upon the differences between forecasts and actual values. A statement about checking forecast performance might seem almost too obvious to be worth making, yet the evidence suggests that some companies never go back to check their forecasting performance. For example, an informal survey conducted by the software company Business Forecasting Systems produced the results shown in Table 1.1. Respondents *who used statistical forecasts* were asked how they did so:

> *Baseline:* Use the statistical forecast as a baseline and then make judgmental adjustments
>
> or
>
> *Reference:* Use the statistical forecast as a reference, or "sanity check," on some other primary forecast.

Respondents were then asked whether they regularly checked the performance of the forecasts. The *percentages* in each category who said yes are given in Table 1.1. The survey was small and relied upon self-reporting, but the figures are striking nevertheless. The baseline group is making adjustments to the statistical forecasts, yet nearly half of them are not checking to see whether those adjustments improve forecast performance. Nearly one-third of the reference group do not even check their principal forecasts, and more than two-thirds fail to check the statistical forecasts.

Table 1.1 Percentages of Respondents Who Regularly Monitor Forecasting Performance

Use Made of Statistical Forecast	Percentages of Respondents Who Checked	
	Adjusted/Principal Forecast	Statistical Forecast
As a baseline	91	57
As a reference	71	29

Source: Trends Business Forecasting Systems Newsletter.

If you don't check from time to time, how do you know whether your approach is any good? Making effective forecasts is as much an art as it is a science, but asking about PIVASE before you start the exercise will help avoid expensive mistakes.

MONITORING FORECAST ERRORS

Always check on forecast performance. Checking helps you avoid repeating mistakes and leads to improved results.

Example 1.1: PIVASE in action

Some years ago, a US state agency asked one of the authors (KO) to develop a short-term forecasting system for state tax revenues. The author used simple extrapolative models (the details don't matter at this stage) to accomplish the job. The *purpose* was to provide the state legislators with a revenue-forecasting model that they could use to compare with the forecasts generated by the executive branch (the governor's office). Because all state representatives are up for reelection every two years, the forecast horizon was set at two years. The monthly state revenues, for each of a half dozen tax categories, were published typically within a month of the end of the collection period; because these figures are a matter of public record, the necessary *information* was freely available. Some additional information, such as upcoming changes in tax rates, was also provided. The *value* of the study lay in the provision of an independent forecasting system for the legislature to serve as a check on the executive's budgetary proposals.

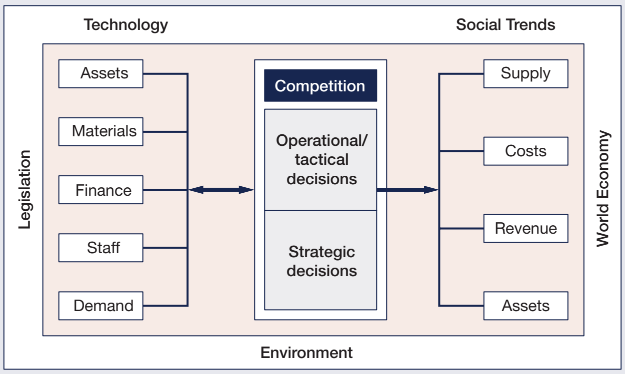
The *analysis* followed the general approach laid out in Chapter 6 and the overall *system* involved monthly updates to the half-dozen series and the generation of new forecasts. As happens all too often with forecasting systems, no follow-up study was performed to *evaluate* performance. However, several years later, it was a pleasant surprise to learn that the system was still in use and that the outputs compared quite favorably with the forecasts from the state's model sponsored by the governor's office. A firm of consultants provided the state's economic statistical (econometric) model for around $100,000 per year (compared with a one-time payment of $10,000 for the short-term model). Doubtless, the econometric model is also used for other purposes, and the fees are probably well spent. ■

DISCUSSION QUESTION: *There was a two-year cycle in the tax revenues data, with a surplus in year one and a deficit in year two. Why might this be?*

1.2 What and Why Do Organizations Forecast?

Organizations — large and small, public, government or private sector — have numerous forecasting needs. To plan their operations, they need to estimate the demand they expect for their products and services. They also need to know how their actions — for example, their marketing plans, including promotions and prices — will affect demand and in turn affect revenues and costs. Figure 1.1 captures some of the complexities that organizations face. Fortunately, many of the variables can, for the most part, be ignored. In the short term, such features as the technology (used by the organization or its customers), legislation, and social trends can usually be regarded as fixed. Operational and tactical (medium term) decisions depend primarily on demand (as it is affected by competition) and the finance needed to support the organization. In the longer term, where the organization's strategy is updated, the variables in the outer box come into play; for example, social and demographic trends can change a market dramatically. In Europe, an aging population requires a different set of services (e.g., health) than a younger market. In the United States, shifts in consumer tastes relating to the changing composition of the population with regard to ethnic origin and geography also affect consumption patterns.

Figure 1.1 An Organization's Forecasting Needs



DISCUSSION QUESTIONS

What might a nuclear electricity supplier or a water company need to forecast with regard to its key assets? Why?

A state or national government provides unemployment benefits for a period of six months following a layoff (or leaving school). It continues for a further six months if the applicant attends a job training program. What variables would the government need to forecast for its annual planning?

Each of these discussion questions points up the critical linkage between forecasting and planning: If the organization or the person does not act on the forecast, it has no value beyond mere curiosity. These issues are pursued further in the end-of-chapter minicases.

1.3 Examples of Forecasting Problems

Forecasting involves using currently available data to make statements about likely future developments. Such data often arise as a time series:

TIME SERIES

A set of comparable observations ordered in time. The values may refer to either a point in time ("the current value of the Dow Jones Index is 17,000") or an aggregate over a period of time ("total sales for the last month were 350 units").

Trend: A time series contains a trend if it shows systematic movements (e.g., increase or decrease) over an extended period.

Seasonality: A time series has a seasonal component if it displays a recurrent pattern with a fixed and known duration (e.g., months of the year, days of the week).

Cycle: A time series has a cyclical component if it displays somewhat regular fluctuations about the trend but those fluctuations have a periodicity of variable and unknown duration, usually longer than one year (e.g., a business cycle).

We usually assume that the observations are equally spaced in time. On occasion, this assumption is incorrect, but we may make some adjustment to the data so that it is a closer approximation. For example, the months of the year differ in length, but we usually treat them as equally spaced. If the time series refers to data such as sales, we may use average sales per day during the month or average sales per trading day (allowing for possible closures on Sundays and public holidays). The revised time series then conforms more closely to the assumption of equal spacing.

An interesting example of such adjustments occurs with stock market data. The New York Stock Exchange (NYSE) is generally open for trading Monday through Friday, 9:30 a.m. to 4 p.m. The NYSE is closed on weekends and public holidays. Nevertheless, it is often possible to ignore the breaks in trading and to treat time series relating to the price of a stock or the trading volume (e.g., the number of shares traded per hour) as being recorded at regular equal intervals.

In the rest of this book, we typically assume that observations are recorded at equally spaced times and effectively ignore the distinction between data either recorded at a point in time or aggregated across time. Such a step poses no problems for the formal development of the forecasting methods, but it needs to be taken into account in applications.

We now examine a few series in detail. Note that we always start out with plots of the data, whether time series or cross sectional. It would be difficult to overemphasize the importance of this first step in any forecasting study. Beyond the initial benefit of identifying unusual observations (whether data entry errors or genuinely strange values), such plots serve to identify patterns in the data and to suggest possible approaches to forecasting.

1.3.1 Retail Sales

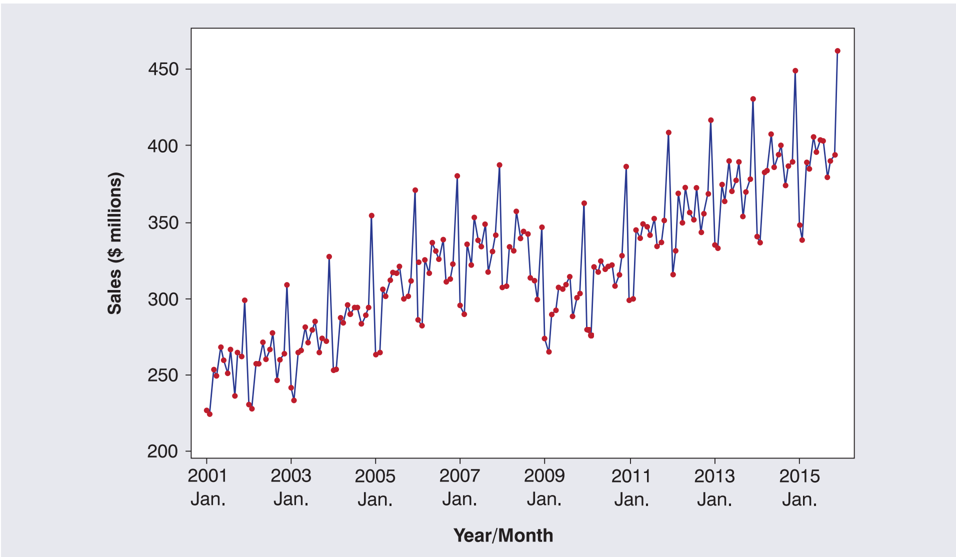
Figure 1.2 shows the monthly level of retailer sales in the United States over the period from January 2001 to December 2015 (see *US_retail_sales_2.xlsx*). The data refer to current dollars, so no allowance has been made for the effects of inflation. The series shows more or less steady growth until late 2007 when the 'Great Recession' hit; with growth resuming in mid-2009. (The National Bureau of Economic Research, NBER, officially dated the start of the recession as December 2007 and the end as June 2009.)

The growth is due to a combination of real growth in the economy and inflation. The inflationary elements could be removed by using a price index and *deflating* the current values to obtain real growth figures, but for the present we will examine the series in current dollars; see Minicase 1.4 for an example.

The second feature that is evident is a regular within-year fluctuation, which we refer to as a *seasonal pattern*. Here, it reflects high sales in December and a drop-off in January. We also observe that the fluctuations are increasing somewhat over time, indicating a proportional effect (up or down by a given percentage) rather than a fluctuation by an absolute amount.

This diagram imparts a lot of information, and it is reasonable to ask whether the different parts of the total information package can be filtered out, rather than trying to read everything from one diagram. The answer is that we can do that by identifying a separate seasonal component and then adjusting the original series by those seasonal factors to produce a seasonally adjusted series. The advantages of doing so are that we can examine not only underlying trends without being distracted by seasonal fluctuations but also the seasonal pattern to look for changes over time. This process of *decomposition* is particularly important for macroeconomic series, and most U.S. and UK government series are published in seasonally adjusted form. This topic is examined in detail in Chapter 4.

Figure 1.2 U.S. Monthly Retailers Sales



Source: U.S. Census Bureau via FRED *https://research.stlouisfed.org/fred2/*
Series ID: RETAILSMNSA; Data: US_retail_sales_2.xlsx

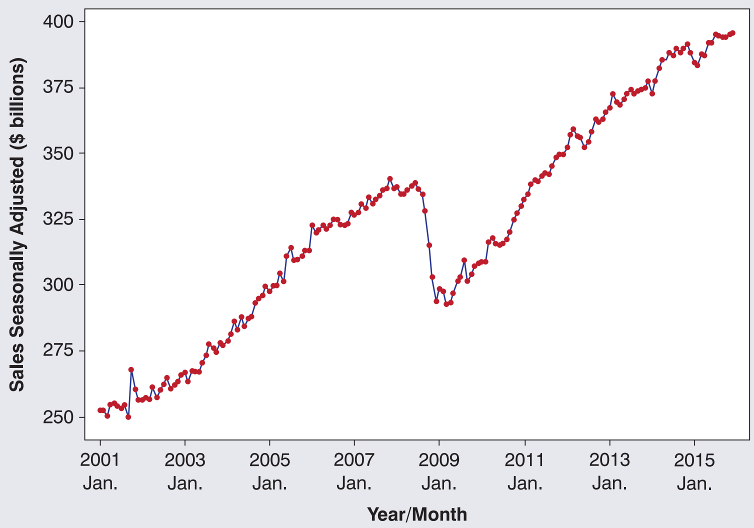
1.3.2 Seasonal Patterns for Retail Sales

We discuss the process of seasonal adjustment more rigorously in Chapter 4, but for purposes of the present discussion we define the seasonal factors indirectly as the ratios of actual sales to the seasonally adjusted sales, as published by the U.S. Census Bureau.

The seasonally adjusted series is shown in Figure 1.3 and provides a clear picture of the decline in sales during the recession. Such a dramatic change in the level (or trend) of a series is called a structural break. The derived seasonal patterns are shown in Figure 1.4. The December peak and subsequent drop for January and February are clearly evident. Looking more closely, we see that the seasonal pattern is fairly stable, although the holiday peaks were lower in 2007 and 2008.

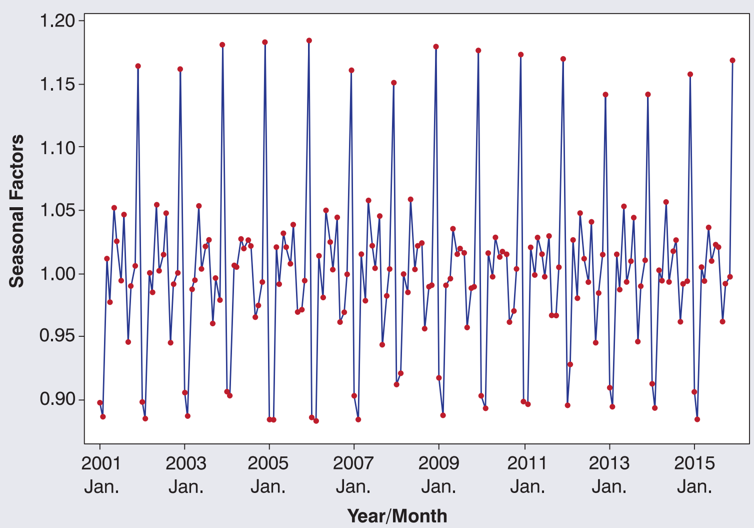
DISCUSSION QUESTION: *Looking ahead two years from the present, how would you expect the series to behave?*

Figure 1.3 Seasonally Adjusted Series for U.S. Monthly Retailers Sales



Source: U.S. Census Bureau via FRED *https://research.stlouisfed.org/fred2/*
Series ID: RETAILSMSA; Data: US_retail_sales_2_SA.xlsx

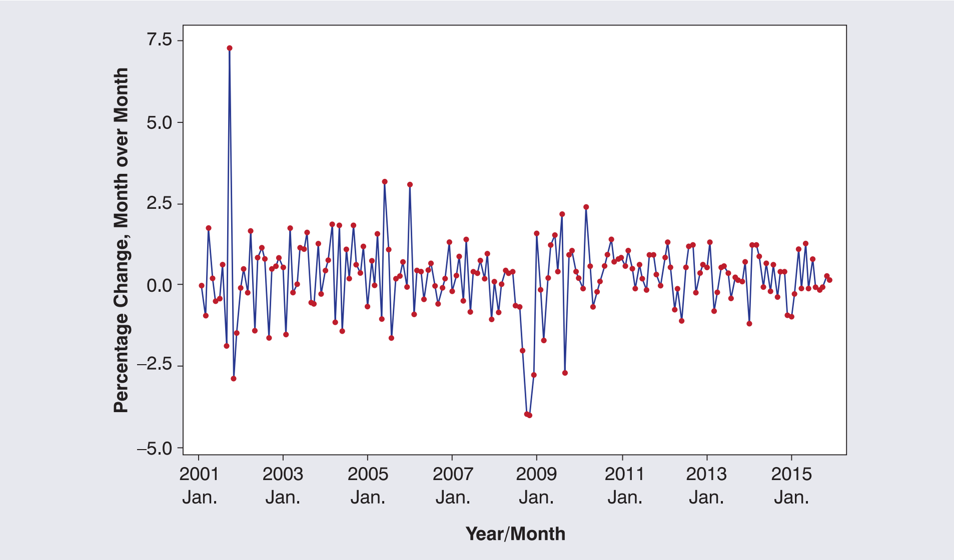
Figure 1.4 Seasonal Factors for U.S. Monthly Retailers Sales



Source: U.S. Census Bureau via FRED *https://research.stlouisfed.org/fred2/*
Derived from Series RETAILSMSA and RETAILSMNSA

Another way to examine the data is to look at month-to-month changes in the seasonally adjusted series, as shown in Figure 1.5. The effect of September 11th 2001 on sales is very clear and the slide into recession at the end of 2007 is also apparent.

Figure 1.5 Month-to-Month Changes in Retailers Sales (seasonally adjusted)



How else could we present this information? One possibility is to look at the changes in the level of sales over time. Better yet, because the retail industry talks about percentage changes for a given month from one year to the next, we can transform the data as follows:

$$\%\text{ change over the year} = \frac{100\ [(\text{Sales in year } t\text{, month } j) - (\text{Sales in year } t-1\text{, month } j)]}{(\text{Sales in year } t-1\text{, month } j)}$$

That is, we look at the percentage change for this month compared with that for the same month last year.

The resulting plot is shown in Figure 1.6. We can see the periods of relative growth and stagnation. Again, the sharp decline starting at the end of December 2007 is very clear, but it is also worth noting the longer periods of somewhat faster or slower growth.

The information contained in these various time series plots enables the forecaster to make statements about the potential December sales peak in conjunction with the likely change relative to the previous year. Such forecasts feed into decisions on overall inventory and staffing levels, as well as financial requirements.

1.3.3 UK Road Accidents

As an example of a series displaying both seasonal and other effects, we consider a time series on injuries caused by road accidents in the United Kingdom over the period from January 1975 to December 1984, examined in detail by Harvey and Durbin (1986). The series is shown in Figure 1.7. The seasonal peak in December is quite evident, although the pattern breaks up somewhat in the later years. The other feature worthy of note is the step-change decline in early 1983, which was when the wearing of seat belts became mandatory. Analysis of the series provides a way to determine the effectiveness of the seat belt legislation, as we shall see in Chapter 9.

Figure 1.6 Year-Over-Year Percentage Change in U.S. Monthly Retailers Sales

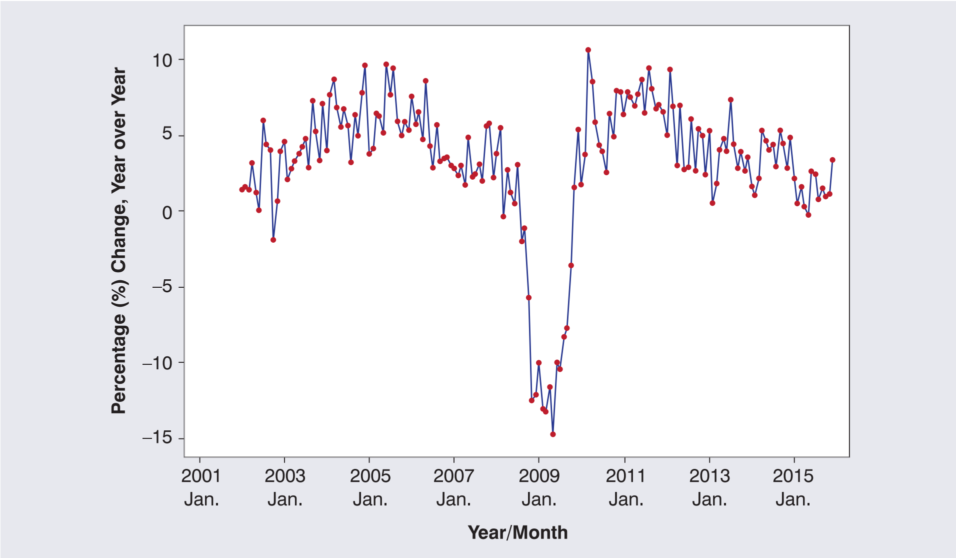
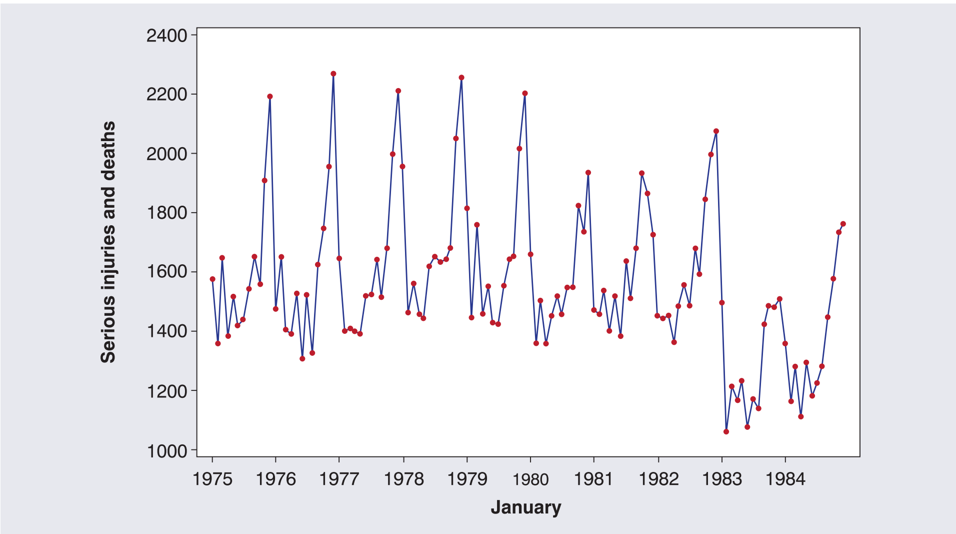


Figure 1.7 Serious Injuries and Deaths on UK Roads, January 1975–December 1984

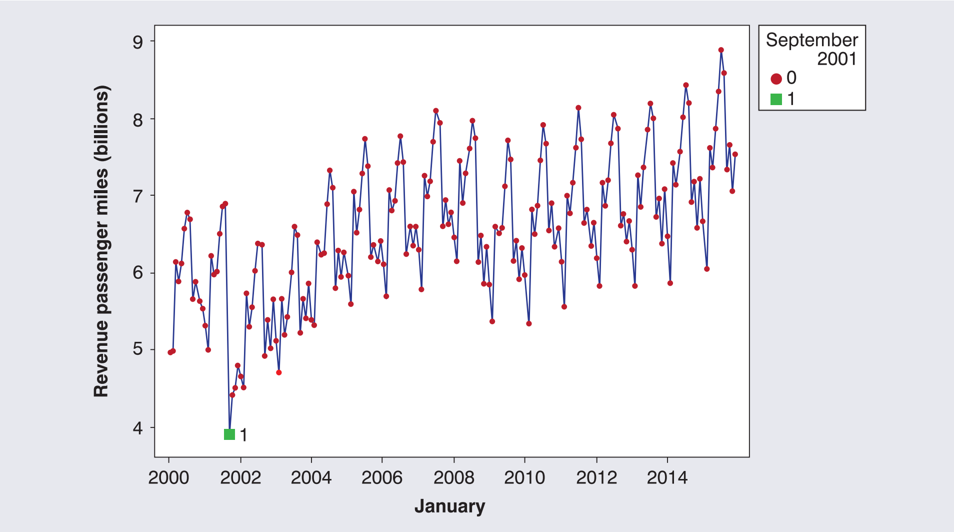


Source: Harvey and Durbin (1986); *Data: Road_accidents.xlsx*

1.3.4 Airline Travel

A major issue in time series analysis is how to deal with unusual observations or shifts (breaks) in the general level. The data on road accidents represent a series in which a change was anticipated; by contrast, an event might be totally unexpected, as with the tragic events in the United States on September 11, 2001. Figure 1.8 shows the number of revenue passenger miles traveled in the United States, by month, from January 2000 to December 2015. The impact of that day upon air travel is clearly seen; September 2001 is marked on the diagram.

Figure 1.8 Monthly U.S Air Revenue Passenger Miles (billions), January 2000–December 2015



Source: Bureau of Transportation Statistics, U.S Department of Transportation via FRED
Data: Revenue_miles_2.xlsx

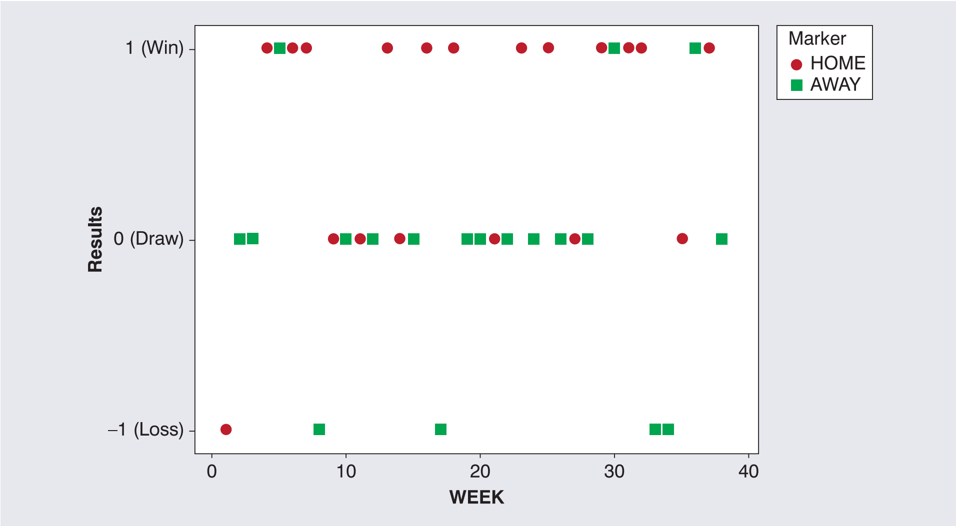
DISCUSSION QUESTION: *How would you estimate the effect of 9/11 upon the airline industry with respect to losses in revenue passenger miles?*

1.3.5 Sports Forecasting: Soccer (a.k.a. Football!)

Sports forecasting is big business. In many countries, betting on the outcome of a game is legal, be it American football, soccer, or tennis. In fact, gamblers go a stage further: They can and do bet on the detailed developments in a game. In soccer, this type of betting can include the number of players cautioned; in tennis, the number of games played in the longest set. Here again, the data are different. Although the outcomes (win, draw, or lose in soccer) are recorded over time, the observations are not observed at regular intervals.

Figure 1.9 shows a time series graph of the data as they fall into these three categories; they are also labeled so as to distinguish between a home game and an away game. What makes this example different from the earlier ones is that the aim here is to forecast the outcome of an event: win, draw, or lose in our example. Sports betting companies use data on events in each minute of the game to predict the outcomes. Their aim is to help set the appropriate betting odds to ensure that they make a profit (Vaughan Williams, 2005). They use the methods of Chapters 7–10 to carry out the modeling and forecasting.

Figure 1.9 The Record of English Soccer Club Manchester United During the 2014–2015 Premier League Season [38 games played, 19 home and 19 away]



DISCUSSION QUESTION: *What conclusions can you draw from Figure 1.9 about the strength of the Manchester United team: Overall? Over time? Home and away?*

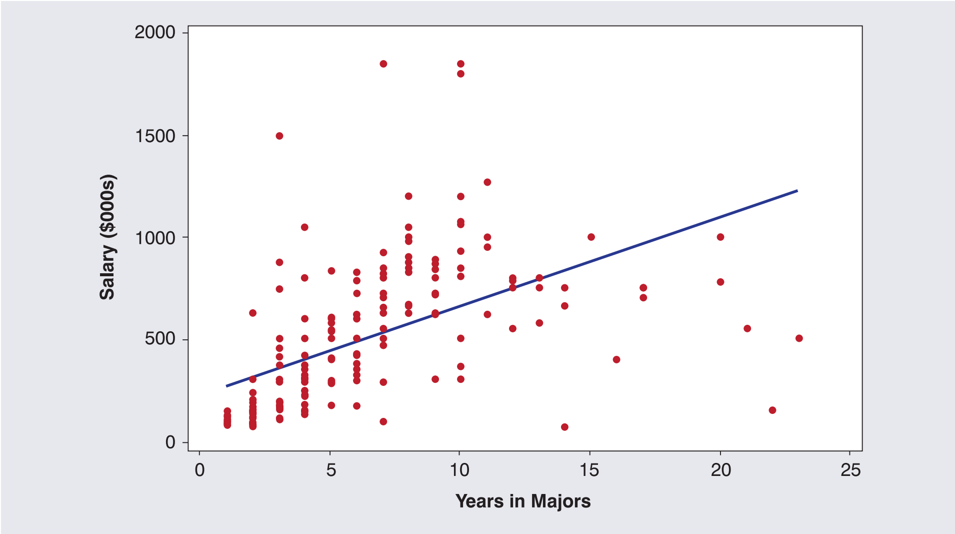
1.3.6 Sports Forecasting: A Cross-Sectional Example – Baseball Salaries

Sports management also relies on forecasting. The manager's decisions on whom to sign, whom to release, and even whom to play, all depend on forecasts of performance. Usually, such forecasts are made intuitively on the basis of the manager's experience. However, in one now famous example, Billy Beane, general manager of the Oakland Athletics baseball team, used statistical methods to identify players who were undervalued (see Lewis, 2003, for a nontechnical description of Beane's management ideas, subsequently filmed in *Moneyball*, 2011, with Brad Pitt). In thinking about the problem of trading players, one relevant relationship is what determines their current salary. Figure 1.10 shows the relationship between salary and the years played in the major leagues. The data here are cross sectional: All observations on players and their salaries were made over the same (relatively short) period. Critically, the time the data were recorded should be irrelevant to interpreting the relationship in the near future.

Cross-sectional data are measurements on multiple units (here, players) recorded over a single period for one or more variables (salaries, years in the major leagues).

Of course, many other variables affect salary, not the least of which are the player's previous season's performance and personal characteristics, such as age. In developing an understanding of whether a player is undervalued, these factors all have to be taken into account. We discuss this question in more detail in Chapter 7.

Figure 1.10 Scatterplot of Baseball Players' Salaries Against the Number of Years Played in the Major Leagues



Source: We wish to acknowledge StatLib of the Department of Statistics at Carnegie Mellon University.
Data: Baseball.xlsx

1.3.7 Random (Stochastic) Processes

Observations may occur over time without forming a time series as we have defined it. For example, consider the flow of customers into and out of a supermarket. An individual customer arrives at a certain time (arrival time), spends some time selecting items for purchase (processing time), then waits in line (waiting time) before being served at the checkout (service time). If we examine the number of arrivals per unit of time or the number of customers served in a given time, we can define a time series by using successive counts. However, from the perspective of the individual customer, the elapsed waiting and service times are important. The purpose of the analysis is typically to manage the service levels cost effectively. These individual times form, not a time series, but rather outputs from a random (or *stochastic*) process. Such processes are beyond the scope of this book; for greater detail see, for example, Winston and Albright (2015).

1.4 How to Forecast

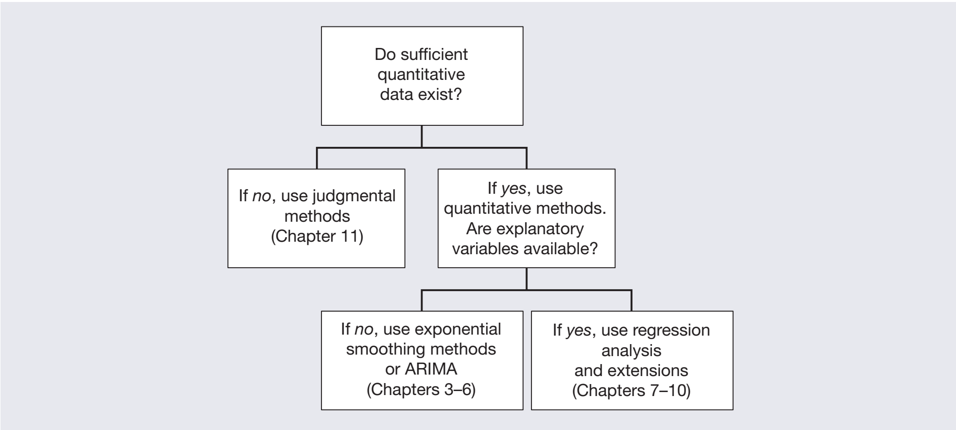
In Section 1.2, we used PIVASE to explain why we should generate forecasts. The next question is how? The how question depends heavily on the components of PIVASE. When we seek to establish a forecasting system, we first need to provide answers to the following questions:

- Do we have the necessary data to enable the use of statistical methods, or will the forecasting process be essentially judgmental? We leave the use of judgment in forecasting and its role in statistical models for later discussion (Chapters 11 and 13).
- If the process is based upon judgment, is the forecaster working alone or is a group involved?

- If the process is statistical, is the analysis to be based upon a single series or are explanatory variables to be included in the study?

The responses to these questions lead to a range of possible approaches, many of which are described in later chapters. The sequence of choices and the available methodologies are summarized briefly in Figure 1.11. A much more detailed discussion of the choice of forecasting procedures is deferred to Chapter 13, because we must first develop an understanding of each principal approach.

Figure 1.11 A Simplified Methodology Tree for Forecasting



1.5 Forecasting Step by Step

On the basis of our preliminary discussion of PIVASE, we can identify seven major steps in the forecasting process:

1. Define the forecasting and planning problem, the forecast horizon and decide the value of better forecasts.
2. Determine the resources to be devoted to providing the forecasts.
3. Collect relevant information, whether from a survey, from company records, or from information generated by other agencies (e.g., government figures).
4. Conduct an initial analysis of the data.
5. Select an appropriate forecasting method.
6. Generate forecasts.
7. Evaluate the forecasting exercise by checking forecasts against actual outcomes.

Ideally, we could use these seven steps as a checklist and work our way through them. Life is not quite that simple, however. For example, we used PIVASE to discuss step 1 in Section 1.1. Although we may be comfortable with our initial definition of the problem, we may have to revisit our understanding. Step 2 is also important: The resources assigned to the forecasting exercise should be commensurate with the potential added value of improved forecasts. If there is little value in achieving improved forecast accuracy, there's no point in using scarce resources for that purpose. By contrast, if the problem is such that small improvements in accuracy have a large payoff, spending substantial sums is worthwhile. Step 3 determines whether the forecasting task is feasible: The relevant data may not be available (either at all or in a timely fashion), so we would have to go back to the drawing board.

Once we have completed steps 1–3, we need to carry out an initial analysis of the data. We describe these steps in detail in Chapter 2, but it is worth noting that the analysis serves two purposes: cleaning the data and understanding the data. *Cleaning the data* involves looking for unusual observations and perhaps correcting data-recording errors; identifying factors that need to be taken into account, such as seasonal patterns or public holidays; and making sure that the data are comparable to each other and represent the phenomenon of interest. For example, shipments to wholesalers do not truly represent sales if they are made on a sale-or-return basis. The second component, *understanding the data*, enables the choice of forecasting method, as well as providing insights about the potential reliability of the resulting forecasts. As in other endeavors, the KISS principle is valuable (keep it simple, statistician).

Once our iterations through steps 1–4 are complete, we can move on to method selection, using the refinements of Figure 1.11. Only at this stage are we in a position to generate the forecasts. Much of the remainder of this book is concerned with the technical aspects of steps 5 and 6. The final step, which must not be overlooked, is to evaluate the forecasting process by comparing actual with forecast values. At the developmental stage, this can be done by means of a so-called *hold-out sample*, in which we split the available data into two parts and then use only the first part (the *estimation sample*) to calibrate the forecasting method. Once everything is set, we use the second part (the hold-out sample) to check the performance of the forecast. When the forecasting system is up and running, periodic checks on performance are essential. Overall, the seven steps should be seen as an integrated process that operates more smoothly as the forecaster gains practical experience; that integration is the subject of the final chapter of the book.

Whenever possible, the development of a forecasting method should include the partition of the data into an *estimation sample* (used to estimate the parameters, etc.) and a *hold-out sample* (used to test the performance of the proposed forecasting method). **Once the method is selected, the entire sample should be used for future forecasting purposes.**

Finally, it must be emphasized that forecasting is not a "once and done" operation. Forecast performance should be monitored on a regular basis and changes made to the system whenever performance falls short. For a detailed discussion see Fildes and Petropoulos (2015).

1.6 Computer Packages for Forecasting

Almost all statistical packages include routines for regression analyses and for time series analysis. The basic procedures for regression are relatively standard, although the quality of the procedures for testing the adequacy of alternative models with the use of diagnostics varies considerably. Computer software for time series analysis is even more variable. Some packages include only the most basic forecasting procedures, and the types of forecast provided vary enormously. In the course of this book, the reader will observe that we have used a number of general purpose software programs: Excel (graphs and macros only), Minitab, SAS, and SPSS. We have also used some specialized programs, notably EViews (for econometric modeling), and PcGive (both econometrics and forecasting). In particular, in this second edition we have introduced R, an open source free software that is thoroughly tested and progressively expanding as new routines become available. These routines include specialist forecasting algorithms. On the book's website, full R routines are available for most of the methods discussed in this book, notably using the forecast package developed by Rob Hyndman (*https://CRAN.R-project.org/package=forecast*), among other packages, some developed by the authors of this book. Each program has its merits, but none provides coverage of all the topics we discuss; we have tried to design the book so that it may be used compatibly with any major program. A brief summary of some of the capabilities of the major forecasting packages is available on the book website in Appendix B.

1.7 Data Sources

Thanks to the Internet, finding and downloading data are now relatively painless tasks. The data sets used in this book are all available on the website for the book (*http://www.wessexlearning.org/pobf2e/*), and they are all contained in Excel 2007 (.xlsx) files. These files may be read directly by any of the programs we have used.

All U.S. government departments have publicly available websites from which data may be downloaded; for example, two major agencies that supply economic data are the Census Bureau (*http://www.census.gov/*) and the Bureau of Labor Statistics (*http://www.bls.gov/*). In the United Kingdom, the UK National Statistics site provides similar data (*http://www.statistics.gov.uk/hub/index.html*). Also, all major companies have websites from which data on their overall operations can be obtained, although these data are not always packaged in a ready-to-download format.

Numerous specialist databases that provide downloadable data series are also available. Some are free, and others require either a personal or an institutional subscription. Following are several such sources of particular interest:

- The Federal Reserve Bank of St. Louis maintains an extensive database of U.S. and international macroeconomic series. The website is known as FRED (*https://fred.stlouisfed.org/*)
- EconStats (*http://www.econstats.com/index.htm*) contains a large amount of economic and financial data on the United States and on several other countries, including the United Kingdom, China, and Japan.
- Data.gov.uk (*http://Data.gov.uk/*) offers easy-to-access government statistical series for the United Kingdom.
- Econdata (*http://econdata.net/*) contains regional socioeconomic data for the United States.
- The Forecasting Principles website (*http://www.forecastingprinciples.com/*) provides a variety of links to other sources and data sets.

1.8 The Rest of the Book

The overall plan of the book broadly follows the different branches of the methodology tree shown in Figure 1.11. After a brief review of basic statistics and a discussion of how to measure forecast accuracy in Chapter 2, we go on to examine extrapolative (single-time-series) methods in Chapters 3 and 4. These methods are introduced in a pragmatic way as "good ideas"; the statistical underpinnings of the methods are then developed in Chapters 5 and 6. In Chapters 7–9, we develop models that may be used for either cross-sectional or time series data that can involve the inclusion of explanatory variables. Chapter 10 describes a variety of more advanced techniques, which have become an integral part of forecasting practice in recent years, while Chapter 11 discusses the role of judgment in forecasting and how it is best used. Judgmental methods are perhaps overused in modern industry, but they do have an important role to play, and it is necessary to understand how they interface with statistical procedures. With the various approaches to forecasting established, in Chapter 12 we go on to discuss their application in a variety of contexts. Finally, in Chapter 13, we examine forecasting in practice and try to provide signposts to guide the forecaster to successful applications of the tools of her trade.

Summary

In this chapter, we have discussed the reasons organizations have for forecasting, and its role in both operational and strategic planning. The why and what of forecasting have been analyzed in detail by using PIVASE: the purpose that underlies the forecasting activity, the forecast horizon, the information available, the value of the forecasting exercise, the analysis and development of a suitable method within the forecasting system, and finally, the need to evaluate forecasts. We then examined several series that showed different data patterns, including trend and seasonal components. This led to the idea of seasonal decomposition as a method for identifying the core characteristics of a time series. Other issues discussed were the examination of cross-sectional data for forecasting and the impact of unexpected events. Finally, as an introduction to the remainder of the book, we briefly explored the how of forecasting, which depends on the aims of the forecasting activity and the data available.

Minicases

Minicase 1.1 Inventory Planning

The Nuts'n'Bolts hardware store stocks some high-value items and a large number of relatively low-value components. Inventory is reviewed weekly, and orders are placed with suppliers when inventory levels indicate that current levels are "too low." Most suppliers deliver the low-value items within two weeks; some of the high-value items may take four weeks for delivery. How would you develop a forecasting system to predict the sales of such items at the individual SKU (Stock Keeping Unit) level? How would you use such forecasts to plan future purchases?

Use PIVASE to go through the necessary steps. For purposes of the discussion, assume that any historical data you might need can be made available.

Minicase 1.2 Long-Term Growth

You work for a consulting firm that is bidding on a contract to forecast the growth prospects of an electronics company over the next five years. The company has developed a specialized computer chip that is used in highly sensitive medical equipment and does not anticipate creating any new product lines, although periodic improvements to the chip are anticipated.

Write a short proposal describing how you would develop such forecasts, using PIVASE as a guide. For purposes of the assignment, assume that any historical data you might need can be made available, but describe carefully the nature of the data needed.

Minicase 1.3 Sales Forecasting

A company sells many hundreds of clothing products, historically through mail order and its own high street shops, but increasingly through the Internet. The firm also offers the customer four different purchase schemes, from pay on purchase through payment over different periods (of 6 and 12 months). The final plan allows a postponement of any payment for 12 months.

The firm has to ensure a supply of its advertised products to meet forecast demand. Adequate inventory has consequences for the distribution network. (The products need to be stocked and then shipped.) The retailer also has to decide which customers to target through mailings to stimulate further purchases.

What variables would the company need to forecast? What strategic financial decisions may need to be linked to the various customer purchase plans? Use the PIVASE framework to guide your thinking.

Minicase 1.4 Adjusting for Inflation

Use the data files for U.S. retail sales (*US_retail_sales_2.xlsx*) and for the U.S. Consumer Price Index (*US_CPI_2.xlsx*) to create a series called "real prices" by dividing sales by the *CPI*; then replicate Figures 1.2, 1.5, and 1.6 for this series. Compare your results with the three figures provided in the text, commenting on major differences.

References

Fildes, R. and Petropoulos, F. (2015). Improving forecast quality in practice. *Foresight,* 36, 5–12.

Harvey, A. C., and Durbin, J. (1986). The effects of seat belt legislation on British road casualties: A case study in structural time series modelling. *Journal of the Royal Statistical Society*, series A, 149, 187–227.

Lewis, M. (2003). *Moneyball: The Art of Winning an Unfair Game.* New York: Norton.

Vaughan Williams, L. (2005). *Information Efficiency in Financial and Betting Markets.* Cambridge, UK: Cambridge University Press,

Winston, W., and Albright, S. C. (2015). *Practical Management Science*, 5th ed. Pacific Grove, CA: Duxbury.

Appendix 1A Model-Based Probability

Forecasting by its nature involves uncertainty and we express our understanding of this uncertainty through the language of probability. The purpose of this appendix is not to provide a primer on probability as there are many sources that offer such introductions. Rather, the aim is to describe how probabilistic statements are generated in the context of time-series based forecasts.

Standard notions of probability may rest upon the notions of equally likely events, relative frequencies in repeated trials, or as a degree of belief but none of these ideas is entirely appropriate for the analysis of time series. A typical time series, such as Gross Domestic Product (*GDP*), has one observation each quarter which represents its observed value. Sampling may be involved in arriving at the final value of *GDP* in a given period but the reported number is unique and is not open to any interpretation of repeated sampling such as we employ in cross-sectional studies.

How then do we proceed? We rely on the concept of *model-based probability*. That is, we formulate a model of the process that generates the time series and derive probabilistic statements from that model. Of course, if the model is wrong, so will be the assessments of uncertainty. For that reason, throughout the book we emphasize careful model selection and checking of the underlying assumptions. For now, such issues do not concern us and we assume the model-based specification is correct. In fairly general terms we might specify a model as:

Observable variable = known component + unknown component.

Algebraically we may write this expression, at time t as:

$$Y_t = m_t + e_t,$$

m_t represents the known (or predictable) component or signal whereas e_t denotes the unknown component (random error or noise). Typically, it is assumed that the random error has a mean of zero so that the observable variable has mean m_t, also written as the expectation of the observable variable, or $E(Y_t) = m_t$. Thus, the *point forecast* for the series will be m_t and the uncertainty is expressed via e_t. Model-based probabilities are formulated using the assumed properties of the random error: on many occasions we will assume that the errors are normally distributed with a constant variance over time and that successive errors are independent of one another. More generally, any specification of the distribution of the errors will provide a framework for making statements about the uncertainty related to the point forecasts.

CHAPTER 2

Basic Tools for Forecasting

Table of Contents

DDD: Draw the doggone diagram!

— In memory of the late David Hildebrand, who stated the matter rather more forcibly!

Introduction

In most of the chapters in this book, we assume that some kind of database is available from which to build numerical forecasts. In some situations, the data may be already available (e.g., government figures for macroeconomic forecasting); in other cases, the data may be collected directly (e.g., sales figures or purchasing behavior from scanning records). Such data may be incomplete or subject to error, they may not relate directly to the key variables of interest, and they may not be available in a timely fashion. Nevertheless, they are all we have, and we must learn to understand and respect them if not to love them. Indeed, such is the basis of any good relationship!

At a conceptual level, we need to understand how the data are compiled and how they relate to the forecasting issues we seek to address. We must establish from the outset that the data are appropriate for the *purpose* (the P in PIVASE) we have in mind, as noted in Chapter 1. Otherwise, we are only living up to the old adage "Garbage in, garbage out." Although we will not keep repeating this message in chapter after chapter, the PIVASE issues must always be examined before embarking on a forecasting project. We must then examine the data to understand their structure and main features and to summarize the available information (the I in PIVASE).

In Section 2.1, we examine the types of data that arise in practice, and then we turn to graphical summaries in Sections 2.2 and 2.3. Section 2.4 describes the basic numerical summaries that are useful, and we then move on to measures of association in Section 2.5. Sometimes the original form of the data is not appropriate, and some kind of transformation or modification is needed. This topic is the focus of Section 2.6. The steps described in all these sections are essentially those of preliminary data analysis, although similar procedures will be undertaken later in the context of model diagnostics.

Methods for the generation of forecasts are the focus of later chapters, but in this chapter we consider the evaluation of outputs from the forecasting process. Thus, in Section 2.7, we examine measures of forecasting accuracy and the evaluation of forecasting performance (the E in PIVASE), after which we turn to prediction intervals in Section 2.8. The chapter ends with a discussion of some underlying principles in Section 2.9.

2.1 Types of Data

A database may be thought of as a table with multiple dimensions, as the following examples illustrate:

- *A survey of prospective voters in an upcoming election:* The measured variables might include voting intentions, party affiliation, age, gender, and address.
- *A portfolio of stocks:* For each company, we would record contact information, market capitalization, opening and closing stock prices, dividend payments over suitable periods, and news announcements.
- *The economy of the United States:* The factors of interest would certainly include gross domestic product (*GDP*), inflation, consumer expenditures, capital investment, unemployment, government revenues and expenditures, and imports and exports.

Data may be numeric, ordered, categorical, or even text-based.

A survey of voters refers to *cross-sectional data* in that what matters is the inherent variation across respondents. For practical purposes, we view the data as being collected in the same (short) time period. Of course, voters may change their minds at a later stage, and such shifts of opinion are a major source of discrepancies between opinion polls and election outcomes.

The closing price for a particular stock or fund, recorded daily over several months or even years represents an example of *time series data*. We are interested in the movement of the price over time. The same applies if we track the movements over time of aggregate macroeconomic variables such as *GDP*; it is their development over time that is important.

TYPES OF DATA

Cross-sectional data are measurements on multiple units, recorded in a single time period.

A *time series* is a set of comparable measurements recorded on a single variable over multiple time periods.

Panel data are cross-sectional measurements that are repeated over time, such as monthly expenditures for a sample of consumers.

From these examples, we see that a database may be cross-sectional, or time-dependent, or both. (Consider tracking voting intentions over time or looking at consumer expenditures for different regions of a country). Although forecasting practice often involves multiple series, such as the sales of different product lines, the methods we examine have the common theme of using past data to predict future outcomes. Thus, our primary focus in the first part of the book is on the use of time series data. For convenience, we begin our development of these ideas in the context of a single time series, even though applications may involve a large number of such series.

As methods of data capture have become more sophisticated (e.g., scanners in supermarkets), it has become possible to develop databases that relate to individuals, be they consumers or machines. Forecasting may then involve the use of cross-sectional data to predict individual preferences or to evaluate a new customer based upon individuals with similar demographic characteristics.

By way of example, consider the data shown in Tables[1] 2.1, 2.2, and 2.3. Table 2.1 shows the weekly sales of a consumer product in a certain market area; the product is produced by a major U.S. manufacturer. This data set will be examined in greater detail in Chapter 3. The data are genuine, but we have labeled the product WFJ Sales to preserve confidentiality. Table 2.2 shows the annual numbers of domestic passengers at Washington Dulles International Airport for the years 1963–2015. Clearly, both data sets are time series, but as may be seen in the time series plots shown later in Section 2.2 in Figures 2.1 and 2.2, the sales figures are fairly level (at least after the first 12 weeks or so) whereas the passenger series shows a strong upward movement, followed by a decline in recent years. Table 2.3, appearing in Section 2.3, involves cross-sectional data on forecasts for the German economy made by different forecasting agencies.

2.1.1 Use of Large Databases

A manager responsible for a large number of product lines may well claim that "the forecasting can all be done by computer, so there is no need to waste time on model building or detailed examinations of individual series." This assertion is half right. The computer can indeed remove most of the drudgery from the forecasting exercise; see, for example, the forecasting methods described in Chapter 3. However, a computer is like a sheepdog: Properly trained, it can deliver a sound flock of forecasts; poorly trained, it can create

1 Here, as throughout the book, we reproduce only a few lines from each data file to illustrate the nature of the data. The complete data set is available in the cited Excel file; it is also included in the chapter data file, e.g., *Chapter2_data_xlsx*. These may be downloaded from the book's website.

mayhem. Even if the task in question involves thousands of series to be forecast, there is no substitute for understanding the general structure of the data so that we can identify appropriate forecasting procedures. The manager can then focus on the products that are providing unusual results.

Table 2.1 Value of Weekly Sales of Product WFJ Sales ($)

Week	Sales
1	23056
2	24817
3	24300
…	…
60	28155
61	28404
62	34128

Week 1 is first week of January; *Data: WFJ_sales.xlsx*

Table 2.2 Washington Dulles International Airport, Domestic Passengers, 1963–2015

Year	Passengers (000s)
1963	641
1964	728
1965	920
…	…
2013	14958
2014	14393
2015	14463

Source: U.S. Department of Transportation, Bureau of Transport Statistics
Data: Dulles_2.xlsx

To build an effective forecasting system, therefore, we need to understand the kind of data we are handling. That does not mean examining every series in detail, or even at all, but rather looking at a sample of series in order to establish a framework for effective forecasting. Thus, we need to understand when and how to use forecasting methods, how to interpret the results, and how to recognize their limitations.

2.2 Time Series Plots[2]

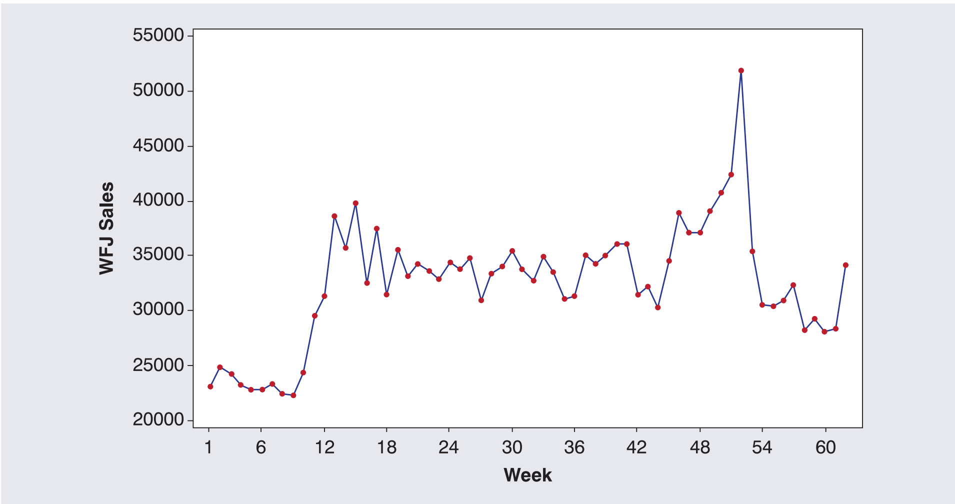
Our aim in the next several sections is not to provide detailed technical presentations on the construction of the various plots; rather, we indicate their application in the current context. All the major statistical and forecasting programs such as Autobox, EViews, Forecast Pro, SAS, SPSS and STATA provide detailed tutorials as well as Help commands. In this text while we use a variety of software, we provide guidance on using Excel, Minitab, SPSS, and R.

2 The plots in this chapter are generated by Minitab, unless stated otherwise.

As its name suggests, a time series plot shows the variable of interest on the vertical axis and time on the horizontal axis; as noted in Section 1.3, we will assume that the data relate either to equally spaced points in time (e.g. daily stock market closing prices) or to periods of equal length (e.g. weeks or months). As these examples indicate, some adjustments may be needed, such as excluding weekends and holidays for stock price series, or adjusting a monthly figure to allow for the number of days.

The time series plot for WFJ Sales is shown in Figure 2.1. Several features are immediately apparent. Sales are low for the first 12 weeks and then remain stable until week 46, when there is an increase over the Thanksgiving-to-Christmas period (weeks 48–51) followed by a peak in the last week of the year. Sales at the start of the next year are lower than for the final weeks of the previous year, but higher than for the corresponding period a year before. We should not make too much of data for one product over little more than a year, but inspection of the plot has revealed a number of interesting patterns that we might check out for similar products. If these patterns were found to persist across a number of product lines, we would need to take them into account in production planning. For example, the company might initiate extra shifts or overtime to cover peak periods and might plan to replenish inventories during slack times.

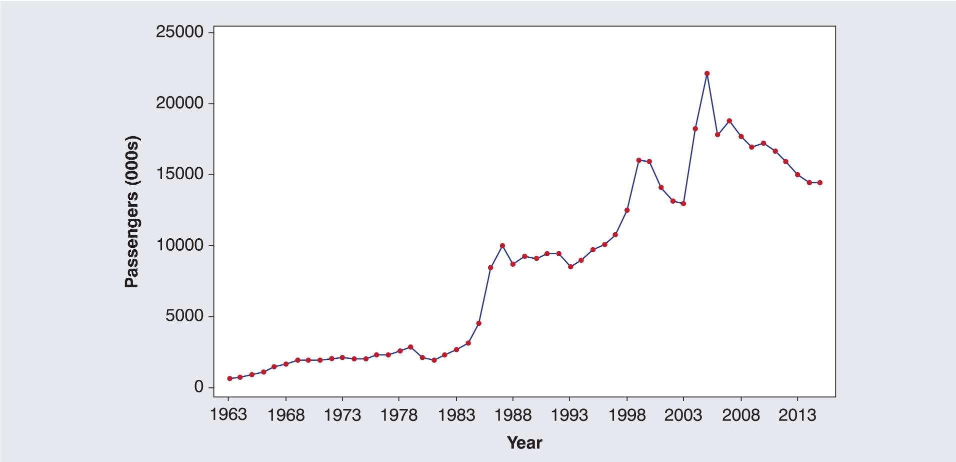
Figure 2.1 Time Series Plot of Weekly WFJ Sales



Data: WFJ_sales.xlsx

The second time plot, in Figure 2.2, presents the data from Table 2.2 on the annual number of airline passengers through Dulles Airport, Washington, DC. The figure shows steady growth from 1963 to 1979 (the airport opened in 1962), then a pause followed by rapid growth in the late eighties (after airport expansion in 1983-84). A further pause in the early nineties was followed by a long period of growth, with peaks in 1999 and 2005, and then a period of decline through to 2015. A detailed explanation of these changes lies outside the present discussion; it would require us to examine airport expansion plans, overall levels of passenger demand, the traffic at other airports in the area, and so on. The key point is that a time series plot can tell us a lot about the phenomenon under study and features that require explanation which often suggest suitable approaches to forecasting.

Figure 2.2 Time Series Plot of Domestic Passengers at Dulles, 1963–2015



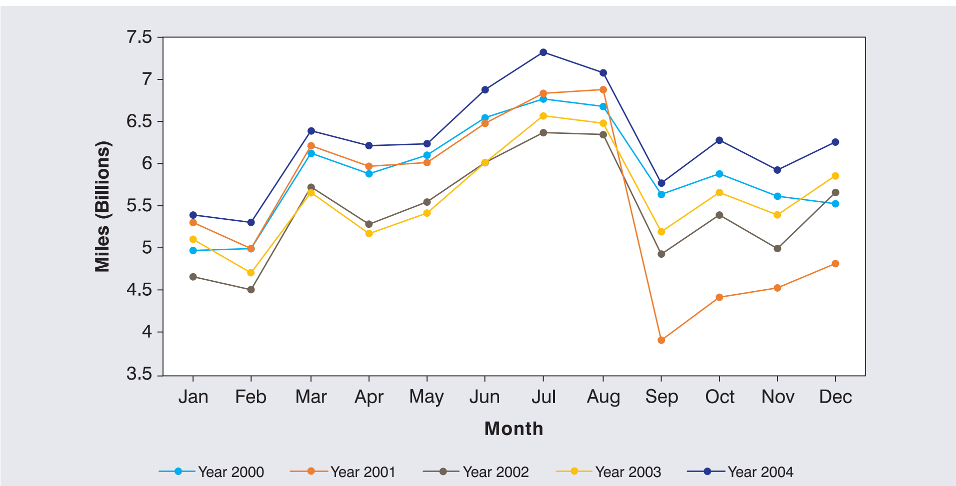
Data: Dulles_2.xlsx

2.2.1 Seasonal Plots

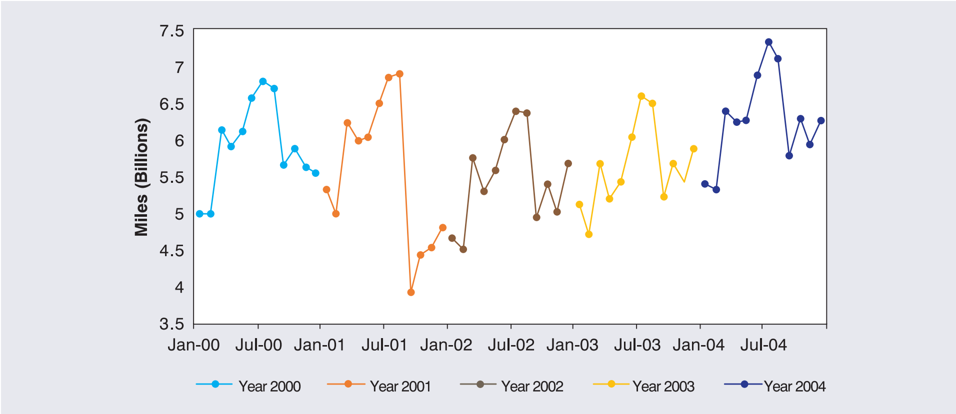
Figure 2.1 has some elements of a seasonal pattern (the end-of-year peak) but just over one year of data from which to identify seasonal behavior. Clearly, Figure 2.2 has no seasonal pattern, because the figures represent complete years. However, seasonal variations are often very important for planning purposes, and it is desirable to have a graphical procedure that allows us to explore whether seasonal patterns exist. For monthly data, for example, we may plot the dependent variable against the months and generate a separate, but overlaid, plot for a succession of years. In Figure 2.3A, we provide such a plot for airline revenue passenger miles (RPM). RPM measures the total number of revenue-generating miles flown by passengers of U.S. airlines, measured in billions of miles. To avoid cluttering the diagram, we use only five years of data, for 2000–2004; such a multicolored diagram that could be created online is more informative and can readily accommodate more years without confusion.

Figure 2.3 Seasonal Plots for Airline Revenue Passenger Miles for 2000-2004

A: Scatterplot by month, with years overlaid, of RPM vs. month



B: Time series plot of RPM, with each year identified as a subgroup



Data: Revenue_miles_2.xlsx

In Figure 2.3A, the line for each year lies above those for earlier years, with only rare exceptions. This configuration indicates the steady growth in airline traffic over the period shown. Figure 2.3B also shows the trend and the seasonal peaks and allows easy comparison of successive seasonal cycles. There is a major seasonal peak in the summer, as well as a lesser peak in March–April depending on the timing of Easter. These plots of the data provide considerable insight into the variations in demand for air travel. "Draw the doggone diagram" (DDD) is indeed wise counsel!

2.3 Scatterplots

The time series plots displayed so far show the evolution of a single series over time. As we saw with the seasonal plot, we can show multiple series on the same chart, although some care is required to make the axes sufficiently similar. Scatterplots may represent either cross sectional or time series data: we may plot one variable against another without any implication of causality. An alternative is to plot a variable of interest against one or more potential explanatory variables, to see how far knowledge of the explanatory variable(s) might improve the forecasts of the variable of interest.

Table 2.3 shows data from 25 forecasting organizations in Germany, each forecasting changes in eight macroeconomic variables, for year 2013. The variables of interest are Gross Domestic Product (*GDP*), Private Consumption (*Privcons*), Gross Fixed Capital Formation (*GFCF*), Exports, Imports, Government Surplus (*Govsurp*), Consumer Prices (*Consprix*) and Unemployment (*Unemp*).

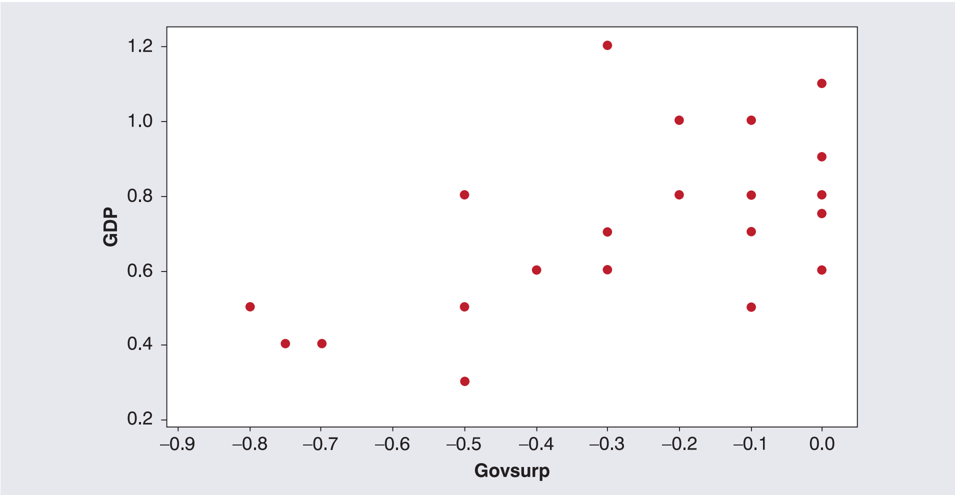
In Figure 2.4, we show a cross-sectional scatterplot for the variables *GDP* and *Govsurp*. In this case, we are not attempting to impute any kind of causal relationship, but rather we seek to identify possible associations across forecasts among the different providers. For example, Figure 2.4 indicates that those organizations who predicted stronger growth in *GDP* tended to be more optimistic about the change in the government's budget surplus (or deficit).

Table 2.3 Forecasts for the German Economy for 2013 Made by Different Forecasting Organizations

Institutions	GDP	Privcons	GFCF	Exports	Imports	Govsurp	Consprix	Unemp
Bundesbank	0.4	1	–0.1	1.9	3	–0.75	1.5	7.2
Commerzbank	0.5	1.3	0.1	2.8	4.1	–0.5	1.9	7.1
Deka	0.7	1.1	–0.3	3.3	3.3	–0.3	1.9	6.9
Deutsche Bank	0.3	0.6	1.1	3.2	4.2	–0.5	1.7	7
…	…	…	…	…	…	…	…	…
UBS	0.8	0.9	1.7	2.6	3.9	–0.2	2.1	7.2
Wirtschaftsweise	0.8	0.8	1.4	3.8	4.2	–0.5	2	6.9

Data: German_forecasts.xlsx; reproduced with permission from Müller-Dröge, Sinclair and Stekler (2016).

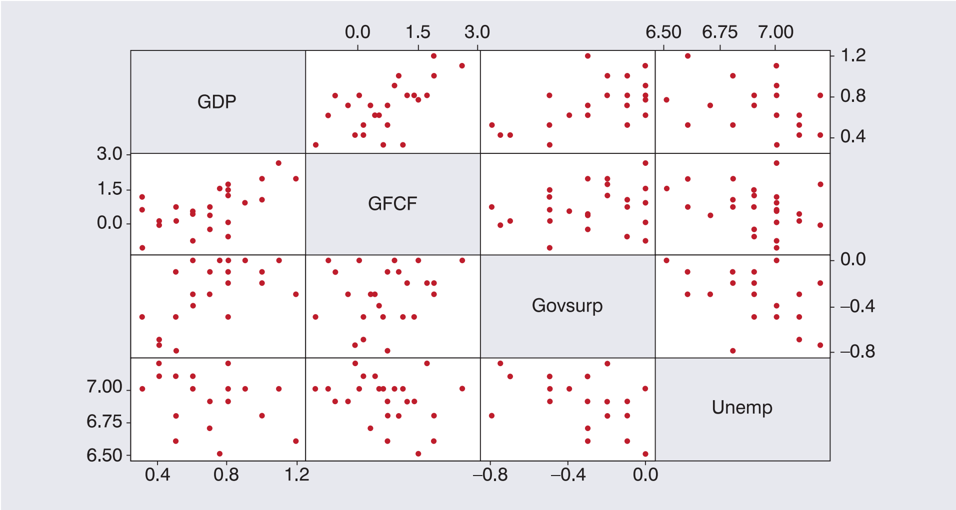
Figure 2.4 Scatterplot of GDP vs. Government Surplus for 25 Forecasting Organizations



We often have multiple variables of interest and want to look for relationships among them. Rather than generate a series of separate scatterplots, we may combine them into a *matrix plot*, which is just a two-way array of plots of each variable against each of the others. The plots in the bottom left are these same as those in the top right, but with the x- and y-axes reversed. The matrix plot provides a condensed summary of the relationships among multiple variables and is a useful screening device for relevant variables in the early stages of a forecasting exercise.

Such plots can become difficult to read as the number of variables increases, so Figure 2.5 shows such a plot for just four of the variables: *GDP*, *GFCF*, *Govsurp*, and *Unemp*. The forecasts for the first three variables tend to show positive associations, and those three factors tend to be negatively associated with forecasts for unemployment, as might be expected, although the associations are rather weak in some cases.

Figure 2.5 Matrix Plot of GDP, GFCF, Govsurp and Unemp for 25 Forecasting Organizations



2.4 Summarizing the Data

Graphical summaries provide invaluable insights. Time plots and scatterplots should always be used in the early stages of a forecasting study to aid understanding. Furthermore, as we will see in later chapters, such diagrams also play a valuable role in providing diagnostics for further model development. Even when we have a large number of items to forecast, plots for a sample from the whole set of series provide useful guidance and insights.

At the same time, we must recognize that, although graphical methods provide qualitative insights, we often need some kind of numerical summary, such as the average level of sales over time or the variability in P–E (Price-Earnings) ratios across companies. These measures are also valuable for diagnostic purposes, when we seek to summarize forecasting errors, as in Section 2.7.

2.4.1 Notational Conventions

At this stage, we need to elaborate upon some notational conventions because we will use the following framework throughout the remainder of the book:

1. *Random variables and observations:* When we speak of an *observation*, it is something we have already recorded, a specific number or category. By contrast, when we talk about *future observations*, uncertainty exists. For example, for tomorrow's closing price of the Dow Jones Index, we face a range of possibilities that can be described by a probability distribution. Such a variable, with both a set of possible values and an associated probability distribution, is known as a *random variable*. Texts with a more theoretical orientation often use uppercase letters to denote random variables and lowercase letters for observations that have already been recorded. By contrast, books that are more applied often make no distinction but rely upon the context to make the difference clear. We will not make such a distinction and generally use the same notation for both existing observations and random variables.

2. *Variables and parameters:* As just noted, *variables* are entities that we can observe, such as sales or incomes. By contrast, *parameters* contribute to the description of an underlying process (e.g., a population mean) and are typically not observable. We distinguish these concepts by using the Roman alphabet for observed variables (sample values), but Greek letters for parameters (population values). Thus, the variable we wish to forecast will always be denoted by Y and, where appropriate, the sample mean and standard deviation by $\bar{Y}$ and S. The corresponding population mean and standard deviation will be denoted by μ (mu) and σ (sigma), respectively.
3. *Probability distributions:* The concept of a probability distribution of future observations can be somewhat elusive in a time series context; we make use of model-based probabilities, as described in Appendix 1A. As an example, consider the daily closing prices for a stock such as IBM. There is only one IBM, and only one closing price each day. Nevertheless, we could define a probability distribution for future closing prices, known as a *predictive* (or *prediction*) *distribution*. That is, probabilities are associated with the set of possible future prices and formulated using a model-based description of the past behavior of IBM stock. The "population" (of possible prices) is conceptual in that only one actual value will ever be recorded from each future closing.

2.4.2 Measures of Average

By far the most important measure of average is the *arithmetic mean*, often known simply as the mean or the average.

ARITHMETIC MEAN

Given a set of n values Y_1, Y_2. …. Y_n, the arithmetic mean is

$$\bar{Y} = \frac{Y_1 + Y_2 + \ldots + Y_n}{n} = \frac{1}{n}\sum_{i=1}^{i=n} Y_i. \tag{2.1}$$

That is, the sum of the observations is divided by the number of values included.

When the range of the summation is clear from the context, such as the index going from 1 to n in the formula, we often write the summation sign without including the limits.

An alternative measure of average is given by the *median*, defined as follows:

MEDIAN AND ORDER STATISTICS

Given a set of n values Y_1, Y_2. …. Y_n we place these values in ascending order to define the *order statistics*, written as $Y_{(1)} \le Y_{(2)} \le \ldots \le Y_{(n)}$. Then the median is the "middle" value:

If n is odd, $n = 2m + 1$ and the *median* is $Y_{(m+1)}$.

If n is even, $n = 2m$ and the *median* is $½[Y_{(m)}+Y_{(m+1)}]$.

The median is the value where 50% of observations lie below it and 50% above – the middle "observation".

Example 2.1: Calculation of the mean and median

Suppose the sales of a popular book over a seven-week period are as follows:

Week	1	2	3	4	5	6	7
Sales (000s)	15	10	12	16	9	8	14

The mean is $\bar{Y} = \dfrac{(15 + 10 + 12 + 16 + 9 + 8 + 14)}{7} = 12.$

The *order statistics* are 8, 9, 10, 12, 14, 15, 16.

Hence, the median is the fourth value in the sequence, which also happens to be 12.

If data for week 8 now becomes available (sales = 16), the mean becomes 12.5 and the median is ½(12 + 14) = 13. However, suppose that sales for week 8 had been 116 (thousands), because of a sudden surge in popularity. Then the mean becomes 25, yet the median remains at 13. In general, the mean is sensitive to extreme observations but the median is not.

Which value represents the "true" average? The question cannot be answered as framed. The median provides a better view of weekly sales over the first eight weeks, but we are more interested in the numbers actually sold. The forecaster has the unenviable task of trying to decide whether future sales will continue at the giddy level of 100,000+ or whether they will revert to the earlier, more modest level. The wise forecaster would enquire into the reasons for the sudden jump, such as a rare large order or a major publicity event. ■

DISCUSSION QUESTION: *How would you describe housing prices in an area if you were (a) planning to buy a home in the area and (b) seeking to forecast revenues from the local property tax?*

2.4.3 Measures of Variation

A safe investment is an investment whose value does not fluctuate much over time. Similarly, inventory planning is much more straightforward if sales are virtually the same in each period. Implicit in both of these statements is the idea that we use some measure of variability to evaluate risk, whether of losing money or of running out of stock. Three measures of variability are in common use: the *range*, the *mean absolute deviation*, and the *standard deviation*. The standard deviation is derived from the *variance*, which we also define here.

RANGE

The *range* denotes the difference between the largest and smallest values in the sample:

$$\text{Range} = Y_{(n)} - Y_{(1)}.$$

The *deviations* are defined as the differences between each observation and the mean. By construction, the mean of the deviations is zero; so, to compute a measure of variability, we use either the absolute values or the squared values. The absolute values are indicated by vertical lines on either side of the variable, such as $|d|$ for the absolute value of the deviation d.

If we use the squares, our units of measurement become squared also. For example, revenues (in \$) become $(\$)^2$, so we reverse the operation after computing the average by taking the square root. These various measures are defined as follows, in terms of the deviations:

MEASURES OF DISPERSION

The *mean absolute deviation* is the average of the deviations about the mean, irrespective of the sign:

$$MAD = \frac{\sum |d_i|}{n}. \tag{2.2}$$

The *variance* is an average of the squared deviations about the mean:

$$S^2 = \frac{\sum d_i^2}{(n-1)}. \tag{2.3}$$

The *standard deviation* is the square root of the variance:

$$S = \sqrt{S^2} = \sqrt{\frac{\sum d_i^2}{(n-1)}}. \tag{2.4}$$

Example 2.2: Calculation of measures of variation

Consider the values for the seven weeks of book sales, given in Example 2.1. From the order statistics, we immediately see that the range is

$$Range = 16 - 8 = 8.$$

However, if week 8 is entered with sales = 116, the range shoots up to 116 – 8 = 108. This simple example illustrates both the strength and the weakness of the range: It is very easy to compute, but it is severely affected by extreme values. The vulnerability of the range to extreme values makes it unsuitable for most purposes in forecasting.

The deviations for the seven weeks are as follows (the mean is 12):

Week	1	2	3	4	5	6	7	Sums
Sales (000s)	15	10	12	16	9	8	14	84
Deviation	+3	–2	0	+4	–3	–4	+2	0
$\|d\|$	3	2	0	4	3	4	2	18
d^2	9	4	0	16	9	16	4	58

From the table, we have MAD = 18/7 = 2.57, S^2 = 58/6 = 9.67 and S = 3.11.

Why do we use $(n - 1)$ rather than n in the denominator of the variance? The reason is that, because we are using the deviations, if we had only one observation, its deviation would necessarily be zero. In other words, we have no information about the variability in the data. Likewise, in our sample of seven, if we know six of the deviations, we can work out the value of the seventh observation from the fact that they must sum to zero. In effect, by subtracting the mean from each observation, we have "lost" an observation. In statistical parlance, this is known as *losing a degree of freedom* (*DF*), and we say that the variance is computed under the assumption that there are $(n - 1)$ degrees of freedom, which we abbreviate to $(n - 1)$ DF. In later chapters, we sometimes lose several DF, and the measure of variability will change accordingly. This adjustment has the benefit of making the sample variance an unbiased estimator of the population variance. (For a more detailed discussion of unbiasedness, see Section A3 of Appendix A, available on the book's website.)

Note: S gives greater weight to the more extreme observations by squaring them, and it may be shown that $S >$ MAD whenever MAD is greater than zero. ■

2.4.4 Assessing Variability

The statement that book sales have a standard deviation of 3.11 (thousand, remember) conveys little about the inherent variability in the data from week to week, unless, like any penniless author, we live and breathe details about the sales of that particular book. To produce a more standard frame of reference, we use *standardized scores*. Given a sample mean $\bar{Y}$ and sample standard deviation S, we define the standardized scores for the observations, also known as Z-scores, as

$$Z_i = \frac{Y_i - \bar{Y}}{S}.$$

Each deviation is divided by the standard deviation. It follows that the Z-scores have zero mean and a standard deviation equal to 1.0. Following our simple example, we obtain the following table:

Week	1	2	3	4	5	6	7
Sales (000s)	15	10	12	16	9	8	14
Deviation	+3	–2	0	+4	–3	–4	+2
Z-score	0.96	–0.64	0	1.29	–0.96	–1.29	0.64

The Z-scores still do not provide much information until we provide a frame of reference. In this book, we typically use Z-scores to examine forecast errors and proceed in three steps:

1. Check that the observed distribution of the errors is approximately normal. (For details, see Appendix A6 on the book's website.)
2. If the assumption is satisfied, relate the Z-score to the normal tables (available through any statistics textbook, e.g. Anderson *et al.*, 2014). The Excel function is called NORM.DIST (value,0,1,1), which calculates the cumulative probability that the Z-score is less than "value" drawn from a normal distribution with mean 0 and variance 1:
   - The probability that $|Z| > 1$ is about 0.32 (or about 3 in 10).
   - The probability that $|Z| > 2$ is about 0.046 (about 5 in 100).
   - The probability that $|Z| > 3$ is about 0.0027 (about 3 in 1000).

   Most software packages include similar functions.
3. Create a time series plot of the residuals (and/or Z-scores), when appropriate, to determine whether any observations appear to be extreme.

At this stage, we do not pursue the systematic use of Z-scores, except to recognize that whenever you see a Z-score whose absolute value is greater than 3, the observation is atypical, because the probability of such an occurrence is less than 3 in 1000 (if the distribution is approximately normal). Often, such large values signify that something unusual has happened, and we refer to these kinds of observations as *outliers*. In cross-sectional studies, it is sometimes admissible to just delete such observations (e.g., a report of a 280-year-old man is undoubtedly a recording error). In time series forecasting, we wish to retain the complete sequence of values and must investigate more closely, often finding special circumstances (e.g., a strike, bad weather, a special sales promotion) for which we had not allowed. Outliers indicate the need for further exploration, not routine rejection. We defer the detailed treatment of outliers to Chapter 9.

2.4.5 An Example: Forecasts for the German Economy

The default summary outputs from Excel for the data in Table 2.3 on forecasts for the German economy are shown in Table 2.4. The output from other programs may have a somewhat different format, but the summary measures included are similar and most programs allow a variety of options. Excel typically produces too many decimal places; for ease of comparison, our output has been edited to produce a reasonable number of decimal places.

Table 2.4 Descriptive Statistics for Forecasts of the German Economy

	GDP	Privcons	GFCF	Exports	Imports	Govsurp	Consprix	Unemp
Mean	0.682	0.868	0.656	3.344	3.692	-0.294	1.894	6.92
Standard Error	0.050	0.060	0.180	0.190	0.194	0.049	0.054	0.036
Median	0.7	0.9	0.7	3.5	4	-0.3	1.9	7
Mode	0.8	1	0.7	3	4.1	-0.5	2	7
Standard Deviation	0.249	0.302	0.900	0.950	0.970	0.247	0.269	0.180
Sample Variance	0.062	0.091	0.810	0.903	0.941	0.061	0.073	0.033
Kurtosis	-0.53	-0.37	-0.21	2.34	5.19	-0.70	4.26	0.17
Skewness	0.19	-0.41	0.07	-0.67	-2.03	-0.53	1.29	-0.70
Range	0.9	1.2	3.7	4.9	4.6	0.8	1.3	0.7
Minimum	0.3	0.2	-1.1	0.6	0.4	-0.8	1.5	6.5
Maximum	1.2	1.4	2.6	5.5	5	0	2.8	7.2
Count	25	25	25	25	25	25	25	25

Data: German_forecasts.xlsx

Given the mean and standard deviation, we proceed to compute the Z-scores for each variable, shown in Table 2.5. We list those institutions that have one or more Z-scores greater than 2.0 in absolute value. It is noteworthy that the forecasters tend to be relatively optimistic or pessimistic across the board. We explore this phenomenon further by computing the sum of the Z-scores for each institution and, finally, computing the Z-scores for this sum. Helaba and DIW stand out as much more optimistic than the others.

Table 2.5 *Z*-Scores for Forecasts of the German Economy

Institutions	GDP	Privcons	GFCF	Exports	Imports	Govsurp	Consprix	Unemp	SUM	Z
Bundesbank	-1.13	0.44	-0.84	-1.52	-0.71	-1.85	-1.46	1.55	-5.53	-2.09
DIW	0.88	0.77	0.27	0.90	0.94	1.19	3.36	0.44	8.75	3.31
Feri	2.08	1.10	1.38	0.80	0.42	-0.02	0.39	-1.78	4.37	1.65
Gemeinschafts	1.28	0.77	1.38	0.48	0.94	0.38	0.76	-0.67	5.32	2.01
Helaba	1.68	1.10	2.16	2.27	1.35	1.19	0.39	0.44	10.59	4.00
ING	0.47	-1.55	0.60	-2.89	-3.39	0.38	0.39	-0.11	-6.09	-2.30
Landesbank Berlin	-0.73	0.44	0.05	-1.20	-0.82	-2.05	-1.09	-0.67	-6.07	-2.30
MM Wartburg	-0.33	-0.56	-1.62	0.90	-2.16	1.19	-1.46	0.44	-3.59	-1.36
RWI	-1.54	-2.21	-1.95	-0.36	-0.51	-0.83	-0.72	0.44	-7.68	-2.90

2.5 Correlation

In the previous section, we produced numerical summaries to complement the graphical analysis of Section 2.2. We now develop a statistic that performs a similar function for the scatterplots of Section 2.3: the *correlation coefficient*. Before defining the coefficient, we examine Figure 2.6; in each case, the horizontal axis may be interpreted as time. The six plots suggest the following:

*Y*1 increases with time and is perfectly related to time.

*Y*2 decreases with time and is perfectly related to time.

*Y*3 tends to increase with time but is not perfectly related to time.

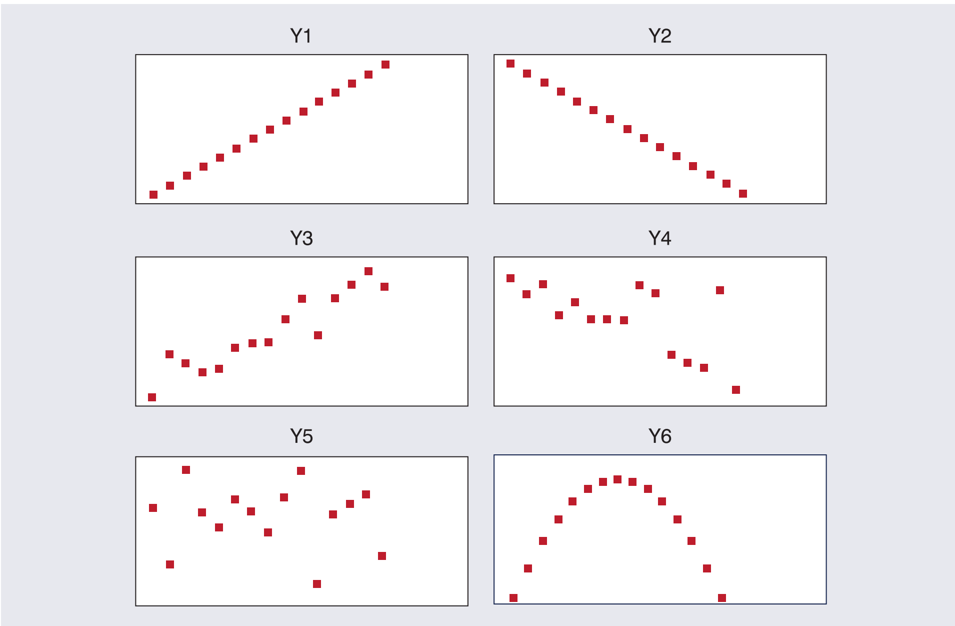
*Y*4 tends to decrease with time, but the relationship is weaker than that for *Y*3.

*Y*5 shows virtually no relationship with time.

*Y*6 is perfectly related to time, but the relationship is not linear.

Our measure should reflect these differences but not be affected by changes in the origin or changes of scale. The origins and scales of the variables are deliberately omitted from the diagrams because they do not affect the degree of association between the two variables. The most commonly used measure that satisfies these criteria is the *(Pearson) product moment correlation coefficient*, which we simply refer to as *the* correlation. We use the letter r to denote the sample coefficient and the Greek letter ρ (rho) to denote the corresponding quantity in the population. Rho is calculated by replacing the sample components by their corresponding population values.

Figure 2.6 Plots of Hypothetical Data Against Time



THE CORRELATION COEFFICIENT

The sample *correlation* between X and Y is defined as

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}. \tag{2.5}$$

The numerator represents the sum of cross products of the deviations, and the denominator terms are sums of squares of the two sets of deviations.

When we divide denominator by $(n - 1)$, the two terms inside the square-root sign become the sample variances of X and Y, respectively; Thus, on taking the square root, the denominator represents the product of the two standard deviations S_X and S_Y. The numerator divided by $(n - 1)$ is known as the sample *covariance* between X and Y, denoted by S_{XY}. Thus, the correlation may be written as

$$\mathrm{r} = \frac{S_{XY}}{S_X S_Y}. \tag{2.6}$$

It may be shown that, for $Y1$ in Figure 2.8, $r = 1$, the maximum value possible. Similarly, $Y2$ has $r = -1$, the minimum possible. The other correlations are, 0.93, –0.66, –0.09, and 0, for $Y3$, $Y4$, $Y5$, and $Y6$, respectively. In general, we see that the *absolute* value of r declines as the relationship gets weaker. At first sight, the result for $Y6$ appears odd: There is a clear relationship with X, but the correlation is zero. The reason for this is that r measures *linear* association, but the relationship of $Y6$ with X is *quadratic* rather than linear. A good example is the relationship between total revenue and price: Charge too much or too little, and total revenue is low.

DISCUSSION QUESTION: *How would you measure the relationship between Y6 and time?*

Example 2.3: Calculation of the correlation

Based on the data from Example 2.1, the detailed calculations for the correlation between sales and time are shown in the table that follows. A spreadsheet could readily be set up in this format for direct calculations, but all standard software packages have a correlation function.

Week, X	1	2	3	4	5	6	7	Sums
Sales (000s), Y	15	10	12	16	9	8	14	
$X - \bar{X}$	– 3	– 2	–1	0	1	2	3	0
$Y - \bar{Y}$	+3	– 2	0	+4	– 3	– 4	+2	0
$(X - \bar{X})^2$	9	4	1	0	1	4	9	28
$(Y - \bar{Y})^2$	9	4	0	16	9	16	4	58
$(X - \bar{X})(Y - \bar{Y})$	– 9	4	0	0	– 3	– 8	6	–10

Thus, $S_{XY} = -10/6$, $S_X = \sqrt{28/6}$ and $S_Y = \sqrt{58/6}$, so $r = -0.248$.

The example shows a weak negative correlation for sales with time; that is, sales may be declining slightly over time. ■

■ Example 2.4: Correlation among forecasts for the German economy

The correlations among the eight sets of forecasts are shown in Table 2.6. The upper number in each pair is the correlation and the lower figure is the *P*-value, enabling a (two-sided) test of whether the coefficient is significantly different from zero. A number of the results are interesting; in particular, it is evident that forecasts of strong growth in *GDP* are associated with other aspects of economic well-being, notably *GFCF*, *Exports*, and *Govsurp*. ■

Table 2.6 Correlations Among Forecasts for the German Economy

	GDP	Privcons	GFCF	Exports	Imports	Govsurp	Consprix
Privcons	0.430 (0.032)						
GFCF	0.601 (0.001)	0.295 (0.153)					
Exports	0.459 (0.021)	0.262 (0.206)	0.192 (0.357)				
Imports	0.209 (0.315)	0.469 (0.018)	0.371 (0.068)	0.641 (0.001)			
Govsurp	0.620 (0.001)	0.033 (0.874)	0.223 (0.285)	0.518 (0.008)	0.085 (0.688)		
Consprix	0.317 (0.123)	0.192 (0.358)	0.336 (0.101)	0.210 (0.313)	0.382 (0.060)	0.179 (0.391)	
Unemp	–0.363 (0.074)	0.073 (0.727)	–0.295 (0.153)	–0.273 (0.187)	–0.087 (0.678)	–0.317 (0.123)	0.063 (0.766)

Data: German_forecasts.xlsx

2.6 Transformations

We now examine the annual figures for the number of passengers on domestic flights out of Dulles airport for the years 1963–2015. The descriptive statistics are as follows:

Descriptive Statistics: Dulles Passengers								
Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Passengers	53	8456	6397	641	2044	8947	14428	22129

Dulles, we have a problem! What does the average of 8456 mean? Such levels were typical of the mid-eighties, but the “average” in a strongly trending series like this one has little or no meaning. Certainly, using either the mean or the median to forecast the next year’s traffic would make no sense.

How should we deal with a series that reveals a strong trend? Everyday conversation provides a clue. We talk of the return on an investment, an increase in company sales over the previous year, or the percentage change in *GDP*. This approach is partly a matter of convenience; some ideas are more readily communicated using (percentage) changes rather than raw figures. Thus, we may regard 3 percent growth in *GDP* as reasonable, 1 percent as anemic, and 7 percent as unsustainable. The same information conveyed in dollars would be hard to comprehend.

From the forecasting perspective, there are two further reasons for considering such alternatives:

- Because we are examining changes, the forecast relates directly back to the previously observed value; such forecasts are unlikely to be wildly off target
- Mean levels in terms of absolute changes or percentage changes are often more stable and more meaningful than averages computed from the original series.

We now explore these ideas in greater detail.

2.6.1 Differences and Growth Rates

The change in the absolute level of a series from one period to the next is known as the *(first) difference*[3] of the series and is written as

$$DY_t = Y_t - Y_{t-1}. \tag{2.7}$$

At time t, the previous value Y_{t-1} is already known. If the forecast for the difference is written as $\hat{D}_t$. The forecast for period t, denoted by F_t, becomes

$$F_t = Y_{t-1} + \hat{D}_t. \tag{2.8}$$

That is, the final forecast is the previous level plus the forecast of the change. Similarly, the rate of growth over time is written as[4]

$$GY_t = 100\,\frac{(Y_t - Y_{t-1})}{Y_{t-1}}. \tag{2.9}$$

Expression (2.9) also defines the one-period return on an investment, given the opening price of Y_{t-1}. Once the growth rate, denoted by $\hat{G}_t$ has been predicted, the forecast for the next period is

$$F_t = Y_{t-1}\left(1 + \frac{\hat{G}_t}{100}\right). \tag{2.10}$$

The time plots for *DY* and *GY* for the Dulles passengers' series are shown in Figure 2.7A and 2.7B. Both series show a fairly stable level over time, so the mean becomes a more useful summary of the transformed series, although *GY* is trending downward, indicating a slowing in percentage growth and then a decline.

Another feature of Figure 2.7A is that the variability in *DY* is much greater at the end of the series than it is at the beginning. By contrast, the *GY* series has more consistent fluctuations. We might claim that *GY* has a stable variance over time, a claim that would be hard to make for *DY*. Which should we use? In part, the choice will depend upon the purpose behind the forecasting exercise, but an often reasonable guideline is a commonsense one: Do you naturally think of changes in the time series in absolute terms or in relative (i.e., percentage) terms? If the answer is "absolute," use *DY*; if it is "relative," use *GY*. In this case,

---

3 Some texts use the Greek capital letter Δ (delta) and others use ∇ (an inverted delta, or "del"), but the use of *D* seems a better mnemonic device for "difference."

4 The use of *G* to describe the growth rate is nonstandard; we use it here for the same reason as before: It is a convenient mnemonic device.

both transformed series show some unusual values, so further investigation is warranted. Out of interest, we note that Dulles airport underwent a significant expansion in 1984.

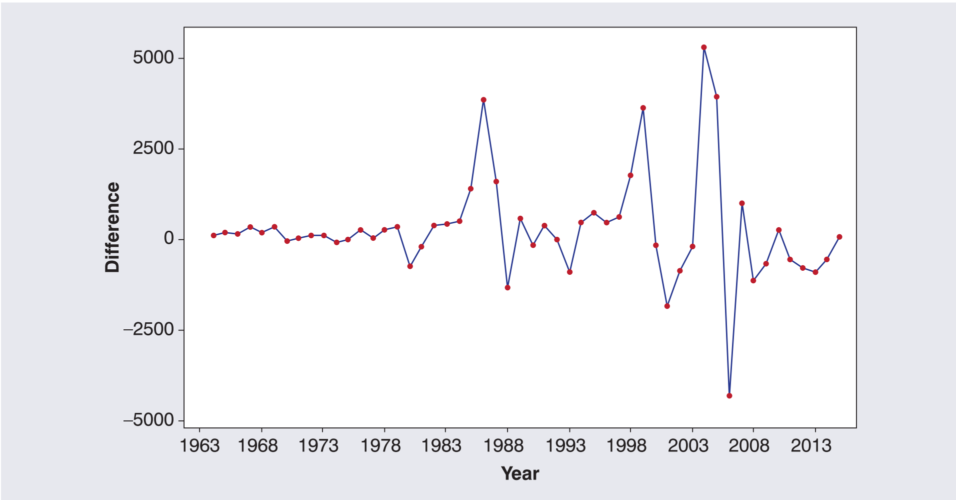
The summary statistics for *DY* and *GY* are as follows:

Descriptive Statistics: Difference, Growth Rate in Dulles Passengers								
Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Difference	52	266	1463	–4342	–213	135	445	5285
Growth Rate	52	7.47	17.87	–26.99	–3.63	4.57	16.97	84.95

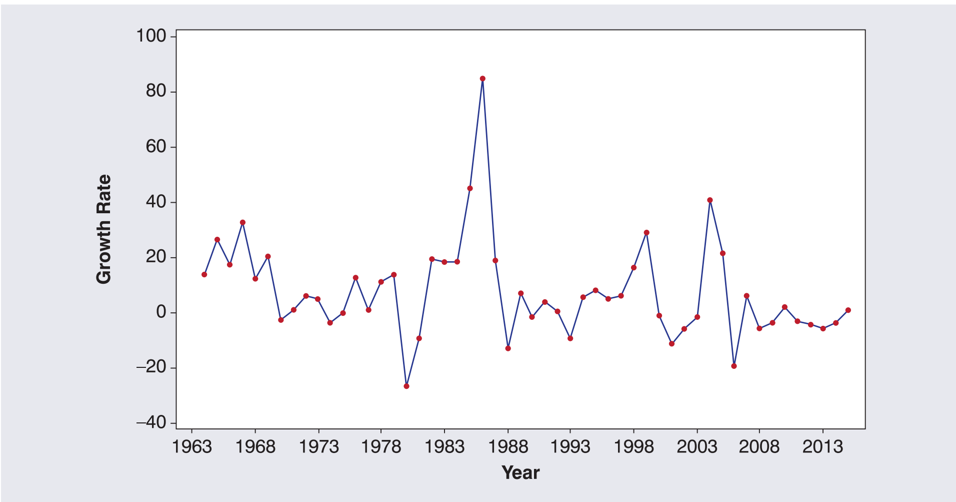
These figures also reflect the considerable fluctuations that appear in each series.

Figure 2.7

A: Time Series Plot for the First Differences of the Dulles Passengers Series



B: Time Series Plot for the Growth Rates for the Dulles Passengers Series



Data: Dulles_2.xlsx

2.6.2 The Log Transform

George Orwell's classic novel *1984* contains a scene in which the chocolate ration is reduced by 50 percent and then increased by 50 percent. The main character, Winston, complains that he does not have as much chocolate as before, but he is sharply rebuked for his remarks. However, Winston is right:

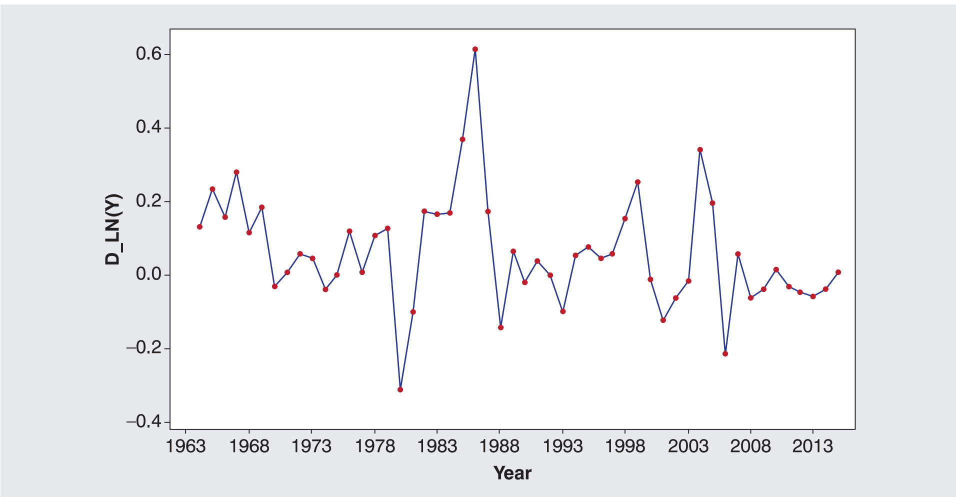
$$\left(1 - \frac{50}{100}\right)\left(1 + \frac{50}{100}\right) = 0.75.$$

So Winston has 25 percent less chocolate than before. To avoid this asymmetry, we may use the logarithmic (or log) transform, usually with the *natural logarithm* defined on the base $e = 2.71828 \ldots$.[5] The log transform may be written as $L_t = \ln(Y_t)$ and any software package has a function such as *ln* to compute the natural logarithm. To convert back to the original units, we use an inverse function (or reverse transformation) called *exp*, the exponential function, so we have $Y_t = \exp(L_t)$. The (first) difference in logarithms represents the logarithm of the ratio:

$$DL_t = \ln(Y_t \,/\, Y_{t-1}) = \ln(Y_t) - \ln(Y_{t-1}). \tag{2.11}$$

The primary purpose of the log transform is to convert exponential (or proportional) growth into linear growth. The transform often has the secondary benefit of stabilizing the variance, as did the use of growth rates. Indeed, the log and growth rate transforms tend to produce similar results, as can be seen by comparing the plot of the log differences for the Dulles passengers' series in Figure 2.8 with Figure 2.7B.

Figure 2.8 Time Series Plot for the First Difference of Logarithms for the Dulles Passengers Series



Data: Dulles_2.xlsx

If we generate a forecast of the log difference, say, $\widehat{DL}_t$, then the forecast for the original series requires that we reverse the log transform by using the exponential function. Given the previous value, Y_{t-1}, the forecast for period t becomes

$$F_t = Y_{t-1} \exp(\widehat{DL}_t). \tag{2.12}$$

5 If $Y = e^X$ or $\exp(X)$, the $\log_e(Y)$ or $\ln(Y) = X$, e being the base for $\log_e Y$.

Example 2.5: Calculation of forecast using log differences

The actual number of passengers using Dulles International Airport in 2014 was 14393 (in thousands). To make a forecast for 2015, we might use the last value for the log difference, which is ln(14393) – ln(14958) = –0.0385. Then equation (2.12) yields

$$F_t = 14393 \times \exp(-0.0385) = 14393 \times 0.9622 = 13849.$$

This short term forecast assumes a continuing decline, whereas the passenger traffic increased slightly. This raises the question of how to measure forecast accuracy, a topic we now address. ■

Note: When x is small, $\exp(x) \approx 1 + x$, which helps to explain why the growth rate and log transform analyses often produce similar results: compare Figures 2.7B and 2.8.

2.7 How to Evaluate Forecasting Accuracy

A key question in any forecasting endeavor is how to evaluate and measure performance – the *E* in PIVASE. Performance measures are of particular value when we come to select a forecasting procedure, because we may compare alternatives and choose the method with the best track record. Then, once the method is being used on a regular basis, we need similar measures to tell us whether the forecasts are maintaining their historical level of accuracy. If a particular set of forecasts is not performing adequately, managerial intervention is needed to get things back on track.

The generation of forecasts and the selection of a preferred method occupy a major portion of the book. Therefore, to discuss issues of evaluating accuracy without the need to develop forecasting methods explicitly at this stage, we consider an example taken from meteorology. Weather forecasts that appear in the media are not directed at a particular audience, and there is no reason to suppose that forecasts of temperature would have any inherent bias. However, we would expect that such forecasts (and this is typically true of all forecasts) would become less accurate as the forecast horizon increases (in this case, the number of days ahead).

We consider a set of local forecasts for daily high temperatures, extracted from the *Washington Post* for the period from June 2 to July 5, 2016. The forecasts are generated by *Accuweather*, a weather-forecasting organization. The forecasts appear for 1–5 days ahead, so the initial data could be summarized as shown in Table 2.7 (first and last few days only). However, this form of presentation is not useful for the evaluation of the forecasts, because, for example, the four-day forecast made on June 2 refers to conditions to be observed on June 6. To match forecasts to actual outcomes, we must slide the columns down, as shown in Table 2.8. We may now compare forecasts along the same row.

In a general sense, it is useful to refer to a forecast for period t as F_t when the time period when the forecast is made is not an issue. However, when timing matters, we need to report both the period being forecast and the number of periods ahead of time that the forecast was made. The notation[6] must reflect this information. Accordingly, we will use $F_t(h)$ to denote the forecast for period t, made h periods previously. Thus, $F_{14}(1)$ refers to the one-step-ahead forecast made for period 14 at period 13, $F_{14}(2)$ to the two-step-ahead forecast

6 The notation for forecasts is not standard. Some texts use F_{t+h} to denote forecasts h steps ahead for Y_{t+h}. Although this notation is simpler than ours, the notation F_{13+2} (not equal to F_{15}!) is potentially confusing and $F_{15}(2)$ is clearer.

made for period 14 at time 12, and so on. These values will be eventually compared with the observed value in period 14, Y_{14}. The general format is shown in Table 2.9; when there is no risk of confusion, we will use F_t to represent $F_t(1)$.

Table 2.7 Temperature Forecasts for Washington, DC, June 1–July 4, 2016

Forecast Origin	Forecast Lead Time (days ahead)				
	1	2	3	4	5
1-Jun	82	82	84	81	79
2-Jun	82	81	83	80	80
3-Jun	83	85	83	82	76
4-Jun	86	86	83	74	76
5-Jun	85	84	74	75	79
6-Jun	85	74	76	75	81
7-Jun	74	78	78	85	83
…	…	…	…	…	…
30-Jun	86	84	85	85	89
1-Jul	83	82	79	84	89
2-Jul	76	78	87	92	93
3-Jul	74	89	95	93	92
4-Jul	88				

Data: DC_weather_2.xlsx
Source: Washington Post: Daily highs at Reagan National Airport

Table 2.8 Temperature Forecasts for Washington, DC, June 2–July 5, 2016

Date Corresponding to Forecast	Forecast Lead Time (days ahead)					Actual Temperature (°F)
	1	2	3	4	5	
2-Jun	82					82
3-Jun	82	82				81
4-Jun	83	81	84			82
5-Jun	86	85	83	81		85
6-Jun	85	86	83	80	79	87
7-Jun	85	84	83	82	80	87
8-Jun	74	74	74	74	76	74
…	…	…	…	…	…	…
1-Jul	86	88	88	88	88	88
2-Jul	83	84	87	87	88	80
3-Jul	76	82	85	85	86	74
4-Jul	74	78	79	85	83	73
5-Jul	88	89	87	84	89	90

Data: DC_weather_2.xlsx
Source: Washington Post: Daily highs at Reagan National Airport

Table 2.9 Structure of Forecasts for 1, 2, 3,… Periods Ahead

Period	Days Ahead that Forecasts are Made				Actual
	1	2	3	…	
t	$F_t(1)$	$F_t(2)$	$F_t(3)$	…	Y_t
$t+1$	$F_{t+1}(1)$	$F_{t+1}(2)$	$F_{t+1}(3)$	…	Y_{t+1}
$t+2$	$F_{t+2}(1)$	$F_{t+2}(2)$	$F_{t+2}(3)$	…	Y_{t+2}

2.7.1 Measures of Forecasting Accuracy for Time Series[7]

Now that we have a set of forecasts and actual values with which to compare them, how should the comparisons be made? In principle, the criteria used should reflect the decision maker's loss function (e.g., minimize cost, achieve a certain level of customer service), but such information may not be available or the forecasts might be used for multiple purposes. In these circumstances, we must develop general-purpose criteria that can be used in a variety of circumstances. A natural approach is to look at the differences between the observed values and the forecasts and to use their average as a performance measure.

ROLLING ORIGIN FORECASTS

Starting at *forecast origin* t, with $t + m$ observations available, generate forecasts successively (one-step-ahead) at time origins $t, t+1, t+2, \ldots, t+m-1$, making m such forecasts in all. This process is known as rolling origin forecasting; an example is provided in Section 3.3.4. The one-step-ahead forecast error at time $t+i$ may be denoted by $e_{t+i} = Y_{t+i} - F_{t+i}$.

These observed error terms may be used to define performance indicators: one possibility is the mean of the errors.

THE MEAN ERROR

The *mean error* (*ME*) is given by

$$ME = \sum_{i=1}^{m} (Y_{t+i} - F_{t+i})/m = \sum_{i=1}^{m} e_{t+i}/m. \tag{2.13}$$

The mean error is a useful way of detecting systematic bias in a forecast; that is, *ME* will be large and positive (negative) when the actual value is consistently greater (less) than the forecast. The forecasts are then said to be biased. When the variable of interest is strictly positive, such as the number of employees or sales revenues, a percentage measure is often more useful; that is, we express the error as a percentage of the actual value and then take an average.

THE MEAN PERCENTAGE ERROR

The *mean percentage error* (*MPE*) is

$$MPE = \frac{100}{m} \sum_{i=1}^{m} \frac{(Y_{t+i} - F_{t+i})}{Y_{t+i}} = \frac{100}{m} \sum_{i=1}^{m} \frac{e_{t+i}}{Y_{t+i}}. \tag{2.14}$$

Note that *ME* is a useful measure for the temperature data, but *MPE* is not because the temperature can fall below zero. More importantly, temperature does not have a natural

7 We discuss measures of forecasting accuracy for categorical predictions in Chapter 10.

origin, so the *MPE* would give different (and equally meaningless) results, depending on whether we used the Fahrenheit or Celsius scale.

■ Example 2.6: Calculation of ME and MPE *(Data: Electricity.xlsx)*

The calculations of *ME* and *MPE* are illustrated in Table 2.10, in which the data represent the monthly electricity consumption (in KWH, kilowatt hours) in a Washington, DC, household; the column of forecasts in Table 2.10A represents the consumption in the corresponding month in the previous year. Consumption is low in the winter and high in the summer because the home uses gas heating and electric air-conditioning.

As noted, the *ME* and *MPE* are useful measures of bias. From Table 2.10A, we see that the household generally reduced its consumption relative to the previous year, so the forecasts tended to be too high. In passing, we note that the year-over-year change is given by comparing the totals for the two years — 13,190, the previous year's total (and this year's forecast), and 11,270 (this year's actual). Such a comparison indicates a 14.6 percent drop. The 18.7 percent average given by the *MPE* reflects month-by-month forecasting performance, not the change in the totals.

A limitation of these measures is that they do not reflect variability. Positive and negative errors could virtually cancel each other out, yet substantial forecasting errors could remain. To see this effect, suppose we used the average monthly figure for the first year (that is, the forecasts in this example) to predict the usage in the second year. The average is 1099 KWH, a figure that seriously underestimates summer consumption and overestimates the rest of the year. Yet the *ME* would be unchanged. The *MPE* expands to –37.2, because the errors are larger in the months with low consumption. Such changes are largely meaningless. For example, a forecast value of 800 KWH per month is clearly not very useful, yet it reduces the *MPE* to –0.1 as we shall see later!

From this discussion, we evidently also need measures that account for the magnitude of the errors regardless of their signs. ■

Table 2.10 Analysis of Forecasting Accuracy for Electricity Consumption in a Washington, DC, Household

(A) One Year Ahead: Forecast = Actual Value for Same Month in Previous Year

Period	Actual	Forecast	Errors	Absolute Errors	Percentage Errors	Absolute Percentage Errors	Squared Errors
Jan.	790	820	–30	30	–3.8	3.8	900
Feb.	810	790	20	20	2.5	2.5	400
Mar.	680	720	–40	40	–5.9	5.9	1600
Apr.	500	640	–140	140	–28.0	28.0	19600
May	520	780	–260	260	–50.0	50.0	67600
Jun.	810	980	–170	170	–21.0	21.0	28900
Jul.	1120	1550	–430	430	–38.4	38.4	184900
Aug.	1840	1850	–10	10	–0.5	0.5	100
Sep.	1600	1880	–280	280	–17.5	17.5	78400
Oct.	1250	1600	–350	350	–28.0	28.0	122500
Nov.	740	890	–150	150	–20.3	20.3	22500
Dec.	610	690	–80	80	–13.1	13.1	6400
	Totals		**ME**	**MAE**	**MPE**	**MAPE**	**MSE**
	11270	13190	–160	163.3	–18.7	19.1	44483.3
						RMSE	210.9

(B) One Month Ahead: Forecast = Previous Month's Actual

Period	Actual	Forecast	Errors	Absolute Errors	Percentage Errors	Absolute Percentage Errors	Squared Errors
Jan.	790	690	100	100	12.7	12.7	10000
Feb.	810	790	20	20	2.5	2.5	400
Mar.	680	810	–130	130	–19.1	19.1	16900
Apr.	500	680	–180	180	–36.0	36.0	32400
May	520	500	20	20	3.8	3.8	400
Jun.	810	520	290	290	35.8	35.8	84100
Jul.	1120	810	310	310	27.7	27.7	96100
Aug.	1840	1120	720	720	39.1	39.1	518400
Sep.	1600	1840	–240	240	–15.0	15.0	57600
Oct.	1250	1600	–350	350	–28.0	28.0	122500
Nov.	740	1250	–510	510	–68.9	68.9	260100
Dec.	610	740	–130	130	–21.3	21.3	16900
	Totals		**ME**	**MAE**	**MPE**	**MAPE**	**MSE**
	11270	11350	–6.7	250.0	–5.6	25.8	101316.7
						RMSE	318.3

Data: Electricity.xlsx

2.7.2 Measures of Absolute Error

The simplest way to gauge the variability in forecasting performance is to examine the *absolute error (AE)*, defined as the value of the error regardless of its sign and expressed as

$$\left| e_i \right| = \left| Y_i - F_i \right|. \tag{2.15}$$

Thus, if we generate a forecast of $F = 100$, the absolute error is 20 whenever the actual value turns out to be either 80 or 120. As before, we may consider various averages, based upon the absolute errors. They can usefully be split into two categories. Absolute measures which are on the same scale as the measurements themselves and relative measures where the (absolute) error is measured relative to some baseline. Those in common use are summarized in the following box.

The Mean Absolute Error (*MAE*) and the Root Mean Square Error (*RMSE*) measure variability in absolute terms, whereas the Mean Absolute Percentage error (*MAPE*) does so in relative terms.

The Relative Mean Absolute Error (*RelMAE*) is the ratio of the *MAE* for the current set of one-step-ahead forecasts to the *MAE* for forecasts made with the random walk, which uses the most recent observation as the forecast for the next period. Similar comparisons for *h*-periods-ahead just involve the use of the two sets of forecasts *h* steps ahead, which we designate *RelMAE*(*h*). When *RelMAE*(*h*) is greater than one, we may conclude that the random-walk forecasts are superior for *h* periods ahead. When *RelMAE*(*h*) is less than 1, the method under consideration is superior to the random walk for *h* periods ahead.

Theil's U is a long-established criterion that is widely used in econometrics; its structure is similar to that of *RelMAE*, but using squares in place of absolute values. As for the *RelMAE*, *U(h)* may be calculated using forecasts made *h* steps ahead.

ONE-STEP AHEAD ERROR MEASURES STARTING AT FORECAST ORIGIN *t*

Mean absolute error:

$$MAE = \sum_{i=1}^{m} |Y_{t+i} - F_{t+i}|/m = \sum_{i=1}^{m} |e_{t+i}|/m. \tag{2.16}$$

Mean absolute percentage error:

$$MAPE = \frac{100}{m}\sum_{i=1}^{m}\frac{|Y_{t+i} - F_{t+i}|}{Y_{t+i}} = \frac{100}{m}\sum_{i=1}^{m}\frac{|e_{t+i}|}{Y_{t+i}}. \tag{2.17}$$

Mean square error (not recommended):

$$MSE = \sum_{i=1}^{m}(Y_{t+i} - F_{t+i})^2/m = \sum_{i=1}^{m} e_{t+i}^2/m. \tag{2.18}$$

Root mean square error:

$$RMSE = \sqrt{MSE}. \tag{2.19}$$

Relative absolute error[8]*:*

$$RelMAE = \frac{\sum_{i=1}^{m}|Y_{t+i} - F_{t+i}|}{\sum_{i=1}^{m}|Y_{t+i} - Y_{t+i-1}|}. \tag{2.20}$$

Theil's *U:*

$$U = \frac{\left(\sum_{i=1}^{m}[Y_{t+i} - F_{t+i}]^2\right)^{1/2}}{\left(\sum_{i=1}^{m}[Y_{t+i} - Y_{t+i-1}]^2\right)^{1/2}}. \tag{2.21}$$

The following comments are in order:

1. *MAPE* should be used only when $Y > 0$; *RelMAE* and U are not so restricted.
2. The *RMSE* is used because the *MSE* involves squared errors so that the *MSE* is measured in dollars squared. Taking the square root to obtain the *RMSE* restores the original units of dollars and makes interpretation more straightforward.
3. The *RMSE* gives greater weight to large (absolute) errors. It can be shown that $RMSE \geq MAE$ for any set of m forecasts; recall the discussion in Section 2.4.3.
4. The measure using absolute values always equals or exceeds the absolute value of the measure based on the errors, so $MAE \geq |ME|$ and $MAPE \geq |MPE|$. Values that are close in magnitude suggest a systematic bias in the forecasts.

8 Hyndman and Koehler (2006) introduced the Mean Absolute Scaled Error (or *MASE*). *MASE* is the ratio of the average out-of-sample absolute error to the average in-sample absolute error using the random walk to make forecasts. In our present notation, this expression may be written as:

$$MASE = \frac{\sum_{i=1}^{m}|Y_{t+i} - F_{t+i}|/m}{\sum_{i=2}^{t}|Y_j - Y_{j-1}|/(t-1)}.$$

MASE may also be used to compare methods by summing across multiple series; this property is particularly useful when only a one or a small number of out-of-sample forecasts is available.

5. *MPE*, *MAPE*, *RelMAE*, and *U* are scale free and so can be used to make comparisons across multiple series by calculating the mean, first over an individual series and then averaged over a number of related series. The other measures are scale dependent and cannot be used to make such comparisons without an additional scaling. Summarizing accuracy over a number of series is helpful for many managerial decisions such as deciding which forecasting method to apply in a retail chain.
6. All these measures may be defined for more than one period ahead simply by replacing the one-step-ahead forecasts by their *h*-step-ahead counterparts, i.e. for *h*-step ahead forecast error measures, in all the above formulae, F_{t+i} is replaced by $F_{t+i}(h)$.
7. For *RelMAE* and *U* we replace the forecasts in *both* the numerator and the denominator by their *h*-step-ahead counterparts.

The mean error measures are defined above. Because errors are often non-normal or non-symmetric, the median is often used instead of or in addition to the mean. This does not change the interpretation of the measures as "the average accuracy." The median is a better measure of accuracy for the relative error measures than the mean.

Example 2.7: Calculation of error measures

Recall the absolute error measures computed in Table 2.10 for the electricity forecasts. The individual terms are shown in the various columns, and *MAE*, *MAPE*, and *MSE* are then evaluated as the column averages. *RMSE* follows directly from equation (2.18) — that is, by taking the square root of *MSE*. Table 2.10B gives the results for the random-walk (one month ahead) forecasts, so *RelMAE* is given by the ratio of the forecast *MAE* to the *MAE* of the random-walk forecasts. Relative to the random-walk method, the forecasts based upon the same month in the previous year provide a 35 percent [= 100 × (250 – 163.3)/250] reduction in the mean absolute error.

Table 2.11 provides a comparison of the three sets of forecasts for electricity consumption:

- Last year's value for the same month, as given in Table 2.10A
- Monthly average for the previous year (= 1099 for all 12 months)
- Forecast set at 800 for all 12 months
- Random walk (one month ahead), as given in Table 2.10B.

Table 2.11 Comparison of Forecasts for Electricity Data

Forecasting Method	Error Measure						
Means	**ME**	**MPE**	**MAE**	**MAPE**	**RMSE**	**RelMAE**	**U**
Last year's values	–160	–18.7	163	19.1	211	0.65	0.66
Monthly average	–160	–37.2	396	51.5	440	1.58	1.38
All F = 800	139	0.1	299	28.8	433	1.20	1.36
Random walk	–6.7	–5.6	250	25.8	318	1.00	1.00
Medians	**MdE**	**MdPE**	**MdAE**	**MdAPE**	**RMdSE**	**MdRelMAE**	**MdU**
Last year's values	–145	–18.9	145	18.9	145	0.69	0.69
Monthly average	–299	–37.4	389	39.7	389	1.85	1.85
All F = 800	0	0	235	29.9	235	1.12	1.12
Random walk	–55	–6.3	210	24.5	210	1.00	1.00

Data: Electricity.xlsx

From the table, we see that the forecasts based upon last year's figures are clearly more accurate. Indeed, some local utilities use this forecasting method to estimate customers' bills when no meter recordings are available. The *ME* and *MPE* values indicate that the forecasts were somewhat biased. Did the household deliberately try to conserve energy? Perhaps, but the lower figures in the summer suggest that the second year had a cooler summer; this is not something that could be reliably forecast at the beginning of January.

We have also included the median measures. Where forecasting errors are seriously asymmetric the medians (of our standard error measures) are usually a more informative measure than the mean. Although the sample size is small, there are several large absolute errors. Thus, while most of the measures are similar the *RMSE* is considerably larger than the *RMdSE*. ■

■ Example 2.8: Comparison of weather-forecasting errors

We may now use the measures we have examined to assess the performance of the various forecasts presented in Table 2.8. The results are given in Table 2.12. Note that we have used only those days for which all five forecasts are available so that comparisons are based on the same set of days. The *MPE* and the *MAPE* are not reported: they are not sensible measures in this case because the temperature scale has no natural origin and the observations can be negative. As is to be expected, the *MAE* and *RMSE* generally increase as the forecast horizon is extended; forecasts become less accurate as we get farther away from the event.

Table 2.12 Summary of Forecast Errors for Weather Data (Values Computed Over June 6 to July 5, 2016)

Measure	Lead Time (days ahead): 30 observations				
	1	2	3	4	5
ME	0.53	0.67	0.90	1.27	1.43
MdE	0	1	0	1.5	1
MAE	2.07	3.07	3.97	4.53	4.63
MdAE	2	3	4	4	3
RMSE	2.79	4.02	5.03	5.53	5.92
RelMAE	0.46	0.46	0.66	1.01	0.97
U	0.44	0.50	0.67	0.92	1.03

Data: DC_weather_2.xlsx; calculations in *Temperature_work.xlsx.*

Theoretically, the *RMSE* for forecasts should increase as the lead time increases, although this characteristic may be violated in practice when a small number of observations is used to estimate the summary measures. Overall, the expert weather forecasts are clearly superior to those produced by a random walk for one to three days ahead. ■

2.8 Prediction Intervals

Thus far, our discussion has centered upon *point forecasts* — that is, future observations for which we report a single forecast value. For many purposes, managers seem to feel most comfortable with a single figure. However, such confidence in a single number is often misplaced.

Consider, for example, the generation of weekly sales forecasts. Our method might generate a forecast of 600 units for next week. The manager who plans for the sale of *exactly* 600 units and never considers the possibility of selling more or fewer units is naïve and will probably become an ex-manager fairly quickly! Why? Because the demand for most products is inherently variable! Some weeks will see sales below the forecast level, and some will see more. When sales fall short of the point forecast, the business will incur holding costs for unsold inventory or may have to destroy perishable stock. When sales would have exceeded the point forecasts, not only will the business lose sales, but disappointed customers may go elsewhere in the future. The best choice of inventory level will depend upon the relative costs of lost sales and excess inventory, which are described by the statistical distribution of possible sales — that is, the *predictive* (*prediction*) *distribution*. The selection of the best inventory level in this case is known as the *newsvendor problem*, because it was originally formulated in the context of selling newspapers. Our purpose here is not to dwell upon the details, which may be found in most management science texts, such as Winston and Albright (2016), but rather to emphasize the fundamental role that the predictive distribution plays in such cases.

If we know the (relative) magnitudes of the costs of lost sales and of excess inventory, we may define an overall cost function and then select the level of inventory to minimize cost. (See Exercise 2.16 for further details.)

Sometimes these costs are difficult to assess, and the manager will prefer to guarantee a certain level of service. For example, suppose that we wish to meet demand 95 percent of the time. We then need to add a *safety stock* to the point forecast to ensure that the probability of a stock-out is no more than 5 percent. There are two principal approaches to this issue, which we discuss in turn:

1. Assume that the predictive distribution for demand follows the normal law (although such an assumption is at best an approximation and needs to be checked.
2. Use an empirical error distribution based upon the errors already observed.

2.8.1 Using the Normal Distribution

If we assume that the standard deviation (SD) of the distribution is known and the distribution is normal, we may use the upper 95 percent point of the standard normal distribution.[9] (This value is 1.645, from Table A.1 in on-line Appendix A2; alternatively, we may use the Excel function NORM.INV.) So the appropriate stock level is

$$\text{Mean} + 1.645 \times (\text{SD}). \tag{2.22}$$

The mean in this case is the point forecast. Thus, if the point forecast is 600, with an associated SD of 25, the manager would stock $600 + 1.645 \times 25$, or 641 units to achieve the desired level of customer service.

Expression (2.22) is an example of a *one-sided prediction interval*: The probability is 0.95 that demand will be equal to or less than 641, under the assumption that our forecasting method is appropriate for the sales of that particular product. Typically, the SD is unknown and must be estimated from the sample that was used to generate the point forecast. In other words, we use the sample *RMSE* to estimate the SD.

9 The normal distribution is by far the most widely used distribution in the construction of prediction intervals. Hence, it is critical to check that the forecast errors are approximately normally distributed. (See on-line Appendix A6 for details on testing for normality.)

In many forecasting applications, employing two-sided prediction intervals is more common. Putting these ingredients together, we define the two-sided 100(1 – α) percent prediction interval as

$$\text{Forecast} \pm z_{\alpha/2} \times (RMSE). \tag{2.23}$$

Here, $z_{\alpha/2}$ denotes the upper 100(1 – α/2) percentage point of the normal distribution. At this point, although we are using the sample value of *RMSE* to estimate the SD, we are not making any allowance for this fact. In Section 5.2, we define prediction intervals more precisely.

2.8.2 Empirical Prediction Intervals

We first form the empirical distribution, derived from the observed errors. In Table 2.13, we consider the $n = 30$ observed errors for the weather forecast data and proceed as follows:

1. Rank the values from smallest to largest.
2. Assign the percentage value $100(i - 0.5)/n$ to the ith smallest observation; in this case, the percentage values are 100/60=1.67, 300/60=5.0, and so on.
3. Form the two-sided prediction interval by selecting an appropriate pair of percentage points. In the example, 80 percent intervals would be a natural choice using the average of the third and fourth smallest (largest) observed errors.

The empirical prediction distribution is then defined as the error distribution shifted to have the mean equal to the point forecast.

■ Example 2.9: Evaluation of prediction intervals

Table 2.13 shows the ordered error terms based upon the forecasts made one to five days ahead. Then, each of the normal ($z = 1.282$) and empirical approaches is used to create the 80 percent prediction intervals. The normal intervals use the means and standard deviations for the 30 errors for each set of forecasts. The empirical intervals use 10th percentile (average of the third and fourth observations) and 90th percentile (average of 27th and 28th observations) to provide the 80 percent coverage. In this example, the two sets of intervals are very similar, but the differences are more pronounced with higher probabilities of coverage: see Exercise 2.13. By construction, the normal intervals are symmetric about the mean whereas the empirical intervals need not be. ■

2.8.3 Prediction Intervals: Summary

The principal reason for constructing prediction intervals is to provide an indication of the reliability of the point forecasts. The limits so derived are sometimes expressed as *optimistic* and *pessimistic* forecasts. Such nomenclature is useful as a way of presenting the concept to others, but a precise formulation of the limits should be used, rather than a vague assessment of extreme outcomes. Finally, note that prediction intervals may be used for retrospective analysis as here, but their primary purpose is to provide assessments of uncertainty for future events.

A detailed discussion of prediction intervals must await the formal development of forecasting models in later chapters. The reader who wishes to preview these discussions should consult Sections 5.2 and 6.7.

Table 2.13 Comparison of 80 Percent Prediction Intervals for Weather Data

Percentile	Errors				
	1-Day	2-Day	3-Day	4-Day	5-Day
1.67	–3	–8	–11	–12	–12
5.00	–3	–5	–7	–11	–10
8.33	–3	–5	–6	–7	–8
11.67	–3	–4	–6	–4	–4
15.00	–2	–4	–4	–4	–2
…	…	…	…	…	…
85.00	2	4	6	7	8
88.33	3	4	7	7	8
91.67	5	7	7	7	9
95.00	6	9	8	8	10
98.33	9	10	11	9	13
Mean: N = 30 observations	0.53	0.67	0.90	1.27	1.43
SD	2.79	4.03	5.03	5.48	5.85
80 Percent Intervals					
Empirical					
Lower	–3.00	–4.50	–6.00	–5.50	–6.00
Upper	4.00	5.50	7.00	7.00	8.50
Normal					
Lower	–3.04	–4.50	–5.55	–5.75	–6.06
Upper	4.11	5.83	7.35	8.29	8.93

2.9 Basic Principles of Data Analysis

This book, like most other texts on business forecasting, tends to devote most of its space to discussions of forecasting methods and underlying statistical models. However, if the groundwork is not properly laid, the best methods in the world cannot save the forecaster from the effects of poor data selection and inadequate preparation. The volume[10] edited by Scott Armstrong (2001) is particularly valuable in suggesting key principles that underlie good forecasting practice. We make extensive use of this source, among others, in formulating our own sets of principles at the end of each chapter. We number these principles in the order discussed within each chapter to facilitate cross-referencing.

[2.1] Ensure that the data match the forecasting situation.
Once the underlying purpose of the forecasting exercise has been specified, the ideal data set can be identified. However, the ideal may not be available for many reasons. For example, macroeconomic data are published with a lag that may be of several months' duration and, even then, may be published only as a preliminary estimate. The forecaster needs to examine the available data with respect to the end use to which the forecasts will be put and to make sure that a match exists.

10 For the latest developments see *www.forecastingprinciples.com*.

[2.2] Clean the data.
Data may be omitted, wrongly recorded, or affected by changing definitions. Adjustments should be made where necessary, but a record of such changes should be kept and made available to users of the forecasts. Data cleaning can be very time consuming, although the plots and numerical summaries described in this chapter will go a long way toward identifying data errors. Failure to clean the data can lead to the familiar situation of "Garbage in, garbage out."

[2.3] Use transformations as required by the nature of the data.
We examined differences, growth rates, and log transforms in Section 2.6. The forecaster needs to consider whether the original measurements provide the most appropriate framework for generating forecasts in the problem context or whether some form of transformation is desirable. The basic pattern of no growth (use original data), linear growth (use differences), or relative growth (use growth rates or log differences) will often provide adequate guidance.

[2.4] Use graphical representations of the data. Highlight key events.
As we have seen in Sections 2.2 and 2.3, plotting the data can provide a variety of insights and may also suggest suitable transformations or adjustments. Graphical analysis should always be the first step in developing forecasting procedures, even if applied to only a small sample from a larger set of series.

[2.5] Adjust for unsystematic past events (e.g., outliers).
Data may be affected by the weather, political upheavals, supply shortages, or other events. Such factors need to be taken into account when clear reasons can be identified for the unusual observations. The forecaster should resist the temptation to give the data a "face-lift" by overadjusting for every minor event.

[2.6] Adjust for systematic events (e.g., seasonal effects).
Systematic events such as weekends, public holidays, and seasonal patterns can affect the observed process and must be taken into account. We discuss these adjustments in Chapter 9.

[2.7] Use error measures that adjust for scale in the data when comparing across series.
When comparing forecasts for a single series, scale-dependent measures such as the *MAE* or *RMSE* are useful. However, when making a comparison across different series, use scale-free measures, such as the *MAPE* (if appropriate), or relative error measures, such as the *RelMAE* or *MASE*.

[2.8] Use multiple measures of performance based upon the observed forecast errors.
If forecasters are able to use different measures to compare performance, they can better assess performance relative to their particular needs. Multiple measures allow users to focus on those attributes of a forecasting procedure which they deem most relevant and also to check on the robustness of their conclusions. For example, one user may be interested in relative reduction in accuracy compared to a more complex method. Another may wish to avoid large errors, in which case the *RMSE* becomes most relevant, because it depends upon squared errors. A third may avoid the *RMSE* purely because it gives such weight to large errors. Instead a median measure might be used.

Summary

In this chapter, we have described the basic tools of data analysis. In particular, we examined the following topics:

- Scatterplots and time series plots for preliminary analysis of the data (Sections 2.2 and 2.3).
- Basic summary statistics for individual variables (Section 2.4).
- Correlation as a measure of association for cross-sectional data (Section 2.5).
- Transformations of the data (Section 2.6).
- Measures of forecasting accuracy (Section 2.7).
- Prediction intervals as a measure of the uncertainty related to point forecasts (Section 2.8).

Finally, in Section 2.9, we briefly examined some of the underlying principles that should be kept in mind when starting out on a forecasting exercise.

Exercises

Time series data in books do not necessarily reflect the latest figures. When it is appropriate, readers are encouraged to go to the original sources quoted in the text and download the latest figures. Readers may then use the extended data sets to redo the exercises, and the two sets of results may be compared.

2.1 The average monthly temperatures for Boulder, Colorado, from January 1991 to December 2015 are given in *Boulder_2.xlsx* (*Source:* U.S. Department of Commerce, National Oceanic and Atmospheric Administration). Plot the time series, and create a seasonal plot for the last four years of the series. Comment on your results.

2.2 The monthly figures for US Retail Sales, from January 2001 to December 2015 are given in *US_retail_sales_2.xlsx*. Plot the time series, and create a seasonal plot for the last four years of the series. Comment on your results.

2.3 The following table contains data on railroad passenger injuries in the United States (*Rail_safety.xlsx*) from 1991 to 2007. "Injuries" represents the number of persons injured in the given year, "train-miles" denotes the millions of miles traveled by trains, and the final column is the ratio that defines the number of injuries per 100 million miles traveled.

a. Create a scatterplot for injuries against train-miles.
b. Plot each of the three time series.
c. Does the level of injuries appear to be changing over time? If so, in what way?

Year	Injuries	Train-Miles	Injuries per Train-Mile
1990	473	72	657
1991	382	74	516
1992	411	74	555
.....	....	....	....
2005	935	90	1,040
2006	761	92	828
2007	938	95	990

Source: U.S. Department of Transportation, Federal Railroad Administration

2.4 An investor has a portfolio consisting of holdings in nine stocks (*Returns.xlsx*). The end-of-year returns over the previous year are, in increasing order,

$$-5.0, -3.7, 0.9, 4.8, 6.2, 8.9, 11.2, 18.6, 25.4.$$

a. Compute the summary statistics (mean, median, *MAD*, and *S*).

b. Just before the close of business in the last trading session of the year, the company that had reported the 5.0 percent drop declares bankruptcy, so the return becomes −100 percent. Recompute the results, and comment on your findings.

c. Are simple summary statistics relevant to this investor? How would you modify the calculations, if at all?

2.5 For the temperature data (*Boulder_2.xlsx*) in Exercise 2.1, compute the summary statistics (mean, median, *MAD*, and *S*) overall and for each month. Comment on your results. Does it make sense to compute summary statistics across all values, rather than month by month? Explain why or why not.

2.6 Compute the summary statistics (mean, median, *MAD*, and *S*) for each of the variables listed in Exercise 2.3 (*Rail_safety.xlsx*). Are these numbers a sensible summary of safety conditions? Explain why or why not.

2.7 Calculate the correlations among the three variables listed in Exercise 2.3. Also compute the correlation of each variable with time. Interpret the results.

2.8 Calculate the correlation between the monthly values of electricity consumption (*Electricity.xlsx*) for year 1 (listed as forecasts in the table) and year 2 (actual) in Table 2.10A. Interpret the result.

2.9 Compute the correlations among the eight variables listed in Table 2.3 using the data provided in *German_forecasts.xlsx*. Identify those coefficients that are statistically significant (different from zero) and discuss why such associations might exist.

2.10 The quarterly revenues and percentage growth figures for Netflix are given in *Netflix_2.xlsx* and illustrated in the table that follows. Produce a time series plot for each of these variables.

a. Are the mean and median useful in this case? Explain why or why not.

b. Calculate the growth rate for each quarter relative to the same quarter in the previous year; that is, for quarter 1 of 2001, we have 100(17.06 – 5.17)/5.17 = 230. After allowing for the start-up phase of the company, do sales show signs of leveling off?

Year	Quarter	Quarterly Revenues	Growth: Absolute	Growth: Percent
2000	1	5.17	•	•
2000	2	7.15	1.97	38.1
2000	3	10.18	3.04	42.5
2000	4	13.39	3.21	31.5
2001	1	17.06	3.67	27.4
....	....	....	....	....
2015	1	1573.13	88.40	6.0
2015	2	1644.69	71.56	4.5
2015	3	1738.36	93.67	5.7
2015	4	1823.33	84.97	4.9

Data: Netflix_2.xlsx; Source: Netflix Annual Reports

2.11 Use the data in *Boulder_2.xlsx* to generate forecasts 12-months ahead; that is, the forecast corresponds to the value for the same month in the previous year, for 1992–2015. Compute the *ME*, *MAE*, *RMSE*, and *RelMAE* for these forecasts as well as the corresponding medians. Repeat the analysis, this time using the monthly averages calculated in Exercise 2.5.

a. Which set of forecasts appears to work better?

b. Is the comparison fair? (*Hint:* What do you know and when do you know it?)

c. Do you see any consistent differences between the mean and median error measures?

d. What conclusions would you draw about the choice between the two methods?

2.12 The average temperatures (in °F) on the day (January 20) of the president's inauguration in Washington, DC (*Inauguration.xlsx*) are shown in the following table:

Year	1937	1941	1945	1949	1953	1957	1961	1965	1969
Temperature	33	29	35	38	49	44	22	38	35
Year	1973	1977	1981	1985	1989	1993	1997	2001	2005
Temperature	42	28	55	7	51	40	34	35	35

Source: www.inaugural.senate.gov/swearing-in/weather

a. Summarize the data numerically and graphically.

b. Create a 95 percent prediction interval for the inaugurations of President Obama in 2009 and in 2013

c. The actual values were 28° and 40°. Does that come as a surprise, given the width of the prediction interval?

2.13 Compute the 90 percent and 95 percent prediction intervals for the weather data (*DC_weather_2.xlsx*), using Table 2.13. How do the normal and empirical intervals compare at these higher levels of coverage?

2.14 Compute 95 percent prediction intervals for the 12-month forecasts for temperature (*Boulder_2.xlsx*) generated in Exercise 2.11, using (a) the normal distribution with the estimated *RMSE* and (b) the empirical distribution with all the observed errors. Find the percentage of the observations that lies outside the limits in each case. Are these figures close to 95 percent?

2.15 Use the data in Table 2.10A (*Electricity.xlsx*) to generate forecasts for electricity consumption for the household in year 3, based on the end of year 2 as the forecast origin. Use the normal distribution to generate 90 percent prediction intervals for these forecasts.

2.16 Following the discussion in Section 2.8 on inventory management, construct a cost function, assuming that the cost of lost sales is C times that of the cost of holding unsold stock. Past records show that the prediction distribution for future sales is

$$P\,(\text{sales} = x) = \frac{1}{21},\ x = 90, 91, \ldots, 109, 110$$

$$P\,(\text{sales} = x) = 0,\ \textit{otherwise}.$$

Find the optimal inventory level, in the sense of minimizing overall cost, when $C = 1$ and when $C = 3$. What should the safety stock be to guarantee a service level of 90 percent?

Minicases

Minicase 2.1 Baseball Salaries

As part of a project with the Graphics Section of the American Statistical Association, Dr. Lorraine Denby compiled data on baseball salaries. We focus on the data that relate only to pitchers and their salaries for the 1986 season. The data are included in the file *Baseball.xlsx*, and we wish to acknowledge StatLib of the Department of Statistics at Carnegie Mellon University for access to these data.

1. Summarize the data on salaries, using the measures discussed in Section 2.4.
2. The data file also includes information on the number of years a player has spent in the major leagues, his career earned run average (ERA), the number of innings he has pitched, and his career wins and losses. Generate scatterplots of these variables with salary, and examine their correlations.
3. Older players often accept short-term lower paid contracts toward the end of their playing careers. To allow for this feature of the data, eliminate players with 12 or more years of experience from the data set and rerun the analysis.

Summarize your conclusions.

MiniCase 2.2 Whither Walmart?

As Walmart has grown, its stock has proved to be a solid investment in both good times and bad. To determine whether a future investment in the stock is worthwhile, we need to consider the plans the company has for future growth. The annual reports provide a considerable amount of information (see *http://walmartstores.com/investors/*).

One aspect of Walmart's future strategy is its investment in different types of retail outlets, known as Walmart stores, Superstores, and Sam's Clubs. As the name suggests, the Superstore may be thought of as an upgrade of the Walmart store, being generally larger in size and carrying a wider range of merchandise. The Sam's Clubs are more oriented toward bulk purchasing. The spreadsheet *Walmart_2.xlsx* provides annual data on the numbers of each type of store in the United States on March 31, the end of the fiscal year, for the period 1995–2015.

Another feature of interest to the potential investor is the growth in sales over time. The spreadsheet also provides quarterly sales figures for the period from the first quarter of 2003 through the fourth quarter of 2015. The layout of the spreadsheet *Walmart_2.xlsx* is illustrated below:

Year	Walmart Stores	Super Stores	Sam's Club	Year	Quarter	Sales ($ billion)
1995	2176	154	453	2003	1	56.7
1996	2218	255	470	2003	2	62.6
1997	1960	344	436	2003	3	62.5
…	…	…	…	…	…	…
2013	508	3288	632	2015	2	119.3
2014	470	3407	647	2015	3	116.6
2015	442	3465	655	2015	4	128.7

Data: Walmart_2.xlsx; Source: Walmart annual reports, 1995–2015.

1. Summarize the changes in types of store over the period. What does your summary say about Walmart's plans for the future?
2. Compute the growth in sales over time. Is there any evidence that the rate of growth is slowing or increasing?

MiniCase 2.3 Economic Recessions

The following data summarize the length of each recession in the period 1929–2015, as determined by the National Bureau of Economic Research (NBER):

Onset	Duration	End	Gap
August 1929	43	February 1933	50
May 1937	13	May 1938	80
February 1945	8	September 1945	37
November 1948	11	September 1949	45
July 1953	10	April 1954	39
August 1957	8	March 1958	24
April 1960	10	January 1961	106
December 1969	11	October 1970	36
November 1973	16	February 1975	58
January 1980	6	June 1980	12
July 1981	16	October 1982	92
July 1990	8	February 1991	120
March 2001	8	October 2001	73
December 2007	19	June 2009	

Source: National Bureau of Economic Research
Data: Recessions.xlsx

Note that the NBER defines the end of a recession as the time at which the national economy shows an upturn. It does not mean that the economy has recovered to its previous level of activity. First update the data if needed to include any recession that has commenced since June 2017.

1. Calculate the average length of a recession, and provide a 95 percent confidence interval for this quantity. Interpret the result.
2. Calculate the average time between recessions, and provide a 95 percent confidence interval for this quantity. Interpret the result.
3. Is there any correlation between the length of a recession and the period of growth immediately preceding it, referred to as "Gap" in the table?
4. Is there any correlation between the length of a recession and the period of growth immediately following a recession?
5. Comment on your findings.

References

Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D. and Cochran, J. J. (2014), *Statistics for Business and Economics*, 12th ed. Mason, OH: Cengage Learning.

Armstrong, J. S. (2001). *Principles of Forecasting: A Handbook for Researchers and Practitioners.* Boston: Kluwer.

Hyndman, R. J., and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22, 679–688.

Müller-Dröge, H-C., Sinclair, T. M. and Stekler, H. O. (2016). Evaluating forecasts of a vector of variables: A German forecasting competition. *Journal of Forecasting*, 35, 495–505.

Winston, W., and Albright, S. C. (2015). *Practical Management Science*, 5th ed. Mason, OH: Cengage Learning.

CHAPTER 3

Forecasting Non-Seasonal Series

Table of Contents

What's past is prologue.

— Shakespeare, *The Tempest*

Introduction

Forecasts may be based upon subjective judgmental assessments, numerical procedures, or some combination of the two. Judgmental procedures are valuable when we have no track record on which to base our forecasts or when circumstances appear to have undergone fundamental changes. However, such forecasts are often time consuming to generate and may be subject to a variety of conscious or unconscious biases. Often, we find that even simple analyses of available data can perform as well as judgmental procedures and that they may be much quicker and less expensive to produce. Careful subjective adjustments to quantitative forecasts may ultimately be the best combination, but we first need to develop an effective arsenal of quantitative methods. Accordingly, we now focus on quantitative methods and defer consideration of judgmental methods to Chapter 11.

We begin in Section 3.1 by drawing a distinction between *methods* and *models,* a distinction that is often ignored but that has important implications for how we approach a forecasting task. Section 3.2 then provides a general overview of extrapolation methods before we move on in Section 3.3 to the use of different weighted means as forecasts. Our aim is to balance the following potentially conflicting directives:

- Use all the data.
- Pay more attention to the recent past.

Or, to restate, "All data are important, but recent data are even more important." These ideas lead directly to the use of time-dependent averages — notably, simple moving averages and exponentially weighted moving averages (or exponential smoothing, as it is usually called).

When there are clear trends in the data, a simple averaging procedure cannot capture the trend and therefore will not work. As a consequence, we must extend our methods to incorporate such systematic movements, the subject of Section 3.4. Sections 3.5 and 3.6 consider damped smoothing methods and various other approaches to forecasting trends. Section 3.7 considers transformations of series to improve the underlying assumptions. Section 3.8 provides prediction intervals for one-step-ahead forecasts, relying implicitly upon models we develop fully in Chapter 5. Section 3.9 briefly discusses the issue of method selection, a topic that is explored more fully in Section 5.3. Finally, in Section 3.10 we explore some of the principles that underlie the use of simple extrapolation methods.

Software

Good software should provide for the automatic operation of a variety of exponential smoothing methods and ways to select from among them. We summarize some of the alternatives in Appendix B, located on the textbook companion site. At this stage, we merely note that the level of support for exponential smoothing methods varies considerably across different software providers. For example, Excel 2010 includes only simple exponential smoothing, and even then its application is far from automatic. However, exponential smoothing in Excel 2016 contains many more features. To achieve some consistency in the numerical results obtained from these methods, we provide an Excel macro, the Exponential Smoothing Macro (ESM), to carry out the estimation and forecasting procedures. The macro is available on the book's website, which also provides a User's Manual. We also provide scripts in R to carry out the smoothing procedures.

3.1 Method or Model?

The development of various quantitative approaches to forecasting occupies the next eight chapters, as we move steadily from heuristic forecasting methods to procedures that rely upon careful modeling of the process being studied. One of the pleasant discoveries we make along the way is that various heuristic forecasting methods often match up to specific statistical models of the data, even though this property was not known at the time the methods were first proposed. As in other areas of research, we often proceed by discovering what works and then try to figure out why. This chapter concentrates on what works, and we emphasize the distinction in our terminology in the text box below, even though the difference between the two terms (method and model) is sometimes ignored in discussions of forecasting.

METHOD OR MODEL

A *forecast function* is an equation for calculating the forecasts over the forecast horizon.

A *forecasting method* is a (numerical) procedure for generating a forecast. It involves the direct use of a forecast function. When such methods are not based upon an underlying statistical model, they are termed *heuristic.*

A *statistical (forecasting) model provides an approximate* description of the data-generating process. The corresponding forecasting function may then be derived from the statistical model. A statistical model is a necessary foundation for the construction of prediction intervals.

3.1.1 A Forecasting Model

We might formulate a simple trend model for a time series as

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t, \tag{3.1}$$

where Y_t denotes the time series being studied; β_0 and β_1 are the level and slope (or trend) *parameters,* respectively; and ε_t denotes a random error term corresponding to the part of the series that remains once the linear trend has been removed. If we make appropriate assumptions about the nature of the error term, we can estimate the unknown parameters β_0 and β_1. The resulting estimates are typically written as b_0 and b_1. Thus, the forecasting *model* gives rise to a forecast *function,* whose estimated version may be written as

$$F_t = b_0 + b_1 t, \tag{3.2}$$

where F_t denotes a forecast for time period t, b_0 is the (estimated) *intercept* that represents the value at time zero, and b_1 is the (estimated) *slope*, which represents the increase in forecast values from one period to the next.

Equation (3.2) is the forecast function, and the statistical model given by equation (3.1) enables the construction of prediction intervals. If we were to proceed directly to employ equation (3.2) for forecasting without such a model, we would indeed have a forecasting method, but one that suffered from two drawbacks:

1. It lacks a formal basis for choosing values for the parameters, although various ad hoc procedures could be developed.
2. There is no way to assess the uncertainty inherent in the forecasts.

Of course, if you choose a poor statistical model, you will get poor forecasts and poor assessments of uncertainty — maybe not always, but on average.

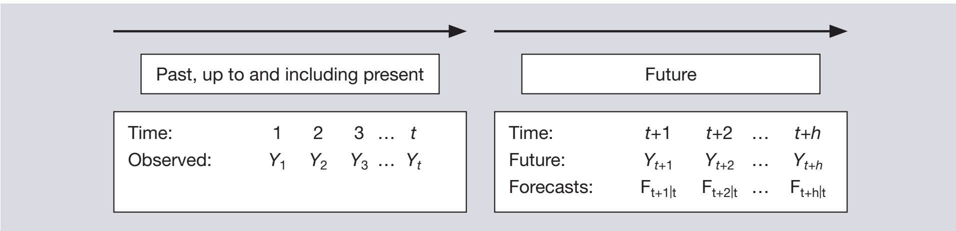
In this chapter, we focus primarily on methods. These methods will be underpinned by the statistical models we introduce in Chapter 5. On the basis of these models, we are able to consider prediction intervals, which we do in Section 3.8.

3.2 Extrapolative Methods

Extrapolative methods of forecasting focus on a single time series to identify past patterns in the historical data. These patterns are then extrapolated to map out the likely future path of the series. The overall structure is shown in Figure 3.1.

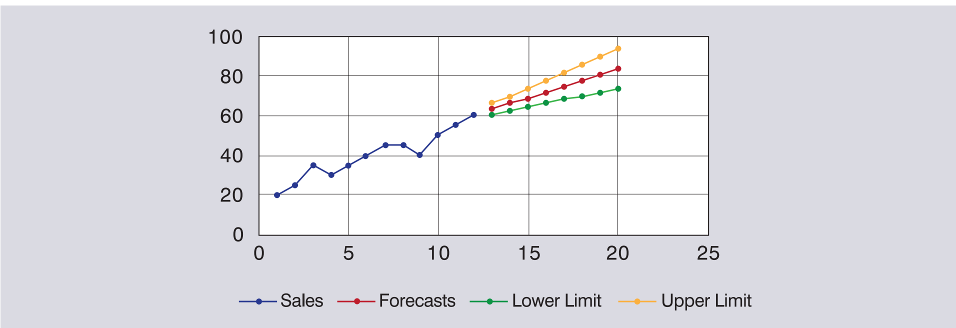
Figure 3.1 is to be interpreted as follows: We denote the particular series of interest (e.g., weekly sales) by Y. We have recorded the observations $Y_1, Y_2, \ldots, Y_t$ over t *weeks,* which represent all the data currently available. Our interest lies in forecasting sales over the next h weeks, known as the *forecasting horizon;* that is, we are interested in providing forecasts for future sales, denoted by $Y_{t+1}, Y_{t+2}, \ldots, Y_{t+h}$.

Figure 3.1 General Framework for Forecasting with a Single Series



Although we use the same notation to describe past, present, and future, there is a key difference: The past and present values are already observed, whereas the future Ys represent random variables; that is, we cannot write down their values, but we can describe them in terms of a set of possible values and the associated probabilities, as discussed in Appendix A to Chapter 1. This concept is illustrated in Figure 3.2, which shows a time series observed for periods 1-12, but to be forecast for periods 13-20, typically with increasing uncertainty in the forecast as the horizon increases.

Figure 3.2 Plot of Actual Sales for Periods 1–12 and a Range of Possible Values for Periods 13–20



The (point) forecasts for future sales, shown in the diagram, are all made at time t (= 12 here), known as the *forecast origin,* so the first forecast (for period 13) will be made one step ahead, the second (for period 14) two steps ahead, and so on. When the results for period 13 are known, that period becomes the new forecast origin and we can compute a new forecast for period 14, which will then be only one step ahead. We need to distinguish these different forecasts, as illustrated in Table 3.1. Suppose that the initial forecast origin is week 12 and we wish to make forecasts in week 12 for weeks 13, 14, and 15. Then in week 13 we would make a new set of forecasts for weeks 14 and 15; finally, in week 14 we would make a forecast for week 15. A quick look at the table indicates that we are considering six forecasts, all of which are based upon different information or relate to different time periods. As introduced in Chapter 2 (Section 2.7.1 and Table 2.7), the notation F_{14} (2) refers to the forecast for Y_{14} made two weeks earlier in week 12. The subscript always indicates which time period is being forecast, and the term in parentheses records how far ahead the forecast is made (the forecast horizon). When no ambiguity arises, we will use F_{t+1} to represent the one-step-ahead forecast $F_{t+1}(1)$ so that $F_{13} = F_{13}(1)$ and so on.

This notation may seem a bit elaborate, but it is important to know both the forecast origin and for how many periods ahead the forecast is being made. That is, the term "forecast for period 15" is ambiguous until we know when the forecast was made, and it would be impossible to evaluate the forecast's accuracy without that information.

Table 3.1 Notation for Forecasts Made at Different Forecast Origins and for Varying Forecast Horizons

Forecast Origin	12	13	14
Forecast for period 13	$F_{13}(1)$		
Forecast for period 14	$F_{14}(2)$	$F_{14}(1)$	
Forecast for period 15	$F_{15}(3)$	$F_{15}(2)$	$F_{15}(1)$

3.2.1 Extrapolation of the Mean Value

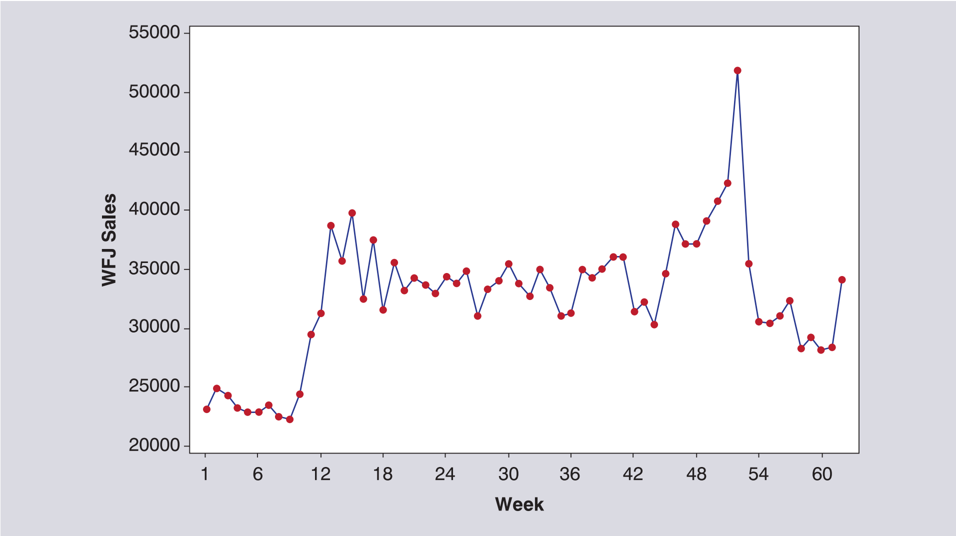
Figure 3.3 (previously shown as Figure 2.1) shows the weekly sales figures for 62 weeks of a product line for a major U.S. manufacturer, which we continue to refer to as WFJ Sales. The series starts at the beginning of the calendar year, picks up after 12 weeks or so, and then stabilizes until a surge in the last few weeks of the year, before dropping back at the beginning of the next year (but at a higher level than a year earlier).

Suppose we traveled back in time and we were back at midyear (week 26) and wished to forecast sales for the next few weeks. A straightforward approach would be to take the average of the 26 weeks to date, which we write as

$$\bar{Y}(26) = \sum_{i=1}^{26} Y_i / 26 = 30102.$$

The bar denotes the operation of taking the mean over the 26 observations. Inspection of Figure 3.4, which shows the first 26 weeks of the series and the mean level, suggests that such a value would be too low a forecast. Somehow, we need to give less weight to the first part of the series and focus on the more recent values.

Figure 3.3 Time Series Plot of Weekly Sales for WFJ



Data: WFJ_sales.xlsx

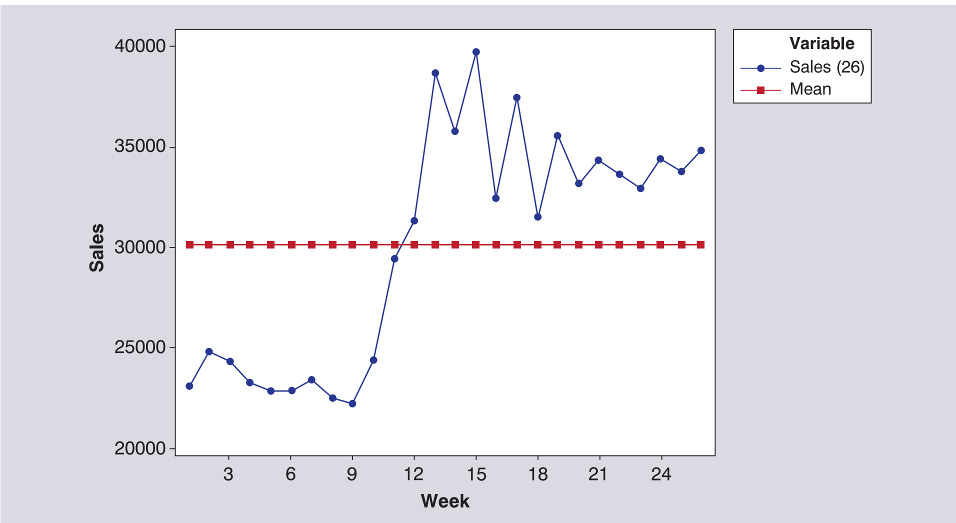
LOCALLY CONSTANT FORECASTS

A series is locally constant if the mean level changes gradually over time but there is no reason to expect a systematic increase or decrease in the future. The forecast function is

$$F_{t+h}(h) = \text{constant};$$

that is, the plot showing future forecasts is a horizontal line. When a new observation becomes available, the constant is updated.

Figure 3.4 Time Series Plot for Weekly Sales for WFJ for the First 26 Weeks, Compared with the Average



3.2.2 Use of Moving Averages

One way to proceed is to use an average of the last few values in the series. In general, we may use the last K terms of the series and update each time to include only the latest K values. We refer to the result as a moving average, which contains K terms, the most recent of which is observation t (the most recent available observation):

MOVING AVERAGE

The *moving average* of order K evaluated at time t is denoted by $MA(t\,|K)$:

$$MA(t\,|K) = \frac{Y_t + Y_{t-1} + \cdots + Y_{t-K+1}}{K}$$

We then forecast the time series at time $(t+1)$, using the formula

$$F_{t+1} = MA(t\,|K).$$

Suppose now that we decide to use a three-week moving average, or $K = 3$. The first such average would be $MA(3|3) = (23056 + 24817 + 24300)/3 = 24058$ as shown in Table 3.2. When a new observation becomes available, the new moving average is $MA(4|3) = (24817 + 24300 + 23242)/3 = 24120$, and so on. Each time, we drop the oldest observation and enter the new one; a slightly quicker way of doing the calculation is to write

$$New\ average = MA(t+1|K) = Old\ average + \frac{(New\ value - Oldest\ value)}{K} \tag{3.3}$$

$$= MA(t\,|K) + \frac{Y_{t+1} - Y_{t+1-K}}{K}$$

Table 3.2 Calculation of Moving Averages for WFJ Sales

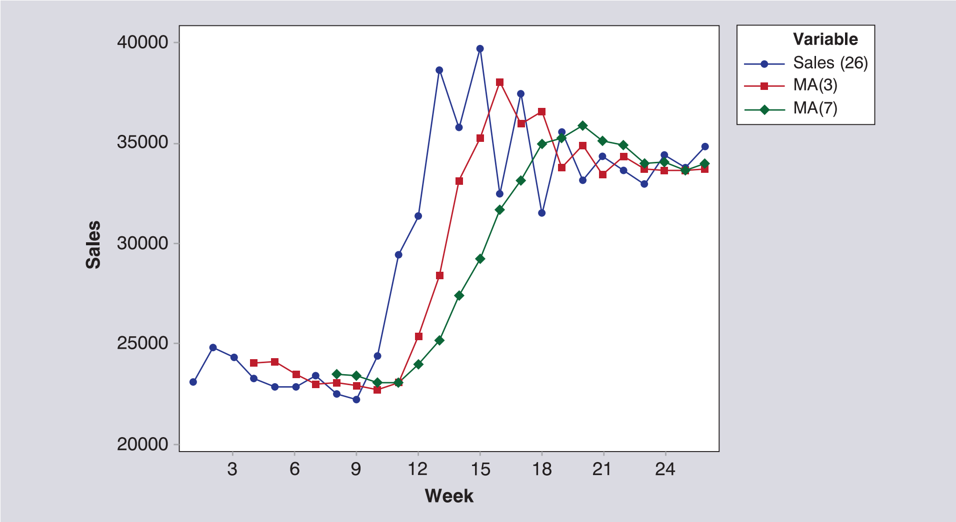
Week	WFJ Sales	MA (3)	MA (7)
1	23056		
2	24817		
3	24300		
4	23242	24058	
5	22862	24120	
6	22863	23468	
7	23391	22989	
8	22469	23039	23504
9	22241	22908	23421
10	24367	22701	23053
...	...	...	...
...	...	...	...
25	33777	33652	33646
26	34849	33701	33975

Data: WFJ_sales_MA.xlsx

A variation of this updating mechanism will feature prominently in later developments. But now, how should we choose K? A simple comparison provides an insight. Table 3.2

illustrates the calculations for $K = 3$ and for $K = 7$, which we respectively refer to as MA(3) and MA(7), and the two averages and the original series are plotted in Figure 3.5. The first forecast that we can calculate for MA(3) is for week 4, which uses the first three observations. This forecast corresponds to the one-step-ahead forecast for week 4 made in week 3. The first forecast for MA(7) is for period 8.

Figure 3.5 WFJ Sales for First 26 Weeks, with Moving Averages of Lengths 3 and 7



From Figure 3.5, we see that the three-term moving average adapts more quickly to movements in the series, but the seven-term average produces a greater degree of smoothing. To decide which method is preferable, we need to evaluate forecast performance, as discussed in Chapter 2. However, before we do so, we introduce another form of adaptive average.

3.3 Simple Exponential Smoothing

The simple moving average, introduced in the previous section, suffers from two drawbacks. First, the averaging process seems rather capricious in that an observation is given full weight one period and none the next, when it reaches the K^{th}, or "oldest," position. Second, if we use a large number of terms, we have to keep all the past values until they are finally removed from the average. This second objection is now of minor importance in practice, although it used to be critical when computer storage was much more limited or when forecasts had to be updated by hand. Ultimately, any method should be judged by its performance in forecasting, although it would be nice to have a technique that adjusted more smoothly over time. Such a method was introduced by Robert G. Brown (often referred to as the father of exponential smoothing), whose 1959 and 1963 books on forecasting are justly recognized as classics.

When a new observation is recorded, the new sample mean of the available data may be expressed as

$$\textit{New mean} = \textit{Old mean} + \frac{(\textit{Difference between new observation and old mean})}{\textit{New sample size } (= \textit{old sample size} + 1)}.$$

We can see from this expression that we need only record the previous mean and the latest value, a useful feature when updating anything from sales reports to batting averages. However, as the series length increases, the mean becomes increasingly unresponsive to fluctuations in recent values because each observation has weight equal to 1/(sample size). The update of the simple MA, given in equation (3.3), avoided this problem and maintained a constant coefficient *(1/K),* but at the cost of dumping the oldest observation completely. The aim is to construct some kind of average that describes the recent behavior of the series, but in a way that adjusts smoothly over time. We refer to such an average as the *local (mean) level.* If we use too many observations, the estimated level may be out of date, whereas if we use too few, the estimate may be inaccurate and jump around a lot. So, instead of discussing the issue in terms of a moving average, we refer to the local level at time t as L_t and resolve these two conflicting elements by using an updating relationship of the form

$$L_{t+1} = L_t + \alpha(Y_{t+1} - L_t). \tag{3.4}$$

That is,

New local level = *Old local level* + α × (*Difference between new observation and old local level*).

EXPONENTIAL SMOOTHING

The basic equation for *exponential smoothing* is

$$L_{t+1} = L_t + \alpha(Y_{t+1} - L_t) = L_t + \alpha e_{t+1}$$

where $e_{t+1} = Y_{t+1} - F_t$ denotes the forecast error for period t+1.

The process involves comparing the latest observation with the previous weighted average and making a proportional adjustment, governed by the coefficient α, known as the *smoothing constant.* By convention, we constrain the coefficient to the range $0 < \alpha < 1$ so that only a part of the difference between the old level and the new observation is used in the updating. Inspection of equation (3.4) indicates that this form of average provides a constant weight *(α)* to the latest observation. Further, the updates require only the latest observation and the previous local level. The average in equation (3.4) is known as an *exponentially weighted moving average* (EWMA), and we now explain the origin of that name. If we start with the expression for period *(t* + 1) and substitute into it the comparable expression for time *t,* we obtain

$$\begin{aligned} L_{t+1} &= (1-\alpha)L_t + \alpha Y_{t+1} = (1-\alpha)[(1-\alpha)L_{t-1} + \alpha Y_t] + \alpha Y_{t+1} \\ &= (1-\alpha)^2 L_{t-1} + \alpha[(1-\alpha)Y_t + Y_{t+1}]. \end{aligned}$$

Continuing to substitute the earlier smoothed means, we eventually arrive back at the start of the series with the expression

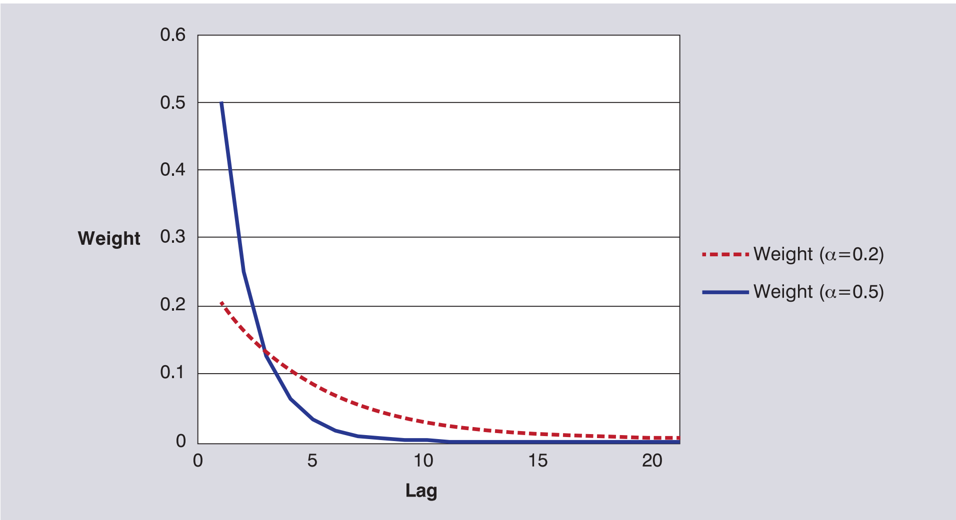
$$L_{t+1} = (1-\alpha)^{t+1}L_0 + \alpha[Y_{t+1} + (1-\alpha)Y_t + (1-\alpha)^2 Y_{t-1} + \cdots + (1-\alpha)^t Y_1]. \tag{3.5}$$

The right-hand side contains a weighted average of the observations, and the weights $\{\alpha, \alpha(1-\alpha), \alpha(1-\alpha)^2, \ldots\}$ decay steadily over time. If the weights are plotted against time and a smooth curve is drawn through the values that curve is exponential — hence the name "exponentially weighted moving average." The decay is slower for small values of α, so we can control the rate of decay by choosing α appropriately. Figure 3.6 illustrates the decay rates for $\alpha = 0.5$ and $\alpha = 0.2$. At $\alpha = 0.5$, over 99 percent of the weight falls on the first seven observations whereas the comparable figure for $\alpha = 0.2$ is only 79 percent. As α gets smaller, so does this percentage. Thus, just as we could choose K to control the rate of adjustment of the simple moving average, we may select α to achieve similar adjustments for the EWMA.

Caution: Equation (3.5) depends on a starting value L_0. When t and α are both small, the weight attached to the starting value may be high, as seen in the following table (in these circumstances, it is recommended that more sophisticated software be used to estimate the optimal value for the starting level):

α	t	Weight on Starting Value after t periods
0.50	10	0.001
	30	0.000
0.20	10	0.107
	30	0.001
0.05	10	0.599
	30	0.215

Figure 3.6 Weights on Different Lags in an Exponentially Weighted Moving Average



3.3.1 Forecasting with the EWMA, or Simple Exponential Smoothing

The next step is to convert these averages into forecasts. When we use the EWMA for forecasting, we refer to the method as *simple* (or *single) exponential smoothing* (SES).[1] The underlying logic of this process is that, although we believe that the process will fluctu-

1 The term is perhaps unfortunate, because "smoothing" is an overused qualifier in time series analysis. Nonetheless, "SES" is in common usage.

ate, we have no evidence to suggest that it is likely to go up, go down, or stay the same. In such circumstances, the average level for future observations is best forecast by our current estimate of the local level. Thus, forecasts made at time t for all future time periods will be the same; that is,

$$F_{t+1}(1) = F_{t+2}(2) = \cdots = F_{t+h}(h) = L_t. \tag{3.6}$$

At first sight, this equation may seem rather strange, but think of the case of stock prices. The latest price is assumed to capture all the relevant (public!) information about the value of the stock, so the current price is the best forecast for all future periods until new information comes along. We must recognize that, as the lead time increases, the forecasts will usually become less accurate.

Equation (3.4) may now be written in the terms of the forecasts as

$$F_{t+1}(1) = F_t(1) + \alpha[Y_t - F_t(1)]. \tag{3.7}$$

That is, the new one-step-ahead forecast is the previous forecast, partially adjusted by the amount the previous forecast was in error. Because this equation considers only the one-step-ahead forecasts, it may also be written as

$$F_{t+1} = F_t + \alpha(Y_t - F_t). \tag{3.8}$$

To set up the calculations that are implicit in the equation, we need to specify the value for α, as well as a starting value $F_1 = L_0$. We have already discussed the effects of different α values. Earlier literature recommended a choice in the range $0.1 < \alpha < 0.3$ to allow the EWMA to change relatively slowly, and values from that range often work well for series such as sales figures. However, relying on an arbitrary preset smoothing parameter is not advisable. Most computer programs now provide efficient estimates of the smoothing constant, based upon minimizing the mean squared error for the estimation sample (or, equivalently, the *RMSE):*

$$\text{MSE} = \frac{\sum_{i=1}^{t} (Y_i - F_i)^2}{t}. \tag{3.9}$$

MSE is computed with the use of the first t values, which constitute the *estimation,* or *fitting,* sample. Forecasts may then be generated for time periods $t + 1$, $t + 2$ … .

CHOOSING THE SMOOTHING PARAMETER

When forecasting with exponential smoothing, estimate the optimal smoothing parameters, rather than using preset values.

Example 3.1: Basic SES calculation

A short series of hypothetical sales data is given in Table 3.3. The first observation is used as the forecast for period 2, and the smoothing constant is set at $\alpha = 0.3$. Thus, from equation (3.7), we have, for $t = 2$,

$$F_3(1) = F_2(1) + \alpha[Y_2 - F_2(1)]$$
$$= 5.00 + 0.3 \times (6.00 - 5.00) = 5.30$$

The forecasts are computed successively in the same way. The forecast errors, their squares, their absolute values, and their absolute values as a percentage of the observed series are shown in succeeding columns. The last row of the table gives the mean error, the *MSE,* the *MAE,* and the *MAPE,* calculated over observations 2 through 12. The calculations may be checked by using the Excel spreadsheet *Table_3_3.xlsx,* available on the book's website. ■

Table 3.3 Illustration of Spreadsheet Calculations for SES Smoothing Constant: Alpha (α) = 0.3

Time	Sales	Forecast	Error	(Error)^2	Absolute Error	Absolute Percentage Error
1	5.00					
2	6.00	5.00	1.00	1.00	1.00	16.67
3	7.00	5.30	1.70	2.89	1.70	24.29
4	8.00	5.81	2.19	4.80	2.19	27.38
5	7.00	6.47	0.53	0.28	0.53	7.61
6	6.00	6.63	–0.63	0.39	0.63	10.45
7	5.00	6.44	–1.44	2.07	1.44	28.78
8	6.00	6.01	–0.01	0.00	0.01	0.12
9	7.00	6.01	0.99	0.99	0.99	14.21
10	8.00	6.30	1.70	2.88	1.70	21.21
11	7.00	6.81	0.19	0.04	0.19	2.68
12	6.00	6.87	–0.87	0.75	0.87	14.48
13		6.61				
		Means	**0.49**	**1.46**	**1.02**	**15.26**
		RMSE = 1.21				

DISCUSSION QUESTION: *If you had to make a subjective choice for the value of the smoothing constant, what value would you choose for (a) a product with long-term steady sales and (b) a stock price index?*

3.3.2 The Exponential Smoothing Macro (ESM)

The Excel macro Exponential Smoothing Macro.xlsm (or simply ESM for short), available on the book's website, provides a flexible tool for fitting SES and a variety of other exponential smoothing models. A comprehensive *Users' Manual* is also available on the website. ESM uses Solver to fit models (in this case, to select the best value for α by minimizing the *MSE.* The macro allows the user to partition the data series into estimation and hold-out samples and computes summary statistics for each subsample. For consistency, we use ESM to estimate the exponential smoothing models described in the book (except where otherwise noted), but the reader should be aware that different software programs use somewhat different fitting algorithms and results differ from one program to another. That said, the estimated values of α will usually be similar, unless we are dealing with a very short series or one that contains some unusual observations, which we will refer to as *outliers*.

Starting Values We must still resolve the choice of starting values. Two principal options are commonly used: (1) to use the first observation as F_1 or (2) to use an average of a number of initial observations; recommendations vary from the first 3 or 4 up to 6 or 12 or even the mean of the whole sample. Gardner (1985) provides a good review of the options but states, "There appears to be no empirical evidence favoring any particular method." The macro allows the choice of the first observation only or an average of the first several observations to estimate the starting value(s). When either the sample size or α is large, the choice of starting value is relatively unimportant and the different approaches yield similar results. When both the sample size and α are small, more sophisticated methods should be used.[2]

Caution: All major forecasting packages include fitting routines for SES, although the particular algorithms employed vary, especially in regard to the estimation of starting values, so the results may differ somewhat from those shown in this book. Some packages are very coy about the methods they employ.

We now explore the effects of changing α for forecasts of WFJ Sales. The macro was used in each case to generate the results.

Example 3.2: SES forecasts for WFJ Sales *(WFJ_sales.xlsx)*

We used the first 26 observations of the WFJ Sales series as the estimation sample. The one-step-ahead *MSE* (taken over observations 2 to 26) is minimized[3] when $\alpha = 0.728$. Table 3.4 shows the first ten one-step-ahead forecasts for WFJ sales, from week 26 as origin, using SES with $\alpha = 0.2$ and 0.5, as well as 0.728, for comparison purposes. Thus, the first out-of-sample forecast is available at week 27, and the *one-step-ahead* forecasts are listed for weeks 27-36. The starting value was taken as the first observation (the ESM default) in each case. The three sets of forecasts correspond to the following:

- SES(0.2): Simple exponential smoothing with $\alpha = 0.2$.
- SES(0.5): Simple exponential smoothing with $\alpha = 0.5$.
- SES(opt): Simple exponential smoothing with $\alpha = 0.728$.

Table 3.4 Actual and One-Step-Ahead Forecasts for WFJ Sales
(For weeks 27-36, starting from week 26 as origin and using ESM)

Week	WFJ Sales	SES (0.2)	SES (0.5)	SES (opt)
27	30986	33881	34346	34580
28	33321	33302	32666	31963
29	34003	33306	32993	32952
30	35417	33445	33498	33717
31	33822	33840	34458	34955
32	32723	33836	34140	34130
33	34925	33614	33432	33105
34	33460	33876	34178	34431
35	30999	33793	33819	33723
36	31286	33234	32409	31739

To illustrate the nature of the calculations, consider the one-step-ahead forecast for week 28 for $\alpha = 0.2$. From equation (3.7), we have

$$F_{28}(1) = 33881 + 0.2(30986 - 33881) = 33302$$

Similarly, the one-step-ahead for forecast for week 29 is $F_{29}(1) = 33306$.

2 In Section 5.1, we discuss fitting a complete model to estimate both the smoothing parameter and the starting value.

3 The observations over which the *MSE* is minimized also varies with the software employed.

Figure 3.7 SES Forecasts for WFJ Sales: The Effects of Different Smoothing Constants

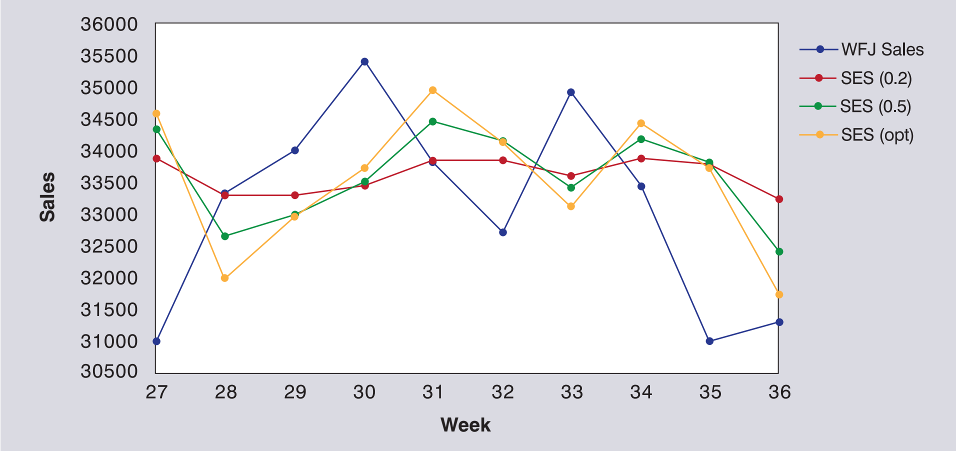


Figure 3.7 shows the one-step-ahead forecasts for weeks 27–36. The forecasts for SES(0.2) are smoother than those for SES(0.5) and SES(opt), which adapt more quickly to the latest observation. ■

3.3.3 The Use of Hold-Out Samples

Recall that in Example 3.2 we partitioned the original series into two parts. That partition provides a framework for an examination of the relative performance of alternative methods. For this purpose, we use the accuracy measures developed in Section 2.7.

When we split the series into two parts, we refer to the first part as the *estimation* sample, used to estimate the starting values and the smoothing parameters. This sample typically contains the first 75–80 percent of the observations, although the forecaster may choose to use a smaller percentage for longer series (a smaller proportion was used in our example to illustrate the results over a relatively stable part of the series). The parameters are commonly estimated by minimizing the mean squared error *(MSE),* although the mean absolute error *(MAE)* and mean absolute percentage error *(MAPE)* can also be used; these alternatives are somewhat more robust to extreme observations, but MSE estimation combined with outlier adjustment is the path more commonly chosen (c.f. Section 9.7).

The *hold-out* sample represents the last 20–25 percent of the observations and is used to check forecasting performance. No matter how many parameters are estimated with the estimation sample, each method under consideration can be evaluated with the use of the "new" observations contained in the hold-out sample. Thus, the hold-out sample provides a level playing field for such comparisons: Relying on the estimation sample advantages more heavily parameterized methods that can overfit the data.

IN-SAMPLE AND OUT-OF-SAMPLE (HOLD-OUT) ERROR MEASURES

Measures of performance based upon the estimation sample are referred to as *in-sample;* measures based upon the *hold-out sample* are referred to as *out-of-sample.*

■ Example 3.3: Comparison of one-step-ahead forecasts

To examine the forecasting performance of the various methods discussed to date, we carried out the experiment described next, using the series WFJ Sales. Following recommended practice, we used an out-of-sample evaluation. In other words, we fit the first part of the series to the data and then examined the performance of the forecasting method by seeing how well it worked on later observations. In this example, we used a hold-out sample of more than 50 percent of the observations because our primary focus was on evaluating forecasting methods; a larger hold-out sample provides a better basis for comparisons. We compared five methods:

1. MA(3): A moving average of three terms
2. MA(8): A moving average of eight terms
3. SES(0.2): Simple exponential smoothing with $\alpha = 0.2$
4. SES(0.5): Simple exponential smoothing with $\alpha = 0.5$
5. SES(opt): Simple exponential smoothing with $\alpha = 0.728$

For each of the SES sets of forecasts, we used the first observation as the starting value, as in the previous example. We then generated the one-step-ahead forecasts for 36 weeks (weeks 27–62), starting at forecast origin $t = 26$. Only method 5 requires any parameter estimation. The summary results appear in Table 3.5. The best performance on each criterion is shown in bold. In this case, the best-fitting optimal SES scheme performs best on all counts. However, it may well happen that the different criteria lead to different conclusions. ■

Table 3.5 Summary Error Measures for One-Step-Ahead Forecasts of WFJ Sales Data *(Hold-out sample, weeks 27–62)*

Method	MAE	RMSE	MAPE
Number of Observations	36	36	36
MA(3)	3067	4320	8.9
MA(8)	3749	4865	11.0
SES(0.2)	3389	4342	9.9
SES(0.5)	2832	3980	8.2
SES(opt)	**2562**	**3915**	**7.3**

Data: WFJ_sales.xlsx

3.3.4 The Use of a Rolling Origin

Some programs allow repeated estimation and evaluation of the forecast error by advancing the estimation sample one observation at a time and repeating the error calculations, a process known as using a *rolling origin*. This process is illustrated in Figure 3.8.

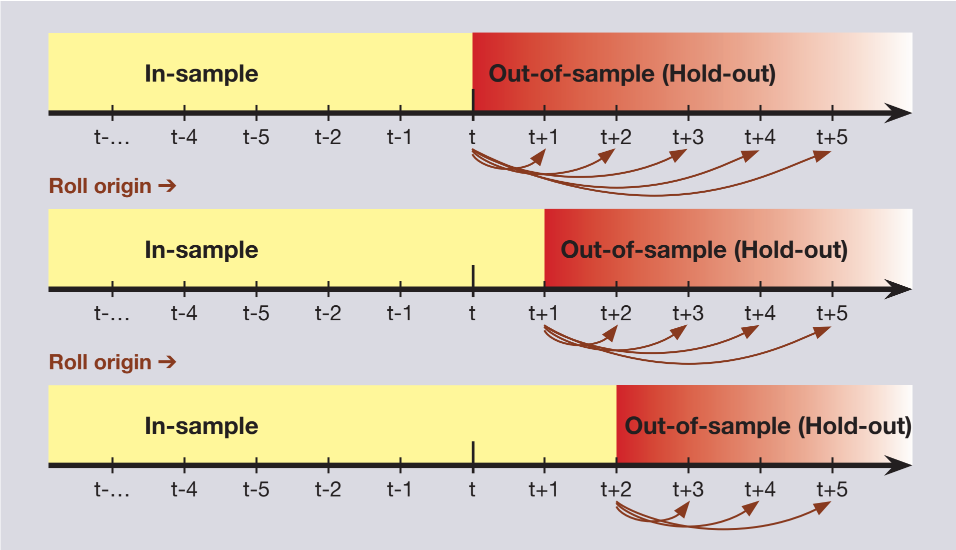
The first set of forecasts is generated with time t as the forecast origin, producing forecasts at time t+1, t+2, …, t+5. The origin is then moved to time t+1 by adding the observation at time t to the estimation sample, to produce forecasts at times t+2, …, t+5. The forecast origin is then moved to time t+2 and the new forecasts generated.

In this schematic example, when the process is complete, we would have five one-step-ahead forecasts, four at two steps ahead and so on, up to one forecast five steps ahead.

The use of a rolling origin provides a more reliable assessment of performance (Fildes, 1992): As forecasts are produced from multiple forecast origins, we collect more evidence of the forecasting performance of a method and avoid relying solely on a single origin, or hold-out that may contain outliers or other irregularities. In addition, a rolling origin reflects practical applications because an organization typically adds the latest observations to the database and reruns the analysis. For further discussion of forecasting accuracy measures, see Tashman (2000) and Davydenko and Fildes (2013).

The ESM provides a rolling origin simulation, which is illustrated in the next example. The reader is encouraged to try the simulation on other series.

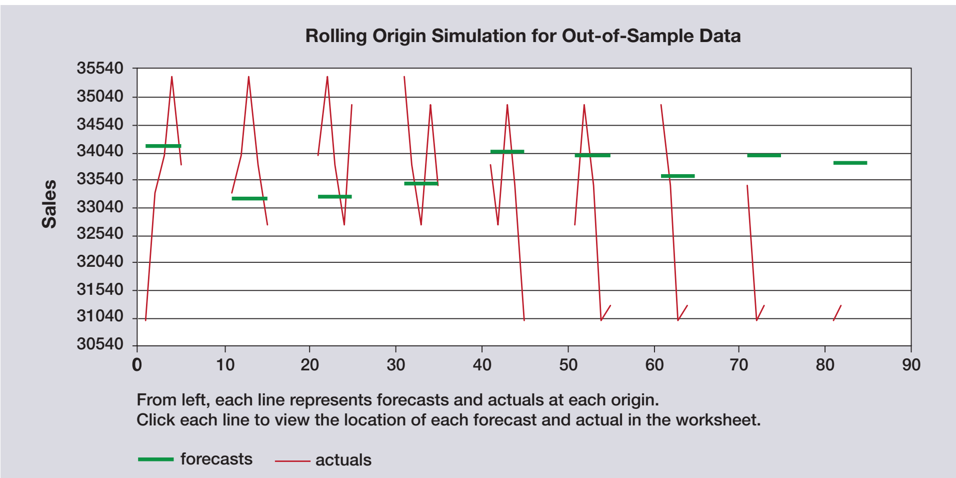
Figure 3.8 The Rolling Origin Process



■ Example 3.4: Rolling origin forecasts for the WSJ series

An illustration of the rolling origin process is presented in Figure 3.9 using the WSJ Sales series. As before, we start with observations 1–26 as the estimation sample. Given that SES is used, the forecasts lie on a horizontal line. Nine forecast origins are shown in the plot from week 26 through week 34. Thus, the first plot, with the origin at week 26, shows forecasts equal to 34149 and the observed sales in week 27 fall to 30986, below the forecast level made in the previous week. Accordingly, the new forecasts are set at the lower level of 33200 and the observed value of sales in week 28 is 33321, producing a slight increase in the next forecast to 33236. The process continues on until week 34 becomes the origin and the forecast is 33839. ■

Figure 3.9 Rolling Origin Analysis for the WSJ Sales Series



3.3.5 Some General Comments

On average, SES tends to outperform MA, as observed in an empirical comparison of their performance in the M3 forecasting competition, as reported by Makridakis and Hibon (2000) and in various other studies. In addition, as shown in Chapter 5, SES corresponds to an intuitively appealing underlying statistical model, whereas MA does not. Given the inferior properties and overall performance of moving-average-based procedures, we do not recommend their direct use for forecasting. However, moving averages have other uses, particularly in the area of seasonal adjustment, as seen in Chapter 4 (Sections 4.3-4.5).

One final question refers to the choice of fitting procedure. We could minimize the *MSE* (or, equivalently, the *RMSE)* or instead choose to minimize the *MAPE* or the *MAE.* To examine the effects of such a choice, we again use the WFJ series. The data through week 26 were used for estimation of the smoothing constant by each method in turn. Different options for the starting values (the first observation and the average of the first four observations) for the level produced essentially identical forecasts for weeks 27-62. Only the results for the starting value based upon the first observation are given in Table 3.6.

Table 3.6 Effects of Fitting SES by Minimization of RMSE, MAE, or MAPE

Estimation Criterion	RMSE		MAE		MAPE	
Sample	Estimation	Hold-Out	Estimation	Hold-Out	Estimation	Hold-Out
Error Measure						
RMSE	3030	3915	3045	3921	3045	3921
MAE	2127	2562	2012	2624	2012	2624
MAPE	6.6	7.3	6.3	7.5	6.3	7.5
Value of α	0.728		0.660		0.660	

Data: WFJ_sales.xlsx

In this example, the choice of fitting method produces only marginal differences in the results; a phenomenon that tends to happen, provided that adequate data are available for estimation. Here, the results when fitting with *MAE* and *MAPE* are identical; these two criteria produce similar results whenever the typical percentage error is small. Finally, we note that the out-of-sample error measures tend to be somewhat higher than those calculated for the estimation sample. This outcome should not come as a surprise, because the smoothing constant was chosen to minimize the appropriate criterion within the estimation sample, not the hold-out sample.

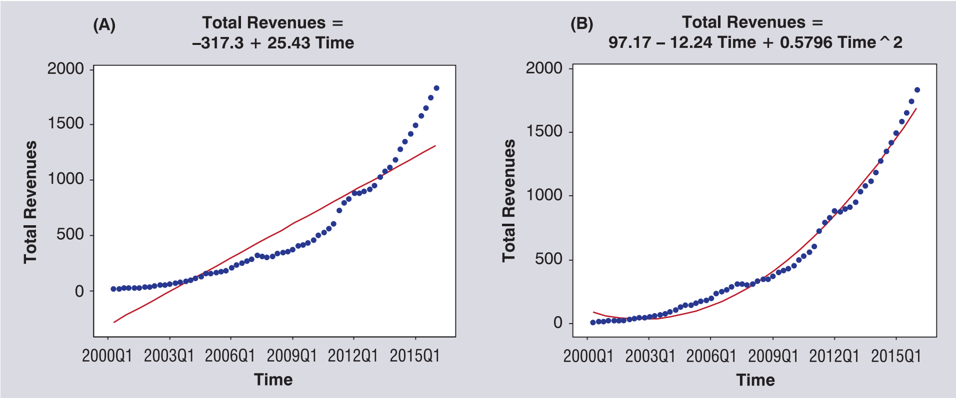
In all further analyses we choose to follow common practice and minimize the *MSE*, as do most forecasting packages.

3.4 Linear Exponential Smoothing

Figure 3.10 shows linear and quadratic trend curves fitted to the Netflix quarterly revenues data, over the period 2000Q1– 2015Q4. Plot (A) shows the linear trend, and plot (B) shows the quadratic trend.[4]

4 The calculations involved in creating these trend lines do not concern us here. A detailed explanation is given in Chapter 8 (Section 8.2).

Figure 3.10 Linear (A) and Quadratic (B) Trend Lines for Netflix Sales Revenues (2000Q1–2015Q4) [Time = 1, 2, …, 64]



Data: Netflix_2.xlsx

The quadratic seems to provide an excellent fit, but we must be wary of such global models — that is, models which assume a never-changing trend into the far future. Indeed, the road to chapter 11 bankruptcy proceedings is littered with the remains of companies that believed such optimistic growth patterns would persist. A successful start-up company will often show dramatic growth in the early years, but the growth moderates as the company and its markets mature and competitors emerge. The Netflix record is impressive, but investors should not get too carried away: The trend is likely to change over time. Accordingly, we introduce methods with a local trend estimate in the next section.

3.4.1 Basic Structure for LES

Global trend models assume that the trend remains constant over the whole time series. Because we are reluctant to rely upon the continuation of global patterns, we now develop tools that project trends more locally, just as we looked at local levels earlier. To understand the approach, we begin with the components of a straight line: the intercept (or starting level at time zero) and the trend (or slope), which we denote by L and T, respectively. To produce forecasts that are sensitive to recent changes in the series, we now define the following variables:

$$L_t = \text{level of series at time } t$$

$$T_t = \text{trend of series at time } t.$$

The level and trend measured at time t represent our current state of knowledge about these quantities. Thus, when we seek to forecast h periods ahead, we may construct a trend line that starts at level L_t and has trend T_t. That is, the forecast function for one-step-ahead is

$$F_{t+1}(1) = F_{t+1} = (\text{level at time } t) + (\text{trend at time } t) = L_t + T_t.$$

More generally, the forecast h steps ahead is

$$F_{t+h}(h) = (\text{level at time } t) + h \times (\text{trend at time } t) = L_t + hT_t. \tag{3.10}$$

Recall that, for SES, the trend is zero, so the forecast reduces to $F_{t+h}(h) = L_t$.

LOCALLY LINEAR FORECASTS

A time series is said to have a local linear trend if the mean level at any point in time is expected to increase (or decrease) linearly over time. The forecast function has the general form

$$F_{t+h}(h) = \text{Intercept} + h \times \text{Trend};$$

that is, the plot showing future forecasts is a straight line. When a new observation is recorded, estimates of the intercept and trend are updated.

3.4.2 Updating Relationships

We now consider updating the level and the trend, using equations like those we used for SES in Section 3.3.

Given the latest observation Y_t, we update the expressions for the level and the trend by making partial adjustments:

$$L_t = L_{t-1} + T_{t-1} + \alpha e_t \tag{3.11}$$
$$T_t = T_{t-1} + \beta(L_t - L_{t-1} - T_{t-1}).$$

As before, we define the observed error (e_t) as the difference between the newly observed value of the series and its previous one-step-ahead forecast:

$$e_t = Y_t - F_t = Y_t - (L_{t-1} + T_{t-1}).$$

Equations (3.11) may be expressed in what is known as the *error correction form* of the updating equations. The first equation is unchanged, but as may be checked by substitution (see Exercise 3.19), the update of the trend can be expressed as

$$T_t = T_{t-1} + \alpha\beta e_t. \tag{3.12}$$

These equations may be explained as follows:

- The new level is the old level (adjusted for the increase produced by the trend) plus a partial adjustment (with weight α) for the most recent error.
- The new trend is the old trend plus a partial adjustment (with weight $\alpha\beta$) for the error.

It is apparent from equations (3.11) that a second round of smoothing is applied to estimate the trend, leading some authors to describe the method as *double exponential smoothing*. Because the forecast function (3.10) defines a straight line, we prefer the name *linear exponential smoothing (LES).* The method is also known as *Holt's Method,* after one of its originators (Holt, 2004). If we set the trend equal to zero at all times, we are back at simple exponential smoothing; in that case, T_t vanishes and the equation for L_t reduces to equation (3.7) on identifying L_t as F_{t+1}.

3.4.3 Starting Values

To set this forecasting procedure in motion, we need starting values for the level and trend, as well as values for the two smoothing constants α and β. The smoothing constants may be specified by the user, and conventional wisdom decrees using $0.05 < \alpha < 0.3$ and $0.05 < \beta < 0.15$. These guidelines are not always appropriate, however; it is better to view them as suggesting initial values in a procedure for selecting optimal coefficients by minimizing the

MSE over some initial sample, as we did for SES. The performance of the resulting forecasting equations should be checked out of sample, as in Example 3.5.

As with SES, different programs use a variety of procedures to set starting values (Gardner, 1985). The ESM uses the starting values

$$T_3 = \frac{(Y_3 - Y_1)}{2} \text{ for the trend}$$

and

$$L_3 = \frac{(Y_1 + Y_2 + Y_3)}{3} + \frac{(Y_3 - Y_1)}{2} \text{ for the level.}$$

These values correspond to fitting a straight line to the first three observations. Once the initial values are set, equations (3.11) are used to update the level and trend as each new observation becomes available.

Caution: All major forecasting packages include fitting routines for LES, although the particular algorithms employed vary, so the results may differ somewhat from those shown in this book. Further, the use of different initial values for estimating the parameters may lead to different final values. Try several values to ensure that a global minimum of the MSE has been reached.

■ Example 3.5: Spreadsheet for LES calculations

A short series of hypothetical sales data is given in Table 3.7. The smoothing constants are set at $\alpha = 0.3$, $\beta = 0.1$, and the calculations were performed using the ESM. Thus, the first three observations were used to set initial values for the level and trend, and the forecast error measures were then computed with the use of observations 4-12, which may be regarded as the hold-out sample in this case. These values are genuine forecasts because no estimation is involved. The entries corresponding to time period 13 represent the one-step-ahead forecast and its components. The forecast errors and their squares, absolute values, and absolute percentage errors *(APEs)* are shown in succeeding columns. The last two rows of the table give the Mean Error, *ME* (as defined in Section 2.7.1), *RMSE, MAE,* and *MAPE,* calculated over the hold-out periods 4-12. ■

Table 3.7 LES Calculations for Hypothetical Sales Data [Alpha (α) = 0.3, Beta (β) = 0.1]

Period	Sales	Level	Trend	Forecast	Error	Error^2	|Error|	APE
1	5.00							
2	7.00							
3	9.00							
4	10.00	9.00	2.00	11.00	–1.00	1.00	1.00	10.00
5	11.00	10.70	1.97	12.67	–1.67	2.79	1.67	15.18
6	12.00	12.17	1.92	14.09	–2.09	4.36	2.09	17.41
7	16.00	13.46	1.86	15.32	0.68	0.46	0.68	4.25
8	17.00	15.52	1.88	17.40	–0.40	0.16	0.40	2.36
9	20.00	17.28	1.87	19.15	0.85	0.73	0.85	4.27
10	17.00	19.40	1.89	21.29	–4.29	18.44	4.29	25.26
11	21.00	20.01	1.76	21.77	–0.77	0.59	0.77	3.66
12	22.00	21.54	1.74	23.28	–1.28	1.63	1.28	5.80
13		22.89	1.70	24.59				
					ME	**RMSE**	**MAE**	**MAPE**
					–1.11	1.83	1.45	9.80

Data: Example 3_5.xlsx. Averages taken over Periods 4-12.

Example 3.6: Linear exponential smoothing for WFJ sales

We now consider the use of LES for the WFJ Sales data given in *WFJ_sales.xlsx*. This table shows a number of interesting features:

- The hold-out sample produces somewhat higher values for the error measures, as would be expected.
- The estimates of β are always zero, so the trend term is not updated. Too much should not be made of a single example, but it is clear that we should retain an adequate number of observations for out-of-sample testing. If there are sufficient data, retain at least 10-12 observations for this purpose.

Finally, a comparison of Tables 3.5 and 3.8 indicates that there is no benefit to using LES in this case; indeed, SES does somewhat better, confirming our initial impression that the series did not show any marked trend, after the initial period when sales were lower. ■

Table 3.8 Estimation Sample and Hold-Out) Sample Results for WFJ Sales *(Fitted by minimizing the MSE over weeks 1–26)*

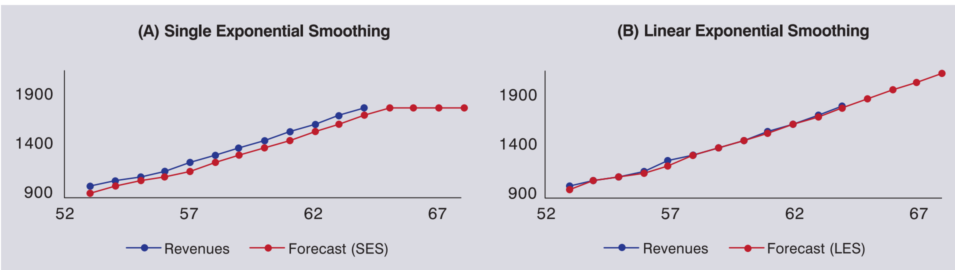
Error Measure	Estimation Sample	Hold-Out Sample	Estimation Sample	Hold-Out Sample
RMSE	3030	3915	3090	4033
MAE	2128	2562	2321	2694
MAPE	6.62	7.33	7.42	7.91
Value of α	0.73		0.70	
Value of β			0.00	

Data: WFJ_sales.xlsx

Example 3.7: Forecasting Netflix sales with LES

The sales figures for Netflix, given in *Netflix_2.xlsx* and plotted in Figure 3.10 exhibit a very strong trend, and we would expect LES to perform much better than SES in this case. The series was fitted to observations 1-52 via the ESM, leaving the last 12 observations as a hold-out sample. From Figure 3.11A, we see that the SES forecast always undershoots the next value of the series because it fails to allow for the upward trend. By contrast, the LES does a much better job; actual and forecast values are barely distinguishable in Figure 3.11 (B).

Figure 3.11 Netflix Revenues and One-Step-Ahead Forecasts for Observations 53–64 from (A) Single Exponential Smoothing and (B) Linear Exponential Smoothing



Data: Netflix_2.xlsx

The results are summarized in Table 3.9. LES is clearly much superior to SES, as we would expect. However, we observe that α is set at its upper level of 1.0, indicating that even LES may have problems capturing the pattern in the data. Again, if we step back from the technical details and look at the plot of the data in Figure 3.10, we see that the growth is exponential rather than linear. In Section 3.7, we examine the use of transformations as a way of dealing with this question. ■

Table 3.9 Summary Measures for SES and LES for Netflix Sales

	SES		LES	
Error Measure	**Estimation Sample**	**Hold-Out**	**Estimation Sample**	**Hold-Out**
RMSE	28.3	75.1	18.9	21.7
MAE	19.1	73.2	10.6	15.2
MAPE	9.5	5.3	4.3	1.2
Value of α	1.00		1.00	
Value of β			0.39	

DISCUSSION QUESTION: *How would you choose among SES, LES, or a fitted straight line for use with a given series?*

3.5 Exponential Smoothing with a Damped Trend

As we saw earlier, the growth rate for Netflix slowed as the company matured. This phenomenon is quite common in time series for sales where a product line matures and sales may then decline unless the product is upgraded in some way. Indeed, such a product life cycle is a standard expectation in marketing. We can accommodate these kinds of life-cycle effects by modifying the updating equations for the level and trend. We anticipate that, in the absence of random errors, the level should flatten out unless the process encounters some new stimulus. In turn, this expectation means that the trend should approach zero. Intuitively, we look for a means of forecasting that damps down the trend component as the forecast horizon is extended. In other words, we assume that the series will level out over time. We achieve this adjustment to our method by introducing a damping factor so that the error correction form of equations (3.11) become

$$L_t = L_{t-1} + \phi T_{t-1} + \alpha e_t \tag{3.13}$$

$$T_t = \phi T_{t-1} + \alpha\beta e_t \, .$$

In equations (3.13), we have inserted a *damping factor* ϕ with a value between 0 and 1 in front of every occurrence of the trend term T_{t-1}. We select ϕ to be positive but less than 1 so that the effect is to shift the trend term toward zero, or to dampen it. The effect is to produce the one-step-ahead forecast

$$F_{t+1}(1) = F_{t+1} = L_t + \phi T_t .$$

That is, we only incorporate a proportion of the trend factor into the forecast. By feeding the forecast values for the level and trend back into this equation we can compute the forecast function for h steps ahead, which has the form

$$F_{t+h}(h) = L_t + (\phi + \phi^2 + \cdots + \phi^h) T_t . \tag{3.14}$$

The derivation is left as Exercise 3.20.

This forecast levels out over time, approaching the limiting value $L_t + \phi T_t/(1 - \phi)$ because the damping factor is less than 1. This limiting value contrasts sharply with LES, where $\phi = 1$ and the forecast keeps increasing because $F_{t+h}(h) = L_t + hT_t$.

The damped trend method was introduced by Gardner and McKenzie (1985) and has proved surprisingly effective (Makridakis and Hibon, 2000).

■ Example 3.8: Forecasting Netflix sales with the damped trend method

The optimal damping factor is estimated to be $\phi = 0.94$; the results are summarized in Table 3.10 including the performance measures for the hold-out sample. The forecasts are marginally inferior to those from LES without damping, but if we believe that company expansion is slowing, the damped form may be preferable for future use: the optimal parameters for past data do not necessarily deliver the most accurate forecasts! ■

Table 3.10 One-Step-Ahead Fitted Values and Forecasts for Netflix Sales Under the Damped Trend Method

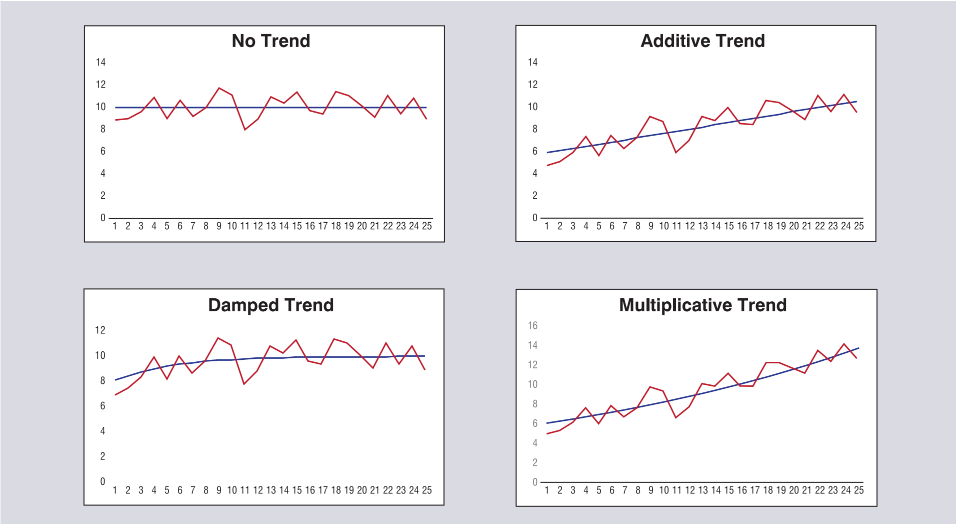
Period	Total Revenues	Level	Slope	Forecast	Forecasting Accuracy Statistics
1	5.17				
2	7.15				
3	10.18	10.01	2.50		
4	13.39	13.39	2.84	12.36	
5	17.06	17.06	3.14	16.06	
6	18.36	18.36	2.18	20.01	
...	...	...	...	...	
60	1484.73	1484.73	68.54	1472.00	RMSE = 24.90
61	1573.13	1573.13	75.69	1549.21	MAE = 18.40
62	1644.69	1644.69	71.37	1644.34	MAPE = 1.44
63	1738.36	1738.36	79.58	1711.83	
64	1823.33	1823.33	79.60	1813.22	α = 1.00
65				1898.21	β = 0.59
66				1968.66	ϕ = 0.94
67				2034.93	
68				2097.27	

Data: Netflix_2.xlsx

3.5.1 Choice of Method

The different forms of exponential smoothing apply to certain types of time series. Figure 3.12 shows a categorization of series by the form of the trend (no trend, additive trend, damped trend, multiplicative trend). In choosing a method, the aim is to match it with the type of data. Often, however, the form of the series is unknown. By implication in our earlier discussion, we recommend choosing the method that performs best on the hold-out sample, using one or more of the criteria discussed. When the series is too short to allow for a hold-out sample of reasonable size, we use measures defined for the estimation sample, as well as our understanding of the time series itself. This approach is developed in Chapter 5.

Figure 3.12 Possible Trend Patterns Showing Trend/No Trend and Additive/Multiplicative Trend



3.6 Other Approaches to Trend Forecasting

Several other methods of trend forecasting are worthy of a brief mention. For a more comprehensive review of recent developments, see the excellent review article by Gardner (2006).

3.6.1 Brown's Method of Double Exponential Smoothing (DES)

As noted earlier, Robert G. Brown was the original developer of exponential smoothing methods and his books (1959, 1963) have become classics. His initial derivation of exponential smoothing was based on minimizing the weighted sum of squared errors, where the past squared errors are discounted by α, α^2, etc. For a local linear trend, his method reduces to the use of LES with $\alpha = \beta$. Unless data are very limited, there is no particular benefit to imposing this restriction, and we do not consider it further. However, the discounted least squares approach is particularly useful when complex nonlinear functions are involved and updating equations are not readily available.

3.6.2 SES with (Constant) Drift: The Theta Method

If we set $\beta = 0$, the updated equations (3.11) become

$$L_t = L_{t-1} + T + \alpha e_t$$

$$T_t = T_{t-1} = T.$$

This version may be referred to as *SES with drift,* because the level increases by a fixed amount each period. Although the method is just a special case of LES, the simpler structure makes it easier to derive an optimal value for T by using the estimation sample, rather than the start-up values we have considered hitherto. This scheme, sometimes called the (simplified) *Theta Method* (Assimakopoulos and Nikolopoulos, 2000), is often surprisingly effective, shown by Hyndman and Billah (2003), who compared it to other methods using the data of the M3 competition (Makridakis and Hibon, 2000). An updated examination of the method with extensions is given by Fiorucci *et al.* (2016).

3.6.3 Tracking Signals

Trigg and Leach (1967) introduced the concept of a tracking signal, whereby not only the level and trend, but also the smoothing parameters, are updated each time. For example, for SES, we would use the updated value for α given by

$$\alpha_t = \frac{|E_{t-1}|}{M_{t-1}}, \tag{3.15}$$

where E_t and M_t are smoothed values of the error and the absolute error, respectively. That is,

$$\begin{aligned} E_t &= \delta e_t + (1 - \delta)E_{t-1} \\ M_t &= \delta\,|\,e_t\,| + (1 - \delta)M_{t-1}. \end{aligned} \tag{3.16}$$

Typically, a value of δ in the range 0.1-0.2 is used. If a string of positive errors occurs, the value of α_t increases to speed up the adjustment process; the reverse occurs for negative errors. Initial values are set to zero.

Over the years, there has been considerable debate over the benefits of tracking signals; for example, Gardner (1985) found no real evidence that forecasts based upon tracking signals provided any improvements. A generally preferred approach is to update the parameter estimate regularly, which is no longer much of a computational problem even for large numbers of series. Fildes *et al.* (1998) found that regular updating produced more consistent gains over all forecast horizons. From the practical viewpoint, a forecasting system may incorporate data on a weekly or monthly basis and then the forecasting analysis rerun at regular intervals to find revised parameters. That way, updating is straightforward and should not be a problem with modern forecasting systems.

3.6.4 Linear Moving Averages

In Section 3.2, we considered a simple moving average as an alternative to SES, albeit not a recommended approach. We could also look at the successive differences in the series $Y_t - Y_{t-1}$ and take a moving average of these values to estimate the trend. The net effect is to estimate the trend by $(Y_t - Y_{t-K})/K$ for a K-term moving average, leading to the forecast function $F_{t+1}(1) = F_{t+1} = Y_t + (Y_t - Y_{t-K})/K$. Again, exponential smoothing for a trend via LES usually provides better forecasts.

3.7 The Use of Transformations

The forecasts derived from the LES method require that the series is locally linear. Intuitively, if the trend for the last few time periods in the series appears to be close to a straight line, the LES method should work well. However, in many cases this assumption is not realistic. For example, suppose we are trying to forecast the *GDP* of a country that has experienced growth of around 5 percent per year in recent years. Then, ignoring statistical variation for the moment, we might describe such growth by the expression

$$Y_t = (1.05)^t Y_0 \text{ or } Y_t = 1.05Y_{t-1}.$$

When plotted, this function is an exponential curve, which increases at an increasing rate in monetary terms. Any linear approximation will undershoot the true function sooner or later, (This observation was first made hundreds of years ago by Malthus regarding linear growth in the world's food supply and exponential population growth, leading to starvation.) For very short-term forecasts, undershooting may not matter, but it becomes serious as the forecasting horizon increases. More complex nonlinear patterns may also exist; for example, many products have a sales history of growth, then stability, then a decline as they are superseded by new (and presumably improved) alternatives. Such series can be forecast in two ways:

1. Transform the series so that the trend becomes linear.
2. Convert the series to growth over time, forecast the growth rate, and then convert back to the original series.

We now examine each approach in turn.

3.7.1 The Log Transform

We defined the logarithmic transformation in Chapter 2 (Section 2.6.2). For the exponential growth equation just given, the log transform yields

$$\ln Y_t = \ln(1.05) + \ln Y_{t-1}.$$

If we now write $Z_t = \ln Y_t$, the reverse transformation is

$$Y_t = \exp(Z_t) = \exp[\ln (1.05) + \ln Y_{t-1}] = 1.05\, Y_{t-1}.$$

Of course, the growth rate is rarely constant. (If it were, we would estimate it by means of the averaging methods of Chapter 2.) Instead, when the log-transform produces a linear trend, we can apply LES. We must then transform back to the original series to obtain the forecasts of interest.

Typically, the effect of the log transformation process is to improve forecasting performance for exponential growth problems.

■ Example 3.9: Forecasting Netflix sales with the log transform

We applied LES to the transformed series, using the ESM. We applied the same procedures for setting starting values and estimating the parameters as before. The summary statistics are presented in Table 3.11. The table gives the results of estimating the parameters by minimizing the *MSE* of (a) the transformed values and (b) the original values. The summary values are computed after transforming back to the original units; otherwise comparisons between methods are not feasible. Also, we are interested in forecasting the original series, not the transformed values.

Because the original series minimizes *MSE* directly, whereas the minimum *MSE* for the transformed series relates to those transformed values, the *MSE* for the estimation sample will be lower for the original series than for the transformed case after transforming back to the original units; this is the case in Table 3.11. In turn, this effect stresses the importance of using a hold-out sample; the transformed series has a lower *RMSE* for the hold-out sample, although the *MAE* and *MAPE* are higher.

Whether to use LES on the original or transformed series or to use SES on growth rates remains a question for further examination in any particular study. ■

Table 3.11 Summary Measures for LES for Netflix Sales, Using Both the Original Data and a Log Transform*

	Original Series		Log Series	
Error Measures*	**Estimation**	**Hold-Out**	**Estimation**	**Hold-Out**
RMSE	18.97	21.72	20.69	20.48
MAE	10.62	15.17	12.20	16.62
MAPE	4.29	1.22	5.11	1.34
Value of α	1		1	
Value of β	0.39		0.43	

*The error measures refer to the original units of the observations.

Data: Netflix_2.xlsx

3.7.2 Use of Growth Rates

We define the growth rate as in Section 2.6.1. Assuming that the variable has a natural origin, we see that the growth rate from one time period to the next is

$$G_t = 100 \times \frac{Y_t - Y_{t-1}}{Y_{t-1}}. \tag{3.17}$$

The growth rate is approximately equal to $\ln Y_t - \ln Y_{t-1}$ as indicated in Exercise 3.18.

Note that growth, by this definition, can be negative. For example, the period-by-period returns on an investment are defined exactly as in equation (3.17). After we compute the single-period growth rates, we may use SES to predict the growth for the next period, which we denote by $g_{t+1}(1) = g_{t+1}$, following our usual convention. The one-step-ahead forecast for the original series is given by

$$F_{t+1}(1) = F_{t+1} = Y_t \times \left[1 + \frac{g_{t+1}}{100}\right]. \tag{3.18}$$

In other words, we unscramble the growth forecast to determine the forecast for the original series.

■ Example 3.10: Forecasting Netflix sales revenue data by means of the growth rate

The quarterly revenues (in $ million) for Netflix are listed in *Netflix_2.xlsx* for the first quarter of 2000 through the fourth quarter of 2015. The growth rates for one quarter over the immediately preceding quarter are then computed with equation (3.17). Selected values are listed in Table 3.12, including the summary measures for the hold-out sample (observations 53–64). The growth rates are plotted in Figure 3.13, which clearly reveals the instability in the early growth phase. The last four one-step-ahead sales forecasts are then given in the last column of Table 3.12, generated from equation (3.18), along with the 1 to 4 step-ahead forecasts for periods 65–68.

The summary measures for the hold-out sample given in Table 3.12 suggest that the log transform (recall Table 3.11) is marginally superior to the growth approach in this case. ■

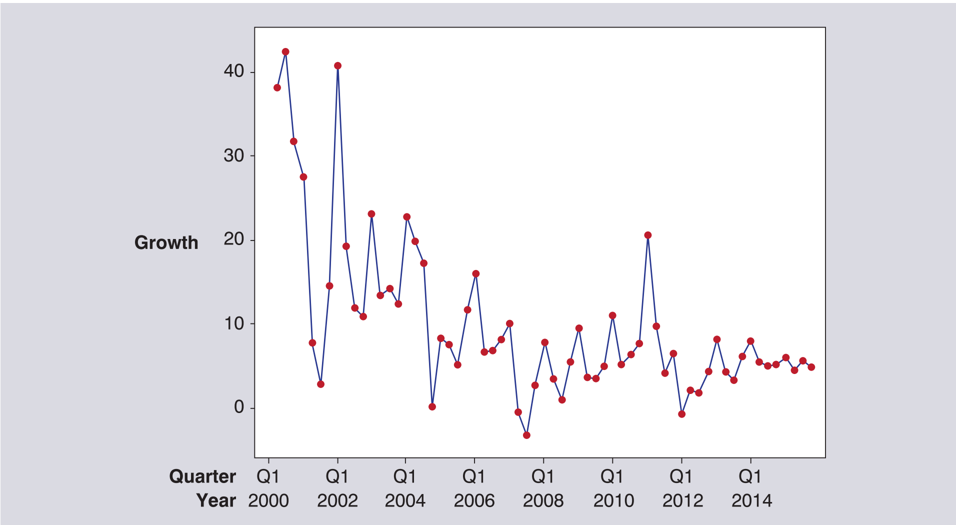
Table 3.12 Growth Rate Analysis of Netflix Quarterly Sales, 2000–2015

Year	Quarter	Total Revenues	Growth Percentage	Growth Forecast	Revenues Forecast	Forecast Accuracy Statistics
2000	1	5.17				
2000	2	7.15	38.13			
2000	3	10.18	42.47	38.13	9.87	
2000	4	13.39	31.52	40.26	14.28	
2001	1	17.06	27.38	35.97	18.21	
2001	2	18.36	7.63	31.76	22.47	
…	…	…	…	…	…	For α = 0.49
2015	1	1573.13	5.95	5.49	1566.25	
2015	2	1644.69	4.55	5.72	1663.08	RMSE = 20.86
2015	3	1738.36	5.70	5.14	1729.31	MAE = 17.35
2015	4	1823.33	4.89	5.41	1832.49	MAPE = 1.40
2016	1			5.16	1917.35	
2016	2			5.16	2016.22	
2016	3			5.16	2120.19	
2016	4			5.16	2229.52	

Quarterly revenue (sales) in $ million; growth in percentages.

Data: Netflix_2 .xlsx

Figure 3.13 Netflix Growth Rates, First Quarter, 2000, Through Fourth Quarter, 2015



3.7.3 The Box-Cox Transformations

The logarithmic transformation is appealing because it reflects proportional rather than absolute change. For many series in business and economics, the notion of proportional, or percentage, change is a natural framework; we often encounter statements, such as "We expect *GDP* to grow by 2.5 percent next year," leading to exponential growth. However, proportional change may project future growth patterns in excess of reasonable expectations. An examination of Figure 3.13 shows that the growth rate for Netflix moderated over time, and became more stable.

How can we allow for the "irrational exuberance" sometimes shown by exponential growth but where the logarithmic transformation fails to induce linearity? We have already discussed a modification of LES to allow for a damped trend; this modification can be applied after the log transform when appropriate. A second possibility is to select a transformation that is more moderate than the logarithmic form.

Instead of restricting the choice to the original units or the logarithmic transform, Box and Cox (1964) suggested using a power transformation, of which the square root and the cube root are the most obvious cases. On occasion, such transformations may have a natural interpretation, as when one is considering the volume of a sphere; the cube root is proportional to the radius of the sphere. It is difficult, however, to find any examples in business or economics in which such transformations have intuitive appeal. Nevertheless, from a purely empirical perspective, these transforms may provide better forecasts. Given the original series Y_t we define the Box-Cox transform[5] as

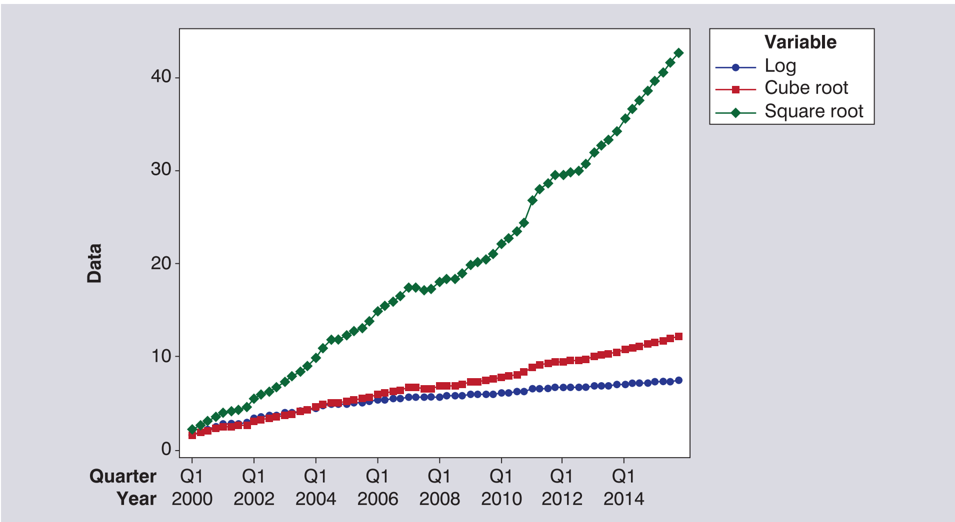
$$Z_t = Y_t^c. \tag{3.19}$$

The reverse transformation is $Y_t = Z_t^{1/c}$.

The parameter c is usually restricted to the range where $-1 \leq c \leq 1$. where $c = 1$ corresponds to the original series and $c = -1$ represents the reciprocal. (Think of miles per gallon versus gallons per mile.) We use only the square and cube root as examples; in general, the choice among different values for c may be made by minimizing the MSE in the usual way; the ESM enables this option.

The transformed data values are plotted in Figure 3.14. From the plots, we can see that the log transform overcorrects, exhibiting a slight curvature or flattening out. By contrast, the cube-root plot suggests a straight-line trend.

Figure 3.14 Plots of Transformed Netflix Data



5 Box and Cox (1964) used the slightly more complex form $Z_t = (Y_t^c - 1)/c$, which has numerical advantages when one is seeking an optimal value for c. For our purposes, the simpler form in equation (3.19) suffices. Optimal values for c may be computed using the ESM.

Example 3.11: Forecasting Netflix sales by means of the Box-Cox transform

We considered the square- and cube-root transformations for the Netflix series. The results given in Table 3.13 relate to fitting by minimizing the *MSE* for the transformed series. When the *MSE* for the original data is used to estimate the parameters, the forecasting results are similar and are omitted.

When we compare the results in Table 3.13, we see that the square-root transformation seems to provide the better performance in the estimation sample, but the cube root does much better in the hold-out sample, presumably because it is produces a more linear pathway. The cube-root also outperforms the logarithm. ■

Table 3.13 LES Summary Measures for Netflix Sales, Using the Box-Cox Transform

Transform	c	α	β	RMSE	MAE	MAPE
Square root	0.50	1.00	0.05			
Estimation				19.13	10.75	4.23
Holdout				27.34	23.98	1.78
Cube root	0.33	1.00	0.09			
Estimation				19.51	11.11	4.50
Holdout				18.93	14.84	1.17

DISCUSSION QUESTION: *Suppose you carry out an extensive analysis and conclude that the best forecasting model involves a Box-Cox transform with $c = 1/3$. Would you feel comfortable using this transform and reporting the results to management? Why or why not?*

WHEN TO TRANSFORM

- Do not use complex transforms unless they are supported by both theory and data.
- Always compare transformed methods with a benchmark by transforming the forecasts back to the data series of interest.

3.8 Prediction Intervals

Our discussion thus far has been purely in terms of point forecasts, yet in Chapter 2 (Section 2.9) we stressed the importance of interval forecasts. At this stage, we restrict attention to the construction of prediction intervals for one-step-ahead forecasts. Intervals for forecasts made two or more periods ahead require the specification of an underlying statistical model, so we defer the formal discussion to Section 5.2. Indeed, the intervals for one-step-ahead forecasting also imply an underlying model, but we can bypass that requirement for now by using the approximate interval given in Section 2.9, based upon the normal distribution:

$$\text{Forecast} \pm z_{\alpha/2} \times (RMSE).$$

In other words, $100(1 - \alpha)\%$ of the time the outcome is expected to lie within the range

$$[\text{Forecast} - z_{\alpha/2} \times (RMSE), \text{Forecast} + z_{\alpha/2} \times (RMSE)].$$

We may also construct a prediction interval empirically, using the alternative approach given in Section 2.9; the interpretation is the same, but the interval need not be symmetric. When a transformation is used, the prediction interval is computed using the transformed values and then transformed back to the original units.

Example 3.12: Prediction interval for Netflix

From Table 3.11, the point forecast for the next period (2016.1) is 1906.8 with estimation sample $RMSE = 18.97$. The resulting (approximate) 95 percent prediction interval is

$$1906.8 \pm 1.96 \times 18.97 = 1906.8 \pm 37.0 = (1869.86, 1944.0)$$

The actual outcome is expected to lie in the interval (1870, 1944) approximately 95 percent of the time. As the forecasting horizon increases, the intervals get progressively wider, Chapter 5 (Section 5.2). ■

Example 3.13: Prediction interval for WFJ sales

The first 51 observations of the series are used as the estimation sample, which provides 50 observed error terms (one being lost in the initialization). To obtain empirical prediction intervals we use the pairs (3, 48), (2, 49), and (1, 50) of the 50 ranked error terms, as shown in Table 3.14A. Recall from section 2.8.3, with 50 observations, 100(1-0.5)/50, i.e. 1% are estimated as lower than the smallest observation with 1% larger than the highest giving a 98% interval. Similarly, 3% are lower (higher) than pairs (2, 49) and 5% from pair (3, 48). The half-widths of the prediction intervals (Table 3.14B) show that the distribution may be somewhat more heavy-tailed than the normal distribution. Finally, the two alternative methods of calculation, normal and empirical, for the 90 percent one-step-ahead prediction intervals for weeks 52-62 are shown in Table 3.14C, along with the actual values of the series. These ten observations occur around the year's end, when the volume of sales changes dramatically, so it comes as no surprise that the coverage of the two sets of prediction intervals is less than the nominal level. Such discrepancies would be a signal to management of changed market conditions: in this case, clearly higher sales in the holiday period. Model extensions to handle such circumstances are discussed in Section 9.1. ■

Table 3.14 One-Step-Ahead Prediction Intervals for WFJ Sales Series

(A) Smallest and largest observed errors			
Smallest Five		Largest Five	
Rank	Value	Rank	Value
1	–6378	46	3765
2	–5146	47	3801
3	–4574	48	5253
4	–3609	49	5605
5	–2715	50	8256

(B) Half-width of prediction interval			
Level	90	94	98
Normal	3436	4168	5506
Empirical	4913	5376	7317

(C) Normal and empirical prediction intervals for observations 51-62

Week	WFJ Sales	Point Forecast	90% Normal PI		90% Empirical PI	
			Lower	Upper	Lower	Upper
52	51914	41801	38365	45236	36887	46714
53	35404	49329	45893	52764	44416	54242
54	30555	38963	35528	42399	34050	43877
55	30421	32704	29269	36140	27791	37617
56	30972	31005	27569	34440	26091	35918
57	32336	30981	27545	34416	26067	35894
58	28194	31990	28554	35425	27076	36903
59	29203	29164	25729	32600	24251	34077
60	28155	29193	25757	32628	24280	34106
61	28404	28420	24985	31856	23507	33334
62	34128	28408	24973	31844	23495	33321

Highlighted values fall outside one or both sets of intervals.

Data: WFJ_sales.xlsx

DISCUSSION QUESTION: *Suppose a revised forecasting method describes the WFJ Sales data for weeks 1-52 pretty well. But four actual observations fall outside the normal prediction interval for weeks 53-62. What possible explanations are there for this occurrence?*

3.9 Method Selection

In the course of this chapter, we have outlined a considerable number of forecasting methods, and more are to come in the next chapter when we discuss seasonality. For any particular forecasting task, one approach may naturally suggest itself in preference to another. For example, it is always reasonable to ask and answer the following questions:

- Does the series display a trend? Do we expect the trend to continue into the future? If so, then
  - ✓ Use a method that includes a trend.
- Do the observations tend to be more (or less) variable over time? If so, then
  - ✓ Transform the data to obtain roughly constant variability.

In addition, we need to check for conformity with the principles outlined in the next section. When there are a number of equally plausible methods, we usually rely on comparing their out-of-sample forecasting performance. When we are dealing with a large number of series, it is usually feasible to examine the preceding questions for a (possibly small) sample of all the series. We may overcome such obstacles by using a common estimation sample to fit each method and then generating a set of out-of-sample forecasts that can be compared with values in the hold-out sample. We then choose the method that corresponds to the best value of the selected criterion (*RMSE*, *MAE*, *MAPE*, *MdAPE*, etc.), calculated over the sample of time series. We return to the issue of method selection in Chapter 5 (Section 5.3), in which the analysis is reinforced by the development of underlying statistical models.

3.10 Principles of Extrapolative Methods

To avoid undue repetition, we assume that the data series being forecast is appropriate for the problem at hand in terms of relevance, timeliness, and reporting accuracy. These assumptions are by no means trivial, but we have discussed them in the previous chapters and they remain critical to any forecasting exercise. We should always recall the maxim "Garbage in, garbage out." If the data do not satisfy the aforementioned criteria, further analysis may be useless. As before, Armstrong (2001) is a valuable resource, and many of the principles quoted reflect his ideas. A few principles are repeated from Chapter 2 because they are an integral part of the forecasting approach described in the current chapter.

[3.1] Plot the series.
Data plotting should be the first step in any analysis. If a large number of series is involved, plot a selection of them. Such plots will often serve to identify data-recording errors, missing values, and unusual events.

[3.2] Clean the data.
Data plots and simple screening procedures (checks for outliers) provide the basis for making adjustments for anomalous values. Make sure that the adjustments are for valid data-recording reasons, and keep a record of all such changes.

[3.3] Use transformations as required by expectations about the process.
Such transformations may involve a conversion of current to real-dollar values, a logarithmic transformation to reflect proportional growth, or a switch to growth rates to account for trends. The intelligent use of knowledge related to the phenomenon being studied helps to avoid "crazy" forecasts.

[3.4] Select simple methods, unless convincing empirical evidence calls for greater complexity.
The set of exponential smoothing methods described in this chapter relies upon only the past history of the series in question. Such extrapolative methods often suffice in the short to medium term, unless measurements on key explanatory variables are available.

[3.5] Evaluate alternative methods, preferably using out-of-sample data.
Methods that use more parameters or are based on a more complicated nonlinear transform often fit better within the estimation sample, but this advantage may be illusory. Out-of-sample testing provides an even playing field for comparing the performance of different methods. Note that the estimation sample can be used to make such comparisons, provided that due care is taken (see Chapter 5).

[3.6] Update the estimates frequently.
Regular updating of the parameter estimates is found to improve forecasting performance because it helps to take into account any changes in the behavior of the series. Once the database has been established and updated with the most recent information, updating the parameter estimates is straightforward.

Summary

Extrapolative forecasts are useful in the short to medium term whenever the recent behavior of the series under study is sufficient to provide a framework for forecasting. When a series does not display marked changes in level over time, simple exponential smoothing (SES), as described in Section 3.3, usually suffices. However, many series do contain systematic trends, and in such circumstances linear exponential smoothing (LES) should be considered, as described in Section 3.4. A series may display nonlinear behavior, either in the growth pattern or because of some kind of life cycle, as is the case for the sales of many products. In Section 3.5, we explored the use of damped trend methods to handle these issues. Other extrapolative methods were reviewed briefly in Section 3.6. Section 3.7 looked at different transformations. The construction of one-step-ahead prediction intervals was examined in Section 3.8. We have introduced quite a number of different variants of exponential smoothing in this chapter so in Section 3.9 we discussed method selection — how to choose among them. Finally, some basic principles for forecasting with extrapolative methods were summarized in Section 3.10.

Exercises

*Topics marked with an * are advanced and may be omitted for more introductory courses.*

3.1 The growth rate in the U.S. gross domestic product *(GDP)* for 1963–2015 is provided in *GDP_change_2.xlsx.*

a. Use three- and seven-term moving averages to generate one-step-ahead forecasts for 2001 to the end of the series. Graph the results, and comment on the differences between the two moving averages.

b. Compare the performance of the two procedures by calculating the *RMSE* and *MAE.* Why is the *MAPE* inappropriate in this case?

3.2 The annual percentage change in the consumer price index *(CPI)* for 1963–2015 is provided in *CPI_change_2.xlsx*.

a. Use three- and seven-term moving averages to generate one-step-ahead forecasts for 2001 to the end of the series.

b. Compare the performance of the two procedures by calculating the *RMSE* and *MAE.*

c. Calculate the *RelMAE* and *MASE* for the one-step-ahead forecasts. Why is the *MAPE* inappropriate in this case?

3.3 Use the data in *GDP_change_2.xlsx* to generate forecasts for *GDP* growth by simple exponential smoothing (SES).

a. With the observed value for 1963 as the starting value, compute the one-step-ahead SES forecasts for 2001–2015, using each of α = 0.2, 0.5, and 0.8 in turn.

b. Compare the forecasting performance for the given values of α by calculating the *RMSE* and *MAE* over the period 2001–2015.

c. How does this method compare with the moving-average procedures used in Exercise 3.1? (Be careful to make comparisons over the same time periods.)

3.4 Rework the analysis in Exercise 3.3, using the optimal value of α. (Use the ESM or other suitable software.)

3.5 Use the data in *CPI_change_2.xlsx* to generate forecasts for changes in the *CPI* by means of simple exponential smoothing (SES).

a. Use the observed value for 1963 as the starting value, and compute the one-step-ahead forecasts for subsequent years for $\alpha = 0.2$, 0.5, and 0.8 in turn.

b. Compare the performance for the given values of α by calculating the *RMSE* and *MAE* over the period 2001–2015.

c. How does this method compare with the moving-average procedures used in Exercise 3.2? (Be careful to make comparisons over the same time periods.)

3.6 Rework the analysis in Exercise 3.5, using the optimal value of α. (Use the ESM or other suitable software.)

3.7 The average annual U.S. landed cost of Saudi Arabian Light Crude Oil (in U.S. dollars per barrel) for 1978–2015 is provided in *SA_oil_prices_2.xlsx*.

a. Use the observed value for 1978 as the starting value, and compute the one-step-ahead SES forecasts for subsequent years for each of $\alpha = 0.2$, 0.5, and 0.8 in turn.

b. With the same starting value, find the optimal level for α, using the data for the period 1978-2007 as the estimation sample. Use the ESM or other suitable software.

c. Generate the forecasts for 2008–2015.

d. Compute the out-of-sample *RMSE, MAE,* and *MAPE* for each case, and contrast the results. Does using the median rather than the mean make any difference?

3.8 Repeat the analysis in Exercise 3.7, using linear exponential smoothing. (Use the ESM or other suitable software.) Compare the forecasting performance of the two methods.

3.9 Repeat the analysis in Exercise 3.8, using 1978–2004 and then 1989–2004 as the estimation samples. Generate the 95 percent prediction intervals for the one-step-ahead forecasts in each case, using the normal approximation. Do the prediction intervals include the observed values? Interpret your results.

3.10 Evaluate the performance of SES and LES for the Netflix series *(Netflix_2.xlsx),* using different time ranges for the estimation sample, and note how the optimal values of the parameters and forecasting performance vary.

3.11 Use LES to forecast the number of domestic passengers at Dulles airport for 2004-2015 (see Table 2.2, *Dulles_2.xlsx),* with the years 1963–2003 as the estimation sample. Then use the complete data set through 2010 to make forecasts for 2011–2015. Use different time ranges for the estimation sample, and note how the optimal values of the parameters and forecasting performance vary.

3.12 Dulles Airport was greatly expanded in the mid-eighties, producing a significant increase in the number of passengers. Rerun the analyses of Exercise 3.11, using the years 1986–2010 as the estimation sample. Compare your results with those for the complete sample. What conclusions do you draw from the comparison?

3.13 Rerun the analyses of Exercise 3.9 for the damped trend and log transform versions of LES, and compare the results of the four methods, using *RMSE, MAE,* and *MAPE.*

3.14 Evaluate the approximate 95 percent prediction intervals for one-step-ahead forecasts for the hold-out sample used in Exercise 3.9.

3.15 Rerun the analyses of Exercise 3.12 for the damped trend and log transform versions of LES, and compare the results of the four methods, using *RMSE*, *MAE*, *RelMAE*, and *MAPE*. What conclusions do you draw?

3.16 Evaluate the approximate 95 percent prediction intervals for one-step-ahead forecasts for the hold-out sample used in Exercises 3.12 and 3.15.

3.17 Use the observed errors from the estimation sample in Exercise 3.12 to generate empirical 90 percent prediction intervals for the one-step-ahead forecasts for 2011–2015.

3.18* Show that $\ln(Y_t / Y_{t-1}) \approx (Y_t - Y_{t-1}) / Y_{t-1}$ when the difference between successive terms is small and, hence, that the log transform and growth rate methods often produce similar results. *(Hint:* $\ln(1 + x) \approx x$ for small x.)

3.19 Demonstrate that the trend equation in (3.11) may be rewritten as equation (3.12) by substituting the expression for the level at time t.

3.20* By repeated substitution of the level and trend forecasts into equations (3.13) show that the h-step-ahead forecast for damped trend is given by equation (3.14).

3.21 Repeat the analysis for the data on oil prices in Exercise 3.7, using the Box-Cox transform. Compare the results with the original forecasts ($c = 1.0$).

Minicases

Minicase 3.1 Job Openings

The monthly employment reports issued by the US Department of Labor are eagerly monitored by investors as indicators of the health of the overall economy. The file *Job_openings.xlsx* contains the monthly figures for total nonfarm openings, from January 2001–December 2015, seasonally adjusted.

To provide investing insights, develop a forecasting procedure for this series, using 2001–2012 as the estimation sample and 2013–2015 as the hold-out sample. You should consider all the options described in the chapter: SES, LES, damped LES, Log and Box-Cox transformations. All these options may be explored via the ESM.

Minicase 3.2 The Evolution of Walmart

Consider the data on Walmart stores described in Chapter 2 (Minicase 2.2) and available in *Walmart_2.xlsx*. In Minicase 2.2, we used basic statistical tools to examine the changing composition of store types and the company's growth. The objective now is to use exponential smoothing methods to generate the following forecasts:

1. Store composition for the next three years.
2. Sales for the next eight quarters.

In each case, plot the fitted and observed values of the series and identify any possible shortcomings in your forecasts. What steps, if any, might be taken to improve the forecasts *without collecting any new data?*

Minicase 3.3 Volatility in the Dow Jones Index

The efficient markets hypothesis (EMH) in finance embodies the notion that financial markets process new information rapidly in order that the best forecast of a stock's price in the next (short) time period be the current price. The EMH is also known as the random-walk hypothesis because any movement from the present state is essentially unpredictable. The percentage errors (or stock returns), which in this context are given by $R_t = 100(Y_t - Y_{t-1})/Y_{t-1}$, are so defined because the last period's closing price is the current forecast. The EMH corresponds to the assumption that the successive errors are independent and unpredictable from publicly available information. There is a vast literature on the EMH in its various forms (see, for example, *http://en.wikipedia.org/wiki/Efficient-market_hypothesis),* and our purpose is not to discuss the ideas in any detail. Rather, we note that, in forecasting terms, the EMH translates into the use of SES with $\alpha = 1$, essentially of no real value to the stock trader.

Despite many efforts to the contrary, it is essentially impossible for any purely statistical forecasting method to beat the random walk for any long period of time. Insider information can, of course, be of considerable value, but those forecasters tend not to be at liberty to talk about their methods. Nevertheless, some interesting questions can be answered that relate to *volatility.* Toward that end, we can examine the pattern of the forecast errors to determine whether the inherent variability in those forecasts varies over time. When variability is high, considerable opportunities exist for traders to buy or sell options (Hull, 2002). Conversely, when variability is very low, there is little room in the market for such contracts.

We may examine the absolute values of the one-step-ahead errors $|e_t| = |R_t - R_{t-1}|$. The spreadsheet *DowJones_2.xlsx* provides daily closing prices for the period 2011–2016, along with the values of the returns and their absolute values.

1. Examine the validity of the EMH for this series.
2. Use the SES to develop a forecasting method for the absolute values of the returns. Is volatility predictable, at least to some extent?

References

Armstrong, J. S., ed. (2001). *Principles of Forecasting: A Handbook for Researchers and Practitioners.* Boston and Dordrecht: Kluwer.

Assimakopoulos, V. and Nikolopoulos, K. (2000). The theta model: a decomposition approach to forecasting. *International Journal of Forecasting,* 16, 521–530.

Box, G. E. P., and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, series B,* 26, 211–252.

Brown, R. G. (1959). *Statistical Forecasting for Inventory Control.* New York: McGraw-Hill.

Brown, R. G. (1963). *Smoothing, Forecasting and Prediction of Discrete Time Series.* Englewood Cliffs, NJ: Prentice-Hall.

Davydenko, A., and Fildes, R. (2013). Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts. *International Journal of Forecasting*, 29(3), 510–522.

Fildes, R. (1992). The evaluation of extrapolative forecasting methods. *International Journal of Forecasting,* 8, 81–98.

Fildes, R., Hibon, M., Makridakis, S., and Meade, N. (1998). Generalising about univariate forecasting methods: Further empirical evidence. *International Journal of Forecasting,* 14, 339–358.

Fiorucci, J. A., Pellegrini, T. R., Louzada, F., Petropoulos, F., and Koehler, A. B. (2016). Models for optimising the theta method and their relationship to state space models. *International Journal of Forecasting*, 32(4), 1151–1161.

Gardner, E. S., Jr. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting,* 4, 1–28.

Gardner, E. S., Jr. (2006). Exponential smoothing: The state of the art — Part II. *International Journal of Forecasting,* 22, 637–677.

Gardner, E. S., Jr., and McKenzie, E. (1985). Forecasting trends in time series. *Management Science,* 31, 1237–1246.

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting,* 20, 5–10. Reprint of Office of Naval Research Memorandum 52/1957, Carnegie Institute of Technology.

Hull, J. C. (2002). *Fundamentals of Futures and Options Markets,* 4th ed. Upper Saddle River, NJ: Prentice-Hall.

Hyndman, R. J., and Billah, B. (2003). Unmasking the theta method. *International Journal of Forecasting,* 19, 287–290.

Makridakis, S., and Hibon, M. (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting,* 16, 451–76.

Tashman, J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting,* 16, 437–50.

Trigg, D. W., and Leach, D. H. (1967). Exponential smoothing with an adaptive response rate. *Operational Research Quarterly,* 18, 53–59.

Appendix 3A Excel Macro

This appendix provides a brief summary of the capabilities of the Exponential Smoothing Macro (ESM) used in this chapter. Full descriptions, along with the Excel files, are available in the Users' Manual on the book's website.

1. The parameters are estimated by minimizing the *RMSE;* the solution is obtained numerically with Solver.
2. The user may specify the estimation and hold-out samples.
3. One-step-ahead forecasts are generated for the hold-out sample and multiple step-ahead forecasts beyond the hold-out sample.
4. ESM enables consideration of SES, LES, damped LES and both logarithmic and Box-Cox transformations. The summary statistics are always reported for the original data, after inverse transformations where necessary.

CHAPTER 4

Seasonal Series: Forecasting and Decomposition

*Topics marked with an * are advanced and may be omitted for more introductory courses.*

Table of Contents

Every season has its peaks and valleys.
What you have to try to do is eliminate the Grand Canyon.

— Andy van Slyke, professional baseball player

Introduction

In the previous chapter, we examined forecasting methods for time series that contain trends. However, there are often other elements in a series that must be taken into account. In particular, the data frequently exhibit a seasonal pattern, such as sales of ice cream being higher in the summer. The term *seasonal* indicates a pattern that occurs within a *known* time frame. For example, the gross domestic product *(GDP)* is measured quarterly, so that the seasonal pattern relates to the calendar year and the series has a periodicity of 4. Likewise, unemployment is reported monthly, so that the seasonal pattern is again related to the calendar year, but now the series has a periodicity of 12. In both cases, seasonal effects arise because of the weather, so there may be lost production in the winter because of poor working and road conditions; in addition, the mix of economic activities varies as the result of more or less predictable changes, including holidays. From our perspective, it is important to emphasize that the period (such as a year) is known and that the patterns within a period (such as a year) tend to be similar from one period (or year) to the next.

However, we use the term *seasonal* more widely in this book, to refer to any pattern that occurs within a known fixed time frame. Thus, days of the week are an important seasonal effect for retail sales, because certain days (e.g., Saturdays) are known to produce higher levels of sales, whereas midweek sales tend to be lower. A further example is provided by the 24 hours within a day, which reveal a regular pattern for electricity consumption. In both these cases, seasonal effects exist in the annual cycle as well, so there are multiple seasonalities. The key feature in all these examples is that the time period(s) involved is (are) of *known and fixed* duration.

Why do we need to concern ourselves with seasonal effects? As always, forecasts underpin planning. A retailer needs to make inventory decisions and to plan deployment of the work force, whereas a utility company needs to plan for capacity that will meet peak demand, and to set lower prices for off-peak demand because the marginal cost of off-peak production is very low.

In Section 4.1, we examine the components of a time series, including trend and seasonality. Then, in Section 4.2, we look at forecasting for purely seasonal patterns — that is, for series that exhibit no trend. Series such as the average monthly temperature in a given location exhibit very little trend relative to seasonal changes (global warming notwithstanding), and such purely seasonal series represent a simple place from which to start our investigations. This starting point provides a natural progression to series that exhibit both trends and seasonal components, and we use a decomposition approach in Sections 4.3 and 4.4 to develop forecasts for such series. This approach leads to the creation of deseasonalized series — that is, series whose values have been adjusted to remove the effects of seasonal changes. Such methods are widely used in the presentation of macroeconomic data; indeed, many series produced by the United States and other governments are published only in a seasonally adjusted form. The most popular approach to the creation of such series is known as the Census X-13 ARIMA method, developed by the U.S. Bureau of the Census and described briefly in Section 4.5.

Most economic and business series display both trend and seasonal elements that evolve over time, and Sections 4.6 and 4.7 develop forecasting methods for such series. As with the exponential smoothing methods described in Chapter 3, we assume that a time series can be represented by a set of components. We then develop a forecast for each component and produce a forecast for the series that is a combination of all these elements. In the examples, the analysis uses the Exponential Smoothing Macro (ESM), so that results can be consistently replicated by the reader. However, more sophisticated programs may produce different results and this question is explored in Section 4.8 using programs available in R. The methods considered for most of the chapter are generally employed for monthly and quarterly series. Weekly series are often treated rather differently, as there may be very few years' worth of data to estimate so many (weekly) components; this issue is examined in Section 4.9. Prediction intervals for seasonal forecasts are developed in Section 4.10. The chapter concludes with a discussion of underlying principles in Section 4.11 and the chapter summary.

4.1 Components of a Time Series

We have introduced the trend and seasonal components of a time series, but there are also other components that are less regular. For example, the economy goes through business cycles of expansion and contraction; such cycles have tended to average around four years in the United States, but vary between two and ten years in duration. The duration of business cycles is not known in advance, nor is it stable from one cycle to the next. Some commentators have suggested that the business cycle is related to the four-year cycle of U.S. presidential elections, presumably on the grounds that a strong economy begets re-election! Similar suggestions apply in the United Kingdom and elsewhere. Even if successive administrations strive to induce such cycles, there is no guarantee that they will be of any predetermined length. One particularly well-established cycle occurs in the number of observed sunspots, with a mean time between peaks of around 10.7 years. The essential point is that such variations are of random duration (though with a predictable mean). We reserve the term *cyclical* for such patterns.

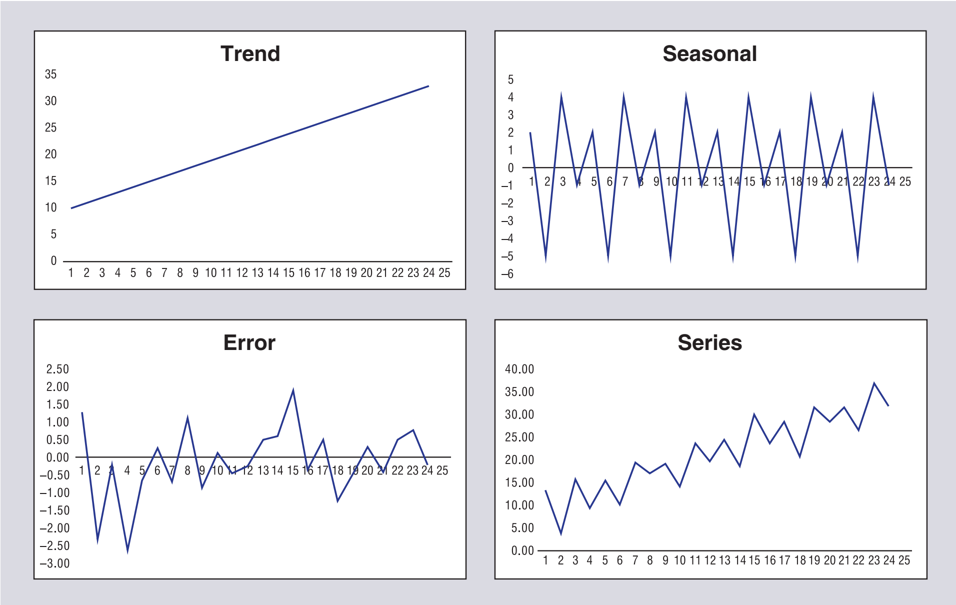
SEASONAL AND CYCLICAL TIME SERIES

A time series is said to have a *seasonal* component if it displays a recurrent pattern with a fixed and known duration.

A time series has a *cyclical* component if it displays somewhat regular fluctuations about the trend but those fluctuations have a periodicity of variable and unknown duration, usually longer than one year (e.g., a business cycle).

In addition to the trend and the seasonal and cyclical components, the series will contain a random error component, which, as before, represents the variations not accounted for by the other components. In common with most introductions to forecasting, we will absorb the cyclical component into the trend term and concentrate upon just three components: *trend (T), seasonal (S),* and *error (E).* Because our current focus is upon short- to medium-term forecasts, this simplification rarely causes any problems in practice. The combination of these components is illustrated in Figure 4.1. In the example, the trend is an upward-sloping straight line and the seasonal pattern corresponds to quarterly data (of periodicity 4). The error terms do not display any regular pattern. The three components are added together to provide the plot of the complete series (series A in Figure 4.1). In any forecasting exercise, we cannot observe the separate components and must find a way to create such elements from the observed (composite) series.

Figure 4.1 The Components of a Time Series



Algebraically, the three components are combined by addition to create the time series *(Y)*:

$$Y = T + S + E. \tag{4.1}$$

Equation (4.1) is known as the *(purely) additive* model for a series. Other models are possible, of which the two best known are the *(purely) multiplicative* model,

$$Y = TSE, \tag{4.2}$$

and the *mixed additive-multiplicative* model,

$$Y = TS + E. \tag{4.3}$$

Whenever the trend and the seasonal component are multiplied together, larger levels in the series will tend to exhibit larger peaks and troughs. In addition, when the error term is also multiplicative, the magnitude of the forecast errors will tend to rise and fall with the level of the series.

Before we consider forecasts derived from these equations, we examine forecasting for purely seasonal patterns — that is, series devoid of any trend. Although this situation is uncommon for business series over periods of a year or more, for short-term forecasts (e.g., retail sales over the next month of a product in a store) such models are often more than adequate. In addition, the approach aids in understanding the methods that examine both trend and seasonality and is useful if we wish to decompose the series into deseasonalized and seasonal components. It is then possible to generate forecasts for the deseasonalized series by using the methods of Chapter 3 and to combine those values with forecasts for the purely seasonal component.

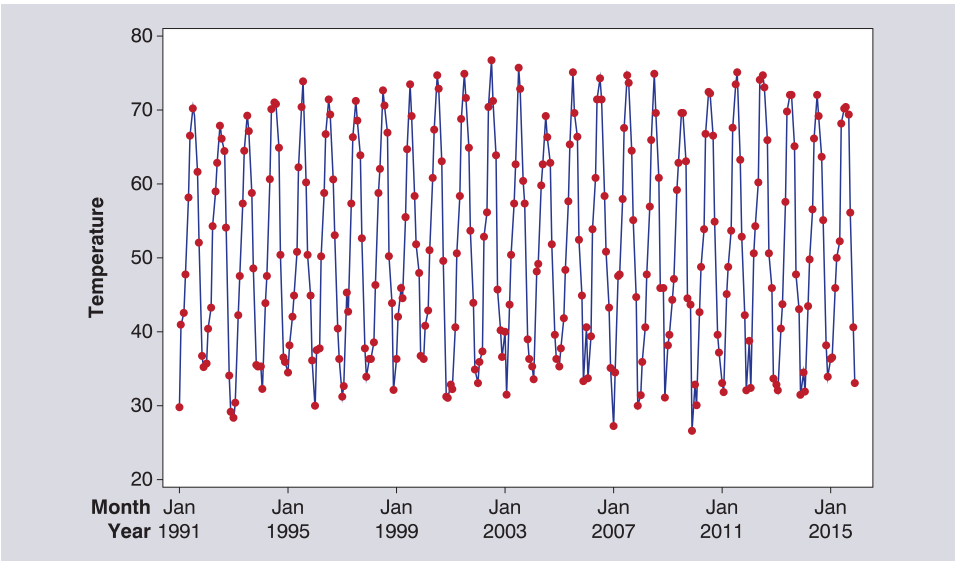
4.2 Forecasting Purely Seasonal Series

Notwithstanding the longer term effects, if any, of global warming, data relating to different aspects of the weather reflect our natural understanding of a strong seasonal pattern with little or no trend. Some other purely seasonal series are specially created, such as the seasonal index for U.S. retail sales shown earlier in Figure 1.3. Cyclical components may also exist, but these are sufficiently long term that they may often be safely ignored when forecasting in the short term.

Example 4.1: Temperatures in Boulder, Colorado

Figure 4.2 is a time plot of monthly average temperatures for Boulder, Colorado, over the period 1991-2015. The data are available in the file *Boulder_2.xlsx*. The seasonal pattern is very clear. Nevertheless, we may also use the seasonal plot, introduced in Chapter 2, as a useful way to detect seasonal patterns when trends also exist. In the present context, we plot temperature against the month of the year, so that there are 12 points on the x-axis; a separate but overlaid plot is then constructed for each year. The results for temperature data for the period 2012–2105 are shown in Figure 4.3; only four years of data were used to avoid cluttering up the graph. The strong seasonal pattern is clear, and the extensive overlapping indicates the absence of any marked trend. An alternative seasonal plot is obtained if we plot temperature against time but color-code each year separately. This version of a seasonal plot is examined in Exercise 4.1.

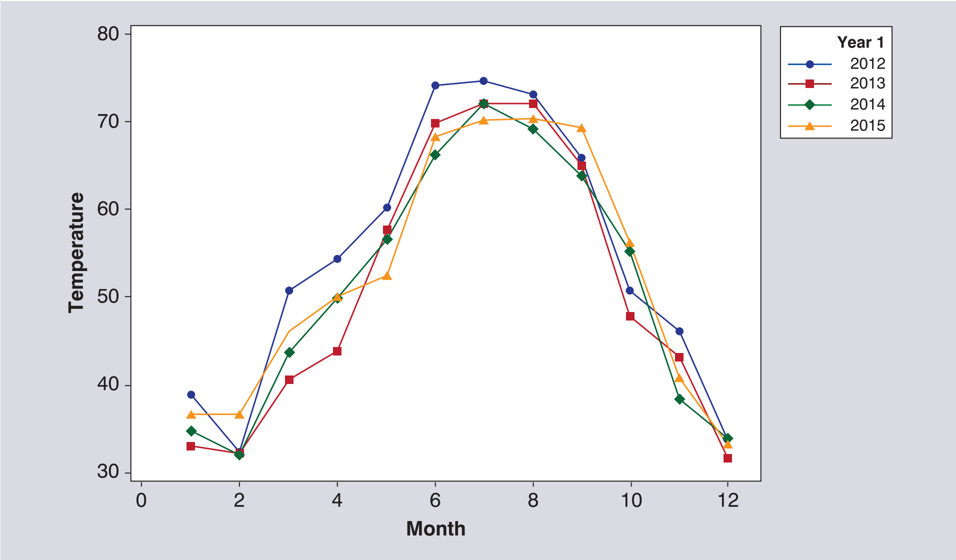
Figure 4.2 Monthly Average Temperatures (In Degrees Fahrenheit) in Boulder, Colorado, 1991–2015



Source: Earth System Research Laboratory, Physical Sciences Division, National Oceanic and Atmospheric Administration (*www.ersl.noaa.gov/psd/boulder/Boulder.mm.html*).

Data: Boulder_2.xlsx

Figure 4.3 Seasonal Plot of Average Temperature (In Degrees Fahrenheit) in Boulder, Colorado *(2012–2015, plotted against month, partitioned by year)*



Source: Earth System Research Laboratory, Physical Sciences Division, National Oceanic and Atmospheric Administration (*www.esrl.noaa.gov/psd/boulder/Boulder.mm.html*).

Data: Boulder_2.xlsx

Suppose we wish to forecast monthly average temperatures 12 months ahead for the year 2015. The simplest approach would be to take an average for each month over the period 1991-2014 and use that as the forecast. The results are shown in Table 4.1. Over the 12 months, the mean absolute error *(MAE)* is 2.28 degrees and the root mean square error (*RMSE*) is 2.64 degrees.

Table 4.1 Actual and Forecast Values for Monthly Temperatures in Boulder, Colorado for 2015

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Temperature	36.5	36.6	46.1	50.1	52.4	68.3	70.3	70.5	69.4	56.2	40.8	33.1
Average	34.2	35.3	43.0	48.8	57.7	66.6	72.7	70.7	63.1	51.4	41.1	33.9
Error	2.3	1.3	3.1	1.3	–5.3	1.7	–2.4	–0.2	6.3	4.8	–0.3	–0.8
Abs Error	2.3	1.3	3.1	1.3	5.3	1.7	2.4	0.2	6.3	4.8	0.3	0.8
Sq error	5.29	1.69	9.61	1.69	28.09	2.89	5.76	0.04	39.69	23.08	0.09	0.64
	ME = 0.98		MAE = 2.48		RMSE = 9.88							

Data: Boulder_2.xlsx

If we now consider forecasts for 2016 after the 2015 data become available, we could recompute the monthly averages. At the other extreme, we could ignore all the past history and use only the values for the most recent year, if we felt that the temperature patterns were changing rapidly. Both options seem rather extreme, given the discussions about smoothing in the previous chapter, so instead we adopt a seasonal smoothing procedure. ■

4.2.1 Purely Seasonal Exponential Smoothing

A seasonal version of simple exponential smoothing (SES) may be written as

New forecast for a given "month" = Old forecast for that "month" + smoothing constant × forecast error.

We use "months" within a "year" to denote the seasonal structure, but in keeping with our earlier comments, the same general idea applies to any purely seasonal series.

PURELY SEASONAL FORECASTS

The forecast for period $t + m$ made m steps earlier, denoted by $F_{t+m}(m)$, is given by the previous m-step-ahead forecast (made at time $t - m$) for period t, adjusted by the error observed for that earlier forecast. Algebraically, the updating relationship may be written as

$$F_{t+m}(m) = F_t(m)+\gamma[Y_t - F_t(m)] = F_t(m)+\gamma e_t \tag{4.4}$$

The smoothing parameter (γ) represents the rate of adjustment to the latest error. The error term in this situation is $e_t = Y_t - F_t(m)$, which leads to the second expression in equation (4.4).

For example, the forecast for July 2016 equals the July 2015 forecast plus a partial error adjustment (based upon the difference between the observed values for July 2015 and its forecast a year earlier on the basis of the data available up to and including July 2014). Because we are focusing purely on the seasonal effects, we do not use data for any months other than previous July values. This prevents the July 2016 forecast from changing (once the July 2015 value has been observed and the forecast updated), despite our having observed August and subsequent values. Another way of looking at this situation is to think in terms of 12 separate series (one for each month) that are updated independently of each other. We update each series as soon as that monthly value is recorded, but we do not use the updated forecast until the next year. This framework is very restrictive and it is usually desirable to include information on recent trends as well. The reason for developing equation (4.4) is that it is a key element in the general framework for forecasting seasonal series, as we shall see in later sections of this chapter.

We could use equation (4.4) to forecast the average temperature but we prefer to keep the analysis within the framework of the Exponential Smoothing Macro (ESM) introduced in the previous chapter. Accordingly, the results in Table 4.1 anticipate later developments and fit a model with both level and seasonal components. The parameters are estimated using data over the period January 1991 to December 2012. The analysis in Table 4.2 considers different forecasts for January 2013 to December 2015. *MAE* and *RMSE* are used to measure forecast accuracy here. In this example, the mean absolute percentage error *(MAPE)* is not sensible, because the variable is not ratio-scaled, as it has no natural origin; consider how the *MAPE* changes if we switch from the Fahrenheit scale to Celsius!

The results are summarized in Table 4.2. The first row represents the error statistics for the estimation sample (with $\alpha = 0.015$, $\gamma = 0.189$), and the remaining rows are the results for the different forecasting methods.

For comparative purposes, we included the forecasts based only on the previous year and those based on the long run average of all years from 1991 – 2012. The results show the better performance of the monthly scheme derived using the ESM. The small value for α indicates that the carry-over from the previous month is relatively slight: the small γ suggests a slowly changing seasonal pattern.

Table 4.2 Summary Error Statistics for Forecasting Average Temperatures in Boulder, Colorado *(Hold-out sample covers from January 2013 to December 2015)*

	ME	MAE	MSE	RMSE
Estimation Sample	0.20	2.63	11.20	3.35
Hold-out Sample				
Model	–0.51	2.12	7.53	2.74
Previous Year	–0.69	3.04	15.67	3.96
Long-run Average	–2.50	3.04	13.26	3.34

Data: Boulder_2.xlsx

4.3 Forecasting Using a Seasonal Decomposition

Most of the series we are likely to encounter contain both trend and seasonal components. Because we observe only the aggregate series and not the components, one approach to forecasting is to decompose the time series into its trend, seasonal, and error components. We could remove the seasonal component and forecast the deseasonalized, or seasonally adjusted, version of the series, using the methods of Chapter 3. We may also forecast the seasonal components and then recombine the two to forecast the overall series.

Our primary interest is in forecasting, and we will approach "seasonal decomposition" from that direction, although decomposition methods are often used in the analysis of macroeconomic series, with particular interest in removing seasonal effects. Over many years, the U.S. Census Bureau has supported decomposition software, the current version of which is known as Census X-13 ARIMA. This program is applied to most major economic series, such as *GDP*, some of which are published only in seasonally adjusted form. We present a brief description of Census X-13 ARIMA in Section 4.5 and a reference to access downloadable versions. Our present focus is upon the fundamental ideas underlying seasonal adjustment. The basic steps in the process are as follows:

1. Generate estimates of the trend component by averaging out the seasonal component.
2. Estimate the seasonal component by removing the trend from the series.
3. Create a deseasonalized series.
4. Forecast the trend and the seasonal pattern separately.
5. Recombine the trend forecast with the seasonal component to produce a forecast for the original series.

DECOMPOSITION

Pure Decomposition. Use all n data points to estimate the fitted value at each time t **(two-sided** decomposition).

Forecasting Decomposition. Use only the data up to and including time t to predict the value at time $t + 1$ and beyond **(one-sided** decomposition).

There are two basic versions of seasonal adjustment. The first version is based upon the purely additive model, originally specified in equation (4.1) as

$$Y_t = T_t + S_t + E_{t,} \tag{4.5}$$

where T denotes the trend term, S designates the seasonal component, and E is the random error (often referred to as the *irregular component* in this context); we have added time subscripts in equation (4.5) for later use. Likewise, the mixed additive-multiplicative scheme given in equation (4.3) is represented by

$$Y_t = T_t S_t + E_t. \tag{4.6}$$

A key element in this approach is the use of moving averages, introduced in Section 3.2.2:

MOVING AVERAGE (for forecasting)

A (simple) *moving average* of order K, denoted by $MA(K)$, is the average of K successive terms in a time series, so the first average is $(Y_1 + Y_2 + \cdots + Y_K)/K$, the second is $(Y_2 + Y_3 + \cdots + Y_{K+1})/K$, and so on. The moving average taken at time t may be denoted by $MA_t(K) = (Y_{t-K+1} + Y_{t-K+2} + \cdots + Y_t)/K$.

The moving average is used to produce an average that is free of seasonal effects. That is, we might average over four consecutive quarters ($K = 4$) or 12 consecutive months ($K = 12$). This average provides an estimate[1] of the trend term, T_t. We use past values of the series to estimate the trend so that, for example, in a quarterly series $K = 4$ and the first trend estimate is $T_5 = (Y_1 + Y_2 + \cdots + Y_4)/4$. We then compute the detrended series Y_t^{DT} as

$$Y_t^{DT} = Y_t - T_t = S_t + E_t. \tag{4.7}$$

We now assume that the seasonal component for each period (such as summer) is stable and estimate that component by averaging Y_t^{DT} over all available observations for that period (e.g., all summers for which data are available). Often, the seasonal components are adjusted to have a zero mean, although this does not affect the forecasts. The seasonal factor for period t is denoted by $S_{(j,t)}$; the subscript indexes the seasons and is included in the parentheses to remind us that there are only 4 (quarterly), 7 (days of the week), or 12 (monthly) such factors. Finally, we produce the deseasonalized series Y_t^{DS} and the error terms:

$$Y_t^{DS} = Y_t - S_{(j,t)}$$

$$E_t = Y_t - T_t - S_{(j,t)}.$$

Now that the seasonal effects have been removed, the deseasonalized series may be forecast with SES or LES, as appropriate, as described in Sections 3.3 and 3.4. The final forecast is then obtained by adding back the seasonal component.

Example 4.2: *(Examples_chapter_4.xlsx)*

The process of additive deseasonalization is illustrated for quarterly data in Table 4.3.

In part A of the table, the third column gives the moving average of terms taken four at a time. To align the forecasts with the actual values, we place the average of the first four terms, $(115 + 90 + 65 + 135)/4 = 101.25$, in the row corresponding to observation 5, and so on. That is, we average the figures for the first year, thereby eliminating seasonal effects, and then use that average as a preliminary forecast of the level for the first period (quarter) of the second year. The fourth column contains the values of the *Detrended series,* computed with equation (4.7); that is, we subtract the initial trend values from the original observations. In part B of the table, we gather together the 12 detrended values (three observations

1 We are working with estimates of the components, not their actual values. This point should be kept in mind, as the distinction is not made in the notation.

for each quarter) and compute their average *(Seasonal Means)*. Next, these quarterly averages are adjusted to have zero mean, yielding the seasonal index values *(Adjusted Means)* with values (20.63, -15.21, -38.54, 33.13). The *Deseasonalized series* is then given in the sixth column of Part A by subtracting the adjusted seasonal component from each observation.

Table 4.3 Additive Deseasonalization and Its Use in Forecasting

Part A: Deseasonalized Series

Obs	Series	MA(4)	Detrended Series	Seasonal Factor	Desea-sonalized Series	Desea-sonalized Forecasts	Series Forecasts	Error	Error Squared	Absolute Error	Absolute Percentage Error
1	115			20.63	94.38						
2	90			–15.21	105.21						
3	65			–38.54	103.54						
4	135			33.13	101.88						
5	130	101.25	28.75	20.63	109.38	109.60	130.22	–0.22	0.05	0.22	0.17
6	95	105.00	–10.00	–15.21	110.21	111.94	96.73	–1.73	2.99	1.73	1.82
7	75	106.25	–31.25	–38.54	113.54	113.28	74.74	0.26	0.07	0.26	0.35
8	150	108.75	41.25	33.13	116.88	115.43	148.56	1.44	2.09	1.44	0.96
9	135	112.50	22.50	20.63	114.38	118.38	139.01	–4.01	16.06	4.01	2.97
10	105	113.75	–8.75	–15.21	120.21	118.30	103.09	1.91	3.64	1.91	1.82
11	85	116.25	–31.25	–38.54	123.54	120.89	82.35	2.65	7.01	2.65	3.11
12	155	118.75	36.25	33.13	121.88	124.43	157.56	–2.56	6.53	2.56	1.65
13	145	120.00	25.00	20.63	124.38	125.39	146.02	–1.02	1.03	1.02	0.70
14	110	122.50	–12.50	–15.21	125.21	126.67	111.46	–1.46	2.13	1.46	1.33
15	85	123.75	–38.75	–38.54	123.54	127.41	88.87	–3.87	14.97	3.87	4.55
16	160	123.75	36.25	33.13	126.88	126.28	159.41	0.59	0.35	0.59	0.37

Part B: Seasonal Calculations

Quarter	Year 1	Year 2	Year 3	Year 4	Seasonal Means	Adjusted Means
Q1		28.75	22.5	25	25.42	20.63
Q2		–10.00	–8.75	–12.5	–10.42	–15.21
Q3		–31.25	–31.25	–38.75	–33.75	–38.54
Q4		41.25	36.25	36.25	37.92	33.13
				Overall	**4.79**	**0**

Part C: Summary Measures

MSE	MAE	MAPE	RMSE
4.74	1.81	1.65	2.18

Data: Examples_chapter_4.xlsx

Finally, the deseasonalized series may be forecast by linear exponential smoothing (LES, described in Section 3.4, because a trend clearly exists). The resulting forecasts are shown in column seven *(Deseasonalized Forecasts)*. The starting values were specified as in Section 3.4.3, and the optimal values of the smoothing parameters turned out to be $\alpha = 0.37$ and $\beta = 0.68$. The seasonal components are then added back in to produce the final forecasts in the eighth column *(Series Forecasts)*. The error analysis then proceeds in the usual way. Note that the values given are *fitted values rather than pure forecasts*, because all the observations were used to construct the seasonal index values and to estimate the smoothing parameters. However, given these estimates, proper forecasts can be constructed for future values. ■

Analogously to the additive model just described, we may construct a mixed additive-multiplicative scheme based upon the model given by equation (4.6). The key difference is that we must divide the original series by the moving average to obtain the detrended series (fourth column in Table 4.3) and then adjust the seasonal indexes to average 1.0 rather than zero. The deseasonalized series (column 6) follows upon dividing the original series by the seasonal indexes. After forecasting this series, we obtain forecasts for the original series (column 8) upon multiplying the deseasonalized forecasts (column 7) by the seasonal indexes. This analysis is left to the reader as Exercise 4.9.

Once the notion of decomposition is recognized, a variety of forecasting methods becomes available. The process of decomposition and the methods for forecasting the deseasonalized and seasonal components must all be determined by the forecaster. We have described two variants of one simple form, but specialized software allows a range of more sophisticated approaches, as we shall see in Section 4.6. In particular, we do not need to assume that the seasonal factors are constant over time.

4.4 Pure Decomposition

As we noted at the beginning of the previous section, time series are sometimes analyzed in order to better understand underlying trends, rather than for forecasting. This orientation is particularly strong with macroeconomic data, for which seasonally adjusted data allow a clearer picture of current developments within the economy. It is therefore worth understanding how such deseasonalized series are developed.

The first key element to recognize is that such seasonal adjustment methods may rely upon both past and future observations, as the focus now is upon understanding rather than pure forecasting. To take an obvious example, if you know the weather on Monday and Wednesday, you have a better understanding of what happened on Tuesday than if you have just Monday's weather. However, Wednesday's weather could not be available for forecasting purposes.

To make this process operational, we use moving averages as before, but now we want to align the average with the current value of the series. Thus, if we are dealing with daily sales figures, we would align the average across Sunday, Monday, ..., Saturday of a given week with Wednesday, the fourth of the seven days. However, a problem arises when the number of periods per season is even (four quarters or 12 months). If we take the average over January–December, the average is a half-period out of phase. That is, the average of the 12 months lands between June and July (the average of the numbers 1 through 12 is 6.5), so the moving-average term would not match up exactly in time with either month. To overcome this difficulty, we take a further average to produce a centered moving average.

CENTERED MOVING AVERAGES (for decomposition)

If K is odd we place $MA_t(K)$ at period t–(K–1)/2. For example the first value of $MA(3)$ would be placed at period 3–(3–1)/2 = 2. With K even, the moving average falls mid-way between periods, i.e. $MA_4(4)$ would be placed at 'period' 2.5 while $MA_5(4)$ placed at 'period' 3.5.

If K is odd, the Centered Moving average, $CMA_t(K)$ is defined as equivalent to $MA_t(K)$ and placed at period t–(K–1)/2. If K is even, $CMA_t(K) = (MA_t(K) + MA_{t+1}(K))/2$ and placed at period t–k/2+1. For example the first $CMA(6)$ is calculated by averaging the first $MA(6)$ being placed at period 6–2.5=3.5 and the second $MA(6)$ at period 7–2.5=4.5. The $CMA(6)$ is then placed at the average of these two times: t = 4.

The effect of using the centered MA is to "lose" $K/2$ observations at the beginning and at the end of the series, if K is even, or $(K–1)/2$ if K is odd. Normally, this does not matter at the beginning of the series, but clearly it is critical at the end of the series, because the most recent observations are usually those of greatest interest. If the aim is only to produce a deseasonalized series, we can estimate the seasonal factors and proceed as in Table 4.3, as we show in Table 4.4. However, if we seek to identify underlying trends, we must find some way to extend the series. Earlier schemes used moving averages that assigned different weights to the last few observations in the series, but current "best practice" is to use a statistical model-based approach. The best-known such technique is Census X-13-ARIMA and is discussed briefly in the next section. A simpler method involves fitting a trend line to estimate the last $K/2$ observations (used, e.g., in the Minitab decomposition routine).

The steps that follow describe the process based upon moving averages for quarterly data, assuming t observations; a similar procedure using 12-term moving averages is followed with monthly data. The description applies to the more commonly used multiplicative scheme; modifications for the additive scheme are shown in square brackets.

1. Calculate the four-term *MA(4)* and then the centered moving average, *CMA(4)*. The first *CMA* term corresponds to period 3 and the last one to period $(t – 2)$; $K = 4$, so we "lose" two values at each end of the series.
2. Divide [subtract] observations 3, ..., $(t – 2)$ by [from] their corresponding *CMA* to obtain a detrended series.
3. Calculate the average value (across years) of the detrended series for each quarter j ($j = 1, 2, 3, 4$) to produce the initial seasonal factors.
4. Standardize the seasonal factors by computing their average and then setting the final seasonal factor equal to the initial value divided by [minus] the overall average.
5. Estimate the error term by dividing the detrended series by the final seasonal factor [subtracting the final seasonal factor from the detrended series].

Note that this process is similar to the forecasting procedure described in the previous section. The key difference is that the moving-average values are now centered on the observations, whereas before they were used (asymmetrically) to predict future values in the forecasting framework.

Example 4.3: Seasonal decomposition — the calculations

(Examples_chapter_4.xlsx)

For purposes of illustration, we revisit the data set considered in Example 4.2 but now apply a multiplicative *Decomposition,* rather than generating forecasts. Following steps 1–5, we arrive at the results shown in Table 4.4. Successive columns of interest in part A show the original series, the moving average, *MA(4),* the centered moving average, *CMA(4),* the *Detrended series,* the *Seasonal factors,* and, finally, the *Deseasonalized series.* The final seasonal factors are calculated in part B of the table and then inserted into part A. For example, the first *CMA(4)* value is $(101.25 + 105.00)/2 = 103.13$, where the original *MA* values may be obtained from Table 4.3. However, it is important to observe that the *CMA* values are now aligned with the corresponding time periods. In Table 4.3, the past moving averages were used to forecast. As in Table 4.3, the seasonal components are considered to be unchanging, and we can produce a deseasonalized series that covers all time periods.

The principal purpose of decomposition is to generate the deseasonalized, or seasonally adjusted, series; such series are commonly used in macroeconomic policy discussions. For example, unemployment is heavily seasonal and the focus of much political debate. Because such issues are of critical importance, the decomposition methods used in practice are much more sophisticated than the simple procedure we have just described. Accordingly, we now provide a brief outline of the most commonly used decomposition method. ■

Table 4.4 Multiplicative Decomposition for Seasonal Adjustment

Part A: Deseasonalized Series

Obs	Series	MA(4)	CMA(4)	Detrended Series	Seasonal Factors	Deseasonalized Series
1	115				1.18	97.65
2	90	101.25			0.88	102.83
3	65	105.00	103.13	0.630	0.67	97.64
4	135	106.25	105.63	1.278	1.28	105.35
5	130	108.75	107.50	1.209	1.18	110.39
6	95	112.50	110.63	0.859	0.88	108.54
7	75	113.75	113.13	0.663	0.67	112.66
8	150	116.25	115.00	1.304	1.28	117.06
9	135	118.75	117.50	1.149	1.18	114.64
10	105	120.00	119.38	0.880	0.88	119.97
11	85	122.50	121.25	0.701	0.67	127.68
12	155	123.75	123.13	1.259	1.28	120.96
13	145	123.75	123.75	1.172	1.18	123.13
14	110	125.00	124.38	0.884	0.88	125.68
15	85				0.67	127.68
16	160				1.28	124.86

Part B: Seasonal Calculations

Quarter	Year 1	Year 2	Year 3	Year 4	Averages	Adjusted
Q1		1.209	1.149	1.172	1.177	1.178
Q2		0.859	0.880	0.884	0.874	0.875
Q3	0.630	0.663	0.701		0.665	0.666
Q4	1.278	1.304	1.259		1.280	1.281
				Overall	**0.999**	**1.000**

Data: Examples_chapter_4.xlsx

4.5* The Census X-13 Decomposition

The original seasonal adjustment program introduced in 1965 by the U.S. Bureau of the Census was known as X-11 (Shiskin, Young, and Musgrave, 1967). The name reflects the many years of improvement and modification (from Method 1 in 1954 and onwards) that elapsed before the method reached a standard form. The X-11 method was among the first such procedures to be computerized, and it became a standard for seasonal adjustment around the world. Its successors, X-12-ARIMA and now X-13-ARIMA continue to use the smoothing operations developed in X-11 but take advantage of the ARIMA modeling framework, which we describe in Chapter 6. The ARIMA model enables us to "fill in the blanks" at the end of the series; the model also allows us to incorporate changing seasonal patterns, as in the forecasting procedures described in Sections 4.6 and 4.7. A full description of X-12-ARIMA[2] appears in Findley, Monsell, Bell, Otto, and Chen (1999) and on

2 A more advanced seasonal adjustment procedure, X-13-ARIMA-SEATS, is now the "gold standard." This new software embraces the modeling approach more comprehensively.

the Census Bureau website. Thus, we will summarize only a few key properties here. The detailed steps are given in appendices in Findley *et al.*

Step 1: The time series is first adjusted to take into account any anomalous observations. (An anomalous observation is an observation that stands out as incompatible with those around it.) This is begun by creating an initial time series model that provides one-step-ahead preliminary forecasts. Anomalous observations can then be identified and appropriate adjustments made to the data. For example, suppose production is interrupted by a strike or severe weather but returns to normal once the event is over. If no adjustment is made, the anomalous observation will distort future estimates of the trend and seasonal factors and so give misleading values for the seasonally adjusted series. The model is also used to extend the series for the trend and seasonal components so that standard moving averages can be applied throughout the series.

Step 2: The multiplicative (or, more rarely, the additive) version of seasonal adjustment is applied to the series, essentially as described in the previous section. However, rather than assume fixed seasonals, preliminary seasonal factors are estimated by taking a moving average of the initial seasonal estimates for each quarter or month.

Step 3: An initial set of seasonal adjustments is applied as in Table 4.4, but using more complex moving averages.

Step 4: The series is extended (using ARIMA; see Chapter 6) so that moving averages can be computed for each time period up to the end of the series, using the seasonally adjusted data. This leads to a revised estimate of the trend and seasonal factors.

Step 5: A final round of detrending is applied, and the error term is then estimated by dividing the observation by the trend and seasonal factors, to complete the decomposition.

DISCUSSION QUESTION: *In retail store sales of a product, what type of events might introduce anomalies into the observed sales figures?*

For readers wishing to conduct "industrial strength" seasonal adjustments, the X-13-ARIMA program is available free of charge from the Census Bureau; the URL is *www.census.gov/srd/www/x13as*. In both the United States and Europe, a heavily used free program is TRAMO-SEATS; the URL is *www.bde.es/bds/en/secciones/servicios/software/* under the Statistics and Econometrics heading.

4.6 The Holt-Winters Seasonal Smoothing Methods

We now examine exponential smoothing methods for forecasting series that include both trends and seasonal patterns. We focus upon two particular schemes, known as the Holt-Winters' additive and multiplicative schemes, respectively, because they were first developed by Holt (2004, original published in 1957) and Winters (1960).

4.6.1 The Additive Holt-Winters Method

We consider a forecast function that combines, by addition, the level, trend, and seasonal components. That is, we add a seasonal component to the linear exponential smoothing

(LES) scheme developed in Section 3.4. As before, we denote the level of the series by L_t, the trend by T_t, and the seasonal factor by S_t. When there are m seasons (e.g., $m = 12$ for monthly data), the forecast for one period ahead may be written as

$$F_{t+1}(1) = F_{t+1} = L_t + T_t + S_{t+1-m}.$$

For example, to forecast the sales in July 2016, we combine the level and trend elements derived up to and including June 2016 with the seasonal component from July 2015. If we wish to forecast h periods ahead, we use the more general expression

$$F_{t+h}(h) = L_t + hT_t + S_{t+h-m}. \tag{4.8}$$

Equation (4.8) takes the current local level of the series modified by h times the current trend and selects the appropriate monthly seasonal corresponding to period $(t + h)$ to estimate the h-step ahead forecast. The notation is designed to maintain the connection with exponential smoothing as developed in the previous chapter. If $h > m$, we select the appropriate seasonal index, cycling through every m periods; for example, the July index is used for future July forecasts, no matter how many years ahead.

The next step is to update these components. The new information we have is the latest observation or, equivalently, the latest error term, written as

$$e_t = Y_t - L_{t-1} - T_{t-1} - S_{t-m} = Y_t - F_t(1). \tag{4.9}$$

The updating expressions for the level and trend are the same as those used for LES in Section 3.4.2:

$$\begin{aligned} L_t &= L_{t-1} + T_{t-1} + \alpha e_t \\ T_t &= T_{t-1} + \alpha\beta e_t. \end{aligned} \tag{4.10}$$

To complete the process, we must also update the seasonal component,[3] relative to the same period in the previous "year." That is, the seasonal update may be written as

$$S_t = S_{t-m} + \gamma e_t. \tag{4.11}$$

This expression is exactly that appearing in equation (4.4), save that we have changed the variable of interest from F to S to recognize that the seasonal component is now one part of the forecast. Each seasonal component will be updated just once per "year," when the corresponding "monthly" value has been recorded. So, if we were dealing with 12 months in a year, we would have 12 such components and each would be updated once every 12 time periods. From equations (4.10) and (4.11), we observe that each component is updated with the use of a partial adjustment for the error. As before, these equations are known as the *error correction* form. The more usual, but also more cumbersome, way of expressing these relationships is to substitute for e_t, using equation (4.9) to arrive at equations (4.12):

$$\begin{aligned} L_t &= L_{t-1} + T_{t-1} + \alpha(Y_t - L_{t-1} - T_{t-1} - S_{t-m}) \\ T_t &= T_{t-1} + \beta(L_t - L_{t-1} - T_{t-1}) \\ S_t &= S_{t-m} + \gamma(Y_t - L_{t-1} - T_{t-1} - S_{t-m}). \end{aligned} \tag{4.12}$$

Demonstrating the equivalence between the two versions is left as an end-of-chapter exercise. At this stage, we should note that the Holt-Winters equations are often written in a slightly different form, with L_t and T_t replacing L_{t-1} and T_{t-1}, which appears in the seasonal

3 An alternate formulation for the seasonal components uses trigonometric functions. A description is provided in Durbin and Koopman (2012, p. 40). A thorough discussion on seasonality is given by Ghysels and Osborne (2001) with various different representations discussed in Chapters 2 and 3.

updating formulas. For the linear case, the two versions produce identical forecasts, although the parameter γ has a different value. We prefer the present formulation because it leads directly to an underlying model (see Section 5.3.2) whereas the more common version does not. The Holt-Winters method allows for a number of special cases, such as the following:

- Fixed seasonal pattern: $\gamma = 0$ (no seasonal updating)
- No seasonal pattern: $\gamma = 0$ and all initial S values are set equal to zero
- Fixed trend: $\beta = 0$
- Zero trend: $\beta = 0$ and $T_0 = 0$
- All fixed components: $\alpha = \beta = \gamma = 0$, a linear trend with fixed seasonal effects.

4.6.2* Starting Values

When the time series is seasonal with m periods, we require a total of $(m + 2)$ starting values: one each for the level and trend, as before, and one for each seasonal component. We again refer to "months" within a "year" for convenience, while recognizing that broader definitions of seasonality are also included. The Exponential Smoothing Macro (ESM) provided on the website uses starting values specified in the way we shall show here, although we should stress that a wide variety of heuristic procedures exists. We use data for the first two years, written as $Y_1, Y_2, \ldots, Y_{2m}$, to initialize the level at period m (L_m), the trend (T_m), and the first m seasonals, $S_1, S_2, \ldots S_m$. The starting values may be written as

$$
\begin{aligned}
L_m &= \frac{1}{m}(Y_1 + Y_2 + \cdots + Y_m) + \frac{(m-1)}{2}T_m \\
T_m &= \frac{1}{m}\left[\frac{1}{m}(Y_{m+1} + Y_{m+2} + \cdots + Y_{2m}) - \frac{1}{m}(Y_1 + Y_2 + \cdots + Y_m)\right] \qquad (4.13) \\
S_{(i)} &= \frac{1}{2m}(Y_i + Y_{i+m}) - \frac{1}{2m}(Y_1 + Y_2 + \cdots + Y_{2m}),\ i = 1, 2, \ldots, m.
\end{aligned}
$$

The initial value for the level is defined as the average across the first year, plus an adjustment using the estimated initial trend to allow for the difference between the middle of the year and its end. Thus, for monthly data, $m = 12$, and we add 5.5 trend increments to move from midyear to December. The trend estimate is the "monthly" average increase for the second year over the first year, and the seasonal estimates are the average value for the month in question taken over years 1 and 2 minus the average for all months taken over the first two years.

It is not too much of an exaggeration to suggest that no two statistical software packages define the start-up values in quite the same way. The net effects will be minor, provided that the series is of reasonable length and the smoothing coefficients are not too small. If greater precision is required, the starting values should be estimated from the full estimation sample; the more technically inclined reader should consult Hyndman, Koehler, Ord, and Snyder (2008, Chapter 5) for details and the software referenced in online Appendix B.

■ Example 4.4: Additive Holt-Winters forecasting — the calculations

(Examples_chapter_4.xlsx)

An illustration of the basic calculations is shown in Table 4.5. We suppose that the time series is quarterly. The first part of the table provides a schematic layout setting up the calculations. Note that these values are aligned in the table in such a way that the forecast

Table 4.5 Illustrative Calculations for the Additive Holt-Winters Scheme
(The numbers in **bold** *indicate the starting values.)*

A: Illustration of Spreadsheet Layout for Forecasting

Series	Level	Trend	Seasonal	Forecast	Error
Y1					
Y2					
Y3					
Y4					
Y5	**L4**	**B4**	**S1**		
Y6	L5	**B5**	**S2**		
Y7	L6	**B6**	**S3**		
Y8	L7	**B7**	**S4**	F8 = L7 + B7 + S4	Y8 – F8
Y9	L8	B8	S5	F9 = L8 + B8 + S5	Y9 – F9
Y10	L9	B9	S6	F10 = L9 + B9 + S6	Y10 – F10
Y11	L10	B10	S7	F11 = L10 + B10 + S7	Y11 – F11

B: Detailed Calculations (Alpha =0.20, Beta = 0.10, Gamma = 0.30)

Period	Series	Level	Slope	Seasonal	Forecast	Error
1	115			15.63		
2	90			–14.38		
3	65			–36.88		
4	135	105.47	2.81	35.63		
5	130	109.50	2.81	15.63	123.91	6.09
6	95	111.73	2.81	–14.38	97.94	–2.94
7	75	114.01	2.81	–36.88	77.66	–2.66
8	150	116.33	2.81	35.63	152.44	–2.44
9	135	119.19	2.82	15.70	134.77	0.23
10	105	121.48	2.76	–15.16	107.63	–2.63
11	85	123.77	2.72	–37.59	87.37	–2.37
12	155	125.06	2.57	33.49	162.11	–7.11
13	145	127.97	2.61	16.19	143.33	1.67
14	110	129.50	2.50	–16.79	115.42	–5.42
15	85	130.12	2.31	–40.41	94.41	–9.41
16	160	131.24	2.19	31.72	165.92	–5.92

Data: Examples_chapter_4.xlsx

is obtained by adding elements in the same row. Thus, the forecast for period 8 is based upon L_4, T_4, and S_{h1} all located in the row corresponding to Y_5. Part B of the table provides a numerical example. We take the smoothing parameters to be $\alpha = 0.2$, $\beta = 0.1$, and $\gamma = 0.3$. The totals for the first two years are 405 and 450 respectively, so the two-year "monthly" average is 106.875. Equations (4.13) provide the starting values:

$$T_4 = \frac{1}{4}\left[\frac{450}{4} - \frac{405}{4}\right] = 2.81$$

$$L_4 = \frac{405}{4} + 1.5 \times 2.81 = 105.47$$

$$S_1 = 0.5 \times (115 + 130) - 106.875 = 15.625$$

$$S_2 = 0.5 \times (90 + 95) - 106.875 = -14.375$$

$$S_3 = 0.5 \times (65 + 75) - 106.875 = -36.875$$

$$S_4 = 0.5 \times (135 + 150) - 106.875 = 35.625.$$

These values are shown in boldface in the table. The updating begins after the first two years. The subsequent values in the table are then calculated with the use of equations (4.8–4.11). Note that although the initial seasonal values are constrained to sum to zero, this constraint is not retained. Fortunately, this does not affect forecasting performance. The seasonal values can be adjusted to sum to zero and an exactly compensating adjustment made to the level; see Archibald and Koehler (2003) for details. ■

The table illustrates the general format used in the ESM spreadsheet and available on the book's website. The user may specify:

- The components of the model: seasonal and/or trend may be included along with the level
- Additive or multiplicative seasonality, or logarithmic transform
- Transformations: either logarithmic or Box-Cox
- The observations to be used in the estimation sample and the hold-out sample

Once we have made the selections, we can identify those values of the smoothing parameters that minimize the fitting criterion *(RMSE),* using Solver in Excel. In all cases, the first two "years" of data are used to establish starting values. Note that the HW models have three smoothing parameters and 6 (quarterly) or 14 (monthly) starting values: level, trend, and seasonals. Solver is not effective in optimizing over so many unknowns, so we always use heuristic starting values.

More sophisticated statistical software enables a search to determine optimal starting values for the states (level, trend and seasonals) as well as for the smoothing parameters. For example, see the R package available on CRAN-R at *https://cran.r-project.org/package=forecast* and described in our on-line supplement. It is used in section 4.8 by way of illustration. The macros provided here serve to illustrate the general principles and are not the final word in computational processes for forecasting! As noted previously, the current ad hoc starting values typically work well unless the series is short and some of the smoothing parameters are close to zero.

The next example relates to quarterly data and uses the ESM. The procedures for monthly data follow essentially the same pattern.

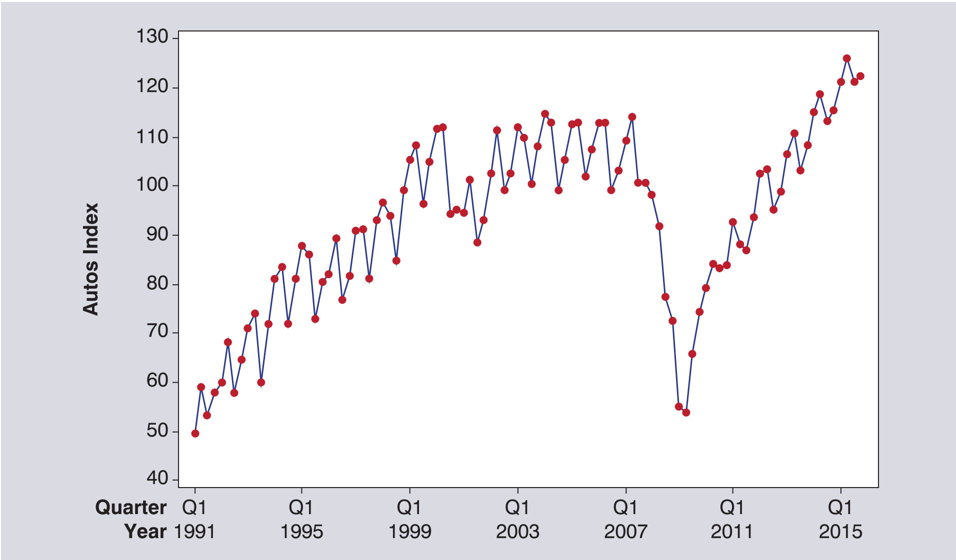
■ Example 4.5: Additive Holt-Winters forecasting *(Autos_index.xlsx)*

Figure 4.4 shows a plot of the total quarterly production of motor vehicles and component parts in the United States from the first quarter of 1991 (1991Q1) through the last quarter of 2015 (2015Q4), recorded as an index with the year 2012 = 100. The mean level clearly changes over time and a seasonal pattern is evident.

An additive Holt-Winters forecasting scheme was fitted using data over the period 1991Q1 to 2009Q4. The performance of the system was then checked by means of the one-step-ahead forecasts over the 24 observations from 2010Q1 to 2015Q4. The results are summarized in panel A of Table 4.6 and the components are plotted in Figure 4.5. It is interesting to note that dropping the trend component has virtually no effect upon forecast performance.

The series reflects smaller seasonal components in later years, as the drop-off in production in the third quarter gradually disappears. In Minicase 4.2, you are encouraged to explore the effects of using smaller estimation samples; recall that a longer series is preferable only if the underlying properties of the series do not change. ■

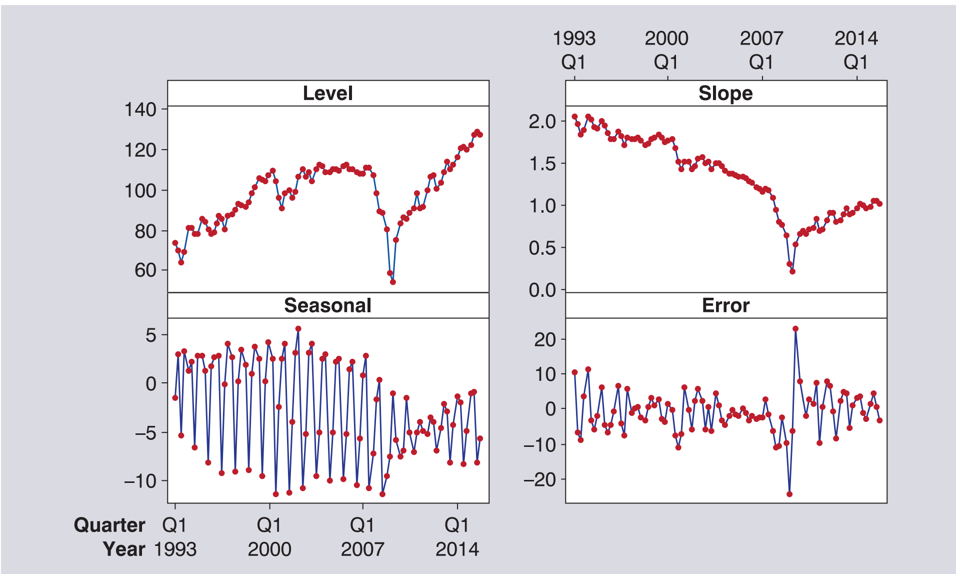
Figure 4.4 Volume of Motor Vehicles and Parts Manufacturing: Quarterly Index, 1991Q1 to 2015Q4 (2012 = 100)



Source: U.S. Federal Reserve Bank of St. Louis (FRED) (*https://fred.stlouisfed.org*)

Data: Autos_index.xlsx

Figure 4.5 Components of the Additive Holt-Winters Scheme for the Autos_index Series



Data: Autos_index.xlsx

Table 4.6 Comparative Performance of Holt-Winters Forecasting Methods *(Autos_Index, 2010Q1–2015Q4)*

	RMSE	MAE	MAPE	α	β	γ
PANEL A: Additive with Trend						
Estimation sample	6.17	4.54	5.60	1.00	0.03	0.33
Hold-out sample	5.93	4.99	4.98			
No trend						
Estimation sample	6.11	4.31	5.32	1.00		0.33
Hold-out sample	5.90	5.01	4.98			
PANEL B: Multiplicative with Trend						
Estimation sample	5.90	4.53	5.50	1.00	0.04	0.32
Hold-out sample	7.11	6.19	6.15			
No trend						
Estimation sample	5.80	4.28	5.22	1.00		0.33
Hold-out sample	6.79	5.96	5.90			
PANEL C: Logarithmic with Trend						
Estimation sample	6.05	4.69	5.70	0.98	0.06	0.32
Hold-out sample	7.12	6.09	6.05			
No trend						
Estimation sample	5.85	4.32	5.27	1.00		0.33
Hold-out sample	6.94	6.09	6.02			

Data: Autos_index.xlsx

4.7 The Multiplicative Holt-Winters Method

An examination of the time series plot sometimes reveals seasonal fluctuations that increase with the level of the series; that is, the movements are proportional rather than absolute. In these circumstances, it makes sense to combine an additive trend with a multiplicative seasonal factor to give as the forecast function

$$F_{t+1}(1) = F_{t+1} = (L_t + T_t)S_{t+1-m}. \tag{4.14}$$

Thus, the one-step-ahead error for period t becomes

$$e_t = Y_t - (L_{t-1} + T_{t-1})S_{t-m}. \tag{4.15}$$

In both equations, the seasonal component is now a multiplicative rather than an additive term. The *error correction* form of the updating equations is then

$$\begin{aligned} L_t &= L_{t-1} + T_{t-1} + \alpha(e_t/S_{t-m}) \\ T_t &= T_{t-1} + \alpha\beta(e_t/S_{t-m}) \\ S_t &= S_{t-m} + \gamma[e_t/(L_{t-1} + T_{t-1})]. \end{aligned} \tag{4.16}$$

The updating relationships involve removing the seasonal element in the error (now by division rather than by subtraction as in the additive scheme) before adjusting the level and trend. Similarly, the trend element is removed from the error before updating the seasonal component. As with the additive case, the seasonals are adjusted only once per "year," so we need $(m + 2)$ starting values in all. Exercise 4.11 provides the version of equation (4.16) that uses the observations and components in place of the error correction form. The ESM may be used to forecast using the multiplicative Holt-Winters scheme; it includes the alternative models with and without a trend term.

4.7.1* Starting Values

As with the additive case, we use data for the first two years, written as $Y_1, Y_2, \ldots, Y_{2m}$. The starting values for the level and trend are unchanged:

$$L_m = \frac{1}{m}(Y_1 + Y_2 + \cdots + Y_m) + \frac{(m-1)}{2}T_m$$

$$T_m = \frac{1}{m}\left[\frac{1}{m}(Y_{m+1} + Y_{m+2} + \cdots + Y_{2m}) - \frac{1}{m}(Y_1 + Y_2 + \cdots + Y_m)\right].$$

The seasonal values are created from the first "year's" values by division rather than subtraction:

$$S_{(i)} = \frac{1}{2} \times (Y_i + Y_{i+m}) \Big/ \left[\frac{1}{2m}(Y_1 + Y_2 + \cdots + Y_{2m})\right], \; i = 1, 2, \ldots, m. \qquad (4.17)$$

■ Example 4.6: Multiplicative Holt-Winters *(Autos_index.xlsx)*

For purposes of comparison, we now examine the motor vehicles series, using the multiplicative Holt-Winters scheme. The results, based upon minimizing the *MSE,* are shown in panel B of Table 4.6. The estimated smoothing parameters are similar. Plots of the components are omitted, but they again show a reduction in the magnitude of the seasonal components from the mid-1970s on. Overall, the multiplicative scheme has a better fit for the estimation sample but produces inferior results for the hold-out sample. The change is probably due to the distorting effects of the "Great Recession" in 2008-09. The results also serve to illustrate the value of the hold-out analysis. Once again, the trend component does not provide an improvement. The additive version might be a safer bet for forecasting.

Note again that the seasonal coefficients start out with an average of 1.0 but later values do not meet this constraint. Archibald and Koehler (2003) show how to adjust the seasonals to retain the average of 1.0; again, their adjustment does not affect the forecasts but makes interpretation easier. ■

4.7.2 Purely Multiplicative Schemes

A third variation is to allow both trend and seasonal components to be multiplicative. In these circumstances, we may make a (natural) logarithmic transformation, as described in Section 2.6.2, and apply the additive Holt-Winters scheme to the transformed data. The optimization step may then use either the *MSE* of the original data or the *MSE* of the transformed data. The ESM on the website uses the transformed data for estimation, but the resulting error measures are reported for the original series because it is the forecasting performance for those data that is of interest. The forecasts for the original series are found by application of the reverse transformation, or exponentiation. That is, if Z denotes the transformed variable, we make the transformation $Z = \ln(Y)$ and then, after generating the forecasts, transform the forecast F_Z back, using the inverse transform $F_Y = \exp(F_Z)$ to obtain forecasts of the original series. The subscripts have been added to the forecast variables to distinguish the transformed and original series.

■ Example 4.7: Purely multiplicative seasonality *(Autos_index.xlsx)*

Application of this scheme to the motor vehicles series yields the results shown in panel C of Table 4.6. The forecasting performance is similar to that of the multiplicative method. Plots of the components are omitted, but they again show a reduction in the magnitude of the seasonal components from 2007 on. Once again, the trend component does not provide an improvement. ■

4.8* Calculations Using R

From time to time we have mentioned that different programs will produce different outputs, even leading to different model choices on occasion. We now explore such differences in more detail by examining the data analysis for the automobiles data examined in the last two sections, using the R package available on CRAN-R at *https://cran.r-project.org/packages=forecast* and described in our on-line supplement. The original results were summarized in Table 4.6 and we provide comparable results in Tables 4.7 and 4.8.

First, we compare Tables 4.6 and 4.7. Both sets of results are based upon the same parameter values so the differences that arise are due to the different starting values used. In general, the R programs use more sophisticated statistical procedures to establish the starting values, so somewhat better results are to be expected.

Table 4.7 Comparative Performance of Holt-Winters Forecasting Methods, Using R Programs with Parameters Estimated by the ESM *(Autos_Index, 2010Q1–2015Q4)*

	RMSE	MAE	MAPE	α	β	γ
PANEL A: Additive with Trend						
Estimation sample	6.06	5.27	5.92	1.00	0.03	0.33
Hold-out sample	5.28	4.39	4.46			
No trend						
Estimation sample	6.13	5.53	6.22	1.00		0.33
Hold-out sample	4.95	4.08	4.13			
PANEL B: Multiplicative with Trend						
Estimation sample	9.16	7.65	8.25	1.00	0.04	0.32
Hold-out sample	8.04	7.10	7.00			
No trend						
Estimation sample	9.18	7.93	8.57	1.00		0.33
Hold-out sample	7.83	6.85	6.72			
PANEL C: Logarithmic with Trend						
Estimation sample	9.06	7.67	8.38	0.98	0.06	0.32
Hold-out sample	6.95	6.15	6.09			
No trend						
Estimation sample	9.41	8.45	9.23	1.00		0.33
Hold-out sample	6.99	5.92	5.82			

Data: Autos_index.xlsx

The results in Table 4.7 are mixed. We are using "R-based" starting values with "ESM-based" parameter estimates, which leads to inferior results for the estimation samples of the multiplicative and log-transform schemes. As might be expected, the results for the hold-out samples are much closer to each other, because the impact of the initial conditions is gradually reduced. Both sets of results point to the additive scheme without trend, although the margin of superiority is greater in Table 4.7.

We now compare Table 4.6 and 4.8. The effect of more careful estimation of the starting values is clearly seen in smaller error measures across the board. The superior performance of models developed using the R programs is evident and carries over to the hold-out samples. The results in Table 4.8 suggest that the best model is now multiplicative without trend. It is also noteworthy that the estimated trend and seasonal parameters are close to zero, but the reasons are rather different. The trend coefficient is small because the trend term does not add value to the analysis. By contrast, the seasonal pattern is important but it is also very stable, so that the updating parameter is small.

The results in Tables 4.7 and 4.8 show the importance of good estimates for the starting values, particularly for models with multiplicative components.

Table 4.8 Comparative Performance of Holt-Winters Forecasting Methods with Parameter Estimation Based Upon R Programs *(Autos_Index, 2010Q1–2015Q4)*

	RMSE	MAE	MAPE	α	β	γ
PANEL A: Additive with Trend						
Estimation sample	5.20	3.60	4.50	1.00	0.00*	0.00*
Hold-out sample	4.42	3.67	3.73			
No trend						
Estimation sample	5.49	3.82	4.84	1.00		0.00*
Hold-out sample	4.77	4.01	4.07			
PANEL B: Multiplicative with Trend						
Estimation sample	5.29	3.80	4.71	0.81	0.00*	0.19
Hold-out sample	4.51	3.78	3.77			
No trend						
Estimation sample	5.28	3.81	4.72	0.88		0.12
Hold-out sample	3.87	3.27	3.31			
PANEL C: Logarithmic with Trend						
Estimation sample	5.19	3.84	4.72	1.00	0.01	0.00*
Hold-out sample	5.15	4.25	4.24			
No trend						
Estimation sample	5.12	3.80	4.64	1.00		0.00*
Hold-out sample	5.42	4.54	4.53			

* Coefficient is non-zero but less than 0.005

Data: Autos_index.xlsx

4.9 Weekly Data

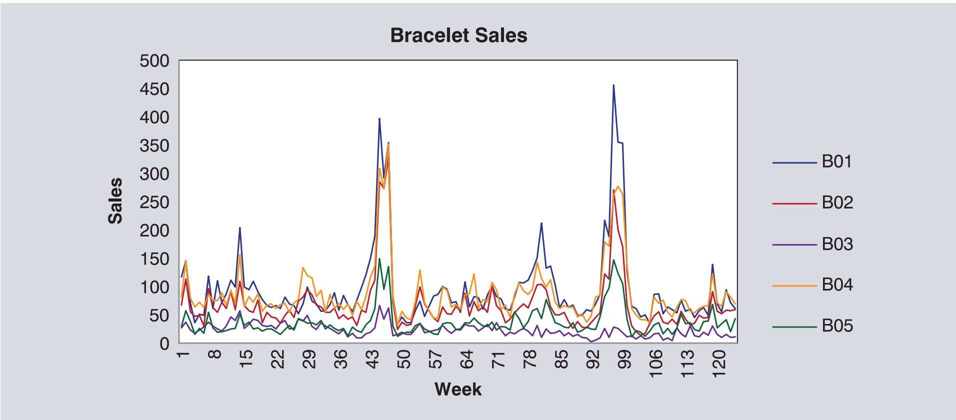
Business time series based upon weekly observations have four distinctive features:

- They are often short, in the sense that each week of the year may be represented only a few times (e.g., two years of data provides 104 observations, but each week appears only twice).
- Key features of the series may fall into different weeks from year to year (e.g., Easter, Thanksgiving).
- There are often a large number of series to be forecast (e.g., stock-keeping units (SKUs) for a retail company or a manufacturer).

- Totals of forecasts for individual items should match up to aggregate forecasts for product groups.

We defer a detailed discussion of forecasting support systems to Chapter 12, but address some of the technical issues here. Figure 4.6 shows the sales figures for five bracelets of a particular line of bracelets sold by a leading costume jewelry company. These data are described and examined in detail in Minicase 12.2 in Chapter 12; the observations run from week 5 in year 1 through week 24 in year 3. The Christmas sales peaks are easy to recognize, and the seasonal patterns are similar for all five products even though the sales volumes differ.

Figure 4.6 Weekly Sales Figures for Five Bracelets *(Year 1, Week 5 – Year 3, Week 24)*



Data: Bracelet_5.xlsx

Clearly, if we are to forecast such series, we must take the seasonal patterns into account. A single series will not provide very reliable information about seasonal patterns, but a well-defined cluster of these series (e.g., all bracelets with Christmas designs) will provide a strong indication of such shifts in sales. Further, we can see that although the sales volumes may change, the seasonal patterns are remarkably consistent. These two features are combined to provide users with a framework for dealing with weekly series; the ideas may apply equally well in other cases where there are a large number of "seasons" but relatively short series.

We adapt the multiplicative Holt-Winters scheme to allow for fixed seasonals, but we calculate the seasonals from the aggregate sales data. This approach has the effect of reducing the randomness in our estimates compared with what would happen if we had just used the individual SKU data. We can drop the trend term because it appears from the data that simple exponential smoothing will suffice once we have taken care of seasonality. Again, it is possible to retain the trend component, but for many series SES will suffice. Particularly in short-term forecasting at an SKU level, the trend has little relevance. Thus, we arrive at the scheme

$$F_{t+1}(1) = L_t S_{t+1-m}$$

$$L_t = L_{t-1} + \alpha(e_t / S_{t-m})$$

$$S_t = S_{t-m}.$$

Further, because the seasonal factors are fixed and prespecified, we may produce a deseasonalized series (Y_t/S_{t-m}), forecast this series by means of SES, and then multiply the results by the seasonal factors to obtain the final forecasts. The detailed calculations are illustrated on hypothetical data in Table 4.9.

Table 4.9 Adjusting for Weekly Seasonality

A	B	C	D	E	F	G	H	I	J	K	L
Week	SKU1 Sales	Group Sales (previous year)	Seasonal Factor	Adjusted SKU1 Sales	One-step-ahead forecasts	SKU1 forecasts	Errors	Absolute Errors	Forecast (no adjust-ment)	Errors	Absolute errors
1	54	557	0.87	62.2	62.1	53.9	0.1	0.1	54	0.0	0.0
2	29	533	0.83	34.9	62.1	51.6	–22.6	22.6	54	–25.0	25.0
3	51	539	0.84	60.7	55.4	46.6	4.4	4.4	46.5	4.5	4.5
4	52	700	1.09	47.6	56.7	61.9	–9.9	9.9	47.9	4.1	4.1
5	113	1100	1.72	65.9	53.9	92.5	20.5	20.5	49.2	63.8	63.8
6	58	672	1.05	55.3	59.9	62.8	–4.8	4.8	68.2	–10.2	10.2
7	79	602	0.94	84.1	58.6	55.0	24.0	24.0	65.3	13.7	13.7
8	48	603	0.94	51.0	65.8	61.9	–13.9	13.9	69.4	–21.4	21.4
9	76	598	0.93	81.5	61.6	57.4	18.6	18.6	63	13.0	13.0
10	32	508	0.79	40.4	67.1	53.2	–21.2	21.2	66.8	–34.8	34.8
Averages		**641.2**	**1.00**	**58.4**	**60.3**	**59.7**	**–0.5**	**14.0**	**58.4**	**0.77**	**19.05**

Data: Table_4_9.xlsx

Example 4.8: Forecasting weekly sales

We consider a single-item *SKU1*, which is part of a larger product group. The sales for *SKU1* are given in column B of Table 4.9, and last year's sales for the entire product group are given in column C. We assume a seasonality of ten weeks. The seasonal factors in column D are created by dividing the total product sales in column C by the weekly average of 609. The adjusted *SKU1* sales are then given by the SKU1 Sales divided by the seasonal factor. Next, these adjusted sales are forecast one step ahead with SES. The one-step-ahead forecasts in column F are then multiplied by the seasonal factors to produce the *SKU1* forecasts in column G, with the corresponding errors shown in column H. The rest of the table provides forecasts based upon SES without any seasonal adjustment and compares the forecast errors.

The data were constructed to show peak sales in week 5; as it happens, both forecasts undershoot, but the error is much more serious when the seasonal pattern is not taken into account. The week-5 error would increase from 20 to 63 units; the overall *MAE* is 15.3 for the adjusted forecasts, as against 21.0 when the seasonal effects are ignored. Further, if seasonal effects appear in different weeks, as in Easter, it is a simple matter to shuffle the seasonal factors accordingly. ■

One final point is worth noting. If we use this approach on a group of individual (sales) series with the same seasonal coefficients and a common value for α, the forecasts for the total group sales will be the same whether we sum the forecasts for the individual items (the so-called bottom-up approach) or whether we forecast the total sales directly. If only the total sales are forecast, it is then necessary to produce forecasts for individual items by partitioning the total (the so-called top-down approach). These ideas are explored further in Chapter 12.

With weekly data, the extension of Holt-Winters exponential smoothing is not straightforward because of initialization. Software has become available in R from *https://cran.r-project.org/package=smooth*.

4.9.1 Multiple Seasonalities

At various places in this chapter we have mentioned weekly, monthly and quarterly seasonal patterns. In forecasting electricity consumption and similar phenomena, time-of-day patterns are also extremely important. Such multiple seasonalities are becoming increasingly common as forecasting methods are extended to cover new situations with more frequently collected data.

We do not have space in this volume to discuss details, save to note that the general principles remain the same: we formulate an equation to describe each seasonal component such as hours within days, days within weeks and weeks within years. More detailed accounts are available in Taylor (2003) and Gould *et al.* (2008).

4.10 Prediction Intervals

The structure of prediction intervals for series with a seasonal component is essentially the same as that of the nonseasonal series discussed in Section 3.8. As in that discussion, we restrict attention, for the time being, to the construction of prediction intervals for one-step-ahead forecasts. (Intervals for forecasts made two or more periods ahead are considered in Chapter 5.) As in Section 2.8, an approximate interval is given by

$$\text{Forecast} \pm z_{\alpha/2} \times (RMSE).$$

Example 4.9: Prediction intervals for automobile production, one quarter ahead

As in Example 4.3, we use the Additive Holt-Winters scheme without the trend term, as this appeared to be the best option based upon the results in Table 4.6. Having made the selection of the method, we use the data through 2014Q4 and then forecast 2015 for one to four quarters ahead. Table 4.10 provides the forecasts and the 95 percent prediction intervals. The *RMSE* was estimated from the fitting sample to be 6.04 and the smoothing parameters are $\alpha = 0.93$, $\gamma = 0.26$.

Two features stand out in the table. First, the intervals appear to be very wide compared with the differences between the actual and forecast values. This discrepancy reflects the greater stability (or lower variability) of the numbers in recent years and the large errors in 2008 and 2009 that inflated the estimate of *RMSE.* At the same time, sales can be quite volatile and we should not be lulled into a false sense of security by a period of relative calm. Second, the forecast for the third quarter is well below that for the other quarters, although little variation is seen among the actual quarterly figures. As we observed earlier, the seasonal pattern has changed somewhat over time and the forecasting procedure may not fully reflect this shift.

Table 4.10 One Step-Ahead Prediction Intervals for Quarterly Autos_index for 2015 [$\alpha = 0.93$, $\gamma = 0.26$]

Quarter	Actual	Forecast	Prediction Interval
2015Q1	121.2	118.7	(106.9, 130.6)
2015Q2	126.1	117.8	(101.7, 134.0)
2015Q3	121.2	111.3	(91.8, 130.9)
2015Q4	122.5	115.2	(92.7, 137.6)

Data: Autos_index.xlsx

Empirical prediction intervals may also be calculated by following the approach given in Section 2.8.2. The details are left as Exercise 4.12. Note that when we use the logarithmic transform, the prediction interval should be calculated for the transformed values and only then converted back to the original series. ■

4.11 Principles for Seasonal Methods

The six principles described in Chapter 3 are equally valid for seasonal modeling. In addition,

[4.1] Carefully examine the structure of the seasonal pattern.
The seasonal component may be additive or multiplicative, and it may also have certain biases that require systematic adjustment. Examples include Easter falling in March or April, some months having more weekend days than others, and certain days being national holidays. In each case, the resulting seasonal pattern could be distorted and preliminary adjustments to the data may be necessary prior to forecasting. Even in the absence of special holidays, some months have more weekend days than others. Programs such as Census X-13-ARIMA allow the user to make "trading day" adjustments so that the different levels of activity (e.g., higher retail sales on Saturday, stock markets closing on weekends) can be taken into account.

[4.2] When there are limited data from which to calculate seasonal components, consider using the seasonals from related data series, such as the aggregate. Alternatively, average the estimates across similar products.

Summary

In this chapter, we have considered time series that possess a seasonal component. There are two principal objectives in the analysis of such series. The first is to take account of the seasonal pattern when making forecasts. The second is seasonal adjustment, whereby we wish to estimate the seasonal effects and then remove them, to produce a seasonally adjusted series. We began by considering a purely seasonal series — that is, one without a trend. Such series are rare in business applications but serve to illustrate how we can examine seasonal patterns, either by using a method that combines trends with seasonality or by isolating the seasonal component. We then developed forecasting procedures based upon decomposition methods and explored decomposition as a tool for describing time series phenomena. The remainder of the chapter developed the Holt-Winters approach to exponential smoothing for series with additive or multiplicative seasonal components. As in Chapter 3, the seasonal and trend components are updated after each new observation is recorded. Finally, we briefly examined the construction of prediction intervals.

DISCUSSION QUESTION: *In many forecasting applications, why is forecasting seasonality critically important?*

Exercises

4.1 Create seasonal plots for the Walmart quarterly sales data *(Walmart_2.xlsx)* by plotting

a. Sales against quarter, classified by year (as in Figure 4.2)

b. Sales against time, with years classified.

4.2 The data in the spreadsheet that follows refer to a quarterly series that is purely seasonal, as in Section 4.2. The values for year 1 are used as starting values (forecasts) for year 2. Use the purely seasonal version of exponential smoothing with $\gamma = 0.5$ to generate forecasts for year 3.

Quarter	Year	Data	Forecasts
1	1	50	
2	1	30	
3	1	20	
4	1	90	
1	2	55	50
2	2	35	30
3	2	15	20
4	2	80	90
1	3		
2	3		
3	3		
4	3		

Data: Exercise_4_2.xlsx

4.3 The file *US_retail_sales_2.xlsx* includes the series *Seasonal Factors* for the monthly series on U.S. retail sales as developed by the Census Bureau. The data cover the period from January 2001 to December 2015.

a. Using the period from January 2001 to December 2013 as the estimation sample, develop a seasonal forecasting method for this series.

b. Generate one-year-ahead forecasts for the hold-out sample for the period from January 2014 through to December 2015 using December 2013 as the forecast origin.

c. Recalibrate the forecasting equation by using the data up to December 2014, and then forecast the next 12 months. Compare the results.

The following exercises may be completed using the Exponential Smoothing Macro (ESM); other software may, of course, be used but the capabilities and estimation procedures vary so that results may differ.

In each case, the aim is to fit a suitable selection of exponential smoothing schemes: with or without a trend term; with none or additive or multiplicative seasonality; with or without transformations. Initial plots help to reduce the number of possibilities that need to be considered. Then examine the performance of the model for both the estimation and hold-out samples. Specific hold-out samples are suggested but it is important to explore the effects of changing the size and coverage of the estimation and hold-out samples.

Conclude the analysis with specific recommendations for the forecasting method to be employed.

4.4 Examine the series on Job Openings (*Job_openings.xlsx*) previously considered in Minicase 3.1. Which seasonal approach works best? Does the inclusion of seasonal factors provide a useful improvement in forecast quality?

4.5 The U.S. Census records monthly Alcoholic Beverage Sales by Wholesalers (*Alcohol_sales.xlsx*). The series is recorded monthly and it is not seasonally adjusted. Sales are measured in $Millions and the series runs from January 2001 to December 2015. Develop an appropriate seasonal model using the last 36 observations as the hold-out sample.

4.6 Use the data in the file *US_retail_sales_2.xlsx w*ith the period from January 2001 to December 2012 as the estimation sample. Develop an appropriate forecasting method for the series.

4.7 The movie industry is naturally very interested in forecasting ticket sales. The spreadsheet *Titanic_box_office.xlsx* provides the daily figures for the number of customers per theater for the period December 19, 1997 to July 23, 1998. Given the greater attendance at weekends, this series is seasonal with a period of seven days. Develop a suitable forecasting method, using the last four weeks as a hold-out sample.

How useful is your analysis to the movie industry? Think about different movie genres and the likely length of their runs before moving to online and DVD sales.

4.8 The spreadsheet *Gas_prices_Ch4.xlsx* gives the average end of month prices "at the pump" of regular-grade gasoline for the period from January 2000 to December 2015. Use the period from January 2000 to December 2012 as the estimation sample, and develop a seasonal forecasting method.

4.9 Modify the spreadsheet corresponding to Example 4.2 in *Examples_chapter_4.xlsx* to forecast, by means of a multiplicative decomposition, the data in Table 4.3.

4.10* Use equation (4.9) to substitute for the error term in the updating equations (4.10-11). Hence, derive the alternative form of the updating equations for the additive Holt-Winters scheme as given in equations (4.12):

$$L_t = L_{t-1} + T_{t-1} + \alpha[Y_t - S_{t-m} - L_{t-1} - T_{t-1}]$$

$$T_t = T_{t-1} + \beta[L_t - L_{t-1} - T_{t-1}]$$

$$S_t = S_{t-m} + \gamma[Y_t - L_{t-1} - T_{t-1} - S_{t-m}].$$

4.11* As in Exercise 4.10, substitute for the error term in the updating equations (4.16). Hence, derive the alternative form of the updating equations for the multiplicative Holt-Winters scheme:

$$L_t = L_{t-1} + T_{t-1} + \alpha[(Y_t/S_{t-m}) - L_{t-1} - T_{t-1}]$$

$$T_t = T_{t-1} + \beta(L_t - L_{t-1} - T_{t-1})$$

$$S_t = S_{t-m} + \gamma[Y_t/(L_{t-1} + T_{t-1}) - S_{t-m}].$$

4.12 Following on from Example 4.9, develop empirical prediction intervals for the automobile data for the four quarters of 2015.

Minicases

Minicase 4.1 Walmart Sales

Consider the data on Walmart sales previously considered in Minicase 3.2 and available in *Walmart_2.xlsx.* Plot the quarterly sales data and verify that the series exhibits a seasonal element. Using the ESM or other suitable software, contrast the additive, multiplicative, and logarithmic forms for forecasting sales, using the last 12 quarters as a hold-out sample. Compare the results, using both approximate normal and empirical prediction intervals, with those obtained for Minicase 3.2.

Minicase 4.2 Automobile Production

Reexamine the automobile production series considered in Example 4.3 and later examples *(Autos_index.xlsx).* Use the last three years (2013–2015) as the hold-out sample, but reduce the size of the estimation sample by starting in 1996 and again in 2001. Compare the results with those obtained using 1991–2012 as the estimation sample. Also, generate prediction intervals for the four quarters of 2015, using both approximate normal and empirical prediction intervals.

What are your conclusions regarding the relevance of the earlier data in the light of possible changes in production plans over time?

Minicase 4.3 U.S. Retail Sales

The spreadsheet *US_retail_sales_2.xlsx* contains the monthly values of U.S. retail sales (measured in $ billions) over the period from January 2001 to December 2016. The series was plotted in Chapter 1. The spreadsheet includes the value of sales (*Sales*), the seasonally adjusted values (*Sales_SA*) and the seasonal factors. Compare the options of

- forecasting the series by using the Holt-Winters approach,
- forecasting the components separately and then combining them.

Use period January 2001–November 2007 as the estimation sample. The onset of the "Great Recession" was determined to have occurred in December 2007 but was not officially confirmed until December 2008; would the one-step-ahead forecasts have provided earlier warning? Generate prediction intervals and see whether the actual values fall within the limits.

If you have access to suitable software, evaluate the ability of the multiple-step-ahead forecasts to predict the downturn. Conduct a rolling origin 1-12 evaluation, forecasting for the years 2008-2010. (If no suitable software is available, use just three forecast origins to compare results.)

Extend the estimation sample to December 2012 and compare the one-step-ahead forecast performance for the period 2013–2015 using the two different estimation samples. [Exercise 4.6]

Minicase 4.4 UK Retail Sales

The spreadsheet *UK_retail_sales_2.xlsx* contains the quarterly values of the UK Retail Sales Index (measured with 2010 set at 100), both unadjusted and seasonally adjusted, over the period from January 1996 to December 2014. Create a purely seasonal series by dividing the unadjusted series by the seasonally adjusted one. Using one-step-ahead forecasts, compare the options of:

- Forecasting the series by using the Holt-Winters approach.
- Forecasting the components separately and then combining them.

It is recommended that you use the period 1996-2011 for estimation and the remaining 12 quarters as the hold-out sample.

Minicase 4.5 Newspaper Sales

A small local bakery purchases a limited number of copies of the *Washington Post* on a "sale or return" basis. Sales and unsold copies were monitored over a six-week period, with the results shown in the table below. The data[4] are available in *Newspapers.xlsx*. The owner seeks advice regarding the number of copies to order in the coming week.

Week	Variable	Sun	Mon	Tue	Wed	Thu	Fri	Sat
1	Sales	15	12	10	12	13	14	16
	Returns	0	3	1	0	2	0	6
2	Sales	16	11	10	13	14	15	15
	Returns	5	0	0	0	6	6	5
3	Sales	15	11	13	16	14	15	14
	Returns	8	6	3	8	2	3	1
4	Sales	15	13	13	13	16	14	13
	Returns	3	7	3	8	7	4	5
5	Sales	14	11	12	14	13	16	12
	Returns	1	0	0	6	3	1	2
6	Sales	13	12	16	13	12	21	12
	Returns	6	7	0	*	4	8	0

Data: Newspapers.xlsx

What additional information would you like to see? How do you treat the days when there were no returns? Is there any seasonal structure? What do you recommend as an ordering policy?

References

Archibald, B.C., and Koehler, A.B. (2003). Normalization of seasonal factors in Winters' methods. *International Journal of Forecasting,* 19, 143–148.

Durbin, J., and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods*. Oxford, U.K.: Oxford University Press, second edition.

Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C., and Chen, B.C. (1998). New capabilities and methods of the X-12-ARIMA seasonal adjustment program (with discussion). *Journal of Business and Economic Statistics,* 16, 127–177. [For recent developments, consult the U.S. Census Bureau website for X1-13-ARIMA: *www.census.gov/srd/www/x13a*]

Ghysels, E., and Osborn, D.R., 2001. The econometric analysis of seasonal time series. Cambridge University Press.

Gould, P.G., Koehler, A.B., Ord, J.K., Snyder, R.D., Hyndman, R.J. and Vahid-Araghi, F. (2008). Forecasting time series with multiple seasonal patterns. *European Journal of Operational Research,* 191, 205–220.

4 The data are genuine, but the name of the bakery is withheld for reasons of confidentiality.

Holt, C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting,* 20, 5–10. Reprint of Office of Naval Research Memorandum 52/1957, Carnegie Institute of Technology.

Hyndman, R. J., Koehler, A. B., Ord, J. K., and Snyder, R. D. (2008). *Forecasting with Exponential Smoothing.* Berlin: Springer.

Shiskin, J., Young, A.H., and Musgrave, J.C. (1967). *The X-11 variant of the Census Method II seasonal adjustment program.* Technical Paper 15, Bureau of the Census, U.S. Department of Commerce, Washington, DC.

Taylor, J.W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society,* 54, 799–805.

Winters, P.R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science,* 6, 324–342.

CHAPTER 7

Simple Linear Regression for Forecasting

Table of Contents

Even trained statisticians often fail to appreciate the extent to which statistics are vitiated by the unrecorded assumptions of their interpreters.... It is easy to prove that the wearing of tall hats and the carrying of umbrellas enlarges the chest, prolongs life, and confers comparative immunity from disease.... A university degree, a daily bath, the owning of thirty pairs of trousers, a knowledge of Wagner's music, a pew in church, anything, in short that implies more means and better nurture... can be statistically palmed off as a magic spell conferring all sorts of privileges.... The mathematician whose correlations would fill a Newton with admiration, may, in collecting and accepting data and drawing conclusions from them, fall into quite crude errors by just such popular oversights as I have been describing.

— George Bernard Shaw, 1906, The Doctor's Dilemma[1]

1 As abstracted in Stuart, A., Ord, J.K. and Arnold, S.F. (1999). *Kendall's Advanced Theory of Statistics, Volume 2A*. London: Arnold, p. 467.

Introduction

So far, we have looked at models that capture patterns within a single time series. A natural next step, in both forecasting and statistical modeling generally, is to look for relationships between the variable of interest and other factors, regardless of whether the data of interest are cross-sectional or defined over time. For example, when trying to predict a company's product sales, we expect such variables as the price of the product and the level of advertising expenditures to affect sales movements. These variables come to mind when we think of consumer choice, because consumers are usually price sensitive and advertising may be needed to bring the product to consumers' attention. Probing more deeply, we recognize that such variables are not absolutes, but must be judged relative to the pricing and advertising actions of the product's principal competitors. Ah, but here lies the rub! Information on our competitors' plans would be very valuable and often helps to explain fluctuations in sales with the benefit of hindsight. However, at the time we make our production and inventory decisions, the plans of our competitors are unknown; at best, we can hope to forecast those plans or to evaluate the impact of their different possible actions. This scenario is not unusual, and it contains the essential questions that plague any forecaster: *What do you know and when will you know it?* We will return to these key questions periodically.

Similar problems arise in cross-sectional studies. We may seek to evaluate an individual's potential for purchasing a particular product, such as a new automobile. We will have demographic information on that person, plus some financial data, but we cannot measure a host of other factors, such as the importance the person attaches to having a new vehicle, other interests that may compete for his or her funds, and information obtained from friends regarding that particular model.

We are now moving away from the purely extrapolative time series methods of the earlier chapters and into the realm of economic and market modeling. Our data may be cross-sectional or time indexed, or even both. The tools we develop in this and the next two chapters are commonly referred to as methods of *regression analysis*. Because our focus is on the development of quantitative models for economic activities, we also use the term *econometric modeling*.

We continue to denote the variable of interest by Y, but now refer to it as the *dependent variable* or the *output*. Our statistical model will then depend upon one or more *inputs*, or *predictors*, or *explanatory variables*[2] that are denoted by X, with subscripts if we have two or more such variables (e.g., X_1 and X_2). In this chapter, we follow the simpler path of assuming that only one such variable is important and we proceed to develop tools for that case before extending the ideas to multiple explanatory variables in Chapter 8. We urge the reader not to be overly impatient, but to think of the chapter as training for the more realistic problems to come.

In Section 7.1, we discuss the relationship between causality and correlation. Two variables may be correlated, and that information may be useful, but it does not imply causation. This discussion serves as a lead-in to Section 7.2, in which we specify a linear relationship between the variable of interest (Y) and the explanatory variable (X) and use the method of ordinary least squares to estimate this relationship; a case study on gasoline prices follows in Section 7.3. In order to determine whether an estimated relationship is useful, we first introduce the standard error and the coefficient of determination in Section 7.4. We then provide a systematic development of statistical inference for the regression model in Sections 7.5 and 7.6, including the use of transformations. In Sections 7.7 and

2 Many texts use the term *independent variable* to describe the X inputs. Because we employ the term "independent" in other, more critical contexts, we will not use it here.

7.8, we discuss forecasting by means of simple linear regression. Principles of regression modeling are left to Chapter 8.

7.1 Relationships between Variables: Correlation and Causation

The ideal basis for forecasting is a *causal model*. In such a scheme, we are able to identify all the key factors that determine future values of Y and to measure them in sufficient time to generate the forecasts. In practice, at least three problems arise:

- We may not be able to identify all the causal variables (e.g., try listing all the factors that affect gasoline prices).
- We may not be able to obtain information to measure these variables (e.g., competitors' plans).
- The measurements may not be available in time to make the forecasts (e.g., measurements on macroeconomic variables such as gross domestic product (*GDP*) are not available until several weeks or even months after the forecasting horizon of interest).

The availability and timing of measurements affect the type of forecast that we can make, as we now describe.

FORECASTING WITH A REGRESSION MODEL

An *ex ante*, or *unconditional*, forecast uses only the information that would have been available at the time the forecast was made (i.e., at the forecast origin).

An *ex post*, or *conditional*, forecast uses the actual values of the explanatory variables, even if these would not have been known at the time the forecast was made.

A *what-if* forecast uses assumed values of the explanatory variables to determine the potential outcomes of different policy alternatives or different possible futures.

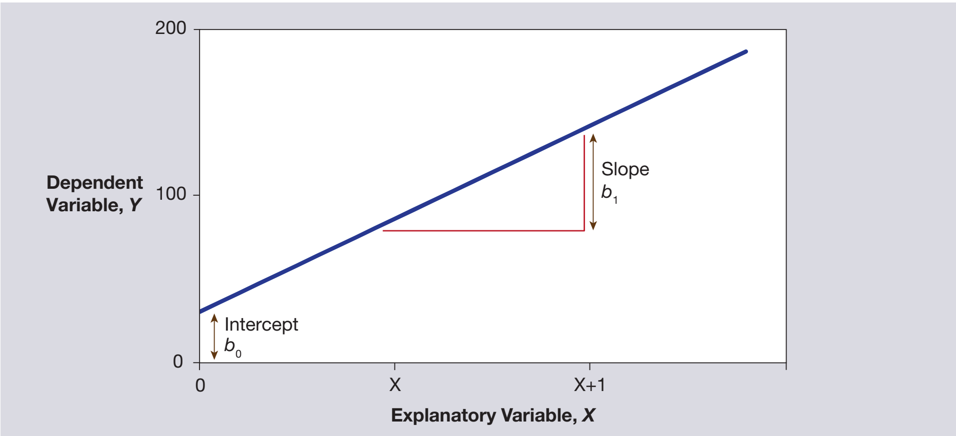
For example, consider the effect upon sales of the difference in price between a supplier and its competition. An *ex post conditional* analysis would use actual price differentials to review forecasting performance over past data, an exercise that is useful for the determination of price sensitivity, but does not provide genuine forecasts. An *ex ante unconditional* analysis would either forecast the price differential and use that forecast in turn to forecast sales. Alternatively, the forecasting model could be based on sufficiently distant lagged values so that all the information is available to produce the forecasts. Finally, a *what-if* analysis would use the model (previously validated by *ex post* and *ex ante* analyses of historical data) to determine the implications of different price-setting policies in the light of possible competitive prices.

Trying to track down all the causal variables in a real economic system is probably a fool's errand, but it is vital that we measure the key factors. If the measurements are not available in time for *ex ante* forecasts, we must seek appropriate substitutes. For example, the variable of interest may be the local unemployment rate, but for short-term forecasting we might use the relative numbers of "Help Wanted" advertisements on the web to gauge the health of a local labor market. Note that, in this last step, we are moving away from a causal model to a model in which an *available* input variable is substituted for the causal input variable. Of course, the available input variable may reasonably be assumed to be correlated with the input variable of interest. The apparent association may be due only to a common dependence upon other variables, but the relationship may still provide useful forecasts. Nevertheless, we will do well to heed George Bernard Shaw's warning; that is, we must beware of interpreting correlation as causation and should recognize that apparent relationships are sometimes spurious.

7.1.1 What Is Regression Analysis?

With this background in mind, we now assume that we can identify a single output variable (Y) that is related to a single input variable (X). In the simplest possible case, the value of X may completely determine the value of Y. For example, consider Figure 7.1, which shows such a *mathematical relationship* between X and Y: A single value of X produces a single value of Y. The intercept, b_0, is the value of Y when $X = 0$, and the slope, b_1 denotes the increase in Y when X is increased by 1 unit. Thus, if $b_0 = 3$ and $b_1 = 1.5$, a value of $X = 4$ yields $Y = 3 + 1.5(4) = 9.0$. Such tidy relationships do not exist in the business world, so we must allow for uncertainty in the relationship between X and Y. The methodology for establishing these relationships is known as *regression analysis*.

Figure 7.1 The Terms of the Line $Y = b_0 + b_1X$

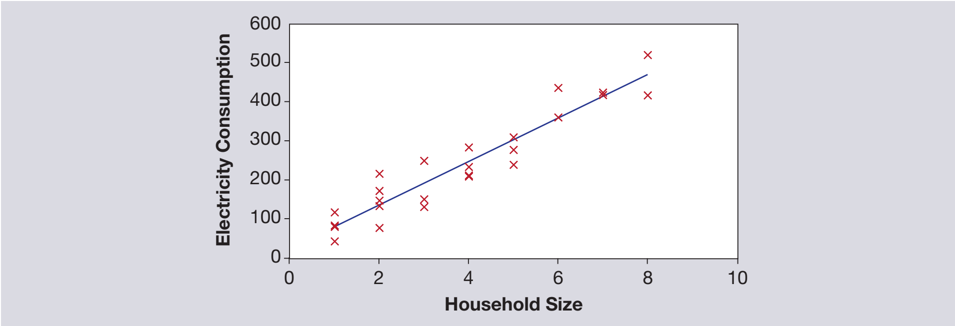


A regression line differs fundamentally from a mathematical relationship. We may observe one of many possible Y-values for a particular value of X, and the regression line reflects the *expected value of Y given X*. For a given observation, Y may be above or below our expectation. For example, in Figure 7.2 X denotes household size and Y represents monthly electricity usage (for this hypothetical data set).

The regression line indicates that the *average* household of size 4 consumes about 250 units of electricity; an individual household may consume more or less, depending on other factors such as house size, mode of heating, lifestyle, and so on. Also, it is worth noting that in this example the intercept value (the value of Y when $X = 0$) does not have any economic meaning: It does not make sense to ask about consumption for a household containing zero people. Regression lines typically have restricted ranges of application (here, $1 \leq X \leq 8$). Also in this example, non-integer values of X are not sensible, but it is still convenient to draw a straight line to represent the relationship between the expected value of Y and a given X.

In our example, the equation of the straight line is $Y = 30.0 + 5.60X$; the approach described in the next section is used to determine the slope and intercept.

Figure 7.2 Plot of Electricity Consumption Against Household Size with a Fitted Regression Line



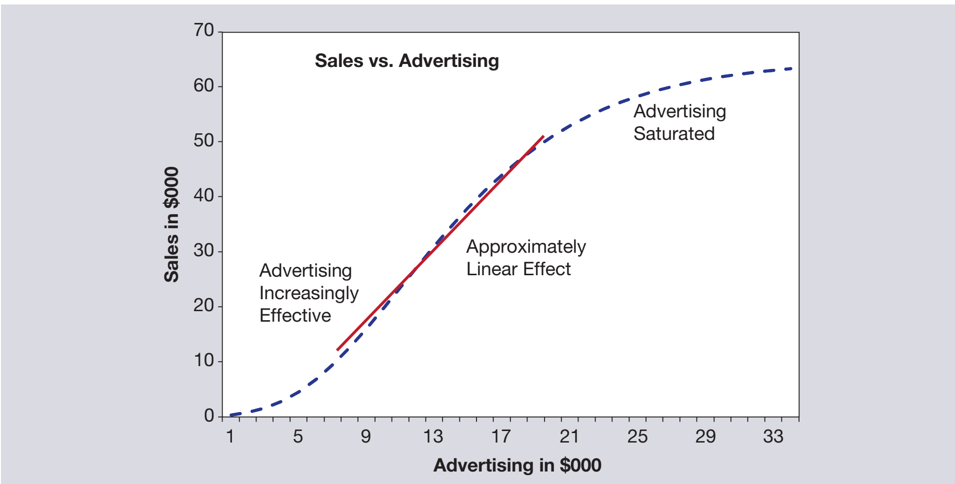
7.2 Fitting a Regression Line by Ordinary Least Squares (OLS)

After we have identified the variables of interest, we create a scatterplot, as in Figure 7.2, to determine whether the relationship between X and Y could reasonably be described by a straight line, at least over the region of interest.

Example 7.1: Sales versus advertising

If a company increases its advertising budget, it expects to see a higher level of sales. In any particular week, that may or may not happen, because sales will be affected by the weather, the state of the economy, and the actions of competitors. Nevertheless, we would *expect* to see an increase in sales. At the same time, there comes a point where further increases in advertising expenditures cease to have much effect. The overall relationship between expected sales and advertising may be like that shown in Figure 7.3.

Figure 7.3 Relationship Between Sales and Advertising



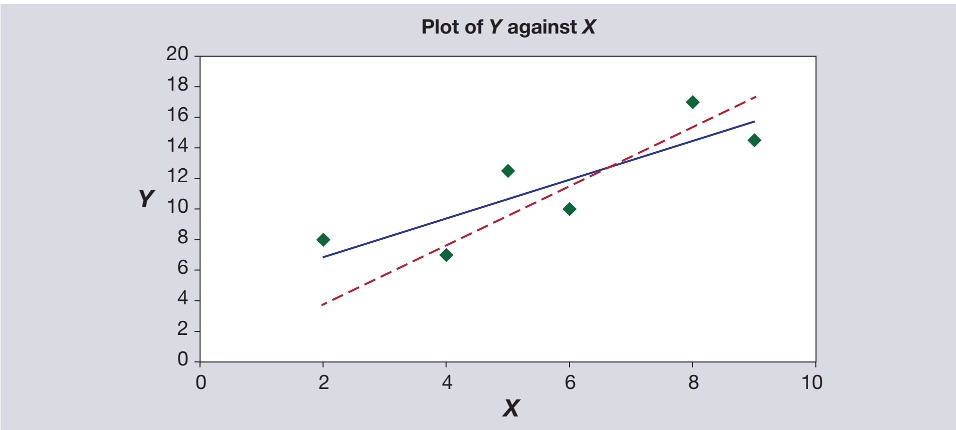
Now consider this figure from the perspective of a brand manager. On the one hand, if that person spends \$8K or less on advertising, he or she is missing a range over which advertising becomes increasingly effective. On the other hand, if the manager spends more than \$21K, the extra expenditures have limited impact. Accordingly, the manager will probably keep the budget in the range $8 \le X \le 21$, and the relationship is very close to a straight line over that range. ■

Two conclusions emerge from this example. First, it may be feasible to use a linear relationship over the range of interest, even when the true relationship is nonlinear; second, the relationship may be valid only over the range of X for which observations are available. If the brand manager estimated the relationship over $8 \le X \le 21$ and then used the resulting line to estimate sales at $X = 25$, he or she would be in for a rude shock! We now assume (for the moment) that a straight-line relationship is appropriate over the range of interest, and we seek to determine a (regression) line that best describes the expected relationship between Y and X. We formulate the relationship as

$$Y = b_0 + b_1X + e = \textit{Explainable pattern} + \textit{Unexplained error}. \tag{7.1}$$

The first two terms on the right-hand side of equation (7.1) show the extent to which Y can be linearly "explained" by X; the third term, e, known as the *error term*, represents the discrepancy between the observed value of Y and the corresponding Y value on the straight line. The errors are the *vertical* distances between the observed and fitted values. As is evident from the figure, we would like to choose values of the coefficients b_0 and b_1 so that the resulting line is a good representation of the relationship between Y and X. One approach would be to draw a freehand line, but this method is unreliable. An individual might draw different lines on different days, and two individuals might be quite unable to agree on whose line seemed better, as can be seen from the two lines in Figure 7.4. To resolve the argument, we must select an objective method of fitting so that, for a given set of data, everyone would arrive at the same result once the method had been agreed upon.

Figure 7.4 Two Possible Regression Lines Relating *Y* to *X*



7.2.1 The Method of Ordinary Least Squares (OLS)

The technique most commonly used to estimate the regression line is the *Method of Ordinary Least Squares*, often abbreviated to OLS.[3] Once we have formulated the nature of the relationship, as in equation (7.1), we define the OLS estimates as those values of $\{b_0, b_1\}$ which minimize the *Sum of Square Errors* (*SSE*): Given n pairs of observed values (X_i, Y_i), we have

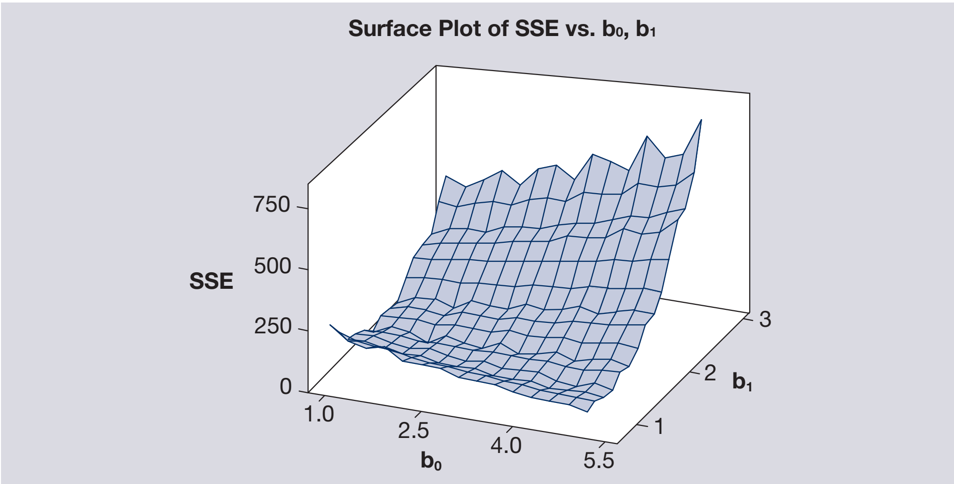
$$SSE = \sum_{i=1}^{n} (Observed - Fitted)^2 = \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2 = \sum_{i=1}^{n} e_i^2. \tag{7.2}$$

ORDINARY LEAST SQUARES

The method of ordinary least squares determines the intercept and slope of the regression line by minimizing the sum of square errors (*SSE*).

We might compute the *SSE* for a range of values of the intercept and slope and search for the best values. Such an approach is illustrated in Figure 7.5, where we plot the *SSE* against b_0 and b_1. Given that we seek to minimize the *SSE*, inspection of that diagram suggests values in the range $2.5 < b_0 < 5.5$ and $1.0 < b_1 < 1.5$. A finer grid on the diagram would allow a more accurate determination of the best values. Such an approach is clearly tedious, but one of the nice features about the method of ordinary least squares is that we can determine exact computational formulas for the coefficients; the derivation is provided in Appendix 7A and the final formulas are presented next. Although the details of these calculations are relatively unimportant in the computer age, unless you are a programmer, details for our simple example are laid out in spreadsheet form in Table 7.1.

Figure 7.5 Minitab Plot of SSE against Different Intercept and Slope Parameters for Data in Table 7.1



3 Other criteria could be — and are — applied, such as minimizing the sum of absolute errors, as defined in Section 2.7. However, the method of ordinary least squares is by far the most popular because of its numerical tractability. The method dates back to Carl Friedrich Gauss in the early nineteenth century. (For some historical insights, see the Wikipedia entry on Gauss at http://en. wikipedia.org/wiki/Carl_Friedrich_Gauss.) The method is labeled "ordinary" because there are a number of extensions, such as "generalized least squares," which we do not cover here.

Table 7.1 Spreadsheet for Regression Calculations Using Equation (7.5)

	X	Y	$(X-\bar{X})$	$(X-\bar{X})^2$	$(Y-\bar{Y})$	$(X-\bar{X})(Y-\bar{Y})$	$(Y-\bar{Y})^2$
	2	8	–3.67	13.44	–3.50	12.83	12.25
	4	7	–1.67	2.78	–4.50	7.50	20.25
	5	12.5	–0.67	0.44	1.00	–0.67	1.00
	6	10	0.33	0.11	–1.50	–0.50	2.25
	8	17	2.33	5.44	5.50	12.83	30.25
	9	14.5	3.33	11.11	0.00	10.00	9.00
Sum	34	69	0.00	33.33	0.00	42.00	75.00
Mean	5.67	11.5	0.00	5.56	0.00	7.00	12.50
		$b_0 = 4.36$ $b_1 = 1.26$					

Data: Example 7_2.xlsx. See Example 7.2 for details of calculations of intercept and slope.

The calculations may be summarized as follows; the summations are taken over all n observations. The sample means are

$$\bar{X} = \frac{\Sigma X_i}{n}, \quad \bar{Y} = \frac{\Sigma Y_i}{n}. \tag{7.3}$$

Now, denote the sum of squares for X by S_{XX} and the sum of cross products for X and Y by S_{XY}, defined respectively as

$$S_{XX} = \sum (X_i - \bar{X})^2 \text{ and } S_{XY} = \sum (X_i - \bar{X})(Y_i - \bar{Y}). \tag{7.4}$$

OLS yields the intercept and slope as

$$b_0 = \bar{Y} - b_1\bar{X} \text{ and } b_1 = \frac{S_{XY}}{S_{XX}}, \tag{7.5}$$

Thus, we have exact solutions and do not need to follow the search process suggested by Figure 7.5. Nevertheless, it is useful to keep the idea of a direct search in mind when models are more complex; indeed, that was the process we followed when fitting the state-space models in Chapter 5 and the ARIMA models in Chapter 6.

Example 7.2: Calculation of slope and intercept

From the spreadsheet given in Table 7.1, we see that

$$\bar{X} = 5.67, \bar{Y} = 11.50, S_{XX} = 33.33, \text{ and } S_{XY} = 42.00,$$

and it follows that $b_1 = 42.00/33.33 = 1.26$ and $b_0 = 11.50 - (1.26)(5.67) = 4.36$.

(Note that the figures have been rounded to two decimal places, so you may get slightly different answers.)

We encourage the reader to work through the details of Table 7.1, if only this once, to aid understanding. Thereafter, we make use of one of the many statistical packages that are available for the calculation of regression lines.

We use Table 7.1 to determine the regression line for Y on X and calculate

$$\hat{Y} = 4.36 + 1.26X. \tag{7.6}$$

The "hat" notation is used to indicate that the line defined by expression (7.6) has been estimated by applying OLS to this data set. The equation refers to the fitted values of Y, not the observations. Note that the equation represents the fitted values of Y, given different values of X, and it is not appropriate to reverse the roles of X and Y. That is why we have included the "hat" only on the dependent variable. ■

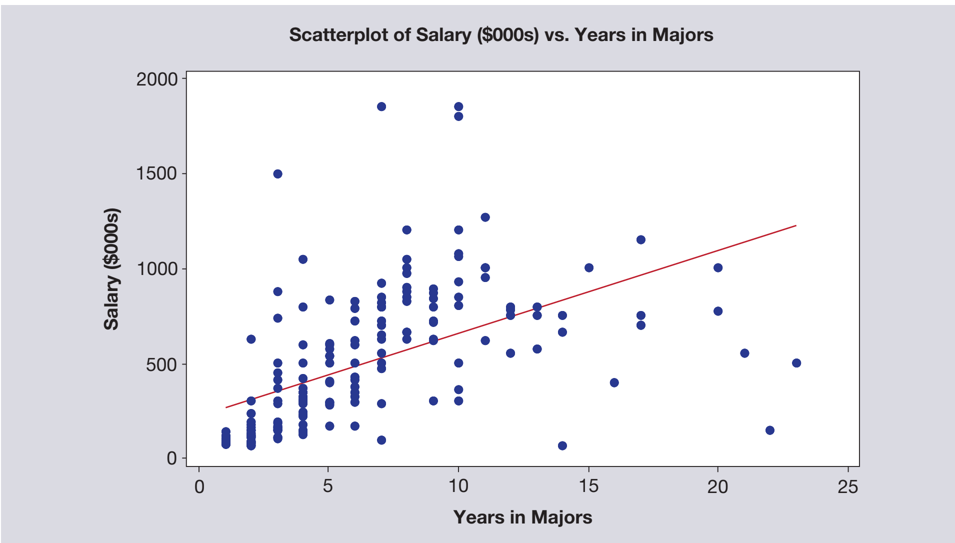
Another way of thinking about this calculation is to return to Figure 7.3. The line is chosen to minimize the *SSE* (Sum of Square Errors), which is based upon the *vertical* distances between the observed and fitted Y values. Finally, leave the mathematics behind. What are you trying to forecast? Y! What are you using to make the forecast? X. Thus, equation (7.6) is the correct version to use.

■ Example 7.3: Baseball salaries

In 2003, a book appeared called *Moneyball: The Art of Winning an Unfair Game* (New York: W. W. Norton), written by Michael Lewis. In 2011 it was made into a film with Brad Pitt in the lead. In the book, Lewis describes how Billy Beane, general manager of the Oakland Athletics baseball team, used statistical methods to identify players who were undervalued. His teams had considerable success on the field despite having a very low budget for player salaries. The essence of Beane's approach was to identify those factors which were important to winning games but were often overlooked by other baseball managers. He then sought out players who could deliver those skills.

As a first step in such an analysis, we consider a simple regression for Y = player's salary (in thousands of dollars) on X = number of years in the major leagues.[4] The data are for the 1986 season, and in this example we use only the data relating to pitchers. The scatter-plot for the data appears in Figure 7.6.

Figure 7.6 Scatterplot of Baseball Players' Salaries Against the Number of Years Played in the Major Leagues



Data: Baseball.xlsx

4 The data are included in the file *Baseball.xlsx*, and we wish to acknowledge StatLib of the Department of Statistics at Carnegie Mellon University for allowing us access to these data. The data were originally compiled by Lorraine Denby as part of a project with the Graphics Section of the American Statistical Association.

The fitted regression line is of the form

$$\textit{Salary}\ (\$000s) = 224 + 43.6 \times \textit{Years in Majors}.$$

One might interpret the line as an average baseline salary of $224K plus $43.6K per year of service, keeping in mind that this is the expected value and that individual salaries vary considerably. Common sense suggests that players tend to become less valuable toward the end of their careers, and we can see from Figure 7.6 that a number of veteran players (say, with 12 or more years in the major leagues) have salaries well below the fitted values. Nevertheless, this simple model suggests that career length is an important determinant of salary. (We examine the model further at the end of the chapter, in Minicase 7.3.). ■

DISCUSSION QUESTION: *How might you modify the analysis to allow for the lower salaries of veteran players?*

7.3 A Case Study on the Price of Gasoline

Fluctuations in gasoline prices have an impact on most sectors of the economy. Although the transportation industry is the first affected, any price increases will eventually work their way through the system to have an impact on the prices of most goods and services. At the same time, price increases in other sectors will have an effect on gasoline prices primarily by changing demand. In addition, decisions by the Organization of the Petroleum Exporting Countries (OPEC) and other world events will have an impact on supply. Our purpose in this section is to illustrate the model-building process by starting to develop a statistical model that can account for the fluctuations in retail gasoline prices. We will revisit the model from time to time and improve it as we develop our model building principles.

Where should we start? Everything seems to be related to everything else! However, recalling Principle 2.1, we start by trying to identify the key variables of interest and come up with the following (albeit incomplete) list of factors and corresponding time series:

- Price of unleaded gasoline to U.S. consumers (variable to be forecast).
- Producer price: price of crude oil
  - The major cost element in gas prices.
- Consumer prices
  - Overall price levels may affect gas prices.
- Consumer demand for gasoline
  - Higher demand leading to higher prices?
- Production, stocks and imports of crude oil
  - Higher availability may well lead to lower prices but may be a leading indicator of expected future prices.
- Consumer income: personal disposable income, total or per capita
  - Consumers with more income more likely to consume and drive.
- Level of economic activity: unemployment; retail sales; new starts in housing
  - More activity in the economy requires more driving.
- Leading indicator of economic activity: S&P 500 Index
  - Future higher levels of economic activity leads to higher prices now?

We also need to decide on the precise way to measure each variable — there are many alternative measures of consumer income and economic activity. Measures of these variables are all readily available from standard sources,[5] and monthly data for January 1996 through December 2015 are included in the file *Gas_prices_1.xlsx*.

DISCUSSION QUESTION: *What other variables might have an impact? How could they be measured effectively, and what possible sources of data might be available?*

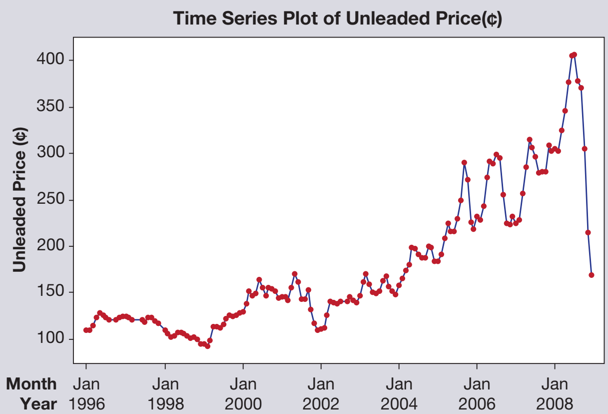
7.3.1 Preliminary Data Analysis

Before we leap into any attempts at model building, we start out by developing a feel for the data. The first step is to take a look at a time plot of the data, as shown in Figure 7.7; prices are quoted in cents per U.S. gallon. Initially we choose to analyze only a part of the data: Observations are plotted through December 2008; January 2009 to December 2010 are used as a hold-out sample. A lot has happened to gas prices since 2008 and Minicase 7.1 explores some of these changes. More experienced automobile drivers will not need to be reminded of the wild fluctuations in gasoline prices over the years since 2008 though settling down towards the end of the time series in 2013!

Even the plot to 2008 is very revealing: Prices were reasonably steady over the first eight years of the period examined and then climbed before dropping sharply in the last six months of the period.

The next step is to think about the variables of interest and how they relate to gasoline prices. A full analysis should include an examination of each variable, using the methods of Chapter 2, plus checks for unusual observations and so on. In the interest of space, we do not report those details here, but encourage the reader to probe in greater depth. Instead, we move to the next phase and look at a matrix plot to relate gasoline prices to some of the explanatory variables. We have restricted the list to the more interesting items to make the plot more readable.

Figure 7.7 Time Plot of U.S. Gasoline Prices, in Cents per Gallon (January 1996–December 2008)



Data: Gas_Prices_1.xlsx; adapted from Minitab output.

5 These series are part of the library of Official Energy Statistics from the U.S. Government. They are located on the website of the Energy Information Administration: *www.eia.doe.gov* which also includes background information. The macro variables are to be found on: *https://fred.stlouisfed.org/*.

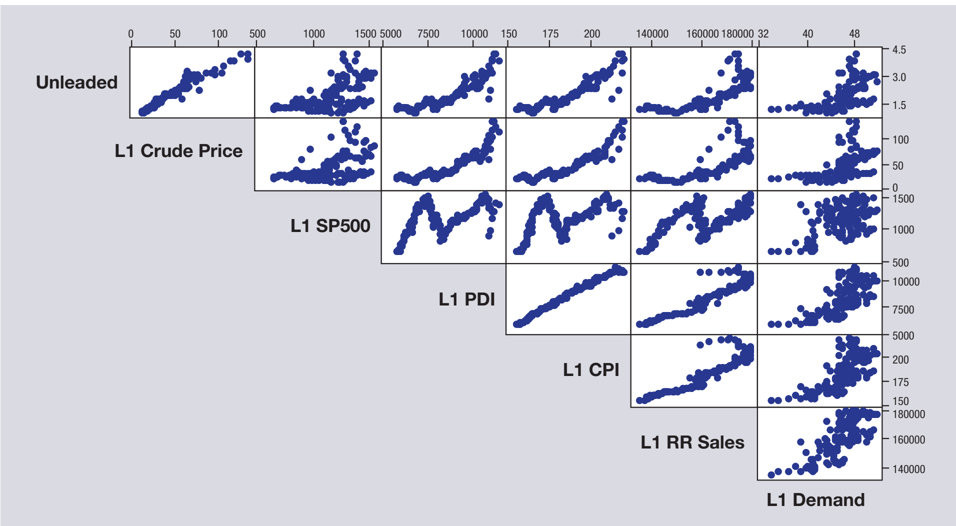
The variables plotted in Figure 7.8 are monthly average gasoline prices for the current month and the one-month lagged values for some of the explanatory variables (with unemployment omitted): we have selected various economic indicators as well as the price of the major component, crude oil. This ensures that the data would be available for forecasting at least one period ahead:

- Price of unleaded gasoline, in cents per U.S. gallon[6] (*"Unleaded"*)
- The price of crude oil, in dollars per barrel (*"L1_Crude_price"*)
- The SP500 Stock Index, end-of-month close (*"L1_SP500"*)
- Personal disposable income, in billions of current dollars (*"L1_PDI"*)
- Consumer price index for all urban consumers, indexed at 100 over 1982–1984, not seasonally adjusted (*"L1_CPI"*)
- Retail Sales: real retail sales in millions of $s, deflated by *CPI* (*"L1_RR_sales"*)
- Unemployment rate, not seasonally adjusted (*"L1_Unemp"*)
- Housing starts, seasonally adjusted annual rate (*"L1_Housing"*)
- Demand for gasoline, in thousands of barrels per day (*"L1_Demand"*).

The prefix *L1* is used to indicate that the variable is lagged by one month. The full data file contains additional variables.

We use the lagged values for two reasons. First, from an economic perspective, increases in production costs and other changes will take some time to pass on to consumers. Second, as forecasters, we need to make sure that the input data are available. The previous month's figures may become available relatively early in the current month, but the current month's value clearly could not be available in time for *forecasting* purposes.[7]

Figure 7.8 Matrix Plot: Gasoline Prices Against Price of Crude Oil, the S&P Index, Personal Disposable Income (current prices), the Consumer Price Index (CPI), Retail Sales (RR_sales), and Demand (all lagged one period)



Data: Gas_Prices_1.xlsx; from the Minitab output.

6 In many countries, tax on gasoline prices is an important component and would have to be considered as a distinct explanatory variable. It is of limited importance in the US.

7 There are a number of different ways in which forecasts might be generated. We defer discussion of this topic until Sections 7.7 and 7.8.

The selective plots shown provide a number of insights:

- Gasoline prices appear to be most closely related to *L1_Crude_price*.
- There is a strong upward sloping relationship between gasoline prices, *PDI* and *CPI* and a somewhat weaker one between gasoline prices and real retail sales (*RR_sales*).
- The relationship with lagged *Demand* and *SP500* appears weak.
- There is an interrelationship between the different variables, perhaps because some trend together, e.g. *Crude Price* and *CPI*.

Finally, we may confirm our visual impressions by computing the correlations between each pair of variables. The full set of correlations is given in Table 7.2. Because correlations measure linear relationships, the values should accord with the intuition gained from the matrix plot, and they do.

Table 7.2 Correlations of Lagged Predictor Variables With Unleaded Price, and Intercorrelations (January 1996–December 2008)

	Lagged							
Lagged	Unleaded	Crude Price	SP500	PDI	CPI	Retail Sales	Housing	Unemp
Crude_Price	0.971							
SP500	0.511	0.462						
PDI	0.890	0.875	0.561					
CPI	0.911	0.904	0.535	0.995				
RR_Sales	0.792	0.736	0.678	0.936	0.912			
Housing	−0.243	−0.348	−0.032	−0.085	−0.141	0.177		
Unemp	0.073	0.129	−0.592	0.197	0.212	0.039	0.006	
Demand	0.544	0.478	0.469	0.711	0.695	0.760	0.200	0.112

Data: Gas_prices_1.xlsx. *P*-value for testing whether the correlations differ from zero are excluded: most values are 'significant' at the 5 percent level.

On the basis of this preliminary analysis, we will focus our attention upon the relationship between gasoline prices and the lagged crude oil price. However, our preliminary investigations suffice to indicate that *L1_Crude_price* alone is unlikely to provide the whole story as both an economic analysis and the correlation matrix suggest other influences. In this chapter, we develop simple regression models, using this relationship as a running example. In Chapter 8, we return to the more general issue of developing a full model for gasoline prices that includes other variables.

Although the price of crude passes the test as the best single predictor in this case, we must recognize that forecasts based upon that model will be useful mainly in the short term. Longer term forecasts would need to capture more details about the infrastructure of the industry and proven reserves, as well as worldwide demand, which in turn affect the price of crude oil.

7.3.2 The Regression Model

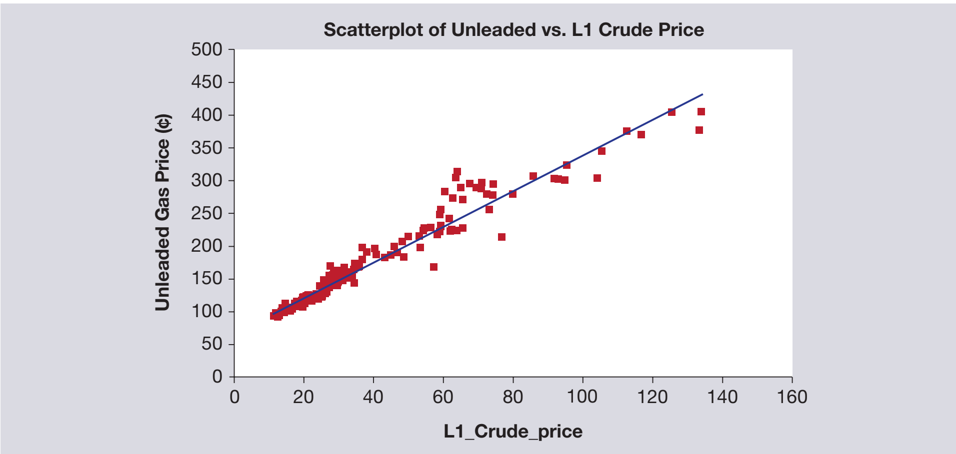
Figure 7.9 shows an enlarged view of the plot for gasoline prices on *L1_Crude_price*, with the fitted regression line superimposed. The relationship appears to be linear, although there does seem to be more scatter at the upper end of the range. The fitted line has the equation

$$Unleaded_t = 68.8 + 2.71\ L1\_Crude\_price_t. \tag{7.7}$$

W often drop subscripts when there is no ambiguity.

We may interpret equation (7.7) as stating that an increase of $1 in the price of a barrel of crude may be expected to increase the price at the pump next month by 2.7 cents per gallon. It is tempting to conclude that the intercept term represents overhead cost plus markup, but this conclusion is highly speculative, because we have no observations that correspond to X near zero. As previously noted in Section 7.1.1 such relationships should be viewed as valid only within the existing range of the explanatory variable, which is about $18 to $130 per barrel. Such a position is theoretically sound, but at times it is too restrictive. Commodity prices such as oil fluctuate dramatically as can be seen from the graph of the full data set to 2015. In addition, many economic aggregates grow over time. Thus, future values of variables such as *L1_Crude_price* are likely to lie outside the "valid range" when it comes to making forecasts. The problem of *extrapolation* is not easily avoided, but we should always ask whether it makes sense to extend the relationship beyond the current range of X. Further, we must recognize that all forecasts are extrapolations, since they use the past to forecast the future. Consequently, we must assume that expressions like equation (7.7) hold both for future time periods *and* for X-values that may lie outside the current range of observations.[8] Because of the need to extrapolate, business forecasters should rightly maintain a certain degree of humility[9] about their craft: There are plenty of assumptions in their models that may well go wrong!

Figure 7.9 Plot of Unleaded Prices Against Lagged Crude Oil Prices (L1_Crude_price)



7.4 How Good Is the Fitted Line?

Once the line has been fitted, the next question is how effective is it as a basis for forecasting? In Section 7.2, we decided to fit the line by minimizing the sum of squares. That is, we

8 At times, we can improve matters somewhat by looking at changes rather than absolute levels (see Section 9.5.1). We should also consider whether to examine prices in real dollars (i.e., dollars adjusted for inflation) rather than current dollars; this question is explored in Minicase 7.1.

9 It is sometimes suggested that shrewd forecasters always ensure that the forecast horizon is longer than they expect to stay in their current job. That way, someone else takes the heat when the forecasts fail. Such people often run for political office.

started from the total sum of squares for Y, measured about its overall mean. We denote this sum by SST or S_{YY}:

$$SST = S_{YY} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2. \qquad (7.8)$$

This quantity is sometimes called the *variation in Y*. We then used OLS to choose the line in order to make the sum of the squares of the errors about the fitted line as small as possible. That sum, denoted SSE, is the variation in Y left unexplained by the model, written as

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2. \qquad (7.9)$$

In this equation, $\hat{Y}_i = b_0 + b_1 X_i$ denotes the fitted value for the ith observation and therefore, $Y_i - \hat{Y}_i$ estimates the error made in predicting the ith observation.

SSE represents the error sum of squares, where $Y_i - \hat{Y}_i$ is the error incurred in using the regression model to calculate the ith fitted value. Therefore, the smaller the value of SSE, the smaller is the average squared difference between the fitted and observed values. Further, we define the *sum of squares explained by the regression equation*, or the reduction in SST, as

$$SSR = SST - SSE.$$

It can be shown that

$$SSR = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2.$$

These sums of squares provide a basis for examining the performance of the model, as we now explain.

7.4.1 The Standard Error of Estimate

In Section 2.4, we introduced the standard deviation as a measure of variability. The average of the squared deviations about the sample mean defined the variance, and the standard deviation is then the square root of the variance. Recall that we then defined the standardized score for an observation X, or Z-score, as

$$Z = \frac{(X - \bar{X})}{S},$$

where S is the standard deviation. Provided the distribution of X is approximately normal, we may then make probability statements such as

The probability that $|Z| > 2$ is approximately 0.046.

We now seek to make similar statements about the errors (deviations from the OLS line), and to do so we define the standard error of estimate, often just called the standard error.

THE STANDARD ERROR OF ESTIMATE

The *standard error of estimate*, denoted by S, is defined as

$$S = \frac{SSE}{\sqrt{n-2}}.$$

The standard error is a key measure of the accuracy of the model and is used in testing hypotheses and creating confidence intervals and prediction intervals.

We may define the (estimated) Z-score for the *ith* observation as

$$Z_i = \frac{Y_i - \hat{Y}_i}{S}$$

and make similar probability statements for these error terms. The assumptions necessary to justify this procedure are listed in Section 7.5.1.

The denominator in the definition of the standard error is $(n - 2)$ rather than the $(n - 1)$ that we used to define the sample variance in Section 2.4. The reason for this choice is that, from a sample of n observations or degrees of freedom (DF), we use 2 DF to specify the intercept and slope. ("A straight line is defined by two points.") Intuitively speaking, this leaves $(n - 2)$ observations from which to estimate the error variance. We say that S is based upon $(n - 2)$ degrees of freedom, or $(n - 2)$ DF. We have "lost" 2 DF in estimating the slope and intercept parameters. Another way to think about this situation is that when $n = 2$, we have a "perfect fit," so it is only with more than two observations that we can hope to estimate the residual uncertainty in the model. In the next chapter, we will use models with more parameters and lose more DF. From the computer output, we find that $S = 17.61$ (recall that the dependent variable is measured in cents).

Thus, unusual values would be about 35 cents (or two standard errors) away from the expected values. Analysis of the differences between observed and fitted values shows a period from April through July 2007 when prices were higher than expected. From August through December 2008, prices were lower than expected.

7.4.2 The Coefficient of Determination

A second measure of performance is defined as the proportion of variation accounted for by the regression line. This measure is known as the *coefficient of determination* and is denoted by R^2:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}. \tag{7.10}$$

Because SSR and SSE are sums of squares, they cannot be negative. It then follows that $0 \le R^2 \le 1$.

Some of its properties are:

1. A value of $R^2 = 1$ means that the regression line fits perfectly ($SSE = 0$).
2. $R^2 = 0$ when $SSE = SST$, an equality that occurs only when every prediction $\hat{Y}_i = \bar{Y}$.
3. This condition in turn means that X is of no value in predicting Y.
4. For *simple linear regression only*, the square of the correlation (denoted by r in Section 2.5) between X and Y is equal to R^2.
5. R is known as the *coefficient of multiple correlation* and represents the correlation between Y and $\hat{Y}$. For simple regression with one explanatory variable, $R = |r|$.

In our gasoline price example, we find that $R^2 = 0.943$, or 94.3 percent. (Both formats are used and referred to as R^2, a notational quirk that needs to be remembered.) Accounting for over 90 percent of the variation seems pretty good, but how good is it? There are several parts to this question:

- Is the relationship meaningful, or could such results arise purely by chance?
- Are the assumptions underlying the model valid?

- Are the predictions from the model accurate enough to be useful?
- Are there other variables that could improve our ability to forecast Y?

R^2, THE COEFFICIENT OF DETERMINATION

R^2, the *coefficient of determination*, measures the proportion (percentage) of the variation in the dependent variable, Y, explained by the regression model.

We will defer discussion of the last three points, but even to examine the first one, we need a more formal framework than we have considered hitherto.

7.5 The Statistical Framework for Regression

In Section 7.2, we specified the fitted line in terms of an explainable (linear) pattern and an unexplained error, as given in equation (7.1). The coefficients for the fitted line were then determined with the use of OLS. No underlying theory was involved; we just applied a numerical recipe. However, in order to make statistical inferences, we must formulate an underlying statistical model, just as we did in Chapter 5 for the exponential smoothing methods developed in Chapters 3 and 4. Once the model is specified, we can use it to make probability statements about the estimated coefficients, the forecasts, and associated assessments of the accuracy of those forecasts by constructing prediction intervals. These statements are predicated on the supposition that the model is correctly and fully specified. Such an ideal state of affairs is unlikely, but we must take steps to ensure that the model we use conforms at least approximately to the assumptions we have made. Thus, statistical model building is a critical element of the forecasting process.

7.5.1 The Linear Model

The method of ordinary least squares is a pragmatic procedure for describing the relationship between variables, based upon the notion that putting a straight line through the cloud of observations provides a convenient numerical summary. We now go further and assume that there truly is an underlying linear relationship between X and Y. We recognize that different samples, or different pieces of a time series record, will lead to different OLS estimates for the slope and intercept. However, we postulate that there is a linear relationship relating the expected value of Y to X, even though individual observations will deviate from that straight line because of random errors. The structure of the assumed relationship between Y and X is illustrated in Figure 7.10.

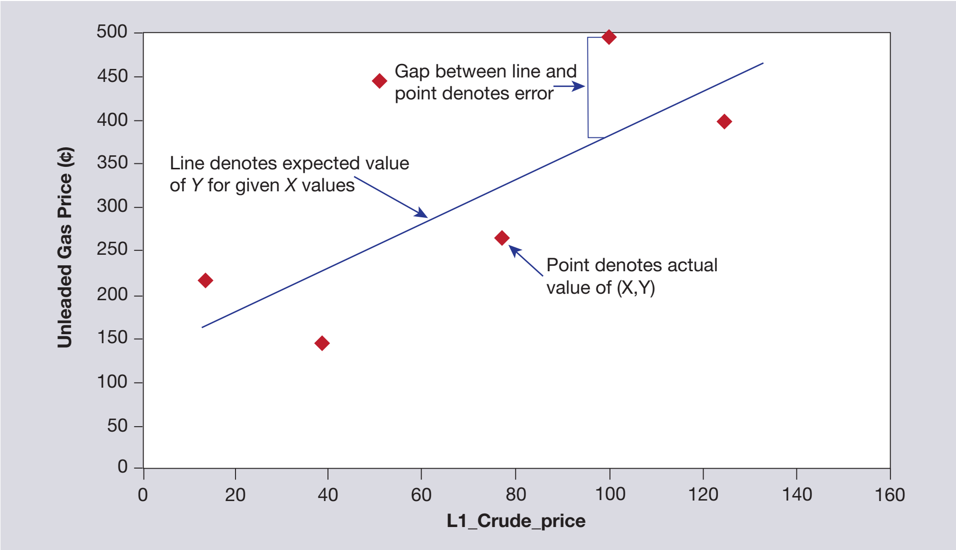
In order to develop a statistical model and use it for data analysis, we must make a set of assumptions. We express these assumptions formally as follows:

Assumption R1: For given values of the explanatory variable X, the expected value of Y given X is written as $E(Y|X)$ and has the form

$$E(Y|X) = \beta_0 + \beta_1 X.$$

That is, $E(Y|X)$ shows how X affects Y and forms the explainable, or predictable, component of the model. Here, β_0 denotes the intercept and β_1 is the slope; the values of these parameters are unknown and must be estimated from the data. The remaining assumptions provide a framework for this estimation process.

Figure 7.10 Line Shows the Expected Value of *Y* for Each *X*
(The vertical distance from the line to the observed value denotes the error.)



Assumption R2: The difference between an observed Y and its expected value is known as a *random error*, denoted by ε (epsilon). The random error represents the unexplained component of the model. Thus, the full model may be written as

$$Y = E(Y|X) + \varepsilon = \beta_0 + \beta_1 X + \varepsilon = (\textit{Expected value}) + (\textit{Random error}). \tag{7.11}$$

This specification applies *before* the observations are recorded; hence, the error term is random and we can discuss its statistical properties. After the value of Y has been recorded for a particular value of X, the corresponding error may be estimated from the fitted model; this estimate is referred to as the *residual*, denoted by e.

Assumptions R1 and R2 constitute a *linear model*. The meaning of this term is not as obvious as would first appear. Intuitively, we would expect the adjective "linear" to indicate that the expected value of Y is a linear function of X. However, in the statistical literature, a linear model is a model that is linear *in its parameters*. We will refer to a model as being "linear in X" or "linear in the parameters" whenever there is any possibility of confusion. For the present, the models we examine are linear in both senses; when we consider transformations of X and Y in Section 7.6.3, the model remains linear in the parameters, but not in X.

The reason for specifying a formal model is to enable us to use the observed data to make *statistical inferences*. Given n observations $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$, we wish to make probability statements concerning the values of the unknown parameters. To do so, we must make some assumptions about the nature of the corresponding random errors. These assumptions could be aggregated into a single statement, but we separate them out because we wish to check them separately later.

Assumption R3: The expected value of each error term is zero.

That is, there is no bias in the measurement process. For example, biases would arise if sales were systematically overreported by counting advance orders that later were cancelled.

Assumption R4: The errors for different observations are uncorrelated with other variables and with one another.

Thus, the errors should be uncorrelated with the explanatory variable or with other variables not included in the model. When we examine observations over time, this assumption implies no correlation between the error at time t and past errors. When correlation exists over time, the errors are said to be *autocorrelated*.

Assumption R5: The variance of the errors is constant.

That is, the error terms come from distributions with equal variances. This common variance is denoted by σ^2, and when the assumption is satisfied, we say that the error process is *homoscedastic*.[10] Otherwise, we say that it is *heteroscedastic*.

Assumption R6: The random errors are drawn from a normal distribution.

If we take assumptions R3-R6 together, we are making the claim that the random errors are independent and normally distributed with zero means and common variance. These four assumptions are exactly the same as those made for the state space models in Chapter 5 and the ARIMA models in Chapter 6. In the next few sections, we suppose these assumptions to be valid so that we can develop our statistical tools. However, in applications of regression analysis, each assumption must be carefully checked, and if any of them are not satisfied, corrective action must be taken. Some, of course, are more critical than others for effective forecasting. We return to the question of checking the assumptions in Chapter 8.

RANDOM ERRORS AND RESIDUALS

The *random error* represents the difference between the expected and observed values.

In a valid model, the random error for a particular observation is *completely unpredictable*, in that it is unrelated to either the explanatory variable(s) or the random errors in other observations. This is the basis of Assumption R4.

Because the intercept and slope are unknown, we must estimate the error from the fitted model; we refer to this estimate as a *residual*.

7.5.2 Parameter Estimates

In the course of our assumptions, we specified three unknown parameters: the intercept, the slope, and the error variance. The method of ordinary least squares may now be invoked in the context of our linear model. The estimates of the intercept and slope are precisely those given in equation (7.5); that is,

$$b_0 = \bar{Y} - b_1\bar{X} \text{ and } b_1 = \frac{S_{XY}}{S_{XX}}. \tag{7.12}$$

Further, the estimate for the error variance is the square of the standard error of the estimate:

$$S^2 = SSE / (n - 2). \tag{7.13}$$

As noted earlier, the denominator in equation (7.13) reflects the loss of two *degrees of freedom* (DF), corresponding to the two parameters that must be estimated to define the regression line.

10 From Greek: homo = same and scedastos = scattered; similarly hetero = different.

To make the distinction clear between the unknown parameters and the sample quantities that are used to estimate them, we summarize the notation in the following table:

Term	Unknown parameter	Sample estimate
Intercept	β_0	b_0
Slope	β_1	b_1
Error variance	σ^2	S^2

So far in this section, we have introduced three new quantities and specified how to estimate them from sample data. We do not appear to have made much real progress! However, we have established a framework that will enable us to answer questions about the value of the proposed model, and it is to these questions that we now turn.

7.6 Testing the Slope

So far, using least squares, we have estimated the regression line's parameters. But these estimates are themselves random (because they are based on random data). This situation leads us back to the question raised in Section 7.4: Is the relationship that we have estimated meaningful, or could such results arise purely by chance? We are now in a position to address that question. First of all, we formulate the question in terms of two hypotheses[11]:

The *null hypothesis* is denoted by H_0 and states that the slope (β_1) in expression (7.11) is zero; that is, there is no linear relationship between X and Y.

The *alternative hypothesis* is denoted by H_A and states that the slope is not zero; that is, there is a linear relationship between X and Y.

These two hypotheses are mutually exclusive and exhaustive; that is, only one of them can be true, and exactly one of them must be true. Algebraically, we write these competing hypotheses as

$$H_0 : \beta_1 = 0 \text{ versus } H_A : \beta_1 \neq 0.$$

One-sided alternatives may also be used if there is prior information that the slope should have a particular sign. Thus, if we wish to test for a positive slope, we consider

$$H_0 : \beta_1 \leq 0 \text{ versus } H_A : \beta_1 > 0.$$

Likewise, for a negative slope we compare the hypotheses

$$H_0 : \beta_1 \geq 0 \text{ versus } H_A : \beta_1 < 0.$$

The null hypothesis always includes an equals sign, which specifies the benchmark value of the parameter to be used in the test.

We assume the null hypothesis to be true and then formulate a test based upon this assumption. We use the test statistic

$$t = b_1 / \mathrm{SE}(b_1) \tag{7.14}$$

11 For a more detailed introduction to hypothesis testing, see Anderson *et al.* (2014, Chapter 9) or any other introductory statistics text.

where $SE(b_1)$ denotes the standard error of the sample slope. (Remember that our estimated regression line moves around, depending on the randomness in the data.) For simple regression, the algebraic form of the standard error is given in Appendix 7A, in equation (7A.7), but there is rarely any need to make direct use of this expression given available statistical software.

We refer to the observed value of t obtained from (7.14) as t_{obs}; this value is to be compared with the appropriate value[12] from Student's t-distribution with $(n - 2)$ DF. If we set the significance level equal to α, we denote the critical value for the two-sided test as $t_{\alpha/2}(n - 2)$ [and for a one-sided test by $t_\alpha(n - 2)$ with the appropriate sign attached]. The classical decision rules for these tests are as follows:

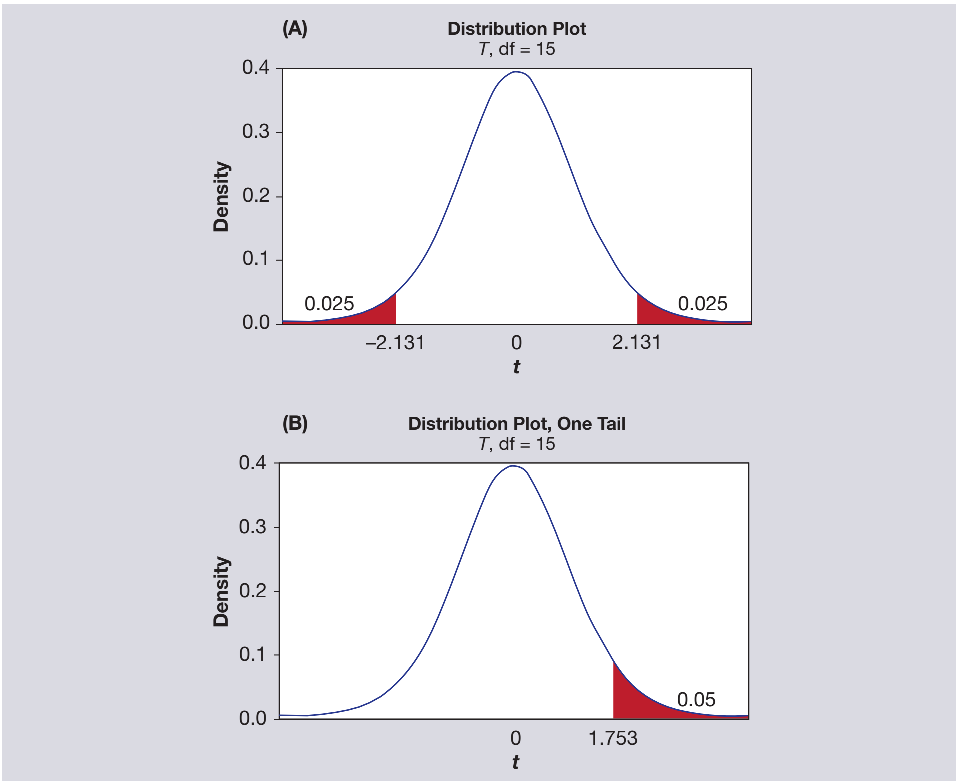
Two sided: If $|t_{obs}| > t_{\alpha/2}(n - 2)$, reject H_0; otherwise, do not reject H_0.

Positive slope: If $t_{obs} > t_\alpha(n - 2)$, reject H_0; otherwise, do not reject H_0.

Negative slope: If $t_{obs} < -t_\alpha(n - 2)$, reject H_0; otherwise, do not reject H_0.

Figure 7.11 shows the rejection region (shaded) for the t-distribution with DF = 15 and significance level $\alpha = 0.05$ for the two-tailed and positive-slope cases; the negative-slope case is just the mirror image of the positive-slope case.

Figure 7.11 The t-Distribution with DF=15: Shaded Areas Corresponding to $\alpha = 0.05$ for (A) A Two-Tailed Test and (B) A One-Tailed Test



12 Most tests are now conducted with the use of P-values, which are described in Section 7.6.1. When needed, values of the percentage points of the t-distribution may be obtained with Excel or any statistical software (see Appendix 7B).

7.6.1 *P*-Values

An equivalent procedure and one that is more convenient for computer usage is to determine the *observed significance level*, denoted by P. The P-value denotes the probability, under H_0, of observing a value of the test statistic at least as extreme as that actually observed. Most statistical software now provides P-values as part of the standard output for regression analysis; see Anderson *et al.* (2014, Section 9.3) for further details. The decision rule for the two-tailed test is then written as

If $P = Prob[|t| > |t_{\text{obs}}|] < \alpha$, reject H_0; otherwise, do not reject H_0.

For a one-tailed test, the rules are also straightforward:

Upper tail: If $P = Prob[t > t_{obs}] < \alpha$ and the sign of the coefficient is compatible with H_A, reject H_0; otherwise, do not reject H_0.

Lower tail: If $P = Prob[t < t_{obs}] < \alpha$ and the sign of the coefficient is compatible with H_A, reject H_0; otherwise, do not reject H_0.

(Note that most computer programs provide only the P-value for the two-sided test. In such cases, when a one-sided test is used, the P-value in the output should be divided by 2 before making the comparison with α.)

Use of the P-value has two advantages; first, we can perform the test by inspecting this value without the need to enter a significance level into the program. Second, the rule just stated is "one size fits all," in that virtually all the tests we consider may be expressed in this format. In view of the importance of this result, we provide the output from both Excel, Minitab and R in the next example, by way of illustration. Other programs produce similar outputs.

HYPOTHESIS TESTING WITH *P*-VALUES

Tests based upon P-values are completely equivalent to those which use the same statistics with reference to tables of the t-distribution.

■ Example 7.4: A test of the slope

The Excel, Minitab, and R outputs for the simple example given earlier as Example 7.2 are shown in Figure 7.12. From the output, we see that $t_{\text{obs}} = 3.10$ and the P-value for the slope is 0.036. (The numbers in the Excel output have been rounded for ease of comparison; in addition, we never need so many decimal places in the final answer!) We set $\alpha = 0.05$ and conduct a two-tailed test. Because $P = 0.036 < \alpha = 0.05$, we reject H_0 and conclude that there is evidence in favor of a relationship between X and Y. In general, it is much easier to use P-values, and we shall usually do so without further comment. ■

Note 1: As may be seen from the computer output, it is also possible to set up a test for the intercept. Such tests, however, rarely provide any useful information, because the value $X = 0$ is usually well outside the range of applications. Further, the value of the intercept can be changed by shifting the origins of the X and Y measurements. Thus, an intercept test is useful *only* when we wish to test that the regression line goes through a prespecified origin.

Note 2: A $100(1 - \alpha)$ percent confidence interval for the slope may be established as $b_1 \pm t_{\alpha/2}(n-2) \times SE(b_1)$. The t value in this expression must be obtained from the t-distribution. Thus for the data in this example, the 95 percent CI for the slope is $1.26 \pm (2.78) \times (0.4069) = [0.13, 2.39]$. Because the sample size is so small, we do not have an accurate assessment of how X affects Y, even though the test tells us that there appears to be a relationship.

Figure 7.12 Excel, Minitab, and R Outputs for Simple Example

(A) Excel Output

	A	B	C	D	E	F	G	H	I
1									
2	*Regression Statistics*								
3	Multiple R	0.84							
4	R Square	0.706							
5	Adjusted R Square	0.632							
6	Standard Error	2.349							
7	Observations	6							
8									
9									
10	ANOVA								
11						*Significance*			
12		*df*	*SS*	*MS*	*F*	*F*			
13	Regression	1	52.92	52.92	9.587	0.036			
14	Residual	4	22.08	5.52					
15	Total	5	75						
16									
17			*Standard*			*Lower*	*Upper*		
18		*Coefficients*	*Error*	*t Stat*	*P-value*	*95%*	*95%*		
19	Intercept	4.36	2.5	1.75	0.156	−2.57	11.29		
20	X	1.26	0.41	3.1	0.036	0.13	2.39		
21									
22									

(B) Minitab Output

Regressional Analysis: Y versus X				
The regression equation is Y = 4.36 + 1.26X				
Predictor	**Coef**	**SE Coef**	**T**	**P**
Constant	4.360	2.498	1.75	0.156
X	1.2600	0.4069	3.10	0.036
S = 2.34947 R-Sq = 70.6% R-Sq(adj) = 63.2%				
Analysis of Variance				
Source	DF	SS	MS	P
Regression	1	52.920	9.59	0.036
Residual Error	4	22.080	5.520	
Total	5	75.000		

(C) R Output and Commands

The R code is as follows (see online Appendix C: *Forecasting in R: Tutorial and Examples*)

fit <- lm (Y ~ X) : #Create Regression Model

summary(fit))$coefficients[,4] : #View Summary Output

with output

```
Call:
lm(formula = Y ~ X)
Residuals:
    1        2        3        4        5        6
  1.12    -2.40     1.84    -1.92     2.56    -1.20
Coefficients:
                 Estimate    Std. Error     t value      Pr(>|t|)
(Intercept)    4.3600        2.4975        1.746        0.1558
X              1.2600        0.4069        3.096        0.0364 *
---
signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 2.349 on 4 degrees of freedom
Multiple R-squared: 0.7056,  Adjusted R-squared: 0.632
F-statistic: 9.587 on 1 and 4 DF, p-value: 0.3635
```

Note 3: For the time being, we ignore the *Analysis of Variance* (ANOVA) table. For simple regression, it provides no new information relative to the two-tailed t-test, as can be seen from the fact that the t-test and ANOVA test have the same *P*-values. We will discuss this further in Section 8.3.

NUMERICAL ACCURACY AND PRESENTATION OF RESULTS

Using a large number of decimal places to ensure numerical precision in the calculations is desirable. However, once the calculations are complete, we must decide how many decimal places to retain for presentation purposes.

■ Example 7.5: Test and confidence interval for gasoline prices: 1996–2008

The simple model of gasoline prices as it depends on the lagged price of crude is:

$$Y_t = \beta_0 + \beta_1 L1\_Crude\_price_t + \varepsilon_t.$$

The relevant estimates for the parameters are as follows:

Predictor	Coef	SE Coef	T	P
Constant	68.760	2.591	26.54	0.000
L1_Crude_price	2.7069	0.05367	50.44	0.000
S = 17.6121 R-Sq = 94.3% R-Sq(adj) = 94.3% N = 155 $\bar{Y}$ = 177.8				

Note that the *P*-value is quoted to only three decimal places, so the output does not imply that *P* is exactly zero; rather, it means that $P < 0.0005$, which rounds down to 0.000 to three decimal places. Here we reproduce the computer output exactly as it appears. Elsewhere in the book for presentation purposes, we simplify the output to an appropriate number of significant figures. In this example, we expect a positive slope, so a one-sided upper-tail test is appropriate. Using $\alpha = 0.05$ (or any other plausible value!), we clearly reject H_0 and conclude that there is a strong positive relationship between the two variables, as expected. As noted earlier the intercept might conceivably be interpreted as the transport and marketing costs of gasoline, but this interpretation is a stretch when the smallest observed value of *L1_Crude_price* is so far from the origin.

To determine the 95 percent confidence interval for the slope, we first obtain the critical value from the *t*-distribution to get $t_{025}(153) = 1.975$, very close to the limiting value of 1.960 obtained from the normal distribution. The confidence interval for the slope is given by $2.7069 \pm (1.975) \times (0.05367) = [2.60, 2.81]$. That is, an increase of $1 in the price of a barrel of crude oil produces an *expected* increase of 2.6 to 2.8 cents per gallon of gasoline at the pump. In this case, the large sample provides a clear idea of the relationship between *X* and *Y*. ■

7.6.2 Interpreting the Slope Coefficient

We have shown that when the slope coefficient is significant we can sensibly regard *X* as a suitable explanatory variable in our model. As before, suppose our linear model for the dependent variable is

$$Y = \beta_0 + \beta_1 X + \varepsilon.$$

That is, *X* has a linear impact on *Y*. Specifically, a unit change in *X* produces an impact of β_1 in *Y*. So, does the size of β_1 measure the importance of *X* in forecasting *Y*? No, because a change of units in either *Y* or *X* will affect the estimate. In our gasoline example, the price of unleaded is measured in cents. If we changed the units of measurement to dollars, this would affect β_1 and give a revised estimate of $\beta_1/100$ (and similarly with β_0). However, the tests in the previous section would again show the same P-value and R^2, and, of course, the same predictions, but this time in dollars.

DISCUSSION QUESTION: *What is the impact on the standard error of a change in the units in which X is measured? In which Y is measured? Does R^2 change?*

An easily interpretable measure that eliminates the effects of the units used in measuring both X and Y is the *elasticity*, which measures the responsiveness of changes in Y to changes in X. For example, if the elasticity of the price of unleaded to the price of crude was 1, then a doubling of the price of crude would lead to a doubling of the expected price of unleaded at the pump.

ELASTICITY

The *elasticity* is defined as the proportionate change in Y relative to the proportionate change in X and is measured by

$$\frac{\Delta E(Y\,|\,X)}{E(Y\,|\,X)} \bigg/ \frac{\Delta X}{X}$$

$$\text{For the linear model} = b_1 \frac{X}{E(Y\,|\,X)} = \frac{b_1 X}{b_0 + b_1 X} \tag{7.15}$$

where ΔY is the change in Y and ΔX is the change in X.

In November 2008, the price of crude was \$57.31 per barrel of oil. From that dollar amount, together with the expected December price of unleaded, \$2.239, the elasticity of the price of unleaded to the price of crude is estimated to be

$$b_1 \frac{X}{E(Y\,|\,X)} = (2.707)\frac{57.31}{223.9} = 0.69.$$

A little mathematical manipulation shows that as X increases, the elasticity also increases, since $b_0 > 0$. This result seems counterintuitive and suggests that we should modify our analysis; accordingly, we now explore a logarithmic transformation, which leads us to a simpler, more intuitive elasticity.

7.6.3 Transformations

The increased fluctuations in the later part of the series, when the price of crude oil is higher, suggest that some kind of transformation is desirable. Two possibilities come to mind:

1. A logarithmic transformation, as in Chapters 3 and 4;
2. Conversion from current dollars to real dollars through division by a price index.

These two options are not mutually exclusive. We pursue the second option and their combination in Minicase 7.1 but now examine the first possibility, letting $\ln Y$ and $\ln X$ denote the logarithms of the two variables of interest. The output is as follows:

Predictor	Coef	SE Coef	T	P
Constant	2.8971	0.03771	76.83	0.000
ln_L1_Crude_price	0.62815	0.01056	59.46	0.000
S = 0.07631 R-Sq = 95.9% R-Sq(adj) = 95.8%				

Transforming back to the original units, we have

$$Unleaded = \exp(2.8971) \times (L1\_Crude\_price)^{0.62815}$$

Using the definition of elasticity given in Equation 7.15, the log-log model has a constant elasticity, given by the coefficient of the slope.[13] That is, the elasticity is $b_1 = 0.628$ for all values of X, implying that an increase of 1 percent in the price of crude leads to an increase of about 0.63 percent in the price at the pump. This value is somewhat lower than that obtained in the previous analysis. The key difference is that for the log transformed model, the elasticity is constant. But before we can make a reasoned choice between the two models, we need to develop diagnostic procedures, a task we defer to Section 8.5.

7.7 Forecasting Using Simple Linear Regression

Once we have established our statistical model, we are ready to use it for forecasting. With X as the explanatory variable in the model, we must first establish values of X for which the model is to be used to forecast future time periods. Similarly, when we forecast from a cross section to new members of the population, again we need to establish the relevant X values. For simplicity, we present the discussion in terms of a single future value, although multiple forecasts will often be needed. The value of X may arise in any of three ways:

a. X is known ahead of time (e.g., the size of a sales force, or demographic details concerning a consumer).

b. X is unknown but can still be forecast (e.g., gross domestic product).

c. X is unknown, but we wish to make what-if forecasts (e.g., the effects of different advertising or pricing policies).

Consider the gasoline prices example that we have used throughout this chapter. We deliberately used X with a one-month lag so that the value of X would be available when we wish to make forecasts *one month ahead*. However, if we forecast gasoline prices two or more months ahead, we must either invoke case (b) and provide forecasts of the price of crude oil or construct new regression models that progressively use X lagged by two, three, or more months. For the present, we stay with forecasts one period ahead. The use of models with longer lags is explored in Section 7.7.4.

Finally, case (c) represents an important use of regression models whereby several alternative values of the input may be considered and the effects on Y examined. For example, we might specify different possible prices for a barrel of crude oil and examine the subsequent changes in price at the pump as a guide to decision making.

7.7.1 The Point Forecast

If we are given the value of the input for the next time period as X_{n+1} and we have estimated the regression line, the point forecast (F_{n+1}) is

$$F_{n+1} = b_0 + b_1 X_{n+1} \tag{7.16}$$

13 For the model, $\ln Y = b_0 + b_1 \ln X$, differentiating gives

$$\frac{\Delta Y}{Y} \bigg/ \frac{\Delta X}{X} = b_1$$

Many people and texts refer to equation (7.16) simply as *the* forecast. However, it is important to recognize that is the *point forecast* and that statistical models provide much more information and greater insights than just one number. In particular, the regression model may be used to generate prediction intervals, as we show in the next section.

Note: We have used $\hat{Y}$ to denote the fitted values in the regression line, as in equation (7.6), and now we use F_{n+1} to denote the forecast in equation (7.16); the two formulas are the same, so why the different notation? The reason is that the *fitted* values correspond to those observations which were used in the estimation process. By contrast, forecasts pertain to new observations that were not used to estimate the model parameters.

■ Example 7.6: Forecasting gasoline prices

The model for gasoline prices was fitted with the use of data from January 1996 through December 2008. We now generate one-step-ahead forecasts, using equation (7.16) for the 24 months from January 2009 to December 2010 so that, for example, the forecast for May 2009 uses the crude price for April 2009 and the calculation is as follows:

$$X_{n+1} = 49.65, \text{ and } F_{n+1} = 68.760 + 2.7069 \times 49.65 = 203.2$$

The first few point forecasts and the overall summary measures are given in Table 7.3. The various error measures are computed in accordance with the formulas in Section 2.7.

Table 7.3 Selected One-Step-Ahead Forecasts for Gasoline Prices, With Prediction Intervals and Overall (*24-Month*) Summary Measures *(January 2009–December 2010)*

Forecast Month	Actual Price	L1_Crude_price	Forecast	95% PI, Lower Limit	95% PI, Upper Limit
Jan-2009	178.8	41.71	180.1	145.2	215.0
Feb-2009	192.3	39.09	181.7	146.8	216.6
Mar-2009	195.9	47.97	174.6	139.7	209.5
Apr-2009	204.9	49.65	198.5	163.6	233.4
…	…	…	…	…	…
	ME	1.9		RMSE	12.6
	MAE	9.2		MAPE	3.73%

Data: Gas_prices_1.xlsx

The full set of forecast and actual values appears in Figure 7.13. The forecasts lagged the upturn in prices from March to June 2009 but otherwise did quite well. The *RMSE* over the hold-out period was 12.6, consistent with the standard error for the model ($S = 17.6$). ■

7.7.2 Prediction Intervals

Given that our forecasts are not going to be exact, what can we say about their likely accuracy? This is the critical question that arises in building a forecasting model. The random variation in the forecasts comes from three sources: the inherent variation of the process (i.e., the error term), the uncertainty caused by needing to estimate the unknown parameters, and misspecification of the model. For the time being, we assume that assumptions R3 through R6 are valid, so we focus only on the first two of these sources.

If we denote the future (unknown) value of the dependent variable by Y_{n+1} and the forecast of that value based upon the estimated regression line by F_{n+1} then the overall mean square error for the forecast in relation to the future value is

$$\text{var}(Y_{n+1} - F_{n+1}) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_{n+1} - \bar{X})^2}{\sum_1^n (X_i - \bar{X})^2} \right]. \tag{7.17}$$

The details are given in Appendix 7A. Note that the term on the left-hand side of equation (7.17) is written as the variance rather than the mean square error because the forecast is unbiased when the linear model is correct. Given that we must estimate σ^2 by S, the estimated variance is

$$\widehat{var}(Y_{n+1} - F_{n+1}) = S^2 \left[1 + \frac{1}{n} + \frac{(X_{n+1} - \bar{X})^2}{\sum_1^n (X_i - \bar{X})^2} \right].$$

Finally, we define the standard error of the forecast as

$$SE(Y_{n+1} - F_{n+1}) = \sqrt{\widehat{var}(Y_{n+1} - F_{n+1})}.$$

Now, recall from Section 7.5 that assumptions R3-R6 combine to assert that the error terms follow a normal distribution with zero mean and variance σ^2. It can be shown that the forecast error then follows a normal distribution with variance given by equation (7.17). Because we must estimate σ by S, we replace the normal distribution by Student's t with the appropriate DF. Putting all this together, we arrive at the prediction interval:

$$F_{n+1} \pm t_{\alpha/2}(n-2) \times SE(Y_{n+1} - F_{n+1}). \tag{7.18}$$

The 100(1 – α) percent prediction interval is a probability statement. It says that the probability that a future observation will lie in the interval defined by equation (7.18) is (1 – α). For example, if we set (1 – α) = 0.95, then, given that the price of crude (X) for April 2009 is 49.65, the prediction interval for May 2009 is 203.2 ± 1.975 × 17.67 = 203.2 ± 34.9. The detailed calculation is shown in the next example. We interpret this result as saying: With probability 0.95, the actual gasoline price will lie in the range from \$1.68 to \$2.38 per gallon when the previous month's crude price was \$49.65.

PREDICTION INTERVAL FOR THE FUTURE OBSERVATION, Y_{n+1}

The *prediction interval* for a forecast measures the range of likely outcomes for the unknown actual observation, for a specified probability and for a given X value.

Example 7.7: Calculation of a prediction interval

We continue our consideration of the forecasts for May 2009, begun in Example 7.6. The various numbers we need are

$$n = 155,\ X_{n+1} = 49.65,\ F_{n+1} = 203.2,\ S = 17.61,\ \bar{X} = 40.44,\ \sum (X_i - \bar{X})^2 = 361163.6$$

We now substitute into equation (7.16) to obtain

$$\begin{aligned} SE(Y_{n+1} - F_{n+1}) &= S \left[1 + \frac{1}{n} + \frac{(X_{n+1} - \bar{X})^2}{\sum_1^n (X_i - \bar{X})^2} \right]^{1/2} \\ &= 17.61 \left[1 + \frac{1}{155} + \frac{(49.65 - 40.44)^2}{361163.6} \right]^{1/2} \\ &= 17.61[1 + 0.00645 + 0.000235]^{1/2} \\ &= 17.67 \end{aligned}$$

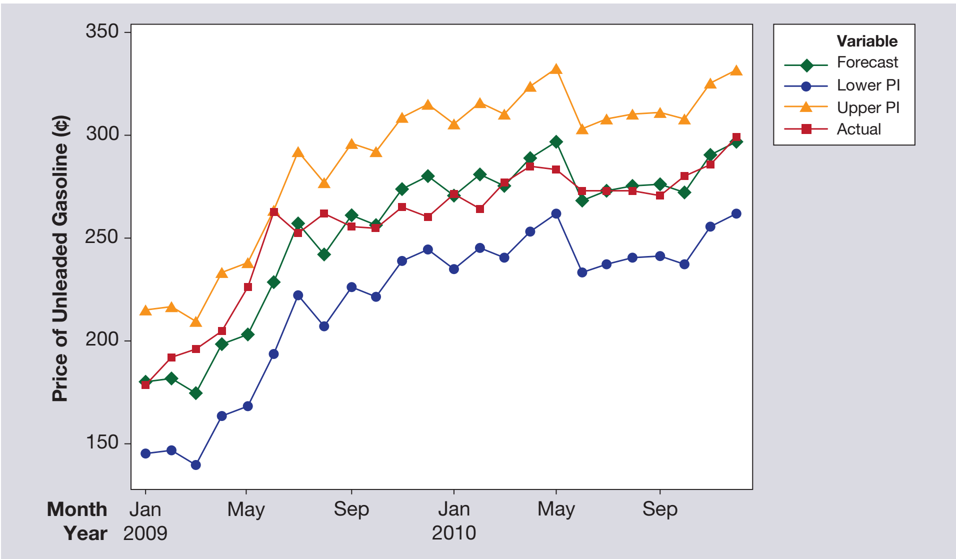
Because the sample is large, the extra terms have a very slight effect on the SE. Returning to equation (7.18), we obtain the 95 percent prediction interval by using $t_{0.025}(153) = 1.975$:

$$203.2 \pm 1.975 \text{ x } 17.67 = 203.2 \pm 34.9 = [168.2, 238.1] \quad \blacksquare$$

In practice, the prediction intervals are evaluated with the use of standard software, which implements the analysis just described. The steps in Minitab, SPSS and R are summarized in Appendix 7C. Excel does not provide the intervals as a standard function.

The actual and one-step-ahead forecast values and the 95 percent lower and upper prediction interval limits for January 2009 through December 2010 are shown in Figure 7.13. All 24 actual values fall within the PI, which might be expected when one is using 95 percent prediction intervals. However, the intervals are of the order of ±35 cents, which is too wide to be very useful. In Chapter 8, we develop improved models that narrow the PIs.

Figure 7.13 Plot of Actual Values With One-Step-Ahead Forecasts and 95% Prediction Intervals for Price of Unleaded Gasoline *(January 2009–December 2010)*



Data: Gas_prices_1.xlsx; adapted from Minitab output.

7.7.3 An Approximate Prediction Interval

The previous example suggests that when n is reasonably large, the term in square brackets in equation (7.17) is close to 1.0. This result is generally true, and as a quick calculation, the 95 percent prediction interval may be reasonably approximated by $F_{n+1} \pm 2S$. This approximation gives us an easy way of interpreting whether our predictions are useful. Effectively, the standard error S measures the predictive accuracy of our model. When $Y > 0$, as is true for most business variables, the comparison of S with $\bar{Y}$ is a rough measure of a model's accuracy. With the gasoline example

$$\frac{S}{\bar{Y}} = \frac{17.67}{177.8} = 0.099$$

which suggests a prediction interval of around 20 percent of the average price level. However, the increases in price in recent years mean that using the overall average is not as useful as relating S to the most recent data.

DISCUSSION QUESTION: *Ideally, prediction intervals should be narrow. How would you decide whether a prediction interval was sufficiently narrow to be useful? How might you recognize that an interval was too narrow?*

7.7.4 Forecasting More than One Period Ahead

Suppose we were living at the end of December 2008 and wished to make forecasts further ahead than January 2009. There are two[14] possible approaches:

1. Generate forecasts for X and apply these to the original model.
2. Reformulate the model so that X is lagged as appropriate.

The first approach is more commonly applied, but suffers from the drawback that the uncertainty in X is not reflected in the prediction intervals for the forecasts. It has the advantage that different what-if paths for X can be formulated and compared using the same model. The second approach is somewhat more tedious, but it will be more valuable when good forecasts for X are unavailable. It has the advantage of providing accurate prediction intervals, especially when forecasts for X are inaccurate. However, the relationship between Y and X will be weaker for longer lags.

■ Example 7.8: Forecasting gasoline prices two periods ahead

To illustrate the preceding two points, we generated two-step-ahead forecasts for gasoline prices by each method:

- We generate one-step-ahead forecasts for the crude oil price series and then insert these forecasts into equation (7.7). The forecasts of crude prices are produced by means of the additive Holt-Winters model (see Section 4.4).
- We fitted a regression model for the crude price, lagged two periods.

The results are summarized in Table 7.4. The following points may be noted:

i. The standard error increases and R^2 falls when we move to two steps ahead, as expected.
ii. Forecast performance also falls off when we move to two steps ahead.
iii. The two sets of forecasts are broadly comparable in performance. (The second method does somewhat better here, but we should not read much into such a finding from just one example.)
iv. The prediction intervals for one step ahead and for two steps ahead with method 2 seem to provide the right level of *coverage*; that is, we would expect 95 percent of future observations to fall within the 95 percent prediction intervals. The prediction intervals for method 1 appear to be too narrow, with 4 out of 24 (16.7% instead of 5%) values lying outside the 95 percent limits for the two-step-ahead predictions.

The first two results are as expected and the third is reasonable. However, point (iv) is often overlooked in applications and can lead to overoptimistic conclusions about forecast accuracy, as assessed by the width of the prediction intervals if the uncertainty in forecasting X is not taken into account. ■

14 We may also use vector autoregressive (VAR) models, which we discuss in Section 10.5

Table 7.4 Summary Results for One-Step-Ahead and Two-Steps-Ahead Forecasts for Gasoline Prices

Statistic	One Step Ahead (from Table 7.3)	Two Steps Ahead: Lag-1 Model + Forecast	Two Steps Ahead: Lag-2 Model
Intercept		68.8	73.0
Slope		2.71	2.62
R^2		94.3	88.6
S		17.6	25.0
ME	1.9	6.4	6.6
MAE	9.2	17.9	14.5
RMSE	12.6	23.5	20.2
MAPE	3.73%	8.64%	6.49%
MWPI	70	70	100
# outside PI	0/24	4/24	1/24

Data: Gas_prices_1.xlsx

[*Note:* MWPI = mean width of the prediction interval, the difference between the upper and lower limits in equation (7.18).]

7.8 Forecasting Using Leading Indicators

When we are developing a forecasting system, we must always ask, "What do we know and when do we know it?" An economic model would naturally relate current sales to the current value of the *gross domestic product* (GDP). However, the figure for current GDP will be available only well after the current quarter is over, so it is of limited value for *forecasting* sales over the next few weeks.

An alternative approach is to seek a suitable *leading indicator* — that is, a series produced early enough to be usable and reflective of likely future movements in sales. One such series is the Index of Consumer Sentiment, produced by the University of Michigan. A second major source of this kind of information is the Conference Board, which produces several such indexes: Consumer Confidence, CEO Confidence, U.S. Leading Indicators, and Help Wanted Advertising. (For further details, see *www.conference-board.org*.). Another natural leading indicator is the stock market, whether a general index like Dow Jones for the economy as a whole or, an industry-level index for a particular sector.

In each case, we are seeking, not a causal relationship, but rather an indicator that is published in timely fashion. Indeed, the causal pattern will often be of the form that some factor (say, Z) causes movements both in the leading indicator (LI) and in the variable of interest (Y). If Z is observed in time to make the forecast, clearly we will use it. If Z is unavailable, but LI is, we use LI. The forecasts may not be as effective as they would have been if Z were available, but LI will generally provide some improvements over forecasting with no knowledge of the inputs. We follow up on these ideas in Minicase 7.2 at the end of the chapter.

Summary

This chapter introduced the notion of modeling the relationship between two variables. After a discussion of the distinction between correlation and causation, we developed an estimation procedure by using the method of ordinary least squares and we went on to develop a statistical framework for the evaluation of regression models. We ended the chapter with a discussion of forecasting methods that provide both point and interval forecasts.

We do not discuss the relevant forecasting principles at this stage, but prefer to wait until the end of Chapter 8, when the full multiple regression model, which can include more input variables, has been developed.

Exercises

7.1 A soft-drink company monitored its television advertising over an eight-week period to evaluate the effect on sales (in millions of dollars) with respect to the number of 30-second spots aired in that week. Estimate the regression equation. The data are as follows:

Week	1	2	3	4	5	6	7	8
No. of spots	8	12	16	10	8	12	16	10
Sales	25	34	39	32	22	30	43	31

Data: Exercise_7_1.xlsx

a. Estimate the regression line for sales on spots.

b. Test whether the slope is significantly different from zero, using $\alpha = 0.10$.

c. Compute S and R^2 and interpret the results.

d. The company has reserved 20 spots for week 9. Forecast the sales and construct a 90 percent prediction interval.

e. Comment on the level of accuracy this analysis provides.

7.2 A company experiments with different price settings over a 12-week period. The weekly sales and revenue figures are shown in the following table:

Week	1	2	3	4	5	6	7	8	9	10	11	12
Price	6	8	10	6	8	10	6	8	10	6	8	10
Sales	28	30	28	30	24	22	34	26	20	36	32	26
Revenue	168	240	280	180	320	220	204	208	200	216	256	260

Data: Exercise_7_2.xlsx

a. Estimate the regression lines for sales on price and for revenues on price.

b. Test whether the slopes are significantly different from zero, using $\alpha = 0.05$.

c. Compute R^2 and interpret the results.

d. Why does the line for sales fit the data better than the line for revenues?

7.3 A linear time trend may be estimated by means of simple linear regression for Y on time.

a. Estimate the time trend for the sales data given in Exercise 7.1.

b. Compute S and R^2 and interpret the results.

c. Test whether the slope is significantly different from zero, using $\alpha = 0.05$.

d. Which model is more useful, the regression of sales on spots or that of sales on time? Give reasons for your answer.

7.4 Estimate a linear time trend for the Netflix data (*Netflix_2.xlsx*), using data for the period 2000Q1 to 2012Q4.

a. Carry out a preliminary analysis of the data and interpret your results.

b. Estimate the time trend.

c. Test whether the slope is significantly different from zero, using $\alpha = 0.05$.

d. Compute S and R^2 and interpret the results.

e. Compute point forecasts for sales in each quarter of 2013 through 2015, and construct 95 percent prediction intervals.

f. Comment upon your results.

7.5 Consider the data on rail safety previously examined in Exercise 2.2 (*Rail_safety.xlsx*).

a. Carry out a preliminary analysis of the data and interpret your results.

b. Estimate the regression line for injuries on train miles.

c. Test whether the slope is significantly different from zero, using $\alpha = 0.05$.

d. Compute S and R^2 and interpret the results.

e. Compute a point forecast for the number of injuries, given an estimated level of train miles equal to 100. Construct a 95 percent prediction interval.

f. Is the forecast likely to be useful? Comment upon your results.

7.6 Develop a simple linear regression model for the median price of a new home as a linear function of time, using the price data in *Housing_2.xlsx*, downloaded from FRED *(https://fred.stlouisfed.org/)* which shows monthly prices from January 1981 through December 2015. (Prices are quoted in dollars and include the price of the land.) Use the data through December 2012 to estimate the parameters and then predict monthly sales for 2013–2015. (Refer to Appendix 7C.) Generate 90 percent prediction intervals for your forecasts and compare your forecasts with the actual values. What other variables might be useful for forecasting purposes?

7.7 Develop a simple linear regression model for the number of new housing starts as a linear function of the mortgage rate, using the data on housing starts in *Housing_2.xlsx*, downloaded from FRED (*https://fred.stlouisfed.org/*). The data relate to monthly housing starts (in thousands, seasonally adjusted) and mortgage rates (end-of-month 30-year percentage rate) from January 1981 through December 2015. Using the data through December 2012, estimate the parameters and then predict monthly sales for 2013 through 2015, using the actual mortgage rates as inputs. Generate 90 percent prediction intervals for your forecasts and compare your forecasts with the actual values. Comment upon your results and then refit the model using the data through December 2015. Compare the two sets of parameters. Interpret the results.

7.8 Reanalyze the gasoline prices data, using the data in *Gas_prices_1.xlsx* after transforming both variables to logarithms, as in Section 7.6.3. Generate forecasts for the hold-out sample, and convert the point forecasts and prediction interval limits back to the original units. Compare your results with the model developed earlier that does not use any transformations. Which model would you recommend?

Minicases

Minicase 7.1 Gasoline Prices Revisited

Use the data in *Gas_prices_1.xlsx*, (or go to the U.S. Department of Energy website, *www.eia.doe.gov*, and download the data on the monthly series for retail gasoline prices and lagged crude oil prices.)

1. Using data up to December 2012, re-estimate the model, contrasting your results with those discussed in the text. Comment on your results.
2. Use the data for 2013–2015 as the hold-out sample to test the forecasting performance of your model.
3. Repeat the analysis, using the log transformed series, and compare your results.
4. Use the Consumer Price Index (*CPI*) to create real (constant-dollar) prices rather than observed prices for both variables. Repeat the analysis described in steps 1 and 2. Does this approach lead to improved forecasts? (This question is harder than it looks!) What if any economic arguments suggest using a constant-price model?
5. If you were to conduct the analysis using weekly data from the same source, how long a lag would you use?
6. In some countries, tax on fuel is a major component of the retail price of gasoline. How would you include the effects of changes in the tax?

Summarize your findings.

Minicase 7.2 Consumer Confidence and Unemployment

Measures of consumer confidence are readily available from sample surveys and often serve as useful leading indicators when data on causal macroeconomic variables are not yet available. Use the data in the spreadsheet *Unemp_conf_2.xlsx* for January 1981 to December 2012 to estimate the regression for U.S. unemployment on the University of Michigan Consumer Confidence Index. Unemployment is seasonally adjusted but the Index is not. Try the current month as the predictor variable and also try the Index of Consumer Confidence at lags one, two and three. Which model provides the best fit?

Generate forecasts for each month of 2013–2015 using each model. Compare the forecasting performance of each model, using the measures developed in Chapter 2. Summarize your findings.

Minicase 7.3 Baseball Salaries Revisited

Use the data in *Baseball.xlsx*. It is evident from Figure 7.6 that salaries start to level off or even decline once players' careers extend beyond 12 years or so. Also, the spread of salaries increases considerably as the number of years played increases. To allow for these features of the data,

1. Eliminate players with 12 or more years of experience from the data set and rerun the analysis. Compare your results with those reported in Example 7.3.
2. Transform the salaries by taking logarithms and determine whether the regression assumptions outlined in Section 7.5 seem to be better approximations to the transformed data.

3. Generate 95 percent prediction intervals for all players, and check to see how many actual salaries fall within those intervals.

What conclusion do you draw from eliminating the unusual observations?

References

Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D. and Cochran, J. J. (2014), *Statistics for Business and Economics*, 12th ed. Mason, OH: Cengage Learning.

Kutner, M., Nachtsheim, C., Neter, J. and Li, W. (2005). *Applied Linear Statistical Models*, 5th ed. New York: McGraw-Hill. [Reissued as paperback in 2013]

Appendix 7A Derivation of Ordinary Least Squares Estimators[15]

In Section 7.2, we stated that the regression line relating the dependent variable Y to the explanatory variable X should be determined by the method of ordinary least squares. That is, we choose the intercept b_0 and the slope b_1 so as to minimize the quantity given in equation (7.2):

$$SSE = (Observed - Fitted)^2 = \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2.$$

Here, SSE denotes the sum of squared errors, with the sum taken over all n observations. A feature of SSE is that it is quadratic in the coefficients $\{b_0, b_1\}$, so that obtaining the partial first derivatives with respect to these coefficients and setting the partial derivatives to zero yields a pair of linear equations. That is, if we set

$$\frac{\partial(SSE)}{\partial b_0} = 0 \text{ and } \frac{\partial(SSE)}{\partial b_1} = 0 \tag{7A.1}$$

we arrive at the equations

$$\begin{aligned} &\sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i) = 0, \\ &\sum_{i=1}^{n} X_i(Y_i - b_0 - b_1 X_i) = 0. \end{aligned} \tag{7A.2}$$

We now define the sample means of the two variables as

$$\bar{X} = \frac{\sum X_i}{n} \text{ and } \bar{Y} = \frac{\sum Y_i}{\text{n}} \tag{7A.3}$$

and the sums of squares and cross products as

$$S_{XX} = \sum(X_i - \bar{X})^2, S_{XY} = \sum(X_i - \bar{X})(Y_i - \bar{Y}), \text{ and } S_{YY} = \sum(Y_i - \bar{Y})^2. \tag{7A.4}$$

15 This appendix provides only a brief outline of the derivation. For a complete treatment, see a standard text on regression analysis, such as Kutner, Nachtsheim, Neter, and Li (2005).

We then simplify equation (7A.4), using equations (7A.2) and (7A.3), to arrive at

$$b_0 = \bar{Y} - b_1\bar{X} \text{ and } b_1 = \frac{S_{XY}}{S_{XX}}. \tag{7A.5}$$

The standard error is estimated as

$$S = \frac{SSE}{\sqrt{(n-2)}}. \tag{7A.6}$$

The estimated standard error of the slope is given by

$$SE(b_1) = \frac{S}{\sqrt{(S_{XX})}}. \tag{7A.7}$$

Further, it may be shown that the solution given in equation (7A.2) produces a minimum for *SSE* and that this minimum value is

$$SSE = S_{YY} - b_1^2 S_{XX}. \tag{7A.8}$$

Equation (7A.8) is just a restatement of the partition of the sum of squares given in Section 7.4 and may be rewritten as $SSE = (1 - R^2)S_{YY}$.

Extension to K explanatory variables

The discussion thus far relates to simple regression, with estimates given by two linear equations in two unknowns, as we can see in equation (7A.2). When we turn to multiple regression in the next chapter, we include *K* predictor variables in the model and *SSE* becomes

$$SSE = (Observed - Fitted)^2 = \sum_{i=1}^{n} (Y_i - b_0 - b_1X_{1i} - \ldots - b_KX_{Ki})^2.$$

The partial first derivatives lead to the $(K + 1)$ linear equations

$$\begin{aligned} &\sum_{i=1}^{n} (Y_i - b_0 - b_1X_{1i} - \ldots - b_KX_{Ki}) = 0, \\ &\sum_{i=1}^{n} X_{ji}(Y_i - b_0 - b_1X_{1i} - \ldots - b_KX_{Ki}) = 0, \quad j = 1, 2, \ldots, K. \end{aligned} \tag{7A.9}$$

Equations (7A.9) reduce to equations (7A.2) when $K = 1$. However, for any value of *K*, we may still write

$$SSE = (1 - R^2)S_{YY}. \tag{7A.10}$$

That is, R^2 denotes the proportion of variance explained by the model. To emphasize, equations (7A.5) through (7A.8) apply only when $K = 1$, but equations (7A.9) and (7A. 10) apply for any value of *K*.

Prediction intervals

The uncertainty in the forecasts depends on the difference between the forecasted value (F_{n+1}) and the actual value of *Y*. If we knew the underlying model, a (future) new observation with the explanatory variable taking on the value X_{n+1} would be given by

$$Y_{n+1} = \beta_0 + \beta_1X_{n+1} + \varepsilon_{n+1}.$$

However, we had to estimate the intercept and slope, so the forecast is

$$F_{n+1} = b_0 + b_1 X_{n+1}. \tag{7A.11}$$

The difference between these two expressions is

$$Y_{n+1} - F_{n+1} = (\beta_0 + \beta_1 X_{n+1} + \varepsilon_{n+1}) - (b_0 + b_1 X_{n+1}) = \varepsilon_{n+1} - (b_0 - \beta_0) - (b_1 - \beta_1)X_{n+1}.$$

The first term in the rightmost expression represents the inherent variation in the process. Even if we knew the parameters of the model, so that the second and third terms were zero, this element would still be present. The second and third terms represent the estimation error; as the sample size increases, they will diminish in importance. Because the error term pertains to the new observation, whereas the estimates are based upon the original sample, it follows that the two terms are statistically independent when the errors are independent. Thus, the variance of the forecast error is

$$var(Y_{n+1} - F_{n+1}) = var(\varepsilon_{n+1}) + var\{(b_0 - \beta_0) + (b_1 - \beta_1)X_{n+1}\}. \tag{7A.12}$$

For simple linear regression, we may derive the variance of the fitted value from the variances for the intercept and slope. It can be shown that

$$var\{(b_0 - \beta_0) + (b_1 - \beta_1)X_{n+1}\} = \sigma^2 \left[\frac{1}{n} + \frac{(X_{n+1} - \bar{X})^2}{\sum(X_i - \bar{X})^2} \right]. \tag{7A.13}$$

We substitute equation (7A.13) into equation (7A. 12) to arrive at the final expression for the *forecast variance in simple linear regression*:

$$var(Y_{n+1} - F_{n+1}) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_{n+1} - \bar{X})^2}{\sum(X_i - \bar{X})^2} \right]. \tag{7A.14}$$

Note that the further X_{n+1} is from the sample mean, $\bar{X}$ the larger the variance. Because the variance is unknown, we use the estimate given in equation (7A.6) to obtain

$$\widehat{var}(Y_{n+1} - F_{n+1}) = S^2 \left[1 + \frac{1}{n} + \frac{(X_{n+1} - \bar{X})^2}{\sum(X_i - \bar{X})^2} \right].$$

The second and third terms in equation (7A.14) decline as the sample size increases, but the first remains constant. Thus, in sufficiently large samples, we can approximate the forecast variance by S^2.

Appendix 7B Computing *P*-Values in Excel and R

In this appendix, we explain how to calculate *P*-values in Excel. Most statistical packages include similar procedures, and further details for select packages may be found on the book's website.

In all cases, begin by clicking on *fx* and select the *Statistical* functions. Continue using the appropriate choices for the *t* or *F* distributions as needed. The table specifies both the function to be used for each tail area and the required inputs. The final column provides numerical examples.

Distribution and Tail Area	Function	Items to Enter	Example
t: Lower tail	T.DIST	*X = Observed t-value* *Deg. freedom = Degrees of freedom* *Cumulative = TRUE*	*X* = –1.58 *Deg. freedom* = 20 *P*-value = 0.0649
t: Upper tail	T.DIST. RT	*X = Observed t-value* *Deg. freedom = Degrees of freedom*	*X* = 1.58 *P*-value = 0.0649
t: Two tails	T.DIST. 2T	*X = Observed t-value* *Deg. freedom = Degrees of freedom*	*X* = 1.58 *Deg. freedom* = 20 *P*-value = 0.1298
F: Upper tail	F.DIST. RT	*X = Observed F-value* *Deg. freedom1 = Numerator degrees of freedom* *Deg. freedom2 = Denominator degrees of freedom*	*X* = 2.50 *Deg. freedom1* = 3 *Deg. freedom2* = 25 *P*-value = 0.0826

The same values are provided in R as follows:

```
#Lower-Tail t-test p-value
pt(q = –1.58, df = 20, lower.tail = TRUE)
#Upper-Tail t-test p-value
pt(q = 1.58, df = 20, lower.tail = FALSE)
#Two-Tailed t-test p-value
2*pt(q = –1.58, df = 20, lower.tail = TRUE)
#Upper-Tail F-test p-value
pf(q = 2.5, df1 = 3, df2 = 25, lower.tail = FALSE)
```

Appendix 7C Computing Prediction Intervals

In practice, prediction intervals are evaluated with the use of standard software, which implements the analysis just described. The steps in Minitab, SPSS, and R are summarized in the box that follows. Excel does not provide the intervals as a standard function.

CALCULATION OF PREDICTION INTERVALS IN MINITAB

1. *Stat > Regression > Regression > Fit Regression Model*
2. Enter the *Y* (Response) and *X* (Predictor) variables.
   - e.g., *Unleaded* and *L1_Crude_price*.
3. Click OK to run the regression.
   - Enter a new column into the data matrix containing the *X* values to be used for forecasting, e.g., Future values of *L1_Crude_price*.
4. *Stat > Regression > Regression > Predict*
5. Specify the confidence level and type of interval (one or two sided) using *Options*.
6. Enter the name of the column with *X* values to be used for forecasting. These can alternatively be entered directly in the command.

Point predictions and confidence and prediction intervals are printed in the output with an option to store in the data matrix.

continues

CALCULATION OF PREDICTION INTERVALS IN SPSS

1. *Analyze > Regression > Linear*
2. Enter the *Y* (Dependent) and *X* (Independent) variables.
   - e.g., *Unleaded* and *L1_Crude_price*.
3. Click on *Save*.
4. Check *Predicted Values–Unstandardized* and *Prediction Intervals–Individual*.
   - e.g., *X* should include the out-of-sample data while *Y* should only include the observations to be used in estimating the model. The regression model is fitted on the in-sample data.
5. Specify the confidence level.

The point predictions (for *Y*, *Unleaded*) and the lower and upper prediction interval limits are saved in the data matrix.

CALCULATION OF PREDICTION INTERVALS IN R

1. Fit the regression model.
   - #Run the Regression
     fit < – lm(*Y* ~ *X*)
     fit < – lm(*Unleaded* ~ *L1_Crude_price*)
2. Store the future values of *X* in a data frame, with the name of the variables in that data frame identical to those used in the regression model.
   - #Extract Values of future *X*, *L1_Crude_price*
     e.g., *L1_Crude__price_future* < – data.frame(*L1_Crude_price* = data$*L1_Crude_price*[157:160])
3. Use the predict function, adding to its argument: (i) the name of the data frame containing the new data, (ii) interval = “prediction”, and (iii) level = “0.95”.
   - #Prediction Intervals
     fit.PI <- predict(fit, newdata =*X*, interval = “prediction”, level = 0.95)

CHAPTER 8

Multiple Regression for Time Series

*Topics marked with an * are advanced and may be omitted for more introductory courses.*

Table of Contents

Why have you included so many variables in your regression model?

— Anonymous Statistician

Why have you included so few variables in your regression model?

— Anonymous Economist

Introduction

One of the key restrictions we faced in Chapter 7 was the inability to consider more than one explanatory variable at a time. Yet both the discussion there and basic common sense indicate that events in the business world are typically affected by multiple inputs. We may not be able to measure all of them, but we do need to identify the main factors and incorporate them into our forecasting framework. A first step toward identifying an appropriate set of variables is to consider any theoretical models of the product market. Basic economics of consumer (or business) behavior can suggest quite a catalog of possibly explanatory driving forces. The next step is to examine plots of the data, which we do in Section 8.1, although we need to proceed with caution because multiple dependencies in the data may make interpretations complex. Then, in Section 8.2, we proceed to formulate a statistical model that incorporates multiple inputs and to interpret the coefficients in that model. Estimation of the parameters follows the method of ordinary least squares developed in Section 7.2 and is extended to cover multiple regression in Section 8.2.1.

Once we have developed a model, we need to know whether it is useful. For simple linear regression, the answer to this question was straightforward. We checked to see whether or not there was a statistically meaningful relationship between X and Y, and that completed the analysis, as in Section 7.6. The question now is more complex. For example, sales of a product may depend upon both advertising expenditures and price. Either variable alone may provide only a modest description of what is going on, whereas the two taken together may give a much better level of explanation. Conversely, a model for national retail sales that includes both consumer expenditures and consumer incomes may be only marginally better than a model that includes only one of them. The reason for this apparent anomaly is that if X_1 and X_2 are highly correlated and X_1 is already in the model, X_2 will not bring much, if any, new information to the table. To resolve such questions, we need to proceed in two steps:

1. We ask, "Is the overall model useful?" If the answer is NO, we go back to the drawing board.
2. If the answer is YES, we check whether individual variables in the model are useful and the model makes sense for forecasting purposes.

These two steps are explored in Sections 8.3 and 8.4. As we explained in Section 7.5, our analysis is based upon a set of standard assumptions. In Section 8.5, we briefly revisit those assumptions and present graphical procedures to determine whether the assumptions are reasonable. Taking action to deal with failures of the assumptions is a more difficult step, which we defer to Chapter 9.

Once the model has been shown to be effective and the assumptions appear to be reasonable, we are in a position to generate forecasts. Point forecasts and prediction intervals are considered in Section 8.6. Finally, in Section 8.7, we consider some of the key principles that underlie the development of multiple regression models.

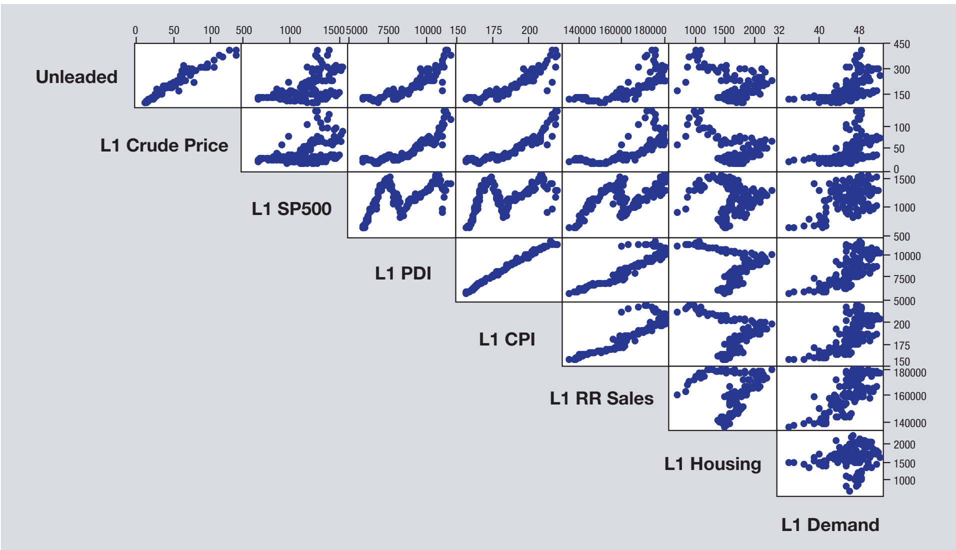
At the end of the chapter, we offer four minicases. Rather than work through "prepackaged" problem sets, these examples provide a more realistic approach to model building using multiple regression methods. The same minicases may be revisited at the end of Chapter 9, to make use of the more advanced skills developed in that chapter.

8.1 Graphical Analysis and Preliminary Model Development

We return to the study of gasoline prices, initially examined in Section 7.3 (see *Gas_prices_1.xlsx*). The matrix plot we considered there is reproduced as Figure 8.1 here for convenience and covers the period January 1996 to December 2008. The variables in the plot (unemployment has been excluded) are as follows:

- Price of unleaded gasoline, in cents per U.S. gallon (*"Unleaded"*) with potential explanatory variables:
- The price of crude oil, in dollars per barrel (*"L1_Crude_price"*)
- The SP500 Stock Index, end-of-month close (*"L1_SP500"*).
- Personal disposable income, in billions of current dollars (*"L1_PDI"*)
- Consumer price index for all urban consumers, indexed at 100 over 1982–1984, not seasonally adjusted (*"L1_CPI"*)
- Retail Sales: real retail sales in millions of $s, deflated by CPI (*"L1_RR_sales"*)
- Unemployment rate, not seasonally adjusted (*"L1_Unemp"*)
- Housing starts, seasonally adjusted annual rate (*"L1_Housing"*)
- Demand for gasoline, in thousands of barrels per day (*"L1_Demand"*).

Figure 8.1 Matrix Plot: Gasoline Prices Against Price of Crude Oil, the S&P Index, Personal Disposable Income (current prices), the Consumer Price Index (CPI), Retail Sales (RR_sales), and Demand (all lagged one period)



Data: Gas_Prices_1.xlsx; adapted from the Minitab output.

Examination of the plot had already revealed that the strongest linear relationship for unleaded gas price appeared to be the lagged value of the price of crude oil. However, we also see a somewhat upward-sloping relationship between price, lagged disposable income, possibly due to the effect of inflation (*CPI*) on both series. The general level of economic activity is reflected in a downward-sloping relationship with unemployment (not shown) and an upward-sloping relationship with the S&P500 Index and real retail sales. None of these last three relationships appears to be nearly as strong as that with *L1_Crude_price* (the lagged crude price); but they all make economic sense and might improve our overall ability to forecast gas prices. Also, as we noted in Table 7.2, all their correlations with gas prices are significantly different from zero, so they may add value to the model.

With this example as background, we now examine the specification of the multiple regression model.

8.2 The Multiple Regression Model

The multiple regression model is a direct extension of the simple linear regression model specified in Section 7.5. Note that we now move directly to the specification of the underlying model, having already motivated the basic ideas in Chapter 7. We consider K explanatory variables $X_1, X_2, ..., X_K$ and assume that the dependent variable Y is linearly related to them through the following model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K + \varepsilon. \tag{8.1}$$

The coefficients in equation (8.1) may be interpreted as follows:

β_0 denotes the intercept, which is the *expected* value of Y when all the $\{X_j\}$ are zero, in which case the equation reduces to $Y = \beta_0$. However, this interpretation only holds *if* the model is meant to apply in such a situation.

β_j denotes the slope for X_j: When X_j increases by one unit *and all the other Xs are kept fixed*, the *expected* value of Y increases by β_j units.

Beyond the extended form of the expected value that now includes all K variables (the explained component of the model), the underlying assumptions are the same as for simple regression presented in Section 7.5.1. That is, we need only extend assumption R1 appropriately.

Assumption R1: For given values of the explanatory variables, $X = (X_1, \ldots, X_K)$ the expected value of Y is written as $E(Y|X)$ and has the form

$$E(Y|X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K.$$

Assumption R2: The difference between an observed Y and its expectation is a random error, denoted by ε. The complete model is

$$Y = E(Y|X) + \varepsilon = [\textit{Expected value}] + [\textit{Random error}]. \tag{8.2}$$

Assumption R3: The errors have zero means.

Assumption R4: The errors for different observations are uncorrelated with one another and with other explanatory variables.

Assumption R5: The error terms come from distributions with equal variances.

Assumption R6: The errors are drawn from a normal distribution.

As in Section 7.5, if we take assumptions R3-R6 together, we are making the claim that the random errors are independent and normally distributed with zero means and equal variances.

8.2.1 The Method of Ordinary Least Squares (OLS)

The method of ordinary least squares (OLS) may be used to estimate the unknown parameters. As with simple linear regression, we choose the sample coefficients, now $\{b_0, b_1, \ldots, b_K\}$ to minimize the sum of squared errors, *SSE*. That is, we choose $\{b_j\}$ to minimize

$$\begin{aligned} SSE &= \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Observed - Fitted)^2 \\ &= \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i} - \ldots - b_K X_{Ki})^2. \end{aligned} \tag{8.3}$$

The technical details were summarized in Appendix 7A, and we will not consider the computational issues further.[1] After we have determined the best fitting model, we use the estimated coefficients to compute the *(least squares) residuals*, defined as

$$e_i = Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i} - \ldots - b_K X_{Ki}. \tag{8.4}$$

The residuals form the basis of many of the tests and diagnostic checks that we employ to validate the model, as we show in later sections.

■ Example 8.1: Multiple regression model for unleaded gasoline prices

(Gas_prices_1.xlsx)

We consider a model driven by the lagged crude price, two economic indicators (Unemployment and Real Retail Sales) and an expectations indicator of future economic activity (S&P 500 Index). We have also considered the compound indicator, Personal disposable income in current prices, which includes both a general price effect and real economic activity. In the previous edition of the book this was used as an alternative to Real Retail Sales; the differences are explored in exercise 8.3. We employ the observations for January 1996–December 2008 as the estimation sample, as before. The OLS solution is

$$\begin{aligned} Unleaded = &-29.0 + 2.4075\ L1\_Crude\_price \\ &- 0.0426\ L1\_SP500 + 0.001430\ L1\_RR\_sales - 14.27\ L1\_Unemp. \end{aligned}$$

Examination of the coefficients indicates that an increase of $1 in the price of a barrel of crude may be expected to increase the price at the pump by about 2.4 cents per gallon, somewhat lower than the figure we got with simple regression. Likewise, an increase in real retail sales (*RR_sales*) produces an increase in the expected price, whereas an increase in unemployment (*Unemp*) reduces the expected gas price. These coefficients have the signs we would expect. The coefficient for the S&P Index is also negative; initially, we might have expected an increase in the S&P to signal increased economic activity and thus exert upward pressure on gas prices. However, the issue of timing is important, and the negative sign could reflect the impact of good news in the crude oil markets, lowering pump prices and boosting the overall economy. Alternatively, the observed nonzero coefficient might have arisen due to chance alone, a topic we pick up in Section 8.4. ■

No matter how good the statistical fit, the forecaster should always check the face validity of the proposed forecasting model. By "face validity" we mean that the model's interpretation conforms to our and other experts' understanding of the product market.[2] If the

1 Standard texts on regression analysis provide the necessary details; see, for example, Kutner et al. (2005, pp. 15–20 and 222–227).

2 Toward that end, imagine standing in front of your boss in his or her office. Can you give a plausible justification of the model and all the coefficients? If not, develop a different model!

model passes the face-validity test, we go ahead and check to see whether it makes sense statistically.

8.3 Testing the Overall Model

We specify the null hypothesis as the claim that the overall model is of no value or, more explicitly, that none of the explanatory variables affects the expected value. Formally, this statement is written as

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_K = 0.$$

When the null hypothesis is true, none of the variables in the model contribute to explaining the variation in Y. The alternative hypothesis, H_A, states that the overall model *is* of value, in that at least one of the explanatory variables has an effect:

$$H_A : \text{Not all } \beta_1 = \beta_2 = \ldots = \beta_K = 0.$$

That is, there is some statistical relationship between Y and at least one of the Xs. If we fail to reject H_0, we conclude that the overall model is without value and we need to start over. If we reject H_0, we may still wish to eliminate those variables which do not appear to contribute, so as to arrive at a more parsimonious model.

When the model is based upon sound theoretical considerations, it makes sense to retain all the variables in it, even if some are not statistically significant, so long as the parameter estimates make sense. This is typically true in econometric modeling. By contrast, if we optimistically include variables on a "see if it flies" basis, we will usually prefer to prune the model to the smaller number of statistically significant and interpretable variables. These are critical issues in building a useful forecasting model and *should not* be resolved by statistical significance testing alone.

8.3.1 The *F*-test for Multiple Variables

The F-test is based upon partitioning the total sum of squares and is known as the *Analysis of Variance*, often referred to as ANOVA. The total sum of squares and the sums of squares resulting from the partition are as follows:

Total sum of squares: $SST = S_{YY} = \sum_1^n (Y_i - \bar{Y})^2$

Sum of squared errors: $SSE = \sum_1^n (Y_i - \hat{Y}_i)^2$

Sum of squares explained by the regression model: $SSR = \sum_1^n (\hat{Y}_i - \bar{Y})^2$

As was true in the case of simple regression, it can be shown that

$$SST = SSR + SSE. \tag{8.5}$$

The ANOVA test is usually summarized in tabular form, and the general framework is presented in Table 8.1. Note that

- The first column describes the partition into the two sources of variation: the sums of squares explained by the regression model and the sum of squared errors.
- The second column gives the number of degrees of freedom (DF) associated with each of the sums of squares; the total DF is $(n - 1)$, because we always start out with a constant term in the model.

- The third column provides the numerical values of the sums of squares.
- Column four gives the mean squares, defined as [sum of squares/DF] for each source: *MSR* and *MSE*, respectively.
- Column five yields the test statistic $F = MSR/MSE$.

The reason for introducing the new term *Mean Square Error* is that, when the null hypothesis is true, both *MSR* and *MSE* have expected values equal to the error variance, σ^2. Thus, the test statistic[3] $F = MSR/MSE$ should have a value in the neighborhood of 1.0, if the null hypothesis is appropriate. When the regression model is useful, the amount of variation explained by the model will increase, so *MSR* will increase relative to *MSE* and *F* will increase. Thus, we reject H_0 for sufficiently large values of *F*.

We refer to the observed value of *F* generated from this table as F_{obs}. The decision rule becomes

$$\text{Reject } H_0 \text{ if } F_{obs} > F_\alpha(K, n - K - 1)\text{; otherwise do not reject } H_0.$$

Table 8.1 General Form of the ANOVA Table

Source	DF	Sums of Squares	Mean Squares	F
Regression	K	SSR	$MSR = SSR/K$	MSR/MSE
Residual error	$n - K - 1$	SSE	$MSE = SSE/(n - K - 1)$	
Total	$n - 1$	SST		

The critical value for *F* depends upon the number of degrees of freedom for both *SSR* (in the numerator, DF = ν_1) and the *SSE* (in the denominator, DF = ν_2).[4]

In keeping with our previous discussions, we use the *P*-value with the following decision rule:

$$\text{Reject } H_0 \text{ if } P < \alpha\text{; otherwise do not reject } H_0.$$

Should it prove necessary to compute the *P*-value for some observed value of the *F* statistic, we follow the procedure outlined in Appendix 7A.2.

■ Example 8.2: ANOVA for unleaded gas prices *(Gas_prices_1.xlsx)*

The ANOVA table for the gas prices model given in Example 8.1 with $K = 4$ is as follows (consistent with our usual convention, we specify $\alpha = 0.05$):

Source	DF	SS	MS	F	P
Regression	4	804090	201022	1278.7	0.000
Residual error	150	32530	217		
Total	154	836620			

Because $P < 0.05$, we reject H_0. Recall that a *P*-value of 0.000 does not mean zero; rather, it signifies that $P < 0.0005$ and the result is rounded down. We conclude that there is strong evidence that the overall model is useful as an explanation of the price of unleaded gasoline. The next step is to determine the contribution made by each of the variables. ■

3 The method of analysis of variance was first derived by Sir Ronald Fisher, the father of modern inferential statistics. The ratio was labeled *F* in his honor.

4 Critical values can be obtained through the Excel function F.INV.RT(α,df1,df2).

The relationship between F and R^2 There is a simple relationship between F and R^2. It can be shown that

$$F = \frac{(n - K - 1)R^2}{K(1 - R^2)}. \tag{8.6}$$

The details are left to Exercise 8.11. From inspection of equation (8.6), it is evident that an increase in R^2 leads to an increase in F, the numerator increasing while the denominator decreases, so the ANOVA test is completely equivalent to a test based upon the coefficient of determination. Either from the ANOVA table or directly from the computer output, we find that, for Example 8.2,

$$R^2 = \frac{SSR}{SST} = \frac{804090}{836620} = 0.961 \text{ or } 96.1\%.$$

Thus, the coefficient of determination shows an increase over that for the single-variable model, which had $R^2 = 94.3$ percent. Indeed, whenever we add a variable to the model, we find that R^2 increases (or, strictly speaking, cannot decrease). However, the value of the F statistic often falls because of the K in the denominator of equation (8.6). There is no inconsistency here, but we need to recognize that the decline in F does not necessarily signal a weakness in the model.

The increase in R^2 seems modest, but more important is the change in S. The single-variable model has $S = 17.61$, whereas the current model has $S = 14.73$, so the prediction interval is about six cents narrower.

8.3.2 ANOVA in Simple Regression

We did not consider the analysis of variance in Chapter 7 because it did not provide any additional information: When there is only one variable in the model, the test of the overall model is formally equivalent to the two-sided test of the single slope. Referring back to the computer output in Figure 7.12, we see that the P-value for ANOVA is identical with that for the t-test of the slope parameter. That is, in *simple linear regression*, the F-test and (two-sided) t-tests provide identical information. Another way of saying this is that, in *simple linear regression*, $F = t^2$, as can readily be verified numerically in Figure 7.12.

8.3.3 S and Adjusted R^2

The steady increase in R^2 as new variables are added is a matter for some concern: If that were all there were to model building we would just add variables! A better guide to the performance of the model is to look at S, the standard error, now defined as

$$S^2 = \frac{\sum_{i=1}^{n} e_i^2}{(n - K - 1)} = \frac{SSE}{(n - K - 1)} = MSE \tag{8.7}$$

If S is smaller, the model has improved as the result of including the extra variable, although the improvement may be marginal. As we argued in Section 7.7.3, S has a straightforward interpretation as the accuracy of the predictions, because it is an estimate of the standard deviation of the error.

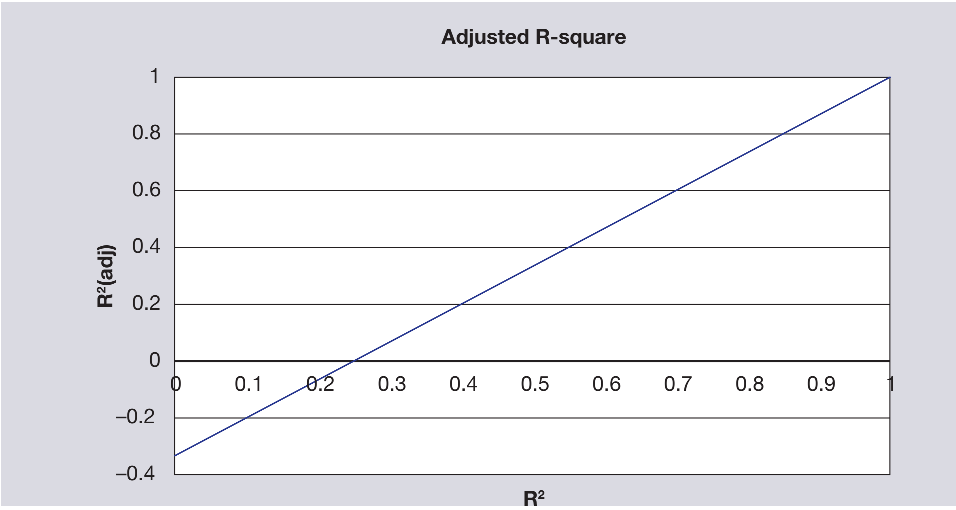
An alternative route to interpreting the overall accuracy of a multiple regression model is through the adjusted form of R^2, which we abbreviate to $R^2(adj)$. The general algebraic expression is

$$R^2(adj) = \frac{(n - 1)}{(n - K - 1)}\left[R^2 - \frac{K}{(n - 1)}\right]. \tag{8.8}$$

The term inside the brackets removes that part of R^2 which could arise just by chance, and the ratio in front of the brackets then rescales the expression so that $R^2(adj) = 1$ when $R^2 = 1$.

The plot of $R^2(adj)$ is shown in Figure 8.2 for $n = 21$ and $K = 5$. For example, when the observed value of $R^2 = 0.20$, it follows that $R^2(adj) = -0.067$, the difference demonstrating that an apparent fit measured by R^2 can always be achieved by including sufficient (even random!) variables, but may prove illusory once adjusted for the degrees of freedom.

Figure 8.2 $R^2(adj)$ Plotted against R^2 for $n = 21$, $K = 5$



An alternative form for $R^2(adj)$ is

$$R^2(adj) = 1 - \frac{(n-1)S^2}{SST}.$$

Thus, when we add a variable to the model, $R^2(adj)$ will increase if and only if S decreases. Hence, any decisions about the choice of model that use $R^2(adj)$ or S will reach identical conclusions.

■ Example 8.3: $R^2(adj)$ and S for the gas prices model *(Gas_prices_1.xlsx)*

We have $K = 4$ and $n = 155$, so

$$S^2 = \frac{32530}{150} = 217 \text{ and it follows that } S = 14.73.$$

We then obtain $R^2(adj) = 1 - \dfrac{155 \times 217}{836620} = 0.960$ or 96.0%.

As we see in this example, when K is small relative to n, the adjusted value is only marginally less than the original R^2.

Computer programs generally provide all the information discussed in this section in summary form, such as

$$S = 14.73,\ R^2 = 96.11\%,\ R^2(adj) = 96.01\%. \ ■$$

8.4 Testing Individual Coefficients

Once we have established that the overall model is of value, we need to determine which variables are useful and which, if any, do not contribute. The process is similar to that described in Section 7.6, but there are some crucial differences. First, because there are K variables in the model, we will perform K separate tests. We describe the procedure for variable X_j, $j = 1, 2, \ldots, K$.

The null hypothesis, now denoted by $H_0(j)$, states that the theoretical slope for X_j in the regression is zero, *given that the other variables are already in the model*. We are not testing for a direct relationship between X_j and Y; rather, we seek a conditional relationship that asks the question: Given that the other variables are already in the model, does X_j add anything? This is captured by the null hypothesis

$$H_0(j) : \beta_j = 0, \text{ given that } X_i, \text{ for all } i \neq j, \text{ are in the model.}$$

The alternative hypothesis is now denoted by $H_A(j)$ and states that the slope, β_j, is not zero, again assuming that the other variables are in the model. That is, there is a relationship between X_j and Y even after accounting for the contributions of the other variables. We write the alternative hypothesis as

$$H_A(j) : \beta_j \neq 0, \text{ given that } X_i, i \neq j, \text{ are in the model.}$$

As before, we assume the null hypothesis to be true and then test this assumption. We use the test statistic

$$t = \frac{b_j}{SE(b_j)}. \tag{8.9}$$

Let t_{obs} denote the observed value of this statistic. The observed value is to be compared with the appropriate value from a table of student's t distribution with $(n - K - 1)$ DF; the number of degrees of freedom is determined by the number of observations available to estimate S. That number is now $(n - K - 1)$, as seen from Table 8.1. If we use a significance level of 100α percent, we denote the value from the t tables as $t_{\alpha/2}(n - K - 1)$. The decision rule for the test is

$$\text{If } |t_{\text{obs}}| > t_{\alpha/2}(n - K - 1), \text{ reject } H_0(j); \text{ otherwise, do not reject } H_0(j).$$

As in Chapter 7, it is more convenient to use the P-value to perform the test. The decision rule is then written as

$$\text{If } P < \alpha, \text{ reject } H_0(j); \text{ otherwise do not reject } H_0(j).$$

A benefit of using the P-value approach is that, once the value of P is available, the decision rule always has this standard form: Reject $H_0(j)$ if $P < \alpha$.

■ Example 8.4: Testing individual coefficients *(Gas_prices_1.xlsx)*

Table 8.2 provides the output for testing the individual coefficients in the gas prices example, carrying out a test on each slope in turn. Standard computer packages typically summarize the set of K tests in a single table.

As in Chapter 7, we ignore the test for the intercept or constant because such a test is not meaningful in this case (the variables are never all zero). All four input variables have $P <$ 0.05, and there is strong evidence that they should be retained in the model. Even though there are also some significant correlations among the explanatory variables (see Figure

8.3), the contribution of each variable is well defined and the signs are consistent with our previous expectations. ■

Table 8.2 Single Variable Tests for the Gas Prices Model

Predictor	Coef	SE Coef	T	P
Constant	–29.0	22.8	–1.27	0.205
L1_Crude_price	2.4075	0.0673	35.76	0.000
L1_SP500	–0.0426	0.0129	–3.30	0.001
L1_RR_sales	0.00143	0.000204	7.02	0.000
L1_Unemp	–14.27	3.37	–4.23	0.000

Figure 8.3 Correlations Between Variables in Gas Prices Model

	Crude_price	SP500	RR_sales
SP500	0.460 (0.000)		
Real Retail (RR_sales)	0.736 (0.000)	0.677 (0.000)	
Unemp	0.125 (0.120)	–0.591 (0.000)	0.031 (0.703)

Data: Gas_prices_1.xlsx. Note: *P*-values appear below the correlation coefficients.

8.4.1 Case Study: Baseball Salaries *(Baseball.xlsx)*

We now revisit the baseball salaries data and examine some other factors that might affect a player's salary. Some will have to do with his overall career performance and others with his recent results. Although baseball is arguably the most statistically oriented of all sports, a few brief explanations are in order for those who are not baseball fans:

1. A pitcher may be recorded as the winner or the loser of a game. He may also record "No decision". Thus, the numbers of wins ("*Career Wins*") and losses ("*Career Losses*") in the course of a career are important factors in a player's remuneration.
2. The quality of a pitcher's performance over the years may be assessed by the average number of runs given up in the equivalent of a full game, known as the earned run average ("*Career ERA*"); the lower the ERA, the better the pitcher is seen to be.
3. The player's recent activity level can be judged by the number of innings pitched in the previous season ("*Innings Pitched*"); the more activity, the more the team is seen to rely on the pitcher.

Baseball fans will be able to suggest a number of other criteria, and some are listed in the file *Baseball.xlsx*, but these four additional factors (wins, losses, earned run average, and innings pitched) will suffice for our purposes. The extended model is summarized in Figure 8.4(B). It is apparent that the "*Career Wins*" and "*Career Losses*" variables do not appear to add much to the overall explanation; further, "*Career Wins*" has the wrong sign. The other variables are highly significant and have the appropriate signs.

Should we drop both variables with high *P*-values? Not necessarily: Pitchers with long careers have time to accumulate a lot of wins, but also a lot of losses, so the variables may be highly correlated, as indeed can be seen from Figure 8.4(A).

Figure 8.4 Correlations and Regression Models for Baseball Players

(A) Correlation Analysis

	Salary ($000s)	Years in Majors	Career ERA	Innings Pitched	Career Wins	Career Losses
Salary ($000s)	1.00					
Years in Majors	0.53	1.00				
Career ERA	–0.34	–0.22	1.00			
Innings Pitched	0.27	0.11	0.09	1.00		
Career Wins	0.51	0.89	–0.21	0.33	1.00	
Career Losses	0.49	0.91	–0.14	0.30	0.97	1.00

(B) Regression Analysis: Five Explanatory Variables

The regression equation is

Salary ($000s) = 547 + 53.6 *Years in Majors* – 152 *Career ERA* + 171 *Innings Pitched* – 0.59 *Career Wins* – 1.12 *Career Losses*

Predictor	Coef	SE Coef	T	P
Constant	547.4	167.2	3.27	0.001
Years in Majors	53.65	13.62	3.94	0.000
Career ERA	–152.37	39.81	–3.83	0.000
Innings Pitched	1.711	0.421	4.07	0.000
Career Wins	–0.590	1.846	–0.32	0.749
Career Losses	–1.123	2.400	–0.47	0.640

S = 292.804 R–Sq = 39.8% R–Sq(*adj*) = 38.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	9639078	1927816	22.49	0.000
Residual error	170	14574798	85734		
Total	175	24213875			

(C) Regression Analysis: Three Explanatory Variables

The regression equation is

Salary ($000s) = 620 + 36.8 *Years in Majors* – 154 *Career ERA* + 1.42 *Innings Pitched*

Predictor	Coef	SE Coef	T	P
Constant	620.2	150.2	4.13	0.000
Years in Majors	36.850	5.023	7.34	0.000
Career ERA	–154.21	36.88	–4.18	0.000
Innings Pitched	1.4156	0.3556	3.98	0.000

S = 292.643 R–Sq = 39.2% R–Sq(*adj*) = 38.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	9483783	3161261	36.91	0.000
Residual error	172	14730092	85640		
Total	175	24213875			

Data: Baseball.xlsx; adapted from Minitab output.

One solution to the question is to drop the variables one at a time and see what happens. We leave that as an exercise for the reader. The other approach is to test the variables as a group, comparing models with and without the "*Wins*" and "*Losses*" variables. We now explore this alternative.

8.4.2 Testing a Group of Coefficients

We have seen how the t-test may be used to evaluate individual variables. Now, we go a step further and consider whether it is necessary to include a group of variables. The question is, "Does the group of extra variables add anything to the predictive power of the model?"

We may solve this problem by comparing two models: M_1 which contains all the variables, and M_0, which contains only a subset:

Model M_1 (with error sum of squares SSE_1)

$$Y = f(X; \beta_0, \beta_1, \beta_2, \ldots, \beta_q, \beta_{q+1}, \ldots, \beta_K), \tag{8.10}$$

and Model M_0 (with error sum of squares SSE_0),

$$Y = f(X; \beta_0, \beta_1, \beta_2, \ldots, \beta_q, 0, 0, \ldots, 0), \tag{8.11}$$

In equation (8.10), M_1 contains the $(K + 1)$ parameters $\{\beta_0, \beta_1, \ldots, \beta_q, \beta_{q+1}, \ldots, \beta_K\}$ while the simpler model M_0 in equation (8.11) contains the $(q + 1)$ parameters $\{\beta_0, \beta_1, \ldots, \beta_q\}$. In each case, the sum of squared errors is

$$SSE_j = \sum_1^n e_{ji}^2 \text{ for Model } j, j = 0, 1.$$

More formally, $(K - q)$ restrictions are placed on model M_1 to obtain the simpler model M_0, so we consider the null hypothesis

H_0: $\beta_{q+1} = \beta_{q+2} = \ldots = \beta_K = 0$, given that $X_1, \ldots, X_q$ are in the model

and the alternative hypothesis

H_A: At least one of the coefficients $\beta_{q+1}, \ldots, \beta_K$ is nonzero when $X_1, \ldots, X_q$ are in the model.

To compare the two models, we just examine their explanatory power through their residuals. We therefore estimate the sum of squared errors from both the extended model M_1 and the simpler model M_0. We define the statistic

$$F = \frac{(SSE_0 - SSE_1)/(K - q)}{SSE_1/(n - K - 1)}.$$

This statistic has an F distribution with $(K - q, n - K - 1)$ degrees of freedom, and the P-value can be found by using the Excel function FDIST.RT.

DISCUSSION QUESTION: *Why is SSE_0 always greater than SSE_1?*

Example 8.5: Testing a group of coefficients *(Baseball.xlsx)*

From Figure 8.4, SSE_1 = 14574798. The analysis without the two variables "*Career Wins*" and "*Career Losses*" yields SSE_0 = 14730092. With $n = 176$, $q = 3$, and $K = 5$, the F statistic has the value

$$F = \frac{(14730092 - 14574798)/2}{14574798/170} = \frac{77647}{85734} = 0.906.$$

The P-value is 0.406, so, clearly, we do not reject the null hypothesis; indeed, whenever $F < 1$, the null hypothesis will not be rejected. The two variables in question do not add to the explanatory power of the model, so we will use the three-variable model summarized in Figure 8.4(C). Tests of the residuals are deferred to Exercise 8.5. ■

The test can also be used to test for nonlinearities (of polynomial form) or any other set of parameter restrictions, so long as the simple model M_0 is a restricted version of the full model.

8.5 Checking the Assumptions

In both this chapter and the previous one we laid out a set of assumptions. However, up to now, we have not attempted to check those assumptions; rather, we have proceeded as though our model was fully and completely specified and all the assumptions were valid. In short, we have been living in a forecasting fool's paradise. In this section, we take the model for gas prices selected in Section 8.4 and try to determine how well it matches up to the assumptions stated in Section 8.2.

We examine these assumptions and devise ways to check the validity of each. Because we have available only a sample, we can never guarantee a particular assumption, but we can check whether it seems plausible. We tend to be very pragmatic: If the data suggest that a particular assumption holds, we stay with that assumption. Given a reasonably sized sample, such evidence suggests that any violation of the assumption is likely to be modest, as will be the likely impact of that violation. However, we should always keep in mind that this argument applies only if we are confident that the system will continue to operate under the same regime as in the past; if major structural changes take place, all bets are off, unless we can incorporate such changes into the model.

If a particular assumption breaks down, the nature of the breakdown will often indicate how the model might be improved. We use the residuals, as defined in equation (8.4), to develop our diagnostics. A comparison with the material in Section 6.5 serves to indicate that many of the analyses suggested here are the same as those for state-space and ARIMA models.

Assumption R1: The expected value of Y is linear in the values of the selected explanatory variables.

Potential violations: We may have missed an important variable, or the relationship may not be linear in the Xs.

Diagnostics:

1. Plot the residuals against the fitted values. Nonlinear relationships will show up as curvature in the plot.
2. Plot the residuals against potentially important Xs not currently in the model. If a particular new X has an impact on Y, it should show up as a nonzero slope on the scatterplot.

Why it matters

Using an inadequate model of the true relationship loses an important element in predicting Y.

Possible actions

1. Include omitted variables in the model.
2. Consider nonlinear models (see Section 9.6).

Assumption R2: The difference between an observed Y and its expected value is due to random error.

Discussion: The assumption states that the error is an "add-on" and serves to justify the least squares formulation for estimating the parameters. The error can always be expressed in this way, but its properties will depend critically upon the next four assumptions. Therefore, we do not check this assumption directly, but examine aspects of it as described next.

Assumption R3: The errors have zero means.

Discussion: Typically, this assumption is not testable, at least when we are looking at a single series. The inclusion of a constant term in the model ensures that the mean of the in-sample errors is zero. However, if a model is intended to apply when all explanatory variables are zero, with Y correspondingly equal to zero, then the constant term can be meaningfully tested. When the model is used for forecasting, the forecast errors may show bias.

Why it matters

Bias in the model suggests the model is missing an important variable or a variable is mismeasured. For example, many macroeconomic series are released in preliminary form and then updated. Thus, the model may have been constructed on a particular set of final figures but then used in forecasting with the preliminary data. Cross-checks between the preliminary and final versions of such variables may reveal biases.

Assumption R4: The errors associated with different observations are uncorrelated with other variables and with one another. Thus, the errors should be uncorrelated with the explanatory variable or with other variables not included in the model. In examining observations over time, this assumption implies that there is no correlation between the error at time t and past errors; otherwise, the errors are autocorrelated.

Possible violations: Assumption R4 lies at the heart of model building and boils down to the claim that the model contains all the predictable components, leaving only noise in the error term. The residuals therefore should not be related to factors omitted from the model, such as nonlinear functions of the input variables. Where the data form a time series, there should be no relationship with past values of the inputs, the dependent variable, or past errors. Such a relationship can obtain if there is a carry-over effect from one period to the next, which could be due to such factors as the weather, brand loyalty, or economic trends. Thus, a positive residual in one time period is likely to be followed by a positive residual in the next period. High-low sequences are also possible, such as a drop in sales after high volumes because of a special promotion. In cross-sectional studies, we should check whether subgroups in the population exhibit correlations, be it from locational, demographic, or other shared properties.

Diagnostics:

1. Plot the residuals against both the predicted value of Y and the input variables included in the model (as well as any other potential explanatory variables that have been excluded).
2. For time series, plot the residuals against time. If positive autocorrelation exists, lengthy sequences of values above zero and then below zero, rather than a random

scatter, will appear. If a negative autocorrelation exists, a saw-tooth pattern will prevail.

3. For time series, plot the sample autocorrelation function (*ACF*) for the residuals (see Section 6.1) and look for departures from a random series by performing tests for the presence of autocorrelation. (The *PACF* discussed in Section 6.2 may also help.) A test that is sometimes recommended for this purpose is based on the Durbin-Watson statistic, described in Appendix 8A. However, this test applies only to first-order autocorrelation so we prefer to use the *ACF* along with the Ljung-Box-Pierce statistic of Section 6.5.1 as useful overall diagnostics. In Chapter 10, we consider more efficient tests.

Why it matters

The diagnostics provide evidence of a predictable element in the error term. Its exclusion means that the error variance is larger than need be and the forecasts less accurate.

Possible actions

1. For cross-sectional data, examine the data collection process carefully. For example, check whether the sample includes multiple respondents from the same company; if it does, company affiliation could be a missing factor that explains differential (segmented) responses.
2. For time series, consider the use of lagged values (see Section 9.3).

Assumption R5: The error terms come from distributions with equal variances.

Possible violations: The most common pattern is that the variability increases as the mean level of Y increases. We naturally talk about percentage movements up or down in GDP, in sales, and in many other series. The implication behind such terminology is that the variations are proportional to the level of the mean, rather than displaying constant variance.

Diagnostics:

1. Plot the residuals against the fitted values. If the errors are heteroscedastic, the scatter will often be greater for the larger fitted values.
2. Various test statistics are available (see Anderson *et al.*, 2014, Chapter 11), but we do not pursue that topic further here; such tests are illustrated in Chapter 10.

Why it matters and possible actions

With errors from one part of the data larger than another, the least squares procedure effectively gives additional weight to the former, leading to miscalibrated prediction intervals. Procedures for dealing with changing variances by transforming the raw data are discussed in Section 9.6.

Assumption R6: The errors are drawn from a normal distribution.

Possible violations: One or more outliers may render the distribution nonnormal, or the whole pattern of the residuals may suggest a nonnormal distribution.

Diagnostics:

1. Plot the histogram of the residuals, and look for a rough bell shape.
2. Use the normal probability plot. A plot that deviates significantly from a straight line indicates nonnormality.
3. Examine the plots of residuals against both time and fitted values for extreme observations.

Why it matters

The normality assumption is used in calculating the prediction intervals so if the assumption fails the prediction intervals are miscalibrated.

NB. It is not needed for tests of the coefficients in a model as these can call on the Central Limit Theorem and the resulting approximately normal distribution of the estimates.

Possible actions

1. There is an interaction between normality assumption R6 and the heteroscedasticity assumption, R5. Often, correcting for extreme observations will lead to the data having an approximately constant variance.
2. Consider transformations of the dependent variable and possibly of the predictors as well (see Section 9.6).

It is evident from the preceding summary that some plots — notably, that of the residuals against fitted values — serve multiple purposes. It is important to keep these several objectives in mind in examining the plots.

8.5.1 Analysis of Residuals for Gas Price Data

Most statistical packages will generate the plots we have just discussed, some more easily than others. In particular, Minitab produces a "Four in One" plot as part of its regression component, a feature that is particularly useful for the analyses we have been discussing.

The four-variable model of unleaded gas prices was identified in Table 8.2, and the residuals are shown in Figure 8.5. We examine these plots in the order they appear in the output:

a. The probability plot (top left) shows several outliers at either end, partially a reflection of the increased volatility in the later part of the series. The histogram (bottom left) tells much the same story: The long tails shown afford some evidence of a departure from the normal curve.
b. As we noted earlier, the plot of residuals against fitted values (top right) may tell several stories. The residuals for fitted values in the range from \$1 to \$2 are tightly bunched, showing low variability. However, the higher fitted values show considerable variation, a clear demonstration of non-constant variances. The plot of residuals against order (bottom right) shows runs of positive values followed by runs of negative values, indicative of autocorrelation. Also, we observe that larger values are clustered together at the end of the series, suggesting a possible change in conditions that should be examined more closely.
c. Figure 8.5(B) explores possible relationships between the residuals and the price level, measured by the *CPI*. Although there is no evidence of a linear relationship, the increased fluctuations in the later part of the series suggest that constant-dollar prices might produce a better model. This figure also checks for a possible relationship with lagged demand (other variables could have been chosen such as lagged production) but there is no evidence of such dependence.

The issue of residual autocorrelation is particularly important because it indicates persistence in the time series that has not been fully captured by the current model. To investigate this phenomenon, we look at the *ACF* of the residuals, shown in Figure 8.6. This is usually done by storing the residuals from the regression command and then using a further time series command to calculate the *ACF*.

The *ACF* indicates a degree of persistence, with a significant positive autocorrelation at lag 1.[5] The spikes at lag 6 and lag 12 suggest possible seasonality, which also merits further examination. (An analysis of the Partial Autocorrelation Function — see Section 6.2.5 — gives no further information.)

Collectively, these plots provide plenty of food for thought and indicate that we have some work ahead of us before we can be satisfied with the model. We return to the model-building endeavor in Chapter 9; for now, we explore the use of such models in forecasting.

Figure 8.5(A) Residual Plots for the Four-Variable Model of Unleaded Gas Prices

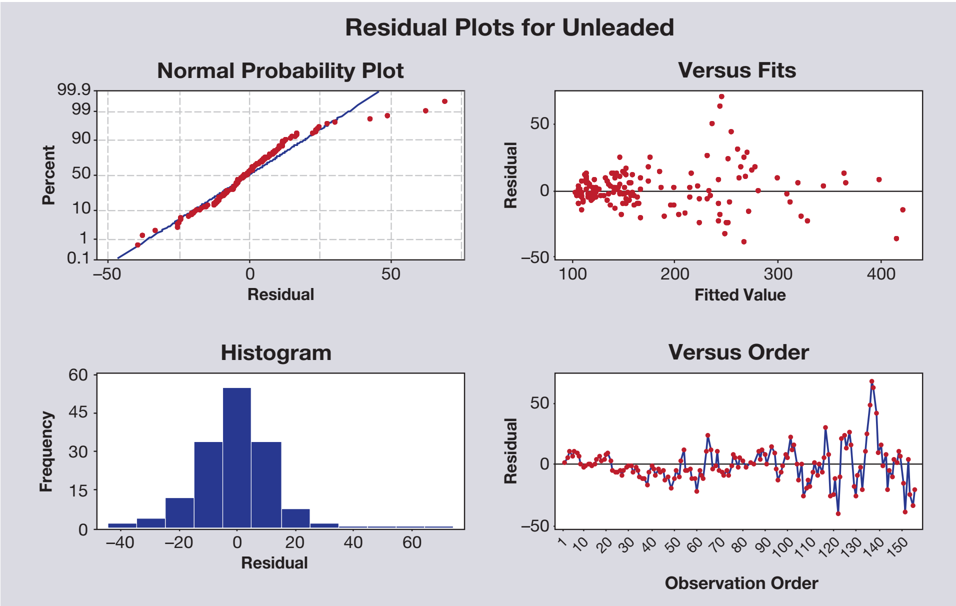
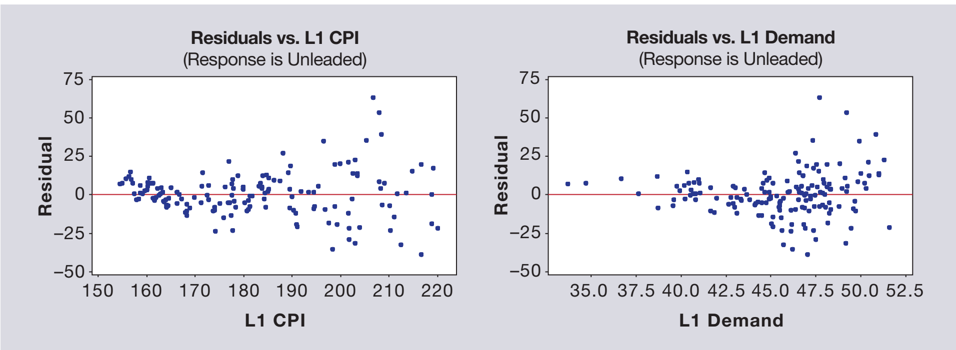
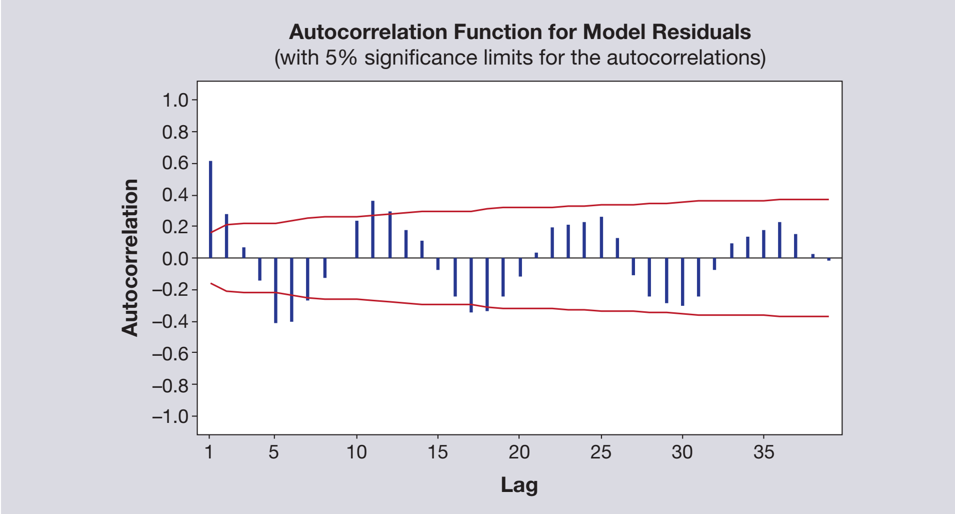


Figure 8.5(B) Residual Plots versus Possible Additional Input Variables for Unleaded Gas Prices



5 The Durbin-Watson statistic, valid for this particular model, of 0.738 is highly significant, suggesting first order autocorrelation. But the *ACF* is more informative.

Figure 8.6 ACF for the Residuals of the Four-Variable Unleaded Gas Prices Model



8.6 Forecasting with Multiple Regression

The general procedure for forecasting with several explanatory variables is essentially the same as that for the single-variable case described in Section 7.7. The interested reader is referred to Kutner *et al.* (2005, pp. 229–232) for further details. The first question we must answer relates to the nature of the explanatory variables. Recall that any particular X may arise in one of three ways:

a. X is known ahead of time.

b. X is unknown but can itself be forecast.

c. X is unknown but we wish to make "what-if" forecasts.

For example, consider a model for sales. Clearly, variables that designate particular seasons are known in advance, as may be substantive variables that have been sufficiently lagged in time. Policy variables — for example, price and advertising expenditures — may be explored with the model to make "what-if" forecasts so that the sensitivity of expected sales to policy changes can be explored. Finally, some variables, such as the price charged by competitors or the level of GDP, will require forecasts themselves. These forecasts are often generated by industry analysts, government sources, or macroeconomic panels (see, e.g., *www.consensuseconomics.com*). Alternatively, time series methods, such as the exponential smoothing methods discussed in Chapters 3 and 4, could be used.

8.6.1 The Point Forecast

We suppose that values for the next time period are available for each of the K variables and denote these values by $X_{n+1,1}$ $X_{n+1,2}$, … $X_{n+1,K}$. Given the estimated regression line, the point forecast is

$$F_{n+1} = b_0 + b_1X_{n+1,1} + b_2X_{n+1,2} + \ldots + b_KX_{n+1,K}. \tag{8.12}$$

As before, we need to distinguish between the fitted values Y and the forecast F_{n+1}. The two formulas are the same, but the fitted values correspond to those observations that were used in the estimation process, whereas the forecasts are based on new observations. These new values may be part of a hold-out sample or values as yet unobserved, but they are not used to estimate the model parameters.

■ Example 8.6: One-step-ahead forecasts for gas prices *(Gas_prices_1.xlsx)*

We use the four-variable model for gas prices as an illustration. One-step-ahead forecasts were generated from equation (8.12), so for example, the forecast for January 2009 uses the December 2008 values of the explanatory variables. The regression model (from Table 8.2) is

$$\begin{aligned} Unleaded &= -29.0 + 2.4075\ L1\_Crude\_price \\ &\quad - 0.0426\ L1\_SP500 + 0.001430\ L1\_RR\_sales - 14.27\ L1\_Unemp. \end{aligned}$$

The values of the explanatory variables for January 2009, which are of course known, are as follows:

$$L1\_Crude\_price\text{: } 41.12 \quad L1\_SP500\text{: } 903.25 \quad L1\_RR\_sales\text{: } 157080 \quad L1\_Unemp\text{: } 7.3.$$

The forecast for January 2009 is then

$$\begin{aligned} F &= -29.0 + 2.4075 \times 41.12 - 0.0426 \times 903.25 + 0.001430 \times 157080 - 14.27 \times 7.3 \\ &= 152.97. \end{aligned}$$

Rounding errors may give a slightly different answer. The forecasts for the 24 months (January 2009 to December 2010) are plotted in Figure 8.7; all the forecasts are one-step-ahead, using the previous month's values as inputs. ■

8.6.2 Prediction Intervals

We now require prediction intervals to provide an indication of the accuracy of the forecasts. We omit the technical details and simply note that, relative to the unknown future value Y_{n+1} the point forecast, when the explanatory variables are known, has an estimated standard error that we write as

$$SE(Y_{n+1} - F_{n+1}) = \sqrt{\text{var}(Y_{n+1} - F_{n+1})}.$$

The prediction error falls into two parts:

$$Y_{n+1} - F_{n+1} = [Y_{n+1} - E(Y_{n+1})] - [F_{n+1} - E(Y_{n+1})].$$

The first term represents the inherent variability in the model even when the parameters are known and the second term measures the uncertainty in the forecast due to parameter estimation. The second term is small when the sample size is large but it cannot be ignored, particularly when the values of the predictor variables are not close to their respective means. (We note in passing that in Chapters 5 and 6 we followed customary procedures and ignored such terms!)

Given assumptions R3-R5, these two factors are independent and the variance of the prediction error may be written as:

$$\text{var}(Y_{n+1} - F_{n+1}) = \text{var}[Y_{n+1} - E(Y_{n+1})] + \text{var}[F_{n+1} - E(Y_{n+1})]. \tag{8.13}$$

This expression approaches its limiting value of σ^2 [(the first term in equation (8.13)] when the sample size is large; it does *not* reduce to zero. Given the additional assumption R6, the forecast error follows a normal distribution, and after allowing for the estimation of σ by S, we may use the Student's t distribution with the appropriate DF to specify the prediction interval.

Prediction interval for the future observation Y_{n+1}:

$$F_{n+1} \pm t_{\alpha/2}(n - K - 1) \times SE(Y_{n+1} - F_{n+1}) \tag{8.14}$$

The 100(1 – α) percent prediction interval is a probability statement: The probability that the future observation will lie in the interval defined by equation (8.14) is (1 – α).

For regression models with more than a single explanatory variable estimating the standard error requires matrix algebra. Fortunately, computer programs such as Minitab, SAS and SPSS calculate these values: they are distinct values corresponding to each individual prediction. Some programs (e.g. Minitab) provide prediction intervals directly but not the standard errors; SAS and SPSS provide both standard errors and prediction intervals.

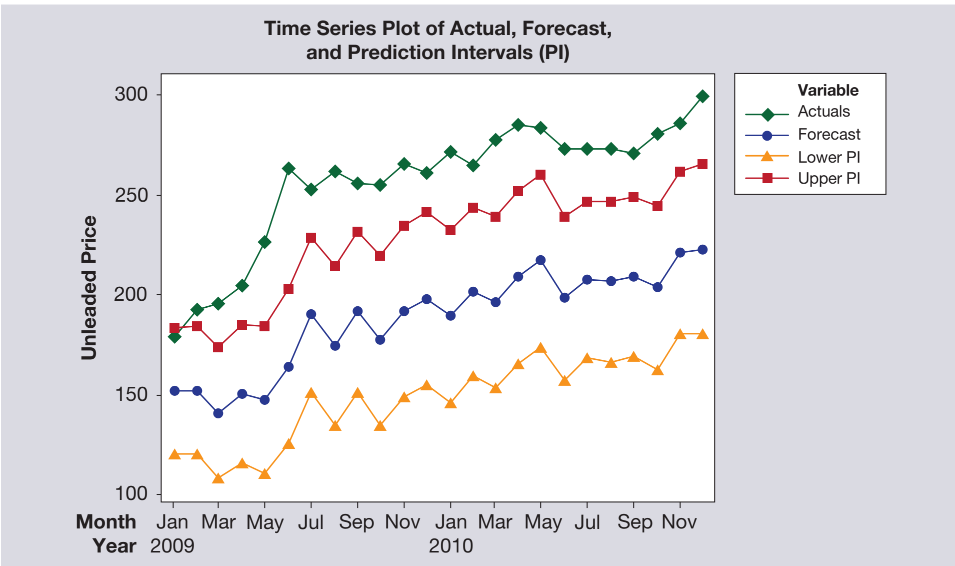
Example 8.7: Construction of a prediction interval

We continue our consideration of the forecasts for January 2009, begun in Example 8.6. With $K = 4$ and $n = 155$, we have DF = 150. The standard error for the point forecast is 15.852. Using $t_{0.025}(150) = 1.976$, we find that the 95 percent prediction interval is

$$152.00 \pm (1.976) \times 15.852 = [120.7,183.3].$$

The one-step-ahead point forecasts for January 2009 to December 2010 are plotted in Figure 8.7, along with the actual values and the 95 percent prediction intervals. From the plot, we can see that the forecasts systematically underestimated the actual values and that this error was not corrected over time.

Figure 8.7 Actual Prices for Unleaded Gasoline, Along with One-Step-Ahead Forecasts and 95 Percent Prediction Intervals (January 2009–December 2010)



As with simple regression, when n is large, an approximate 95 percent prediction interval is given by $F \pm 2S$. From Example 8.3 with $S = 14.73$, we get an approximate prediction interval of ± 29.46, a reasonable approximation, although as the explanatory variables move away from their averages, this becomes less accurate. For multiple regression, the exact formula is complex and we rely on the programmed values from Minitab, R, or SPSS. Here, for example, the width of the exact prediction interval for 2009 Jan is ± 31.3 which increases to ± 42.3 for 2010 Dec because the values of the explanatory variables are much further from their historic mean values in 2010 Dec. Better to rely on the programmed calculations unless the sample size is very large. ■

8.6.3 Forecasting More than One Period Ahead

When we wish to forecast more than one period ahead, we must provide values for all the predictor variables over the forecasting horizon. As we discussed in Section 7.7.4, there are two possible approaches:

1. Generate forecasts for all the *X*s and apply the forecasts to the original model.
2. Reformulate the model so that all unknown *X*s are lagged by two (or more) periods, as appropriate.

The first approach is more commonly applied, but suffers from the drawback that the uncertainty in *X* is not reflected in the prediction intervals for the forecasts. This approach has the advantage that only a single model need be used to compare different "what-if" paths formulated for *X*. The second approach is somewhat more tedious, but it will be more valuable when good forecasts for *X* are unavailable and it will provide more accurate prediction intervals. However, there is the implicit assumption that as the lag order of the *X*s increases there is still a useful predictive connection with *Y*.

As before, neither approach is necessarily better than the other. Where forecasts of the *X*s are unreliable, it will usually prove better (and easier) to adopt the second approach. Exploration of the gas prices model for multiple steps ahead is left as Exercises 8.9 and 8.10.

8.7 Principles of Regression

[8.1] Aim for a relatively simple model specification.
The researcher must strike a balance between failing to include key variables and cluttering the model with variables that have very little effect upon the outcome. For example, the number of consumers in a market area clearly has an impact on the level of sales. However, for a particular market area, that figure is not going to change much over the course of a few months. Accordingly, we would not bother to include a variable for population in a short-term model for sales forecasting (Allen and Fildes, 2001).

[8.2] Tailor the forecasting model to the horizon.
As noted in Principle 8.1, we need to identify those variables which are important for the forecasting horizon under consideration (Armstrong, 2001).

[8.3] Identify important causal variables on the basis of the underlying theory and earlier empirical studies. Identify suitable proxy variables when the variables of interest are not available in a timely fashion.
Expertise and earlier research should be used to formulate a model whenever possible. Statements such as "The stock market goes up when the AFC team wins the Super Bowl" may be factually correct over a period of years, but they are not a reliable guide to investment! (Adapted from Allen and Fildes, 2001)

[8.4] If the aim of the analysis is to provide pure forecasts, you must either know the explanatory variables in advance or be able to forecast them sufficiently well to justify their inclusion in the model.

This principle is a more formal statement of the necessary response to the questions "What do you know?" and "When will you know it?" (Adapted from Allen and Fildes, 2001)

[8.5] Use the method of ordinary least squares to estimate the parameters.

The method of ordinary least squares is strictly valid only when Assumptions R3-R5 apply, but it is often a good place to start, and the basic multiple regression model presented in this chapter can be extended to deal with more complex models (Allen and Fildes, 2001).

[8.6] Update the estimates frequently.

Frequent updating involves little effort beyond recording the latest data. The new parameter estimates will better reflect the relationships among the variables and also help to alert the modeler to any lack of stability in those values (Armstrong, 2001), a point which we take up in Chapter 9.

Summary

In this chapter, we have extended regression models to multiple explanatory variables, thereby greatly increasing the range and value of models that we may use for forecasting purposes. We have also provided the basic inferential framework in terms of parameter estimation and model testing, as well as identifying the key assumptions underlying the models. This structure will enable us to check assumptions and refine the models in the next chapter.

Exercises

8.1 Exercise 7.1 (in *Exercise_7.1.xlsx*) provided data on television advertising and sales for a soft-drink company over an eight-week period. The purpose was to evaluate the effect on sales (in millions of dollars) with respect to the number of 30-second spots aired during that week. In further analysis, additional price data were collected (also in *Exercise_7_1.xlsx*). Conduct a multiple regression analysis for sales on spot and price.

a. Carry out tests on the overall model and on the individual coefficients. Summarize your conclusions.

b. Compare the performance of the two models. Which model would you recommend?

8.2 The file *Exercise_7_2.xlsx* contains additional data on a company's sales on advertising expenditures, beyond the data quoted in Exercise 7.2:

a. Conduct a regression analysis of sales on advertising and price.

b. Carry out tests on the overall model and on the individual coefficients. Summarize your conclusions.

c. Compare the performance of the overall model and the price-only model. Which model would you recommend?

8.3 Using the unleaded gas prices data (*Gas_prices_1.xlsx*) compare the results of using Personal Disposable Income (rather than Real Retail sales) in a model of *Unleaded price*. Comment on any differences in interpretation and the one-step ahead forecasts from the two alternative models for 2009. Which one would you prefer to have used for 2008?

8.4 Create constant-dollar series for the unleaded gas prices data [*Gas_prices_1.xlsx*] by dividing *Unleaded*, *Crude_price*, *SP500*, and *PDI* by the *Consumer Price Index* (*CPI*). Repeat the analysis, using these new series and *Unemp*, and check the assumptions underlying this model. Finally, comment on the model's adequacy for explaining the price of unleaded gasoline.

8.5 Use the model developed in Exercise 8.4 to generate forecasts for the period from January 2009 through December 2010. Compare the performance of this model with that developed in Section 8.2. (This is not as easy as it looks!)

8.6 Use the file *Baseball.xlsx* to carry out the analysis of residuals for the three-variable baseball model developed in Section 8.4.1. Summarize your conclusions.

8.7 Use the file *Baseball.xlsx* and logarithms to transform the salaries data, and repeat the model development process. Compare your results with those given in Figure 8.4, and comment upon your results.

8.8 Use the file *Baseball.xlsx*. Modify the database by removing all the veterans, defined as players with ten or more years in the major leagues. Repeat the analyses of Exercises 8.5 and 8.6 and compare results.

8.9 Use the four-variable gas prices model of Example 8.1 to generate forecasts three periods ahead for the period from March 2009 to December 2010 by first generating forecasts for both the lagged crude oil price and Real Retail Sales (assume *SP500* and *Unemp* are known). Compare the estimates with those for the model developed in the chapter. Compute the forecast accuracy measures and generate the 90 percent prediction intervals, using equation (8.14) and S to estimate $\mathrm{SE}(Y_{n+1} - F_{n+1})$. How many of the actual values fall inside these intervals?

8.10 Develop the four-variable gas prices model, using explanatory variables with lag = 3 to allow for direct prediction of the prices three periods ahead over the period from March 2009 to December 2010.

a. Compute the forecast accuracy measures and evaluate the results.

b. if your computer program provides them, generate the 90 percent prediction intervals. How many of the actual values fall inside these intervals? Compare the results with those obtained for Exercise 8.9.

8.11* Verify the relationship between F and R^2 given in equation (8.6).

Minicases

The purpose of these minicases is to provide opportunities for you to use data analysis to tackle important real-world problems. The format is essentially the same in each: A dependent variable of interest is identified, along with a plausible set of explanatory variables. The aim is to develop a valid forecasting model for at least one period ahead, and multiple periods ahead should also be considered. The full set of modeling steps should be examined:

- Create plots of the data to look for relationships and possible unusual observations.

- Perform basic data analysis.
- Develop a multiple regression model and check the model's adequacy with regard to the regression assumptions.
- Evaluate your chosen model's forecasting performance, preferably using a hold-out sample.

Keep in mind that there are no "right" answers, but some solutions will be more effective than others. After you complete your statistical analysis, do not fail to ask the following questions:

- Would the data be available to enable me to make timely forecasts?
- Are there other variables that should be included in the model?
- Would you feel able to justify your model to a senior manager?

If the answer to any of these questions is NO, you have more work to do!

Minicase 8.1 The Volatility of Google Stock

[Contributors: Christine Choi, Alex Dixon, Melissa Gong, Michael Neches, and Greg Thompson. Data for Minicase 8.1 can be found in file *Google_Volatility.xlsx*.]

Volatility is a measure of the uncertainty of the return realized on an asset (Hull, 2009). Applied to financial markets, the volatility of a stock price is a measure of how uncertain we are of future stock price movements. As volatility increases, the possibility that the stock price will appreciate or depreciate significantly also increases. The volatility measure has widespread implications, particularly for stock option valuation and also for volatility indices (VIX), portfolio management, and hedging strategies. Since its initial public offering, Google, Inc. (GOOG: NASDAQ), stock has become one of the most sought-after and popular investment opportunities. The search engine giant's stock price has fluctuated from an IPO price of $85/share, to a high of $741/share (adjusted close, Nov. 27, 2007), down to a low 2009 closing price of $345/share (adjusted close, Feb. 25, 2009). This fluctuation reveals the uncertainty associated with any stock especially that of high-tech companies with Web-based models for which the monetization of services can confuse even the most sophisticated investor.

The aim of the project is to develop a multiple regression model to forecast the volatility of Google's stock price over the next three months. After an initial review of Google-specific and macroeconomic data, we identified the following potential explanatory variables:

STDEV: the volatility measure for Google stock

VOLUME: amount of trading in Google shares

P/E: the price-to-earnings ratio of Google stock

GDP: quarterly growth in GDP at an annualized rate (quarterly, repeated for each month)

VIX: the market volatility index

CONF: the Conference Board Consumer Confidence Index®

JOBLESS: the number of claims posted for benefits

HOUSING: the number of new housing starts.

Monthly data are available for the period from March 2006 through January 2009 on the preceding variables. The data were downloaded from Bloomberg. A possible analysis would be to analyse the data pre-recession (through December 2007) and then check forecasting performance over the next seven months.

Minicase 8.2 Forecasting Natural Gas Consumption for the DC Metropolitan Area

[Data for Minicase 8.2 can be found in file *Natural_Gas_2.xlsx*.]

This minicase comes in two forms so that the effects of different time periods upon the modeling process can be examined.

The intent of this project was to develop a model to forecast natural gas consumption for the residential sector in the Washington, DC, metropolitan area. The level of natural gas consumption is influenced by a variety of factors, including local weather, the state of the national and the local economies, the purchasing power of the dollar (because at least some of the natural gas is imported), and the prices for other commodities.

Quarterly data

[Contributors: Sameer Aggarwal, Prashant Dhand, Yulia Egorov, and Natasha Heidenrich. Data: *Natural_Gas_2.xlsx, Sheet labelled 'Quarterly'*]

The following variables have been identified and recorded on a quarterly basis (the data cover the period 1997Q1 through 2008Q3, with all data measured quarterly); the series could be truncated at 2007Q3 to avoid the effects of the Great Recession and the last four observations used for model checking.

GASCONS: consumption of natural gas in DC metro area (million cubic feet)

AVETEMP: average temperature for the period in the DC metro area

GDP: annualized percentage change in GDP

UNEMP: percentage unemployment in the DC metro area (not seasonally adjusted)

GAS_PRICE: price of natural gas ($/100 cubic feet)

OIL_PRICE: price of crude oil ($/barrel)

RESERVES: reserves of natural gas (index, 2012 = 100)

DISTRIB: natural gas production distributed (not seasonally adjusted, index, 2012 = 100).

Monthly data

[Data: *Natural_Gas_2.xlsx, Sheet labelled 'Monthly'*]

The following variables have been identified and recorded on a monthly basis (the data cover the period January 2001–December 2015). The data may be analysed pre-recession (up to November 2007) or for the complete period.

GASCONS: consumption of natural gas in DC metro area (million cubic feet)

AVETEMP: average temperature for the period in the DC metro area

GDP: annualized percentage change in GDP

UNEMP: percentage unemployment in the DC metro area (not seasonally adjusted)

GAS_PRICE: price of natural gas ($/100 cubic feet)

CRUDE_OIL_PRICE: price of crude oil ($/barrel)

RESERVES: reserves of natural gas (index, 1972 = 100)

PRODN: natural gas production distributed (not seasonally adjusted, index, 2012 = 100).

Minicase 8.3 U.S. Retail & Food Service Sales

[Contributors: Doug Goff, Rich Marsden, Jeff Rodgers, and Masaki Takeda. Data for Minicase 8.3 can be found in file *Retail _food_sales.xlsx*.]

The purpose of this project is to forecast how U.S. retail and food sales will fare over the coming months. The variables considered include personal income and savings, consumer sentiment, and various macroeconomic variables. Because manufacturing costs and levels of activity are clearly important, these factors are also included. Considered in the analysis as well were three seasonal factors associated, respectively, with the Easter, Thanksgiving, and Christmas holidays. The data set includes monthly figures for the period January 2000–December 2008.

RSALES: U.S. retail and food service sales ($millions)

CONSENT: University of Michigan Index of U.S. Consumer Sentiment

PRICE_OIL: spot price of oil ($/barrel)

IND_PROD: Index of U.S. Industrial Production

PERSINC: U.S. personal income ($ per capita)

PERSSAV: net U.S. personal savings ($ per capita)

POPULATION: total U.S. population (thousands)

UNEMP: U.S. unemployment rate

CPI: U.S. Consumer Price Index

TGIVING:* indicator for Thanksgiving (November)

EASTER:* indicator for Easter (March or April)

XMAS:* indicator for Christmas (December).

* Denotes indicator variables, which are introduced in Section 9.1. These variables are set = 1 for the month in which the event happens, and = 0 for all other months.

Minicase 8.4 U.S. Automobile Sales

[Data for Minicase 8.4 can be found in file *Auto_sales_mc.xlsx*.]

Automobiles are regarded as essential to the American way of life, but people tend to delay replacing an older vehicle when economic conditions deteriorate, as happened most vividly during the Great Recession. Economic stress can be measured in a variety of ways including unemployment levels and declining consumer sentiment. Aside from unusual economic conditions the general state of the economy is reflected in such measures as the level of consumer expenditure, the price level and the performance of the stock market. The demand for new automobiles is also affected by the price of gasoline and the level of interest rates for auto loans.

Monthly data are available for the period January 2000–December 2011 on the following variables.

AUTO_SALES: Seasonally adjusted annual rate (millions)

CRUDE: Price of crude oil ($/barrel)

CPI: Consumer Price Index

CONSUMPTION: Personal consumer expenditures ($billions)

UNEMP: Number of people unemployed (thousands)

UNEMP_PC: Unemployment level (percentage)

UNEMP_DUR: Mean length of time unemployed (months)

CON_SENT: Consumer sentiment (index)

S&P 500: S&P 500 Index

LIBOR: One month LIBOR (London Inter-Bank Offered Rate) interest rate

TREASURY_3: Interest rate on 3-year U.S. Treasury bonds

TREASURY_10: Interest rate on 10-year U.S. Treasury bonds

RECESSION: The timing of the Great Recession, as defined by the NBER. (This is called an indicator variable and is set = 1 during the period of the recession, and = 0 at all other times.)

Data Sources for Minicases

1. National Climatic Data Center (National Oceanic and Atmospheric Administration, Department of Commerce), *http://lwf.ncdc.noaa.gov/oa/climate/research/cag3/ md.html*
2. Bureau of Labor Statistics (U.S. Department of Labor), *www.bls.gov*
3. Energy Information Administration (U.S. Department of Energy), *www.eia.doe.gov/overview_hd.html* and *http://tonto.eia.doe.gov/dnav/ng/ng_stor_sum_ dcu_nus_m.htm*
4. Bureau of Economic Analysis (U.S. Department of Commerce), *www.bea.gov*
5. University of Michigan Index of Consumer Sentiment, *www.sca.isr.umich.edu*

As noted previously, a valuable general source for many macroeconomic series is FRED, provided by the Economic Research Division of the Federal Reserve Bank of St. Louis, *https://fred.stlouisfed.org/*.

References

Allen, P. G., and Fildes, R. (2001). Econometric forecasting. In J. S. Armstrong (ed.) *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Boston and Dordrecht: Kluwer, pp. 300–362.

Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D. and Cochran, J. J. (2014), *Statistics for Business and Economics*, 12th ed. Mason, OH: Cengage Learning.

Armstrong, J. S., ed. (2001). *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Boston and Dordrecht: Kluwer.

Hull, J. C. (2009). *Options, Futures, and Other Derivatives*, 7th ed. Upper Saddle River, NJ: Pearson Prentice Hall.

Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2005). *Applied Linear Statistical Models*, 5th ed. New York City: McGraw-Hill. [Re-issued as paperback in 2013]

Appendix 8A The Durbin-Watson Statistic

The Durbin-Watson test is often used to check for first-order autocorrelation in the random error terms in time series regression. The test compares the null hypothesis:

H_0: error terms are not autocorrelated

to the alternative

H_A: error terms display first order autocorrelation.

Given a sample of size n and residuals $\{e_t, t = 1, \ldots, n\}$ the Durbin-Watson statistic has the form:

$$DW = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}.$$

When H_0 is true, DW is close to 2.0. The minimum value is 0.0 and the maximum value is 4.0; small values indicate positive autocorrelation and large values indicate negative autocorrelation. The exact test requires detailed tables, which may be found in texts such as Kutner *et al.* (2005, pp. 1330–1331). The test is invalid when lagged values of the dependent variable are included in the regression model.

Because we use the sample autocorrelations as diagnostics and are often interested in autocorrelations at other lags as well, we prefer to use the sample *ACF* and *PACF*, as in Section 8.5. However, many software packages provide the value of the *DW* statistic but not the sample *ACF*. To make use of this information, we match up the values of *DW* with the test for first-order autocorrelation shown in Figure 8.6; this leads to the decision rule:

$$\text{Reject } H_0 \text{ with } \alpha = 0.05 \text{ if } |DW - 2.0| > \frac{3.92}{\sqrt{n}} \text{ ; do not reject otherwise.}$$

The right-hand side is $1.96 \times 2/\sqrt{n}$. The expression is based upon the assumption that *DW* is approximately normally distributed in large samples; the number 1.96 is the percentage point from normal tables and the other part represents the approximate standard error of *DW*.

Example

The gas prices model in Section 8.5.1 is found to have $DW = 0.738$. The sample size is $n = 155$ so the lower critical value is $2.0 - 3.92/\sqrt{155} = 2.0 - 0.315 = 1.685$. The observed value of 0.738 is well below this, so we reject the null hypothesis of no first order autocorrelation, reaching the same conclusion as we did from Figure 8.6.

CHAPTER 12

Putting Forecasting Methods to Work

*Topics marked with an * are advanced and may be omitted for more introductory courses.*

Table of Contents

Production is not the application of tools to materials, but logic to work.

— Peter F. Drucker

Introduction

Forecasting applications are many and are implemented in a wide variety of organizational settings. In this chapter, we focus on how the methods we have introduced can be put to work. First, in Section 12.1, we discuss a problem that is common to any organization wishing to appraise its current forecasting methods with a view to improving those methods. We seek also to establish how such an appraisal is best carried out. Most organizations employ a forecasting support system to produce their forecasts. (Such a system is computer-based and consists of data, user interventions, and forecasting methods.) Next, in Section 12.2, we examine the characteristics of the system and how users interact with it. Then, in Section 12.3, we examine perhaps the most common applications area for quantitative forecasting methods: forecasting to support operations in manufacturing, services, government, and retail. The key features here are that the forecasts are regularly updated and that the data are readily available. Of course, judgment is also used in arriving at the final forecasts. Next, in Section 12.4 we are concerned with marketing, where there is a requirement for longer term forecasts that includes various features of the market, such as whether the product is new, the competition the product faces, and the effects of various marketing drivers, such as price. Then, Section 12.5 considers how the methods we have introduced can be applied to the choices made by individuals to predict their behavior — in particular, their behavior as consumers. We round out the chapter with a discussion of other key areas in which quantitative methods are applied, such as macroeconomic forecasting. As with Chapters 10 and 11, we include many references in order to justify or amplify our argument so that the methods we describe can be more easily operationalized by the reader.

12.1 Evaluating a Forecasting Process

In earlier chapters, we introduced a variety of different forecasting methods, both simple and complex (e.g., simple exponential smoothing models, in contrast to regression models or neural networks). How should a forecasting manager with the responsibility of maintaining and developing a forecasting process appraise its current success and its potential for improvement? To answer this question, we need to define the criteria to be used in judging success: It can't be just a question of forecasting accuracy, which we've focused on so far. Typically, the criteria will be organization specific and will certainly include accuracy, but also to be taken into account are the costs and resources needed, the availability of the data, and the speed with which the forecasts are produced. A critical issue is the impact on the organization's bottom line, be it profit, service level, etc. In addition, users often require their forecasts to be produced by a process that they understand and that captures the key features of the activity system being modeled, such as its ability to include the effects of promotional activities and special events. In sum, we need to consider all the components of PIVASE, as defined in Chapter 1.

Forecasts often have a motivational element to them. For example, when a sales force is asked to produce forecasts of its expected sales, before a figure is offered back to the head office, it will certainly be examined for its effect on the target sales that might be set by senior management and on any resulting bonuses. Similarly, the sales forecasts need to be developed prior to the annual budget and must remain distinct from budgeted figures through the year. Whereas the budgeted financial figures depend critically on sales (and associated costs), the regularly updated forecasts through the year should reflect *current* market conditions and may well be incompatible with a budget set months earlier. Budgets also tend to have a motivational aspect to them that conflicts with our concept of a forecast as our best estimate of the most likely (or expected) outcome. In appraising the forecasting process, the interaction between setting targets and budgeting within an organizational setting may lead to distorted "best estimates" that are biased, less accurate forecasts, and result in poor decisions (Galbraith and Merrill, 1996). This topic is discussed in detail in Section 13.2.3. However, it is important to put in a proviso here: In this book, we assume that the aim of forecasting is accuracy (and authors of similar books assume the same thing). However, in some forecasting situations, the aim may well be to affect policy or to motivate, when scenarios might be more effective (Section 11.6). In that case, accuracy would be subservient to these other organizational demands.

One popular route to appraising forecasting performance is through benchmarking studies, whereby an organization compares its practices with those of its peers. The Institute of Business Forecasting and Planning (*www.ibf.org*) regularly conducts surveys of attendees at the organization's training workshops. However, such benchmarks have to be accepted with a large grain of salt because they have many problems, as described by Kolassa (2008). For instance, their samples often are small and not random within any given sector, and the answers obtained from them are not based directly on empirical data and are probably biased. But if such surveys are to be distrusted, and yet benchmarking is desirable, what is the alternative? In the next section, we discuss how to carry out a benchmarking study through a so-called forecasting competition. The aim of such a competition is to use an organization's own data to compare various methods with the organization's current approach. This strategy focuses effectively on the key benchmarking question "How are we doing?" Answering this question also points a way forward toward an improved forecasting system.

Forecasting competitions are most often used in organizations that are considering replacing their current forecasting software or forecasting process. An organization seeking to embark on changes will wish to ensure that the replacement software system is at least as good as the current system in terms of forecasting accuracy.

In the next section, we will focus on just this one aspect of the appraisal: the comparison of forecasting accuracy. However, there are many other aspects that need to be taken into account, and these are discussed later, in Section 12.2.

12.1.1 Evaluating Forecasting Methods: Forecasting Competitions

Most manufacturing and retailing companies need to forecast the sales they expect of the many hundreds, if not thousands, of products they sell. Some companies, such as electricity suppliers or call center operators, focus on just the small number of variables that interest them. For government departments, there is a similar limited forecasting need, although local, regional, and national forecasts of a variable such as unemployment benefit claimants often are required. Periodically, the forecaster needs to evaluate the forecasting accuracy of the chosen method over a number of series and to compare the method with plausible alternatives. As noted earlier, the need is most acute when a new forecasting system or process is being considered.

Carrying out the Evaluation

Step 1: Specify the population of time series to be considered. For a supermarket, for example, this population could include the products sold (in stock keeping units, or SKUs — the most detailed information about a product, including brand, pack size, packaging, and more) in a number of product categories, such as fresh vegetables, beer and lagers, cleaning products, and health care. We may then need to choose a sample from various important subpopulations or segments (the whole population being too large to analyze fully in a cost-effective manner).

Step 2: Define the forecasting task precisely. We can refer back to the PIVASE framework introduced in Chapter 1. The specification includes the information that is available in producing the forecasts, whether human intervention is permitted (or whether the method should be automatic), the forecast horizon(s), whether the method's parameters (such as the smoothing parameter) are reestimated, and the frequency with which the forecasts are updated. Also, we need to consider the estimated value of improved accuracy.

Step 3: Specify the forecasting methods to be considered. The range of methods should include a standard benchmark such as damped trend smoothing (discussed in Section 3.5), as well as a naïve method (e.g., random walk). If the data may be seasonal, a simple seasonal benchmark based on decomposition (Chapter 4) should be included. Although some of the methods used in a competition might not be practical for the organization (i.e., they may be too complex), including them in the competition may give some insight into the level of improvement possible.

Step 4: Define the measures of performance to be used in the evaluation. These measures should correspond as closely as possible to those used by the organization (including measures of profitability or service level) and should incorporate both standard measures and measures that are organization specific. Sometimes however, organizational measures in use are flawed (e.g., overly influenced by outliers or they can be manipulated) so using a robust measure such as Relative Mean Absolute Error (*RelMAE*) should be also included (see Section 2.7).

Step 5: Specify the data to be used in parameterizing the methods under consideration (the fit, or in-sample, data), as well as the data to be used in the out-of-sample (or hold-out) evaluation.

Step 6: Calculate the error measures derived from the various forecasting methods for the sub-populations of time series in the study. We may then identify the best methods and the relative importance of choosing between methods rather than sticking with the benchmark.

We have already introduced various measures of accuracy in Section 2.7; in particular, mean error (*ME*) and mean percentage error (*MPE*) as measures of bias, and mean absolute error (*MAE*) and mean absolute percentage error (*MAPE*) as measures of accuracy. When we wish to measure performance across a population (or sample), we need to decide what weight to give to each of the series in the sample. Table 12.1 illustrates the problem; in the table, we examine monthly electricity consumption by two households (extending the data in *Electricity.xlsx*, analyzed in Table 2.11).

Two forecasting methods — Smooth and Random Walk — have been used to produce forecasts of the two time series on household electricity consumption. In practice, there are often many more. In household 1, Smooth outperforms Random Walk, whether measured by *MAE* or *MAPE*. For household 2, Random Walk is the consistent winner. When the two measures are averaged across households, they give conflicting results. The conflict arises because of the difference in the mean consumption level of the two households and, consequently, the difference in scale of the *MAE*: 163 compared with 6.6. Thus, household 1 counts much more in the overall comparison when the *MAE* is used. However, when *MAPE* is the chosen error measure, the two households are of almost equal importance and Random Walk is identified as the better forecasting method (although the differences are small). The electric utility company would want to use *MAE* when planning production, but might prefer *MAPE* when forecasting consumption by individual customers. (Many companies charge customers monthly on the basis of estimated consumption, but check meters only every two or three months.)

Table 12.1 Comparing Two Forecasting Methods for Two Time Series

	Household 1 Household Mean = 939.2		Household 2 Household Mean = 46.3		Overall Overall Mean = 492.7	
	Smooth	Random Walk	Smooth	Random Walk	Smooth	Random Walk
Mean Error (ME)	–160.0	–149.0	2.6	0.2	–78.7	–74.5
MPE	–18.7	–17.3	3.7	–0.4	–7.4	–8.9
MAE	163.3	169.2	6.6	5.2	85.0	87.2
MAPE	19.1	19.8	13.8	11.1	16.4	15.5

Data: Competition.xlsx

This simple example, of course, generalizes: When error measures are aggregated across time series, the scale of the series might matter, implicitly affecting the weighting given to each series. So, how should a forecaster choose between conflicting measures? The differences can be substantial. In Step 4, in which we define a forecasting competition, we underlined the need to choose a measure that fits an organization's requirements such as profit or service level. In the preceding example, with the aim of forecasting overall electricity consumption, the *MAE* (or the *RMSE*) would be the more appropriate measure because the costs of generation or the marginal profit derived from selling electricity are proportional to the level of consumption. If our aim is to choose between a benchmark method and an alternative, *RelMAE* is the best choice.

Other approaches can be used to weight the individual error measures, such as weighting by product profit margin. The danger is that a forecasting method will be chosen that suits only a few products (or time series). The remedy is to segment the time series into more homogeneous subgroups and choose a (potentially different) method for each separate group (we explore this idea more fully in Section 13.1.5). As large a sample of series as is practical should be used in evaluating performance, particularly when the population being examined has to be segmented into many important subgroups.

In general, in carrying out a forecasting competition, a variety of error measures should be used. They should all conform to the requirements of a reliable and valid error measure. Key criteria to consider are the following:

i. Sensitivity to outliers in the errors (or relative errors)
   - ✓ *RMSE* is based on squared errors and therefore accentuates extreme outlying errors.

ii. Sensitivity to performance on individual series
   - ✓ *MAE* can be dominated by large errors because the data values are large, as are those in Table 12.1.
   - ✓ *MAPE* can be affected by series in which the actual value is close to zero.

iii. Effect of scale
   - ✓ *MAPE* and percentage errors are scale-independent.
   - ✓ Relative error measures, which measure the performance of one method relative to another, are scale-independent (for example, *RelMAE*; see Section 2.7).

iv. The measure's interpretability

✓ *RMSE* is not easily interpretable unless it is translated into inventory holding. Most forecasters find that *MAPE* conveys their understanding of accuracy, although relative error measures are also easily interpretable with regard to the extent to which one method is better than another.

A number of major forecasting competitions have been described in the research literature. Notable among them are those discussed by Newbold and Granger (1974), Makridakis and Hibon (1979), and the M1 and M3 Competitions (Makridakis *et al.*, 1982; Makridakis and Hibon, 2000). All these publications have stimulated sometimes critical commentary; see, for example, the M3 Competition based on 3000 time series and 24 methods.[1] According to Fildes, Hibon, Makridakis, and Meade (1998), key conclusions coming out of this research are as follows:

a. Statistically sophisticated or complex methods do not typically produce more accurate forecasts than simpler ones.
b. The rankings of the performance of the various methods vary with the error measures used.
c. The relative performance of the various methods depends upon the length of the forecasting horizon.
d. The characteristics of the data series are important factors in determining relative performance between methods; therefore, a method designed to capture the particular characteristics of a forecasting situation may well outperform standard benchmarks.
e. The sampling variability of performance measures renders comparisons based on a single time series and a single forecast origin unreliable; thus, comparisons based on multiple time series and on multiple forecast origins are recommended.

In addition, the damped trend method of exponential smoothing (see Section 3.5) has proved to be a benchmark that is very hard to beat, so a forecaster using only time series data should always compare performances with a damped trend (as well as a naïve) benchmark. These principles for carrying out a forecasting competition and the general findings just presented suggest that a new method should be adopted (or an established method continued in use) only if, in a forecasting competition that includes standard benchmarks, its performance is no worse than the benchmarks' performance when evaluated across a wide range of series and appropriate error measures. The same principles apply in appraising a new causal method. Here, the forecaster should take care to ensure that the same data set is used in building and evaluating each model.

However, there is an alternative approach to selecting a forecasting method; we consider this approach in the next section.

12.1.2 Combining Forecasting Methods or Choosing among Methods

Suppose we have developed two separate forecasting methods, each of which delivers one-step-ahead forecasts of Y_t, which we denote F_{1t} and F_{2t}. We are faced with a choice: do we select one method to use for our data series or, instead of trying to decide which of the two methods is better, should we produce a combined forecast? The simplest method of combining the two forecasts for Y_t is by averaging the two individual forecasts:

$$F_t = 0.5F_{1t} + 0.5F_{2t}. \tag{12.1}$$

1 The discussion of the M3 Competition appears in the *International Journal of Forecasting*, 2001,

More generally, we can try to weight the forecasts optimally so that the combined forecast produces the best forecast of Y_t by using least squares regression to estimate the weight w in the formula

$$F_t = wF_{1t} + (1 - w)F_{2t}. \tag{12.2}$$

When there are more than two forecasts available, the formulas generalize to

$$F_t = average(F_{1t}, F_{2t}, \ldots, F_{Kt})$$

and

$$F_t = w_1F_{1t} + w_2F_{2t} + \ldots + w_KF_{Kt},$$

where it is sometimes assumed that

$$\sum_{i=1}^{K} w_i = 1.$$

In addition, some forecasters prefer to restrict attention to nonnegative weights, an approach that has some intuitive appeal.[2]

Example 12.1: Combining forecasts

Table 12.2 (A and B) shows four sets of out-of-sample forecasts of call center data: a random-walk benchmark (*RW*), an exponentially smoothed forecast (*Smooth*), a causal model (*Causal*), and the company's judgmental forecast (*Company*). The table shows (some of) the raw data included in spreadsheet *Combination.xlsx* and some of the corresponding error statistics.

Table 12.2(A) The Combination of Forecasts Compared
(Absolute percentage error statistics for four forecasting methods)

		Forecasts from the Forecast Methods				Absolute Percentage Error			
Period	**Actual**	**RW**	**Smooth**	**Causal**	**Company**	**RW**	**Smooth**	**Causal**	**Company**
1	32.25	25.00	35.00	44.77	30.00	22.48	8.53	38.83	6.97
2	36.62	32.25	34.45	46.70	34.45	11.93	5.92	27.55	8.92
3	50.52	36.62	34.88	48.98	60.81	27.52	30.95	3.04	20.37
4	43.43	50.52	38.01	46.98	44.26	16.32	12.48	8.19	1.92
.									
.									
18	40.38	42.90	50.45	48.39	51.00	6.24	24.94	19.83	26.30
19	50.48	40.38	48.4	54.35	44.41	20.01	4.05	7.66	12.03
20	47.70	50.48		53.00	50.92	5.84	2.41	11.10	6.75
					MAPE	11.29	12.27	9.97	11.51
					MdAPE	9.34	12.98	7.13	8.80

Data: Combination.xlsx

2 As for the creation of neural network ensemble forecasts, in section 10.4, when there are several forecasts available, using the median to combine them, instead of the average, can perform very well, as it eliminates the impact of potentially extreme forecasts. This has been validated for the all methods of forecasting we have discussed (see Barrow and Kourentzes, 2016).

Table 12.2(B) The Combination of Forecasts Compared
(Absolute percentage error statistics for the forecast combinations)

	Absolute Percentage Errors Combination			
Period	**RW+Smooth**	**RW+Causal**	**RW+Company**	**All**
1	6.97	8.18	14.72	4.48
2	8.92	7.81	10.42	0.20
3	29.23	15.28	3.58	10.29
4	1.92	12.26	9.12	3.49
.				
.				
18	15.59	13.04	16.27	19.33
19	12.03	6.18	16.02	7.11
20	4.12	8.47	6.30	6.53
MAPE	10.37	7.30	9.61	7.02
MdAPE	8.80	6.99	9.13	6.57

Data: Combination.xlsx

Of the four individual methods, Causal performed best (9.97 percent *MAPE*). Combining by simple averaging improves the performance when the random walk (*RW*) is combined with the causal forecasting model (*RW + Causal*), but also when it is combined with the quite inaccurate company forecasts. However, combining all four methods outperforms not only all of the individual methods, but also the paired combinations. The improvement from 7.30 percent down to 7.02 percent in *MAPE* (and also in *MdAPE*) and this improvement is probably worth the extra effort. ■

Equation (12.2) for combining unequally weighted forecasts can be transformed to

$$Y_t - F_{2t} = w(F_{1t} - F_{2t}). \tag{12.3}$$

To estimate the unknown weight parameter w, the preceding regression is run with dependent variable $Y_t - F_{2t}$; note the constant term is suppressed. Using the regression estimate for w, however, does not improve the results much at all in the foregoing example — a finding of other researchers as well (see Exercise 12.1).

In general, the combination of a set of forecasts, based on a simple weighting scheme such as equal weights, often produces more accurate forecasts than trying to select the best individual method, series by series. This conclusion has been reached in many studies, including the competitions discussed in the previous section, as well as the many studies summarized by Clemen (1989), Armstrong (2001a) and Graefe, Armstrong, Jones and Cuzán (2014). But why should combining methods often work well when, surely, identifying the correct method would prove more accurate? The fundamental reason that combining is so effective is that all methods (and forecasts) suffer from deficiencies. Statistical approaches and judgmental approaches are all based on simplified understandings of the system being forecast, with nonlinearities and variables omitted. Even if a method is satisfactory for a period of time, the market or the economy will change, and then the method and its parameters become suboptimal. Each of the alternative methods is likely to contain some information that is helpful in improving the overall combined forecast. In addition, there is an element of insurance to using a number of methods, so that when market circumstances change, one method takes over from another as offering the best forecasts. Ord (1988) provides a simple model that explains this effect.

Intuitively, the combination will work best when the error from one method is counterbalanced by the error from another; in statistical terms, they are negatively correlated. If the errors are positively correlated, they accumulate, which is less helpful. Typically, errors from similar methods are positively correlated while errors from very different approaches have a low or even a negative correlation. Thus, an extrapolative method that exploits the autocorrelation structure in the time series can best be combined with a causal model (that includes some of the key explanatory variables) or managerial judgment. By including managerial judgment, the combination may have access to new information that could not have formed a part of the historical data. In general, however, forecast errors from different sources tend to be positively correlated, even when different methods are used (they tend to make the same mistakes), so the benefits that may be achieved from combining are sometimes limited.

In most practical situations, only a limited number of forecasting methods can contribute to the combination. When there are many alternative forecasts (see Exercise 12.2), it is often better (as in the preceding example) to use them all, although the gains from using more than five are usually limited. However, including inappropriate methods in the combination mix is not to be recommended. Therefore, benefits may be derived from excluding the extreme forecasts through trimming (removing, say, 20 percent of the extreme forecast errors in the calculation of the average; see Jose and Winkler, 2008) or using the median rather than the mean.

When forecasts are consistently biased with a nonzero mean error, combining is unlikely to prove beneficial if the foregoing formulas are used. Instead, the forecasts require modification to remove the bias; see Appendix 11B and Exercise 12.1(b).

DISCUSSION QUESTION: *How is the process of combining forecasts similar to portfolio diversification in finance? Why might such procedures reduce the forecast error (or investment risk)?*

What of the alternative to combining: choosing the best performing model from the set of alternatives? This is the implicit basis of the various forecasting competitions, which aim to identify the best performer over the population under study. The 'winner' then would be the method to adopt; it is also the approach used in a variety of software packages (although the selection algorithms employed in these commercial packages often perform poorly). Fildes and Petropoulos (2015) have investigated model selection compared to combining and found that in many circumstances selection is the better option, in particular selection works best when specific sub-populations of data are considered (e.g., trended or seasonal series), but also when the alternative methods' comparative performance is stable over time. Although more demanding to implement, the idea of first segmenting the data series, examining the comparative performance of plausible methods to see if their ranked performance on the in-sample or validation data has remained much the same and then adopting the winning methods, does lead to worthwhile gains.