Anderson Sweeney Williams Camm Cochran Fry Ohlmann

An Introduction to Management Science

Quantitative Approaches to Decision Making

CHAPTER 15

Time Series Analysis and Forecasting

CONTENTS

The purpose of this chapter is to provide an introduction to time series analysis and forecasting. Suppose we are asked to provide quarterly forecasts of sales for one of our company's products over the coming one-year period. Production schedules, raw materials purchasing, inventory policies, and sales quotas will all be affected by the quarterly forecasts we provide. Consequently, poor forecasts may result in poor planning and increased costs for the company. How should we go about providing the quarterly sales forecasts? Good judgment, intuition, and an awareness of the state of the economy may give us a rough idea or "feeling" of what is likely to happen in the future, but converting that feeling into a number that can be used as next year's sales forecast is challenging. The Management Science in Action, Forecasting Energy Needs in the Utility Industry, describes the role that forecasting plays in the utility industry.

A forecast is simply a prediction of what will happen in the future. Managers must accept that regardless of the technique used, they will not be able to develop perfect forecasts.

Forecasting methods can be classified as qualitative or quantitative. Qualitative methods generally involve the use of expert judgment to develop forecasts. Such methods are appropriate when historical data on the variable being forecast are either unavailable or not applicable. Quantitative forecasting methods can be used when (1) past information about the variable being forecast is available, (2) the information can be quantified, and (3) it is reasonable to assume that past is prologue (i.e., the pattern of the past will continue into the future). We will focus exclusively on quantitative forecasting methods in this chapter.

If the historical data are restricted to past values of the variable to be forecast, the forecasting procedure is called a *time series method* and the historical data are referred to as a time series. The objective of time series analysis is to uncover a pattern in the historical data or time series and then extrapolate the pattern into the future; the forecast is based solely on past values of the variable and/or on past forecast errors.

In Section 15.1 we discuss the various kinds of time series that a forecaster might be faced with in practice. These include a constant or horizontal pattern, a trend, a seasonal pattern, both a trend and a seasonal pattern, and a cyclical pattern. In order to build a quantitative

MANAGEMENT SCIENCE IN ACTION

FORECASTING ENERGY NEEDS IN THE UTILITY INDUSTRY*

Duke Energy is a diversified energy company with a portfolio of natural gas and electric businesses and an affiliated real estate company. In 2006, Duke Energy merged with Cinergy of Cincinnati, Ohio, to create one of North America's largest energy companies, with assets totaling more than $70 billion. As a result of this merger the Cincinnati Gas & Electric Company became part of Duke Energy. Today, Duke Energy services over 5.5 million retail electric and gas customers in North Carolina, South Carolina, Ohio, Kentucky, Indiana, and Ontario, Canada.

Forecasting in the utility industry offers some unique perspectives. Because energy is difficult to store, this product must be generated to meet the instantaneous requirements of the customers. Electrical shortages are not just lost sales, but "brownouts" or "blackouts." This situation places an unusual burden on the utility forecaster. On the positive side, the demand for energy and the sale of energy are more predictable than for many other products. Also, unlike the situation in a multiproduct firm, a great amount of forecasting effort and expertise can be concentrated on the two products: gas and electricity.

The largest observed electric demand for any given period, such as an hour, a day, a month, or a year, is defined as the peak load. The forecast of the annual electric peak load guides the timing decision for constructing future generating units, and the financial impact of this decision is great. Obviously, a timing decision that leads to having the unit available no sooner than necessary is crucial.

The energy forecasts are important in other ways also. For example, purchases of coal as fuel for the generating units are based on the forecast levels of energy needed. The revenue from the electric operations of the company is determined from forecasted sales, which in turn enters into the planning of rate changes and external financing. These planning and decision-making processes are among the most important managerial activities in the company. It is imperative that the decision makers have the best forecast information available to assist them in arriving at these decisions.

*Based on information provided by Dr. Richard Evans of Duke Energy.

forecasting model it is also necessary to have a measurement of forecast accuracy. Different measurements of forecast accuracy, and their respective advantages and disadvantages, are discussed in Section 15.2. In Section 15.3 we consider the simplest case, which is a horizontal or constant pattern. For this pattern, we develop the classical moving average, weighted moving average, and exponential smoothing models. Many time series have a trend, and taking this trend into account is important; in Section 15.4 we provide regression models for finding the best model parameters when a linear trend is present. Finally, in Section 15.5 we show how to incorporate both a trend and seasonality into a forecasting model.

15.1 TIME SERIES PATTERNS

A **time series** is a sequence of observations on a variable measured at successive points in time or over successive periods of time. The measurements may be taken every hour, day, week, month, or year, or at any other regular interval.[1] The pattern of the data is an important factor in understanding how the time series has behaved in the past. If such behavior can be expected to continue in the future, we can use it to guide us in selecting an appropriate forecasting method.

To identify the underlying pattern in the data, a useful first step is to construct a time series plot. A **time series plot** is a graphical presentation of the relationship between time and the time series variable; time is represented on the horizontal axis and values of the time series variable are shown on the vertical axis. Let us first review some of the common types of data patterns that can be identified when examining a time series plot.

Horizontal Pattern

A horizontal pattern exists when the data fluctuate randomly around a constant mean over time. To illustrate a time series with a horizontal pattern, consider the 12 weeks of data in Table 15.1. These data show the number of gallons of gasoline (in 1000s) sold by a gasoline distributor in Bennington, Vermont, over the past 12 weeks. The average value or mean for this time series is 19.25 or 19,250 gallons per week. Figure 15.1 shows a time series plot for these data. Note how the data fluctuate around the sample mean of 19,250 gallons. Although random variability is present, we would say that these data follow a horizontal pattern.

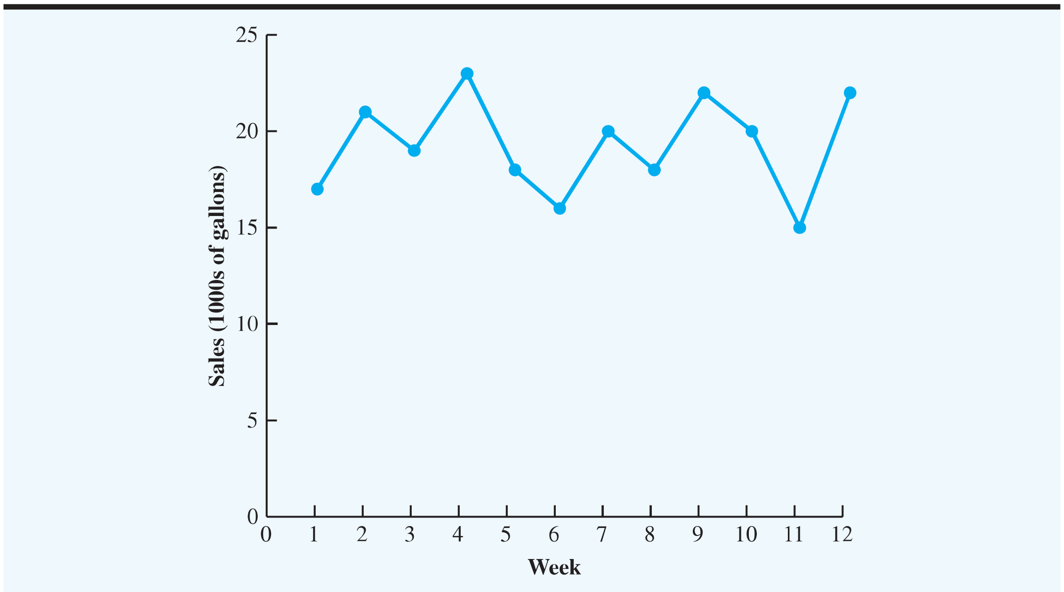
TABLE 15.1 GASOLINE SALES TIME SERIES

DATA file
Gasoline

Week	Sales (1000s of gallons)
1	17
2	21
3	19
4	23
5	18
6	16
7	20
8	18
9	22
10	20
11	15
12	22

[1]We limit our discussion to time series for which the values of the series are recorded at equal intervals. Cases in which the observations are made at unequal intervals are beyond the scope of this text.

FIGURE 15.1 GASOLINE SALES TIME SERIES PLOT



The term **stationary time series**[2] is used to denote a time series whose statistical properties are independent of time. In particular this means that

1. The process generating the data has a constant mean.
2. The variability of the time series is constant over time.

A time series plot for a stationary time series will always exhibit a horizontal pattern with random fluctuations. However, simply observing a horizontal pattern is not sufficient evidence to conclude that the time series is stationary. More advanced texts on forecasting discuss procedures for determining if a time series is stationary and provide methods for transforming a time series that is nonstationary into a stationary series.

Changes in business conditions often result in a time series with a horizontal pattern that shifts to a new level at some point in time. For instance, suppose the gasoline distributor signs a contract with the Vermont Sate Police to provide gasoline for state police cars located in southern Vermont beginning in Week 13. With this new contract, the distributor naturally expects to see a substantial increase in weekly sales starting in Week 13. Table 15.2 shows the number of gallons of gasoline sold for the original time series and the 10 weeks after signing the new contract. Figure 15.2 shows the corresponding time series plot. Note the increased level of the time series beginning in Week 13. This change in the level of the time series makes it more difficult to choose an appropriate forecasting method. Selecting a forecasting method that adapts well to changes in the level of a time series is an important consideration in many practical applications.

Trend Pattern

Although time series data generally exhibit random fluctuations, a time series may also show gradual shifts or movements to relatively higher or lower values over a longer period of time. If a time series plot exhibits this type of behavior, we say that a **trend pattern** exists. A trend is usually the result of long-term factors such as population increases or decreases, shifting demographic characteristics of the population, improving technology, and/or changes in consumer preferences.

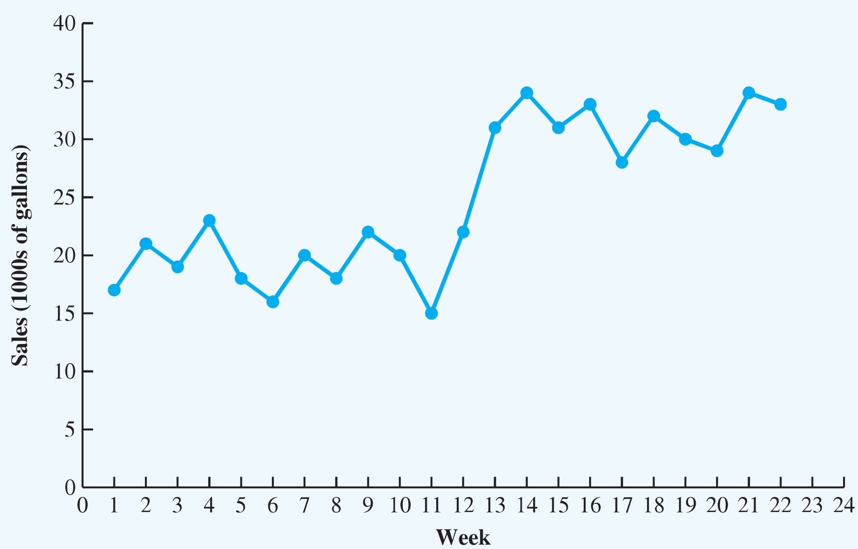
---
[2]For a formal definition of stationarity, see K. Ord and R. Fildes (2012), *Principles of Business Forecasting*. Mason, OH: Cengage Learning, p. 155.

TABLE 15.2 GASOLINE SALES TIME SERIES AFTER OBTAINING THE CONTRACT WITH THE VERMONT STATE POLICE

DATA file
GasolineRevised

Week	Sales (1000s of gallons)	Week	Sales (1000s of gallons)
1	17	12	22
2	21	13	31
3	19	14	34
4	23	15	31
5	18	16	33
6	16	17	28
7	20	18	32
8	18	19	30
9	22	20	29
10	20	21	34
11	15	22	33

FIGURE 15.2 GASOLINE SALES TIME SERIES PLOT AFTER OBTAINING THE CONTRACT WITH THE VERMONT STATE POLICE



To illustrate a time series with a linear trend pattern, consider the time series of bicycle sales for a particular manufacturer over the past 10 years, as shown in Table 15.3 and Figure 15.3. Note that 21,600 bicycles were sold in Year 1, 22,900 were sold in Year 2, and so on. In Year 10, the most recent year, 31,400 bicycles were sold. Visual inspection of the time series plot shows some up and down movement over the past 10 years, but the time series seems also to have a systematically increasing or upward trend.

The trend for the bicycle sales time series appears to be linear and increasing over time, but sometimes a trend can be described better by other types of patterns. For instance, the data in Table 15.4 and the corresponding time series plot in Figure 15.4 show the sales revenue for a cholesterol drug since the company won FDA approval for the drug 10 years ago. The time series increases in a nonlinear fashion; that is, the rate of change of revenue does not increase by a constant amount from one year to the next. In fact, the revenue appears to be growing in an exponential fashion. Exponential relationships such as this are appropriate when the percentage change from one period to the next is relatively constant.

TABLE 15.3 BICYCLE SALES TIME SERIES

DATA *file*
Bicycle

Year	Sales (1000s)
1	21.6
2	22.9
3	25.5
4	21.9
5	23.9
6	27.5
7	31.5
8	29.7
9	28.6
10	31.4

FIGURE 15.3 BICYCLE SALES TIME SERIES PLOT

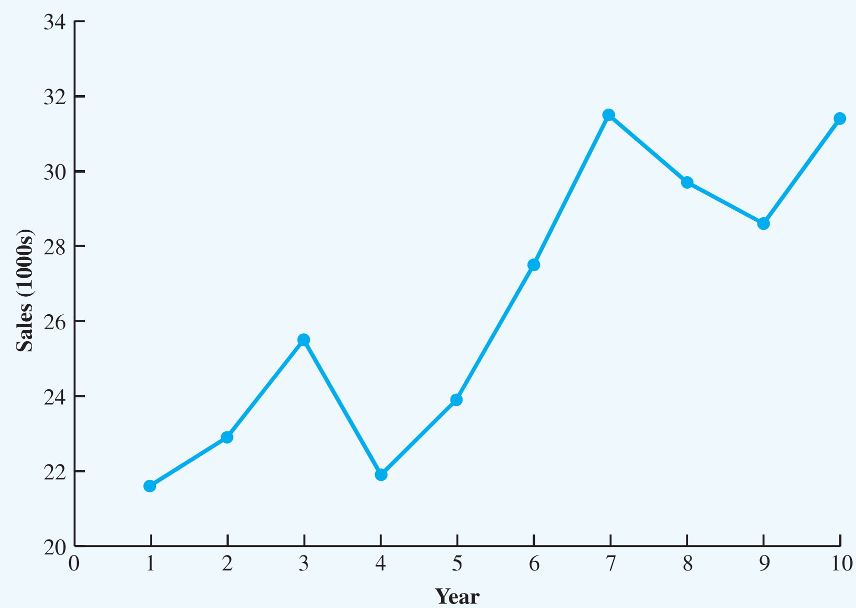
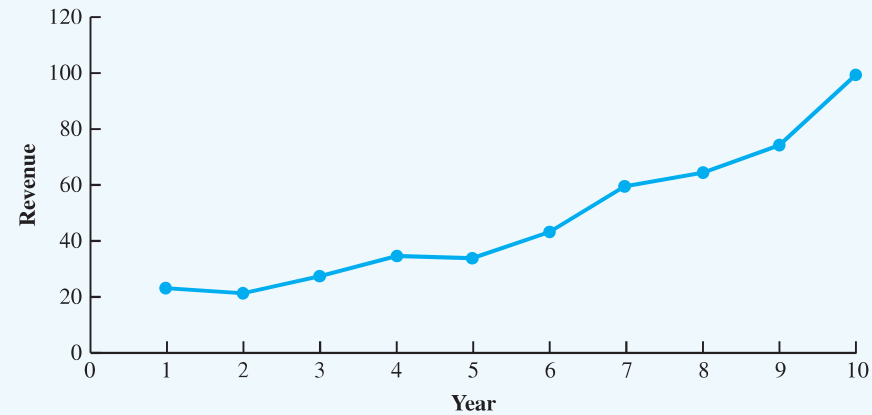


TABLE 15.4 CHOLESTEROL DRUG REVENUE TIME SERIES ($ MILLIONS)

DATA *file*
Cholesterol

Year	Revenue
1	23.1
2	21.3
3	27.4
4	34.6
5	33.8
6	43.2
7	59.5
8	64.4
9	74.2
10	99.3

FIGURE 15.4 CHOLESTEROL DRUG REVENUE TIME SERIES PLOT ($ MILLIONS)



Seasonal Pattern

The trend of a time series can be identified by analyzing movements in historical data over multiple years. **Seasonal patterns** are recognized by observing recurring patterns over successive periods of time. For example, a manufacturer of swimming pools expects low sales activity in the fall and winter months, with peak sales in the spring and summer months to occur each year. Manufacturers of snow removal equipment and heavy clothing, however, expect the opposite yearly pattern. Not surprisingly, the pattern for a time series plot that exhibits a recurring pattern over a one-year period due to seasonal influences is called a seasonal pattern. While we generally think of seasonal movement in a time series as occurring within one year, time series data can also exhibit seasonal patterns of less than one year in duration. For example, daily traffic volume shows within-the-day "seasonal" behavior, with peak levels occurring during rush hours, moderate flow during the rest of the day and early evening, and light flow from midnight to early morning. Another example of an industry with sales that exhibit easily discernable seasonal patterns within a day is the restaurant industry.

As an example of a seasonal pattern, consider the number of umbrellas sold at a clothing store over the past five years. Table 15.5 shows the time series and Figure 15.5 shows the corresponding time series plot. The time series plot does not indicate a long-term trend in sales. In fact, unless you look carefully at the data, you might conclude that the data follow a horizontal pattern with random fluctuation. However, closer inspection of the fluctuations in the time series plot reveals a systematic pattern in the data that occurs within each year. That is, the first and third quarters have moderate sales, the second quarter has the highest sales, and the fourth quarter tends to have the lowest sales volume. Thus, we would conclude that a quarterly seasonal pattern is present.

Trend and Seasonal Pattern

Some time series include both a trend and a seasonal pattern. For instance, the data in Table 15.6 and the corresponding time series plot in Figure 15.6 show quarterly smartphone sales for a particular manufacturer over the past four years. Clearly an increasing trend is present. However, Figure 15.6 also indicates that sales are lowest in the second quarter of each year and highest in quarters 3 and 4. Thus, we conclude that a seasonal pattern also exists for smartphones sales. In such cases we need to use a forecasting method that is capable of dealing with both trend and seasonality.

Cyclical Pattern

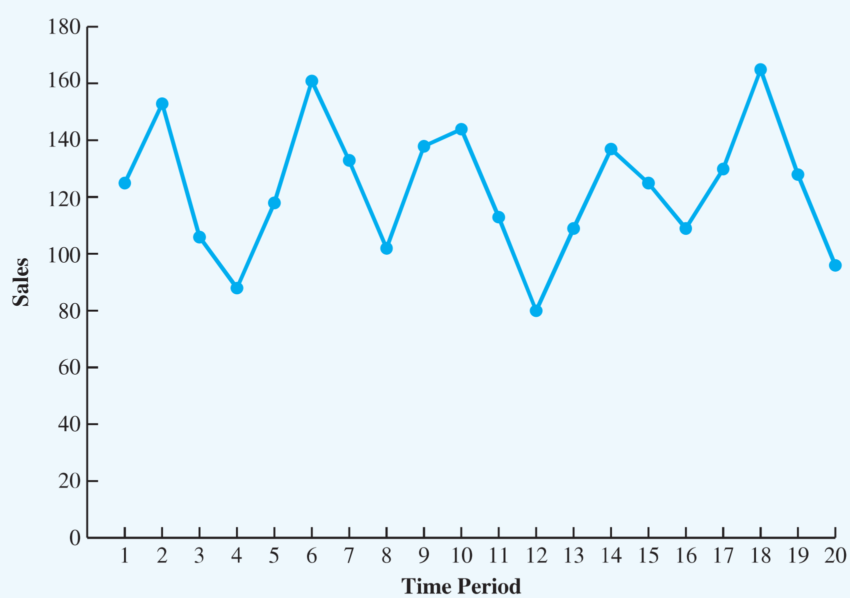
A **cyclical pattern** exists if the time series plot shows an alternating sequence of points below and above the trend line that lasts for more than one year. Many economic time series exhibit

TABLE 15.5 UMBRELLA SALES TIME SERIES

DATA file
Umbrella

Year	Quarter	Sales
1	1	125
	2	153
	3	106
	4	88
2	1	118
	2	161
	3	133
	4	102
3	1	138
	2	144
	3	113
	4	80
4	1	109
	2	137
	3	125
	4	109
5	1	130
	2	165
	3	128
	4	96

FIGURE 15.5 UMBRELLA SALES TIME SERIES PLOT



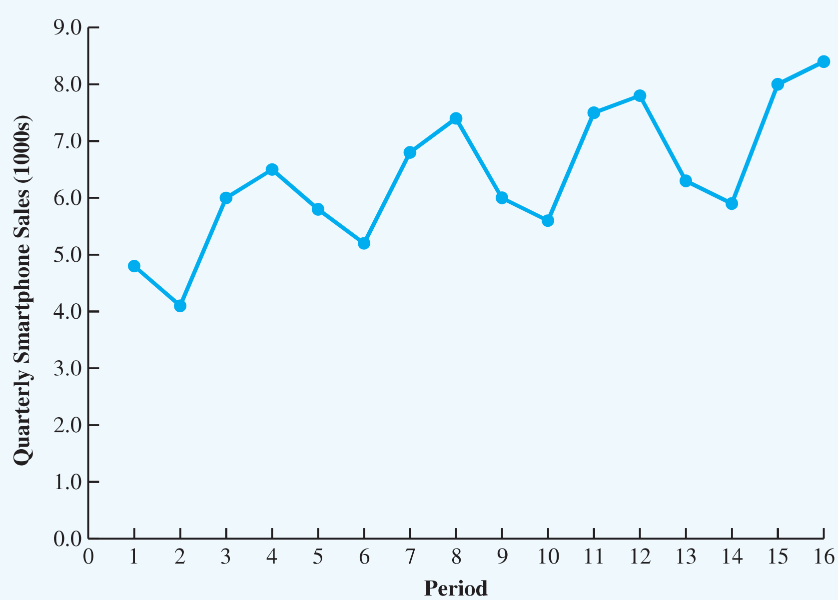
cyclical behavior with regular runs of observations below and above the trend line. Often the cyclical component of a time series is due to multiyear business cycles. For example, periods of moderate inflation followed by periods of rapid inflation can lead to a time series that alternates below and above a generally increasing trend line (e.g., a time series for housing costs). Business cycles are extremely difficult, if not impossible, to forecast. As a result, cyclical effects

TABLE 15.6 QUARTERLY SMARTPHONE SALES TIME SERIES

Year	Quarter	Sales (1000s)
1	1	4.8
	2	4.1
	3	6.0
	4	6.5
2	1	5.8
	2	5.2
	3	6.8
	4	7.4
3	1	6.0
	2	5.6
	3	7.5
	4	7.8
4	1	6.3
	2	5.9
	3	8.0
	4	8.4

DATA file
SmartPhoneSales

FIGURE 15.6 QUARTERLY SMARTPHONE SALES TIME SERIES PLOT



are often combined with long-term trend effects and referred to as trend-cycle effects. In this chapter we do not deal with cyclical effects that may be present in the time series.

Selecting a Forecasting Method

The underlying pattern in the time series is an important factor in selecting a forecasting method. Thus, a time series plot should be one of the first analytic tools employed when trying to determine which forecasting method to use. If we see a horizontal pattern, then we need to select a method appropriate for this type of pattern. Similarly, if we observe a trend in the data, then we need to use a forecasting method that is capable of handling a trend effectively. In the next two sections we illustrate methods for assessing forecast accuracy and consider forecasting models that can be used in situations for which the underlying pattern is horizontal; in other words, no trend or seasonal effects are present. We then consider methods appropriate when trend and/or seasonality are present in the data. The Management Science in Action, Forecasting Demand for a Broad Product Line of Office Products, describes the considerations made by ACCO Brands when forecasting demand for its consumer and office products.

MANAGEMENT SCIENCE IN ACTION

FORECASTING DEMAND FOR A BROAD PRODUCT LINE OF OFFICE PRODUCTS*

ACCO Brands Corporation is one of the world's largest suppliers of branded office and consumer products and print finishing solutions. The company's widely recognized brands include AT-A-GLANCE*, Day-Timer*, Five Star*, GBC*, Hilroy*, Kensington*, Marbig*, Mead*, NOBO, Quartet*, Rexel, Swingline*, Tilibra*, Wilson Jones*, and many others.

Because it produces and markets a wide array of products with a myriad of demand characteristics, ACCO Brands relies heavily on sales forecasts in planning its manufacturing, distribution, and marketing activities. By viewing its relationship in terms of a supply chain, ACCO Brands and its customers (which are generally retail chains) establish close collaborative relationships and consider each other to be valued partners. As a result, ACCO Brands' customers share valuable information and data that serve as inputs into ACCO Brands' forecasting process.

In her role as a forecasting manager for ACCO Brands, Vanessa Baker appreciates the importance of this additional Information. "We do separate forecasts of demand for each major customer," said Baker, "and we generally use twenty-four to thirty-six months of history to generate monthly forecasts twelve to eighteen months into the future. While trends are important, several of our major product lines, including school, planning and organizing, and decorative calendars, are heavily seasonal, and seasonal sales make up the bulk of our annual volume."

Daniel Marks, one of several account-level strategic forecast managers for ACCO Brands, adds:

> The supply chain process includes the total lead time from identifying opportunities to making or procuring the product to getting the product on the shelves to align with the forecasted demand; this can potentially take several months, so the accuracy of our forecasts is critical throughout each step of the supply chain. Adding to this challenge is the risk of obsolescence. We sell many dated items, such as planners and calendars, which have a natural, built-in obsolescence. In addition, many of our products feature designs that are fashion-conscious or contain pop culture images, and these products can also become obsolete very quickly as tastes and popularity change. An overly optimistic forecast for these products can be very costly, but an overly pessimistic forecast can result in lost sales potential and give our competitors an opportunity to take market share from us.

In addition to looking at trends, seasonal components, and cyclical patterns, Baker and Marks must contend with several other factors. Baker notes, "We have to adjust our forecasts for upcoming promotions by our customers." Marks agrees and adds:

> We also have to go beyond just forecasting consumer demand; we must consider the retailer's specific needs in our order forecasts, such as what type of display will be used and how many units of a product must be on display to satisfy their presentation requirements. Current inventory is another factor—if a customer is carrying either too much or too little inventory, that will affect their future orders, and we need lo reflect that in our forecasts. Will the product have a short life because it is tied lo a cultural fad? What are the retailer's marketing and markdown strategics? Our knowledge of the environments in which our supply chain partners are competing helps us to forecast demand more accurately, and that reduces waste and makes our customers, as well as ACCO Brands, far more profitable.

*The authors are indebted to Vanessa Baker and Daniel Marks of ACCO Brands for providing input for this Management Science in Action.

15.2 FORECAST ACCURACY

In this section we begin by developing forecasts for the gasoline time series shown in Table 15.1 using the simplest of all the forecasting methods, an approach that uses the most recent week's sales volume as the forecast for the next week. For instance, the distributor sold 17,000 gallons of gasoline in Week 1; this value is used as the forecast for Week 2. Next, we use 21, the actual value of sales in Week 2, as the forecast for Week 3, and so on. The forecasts obtained for the historical data using this method are shown in Table 15.7 in the column labeled Forecast. Because of its simplicity, this method is often referred to as a naïve forecasting method.

How accurate are the forecasts obtained using this naïve forecasting method? To answer this question we will introduce several measures of forecast accuracy. These measures are used to determine how well a particular forecasting method is able to reproduce the time series data that are already available. By selecting the method that is most accurate for the data already known, we hope to increase the likelihood that we will obtain more accurate forecasts for future time periods.

The key concept associated with measuring forecast accuracy is **forecast error**. If we denote Y_t and F_t as the actual and forcasted values of the time series for period t, respectively, the forecasting error for period t is

$$e_t = Y_t - \hat{Y}_t \tag{15.1}$$

That is, the forecast error for time period t is the difference between the actual and the forecasted values for period t.

For instance, because the distributor actually sold 21,000 gallons of gasoline in Week 2 and the forecast, using the sales volume in Week 1, was 17,000 gallons, the forecast error in Week 2 is

$$\text{Forecast Error in Week 2} = e_2 = Y_2 - \hat{Y}_2 = 21 - 17 = 4$$

The fact that the forecast error is positive indicates that in Week 2 the forecasting method underestimated the actual value of sales. Next we use 21, the actual value of sales in Week 2, as the forecast for Week 3. Since the actual value of sales in Week 3 is 19, the forecast error for Week 3 is $e_3 = 19 - 21 = -2$. In this case, the negative forecast error indicates the forecast overestimated the actual value for Week 3. Thus, the forecast error may be positive or negative, depending on whether the forecast is too low or too high. A complete summary of the forecast errors for this naïve forecasting method is shown in Table 15.7 in the column labeled Forecast Error. It is important to note that because we are using a past value of the time series to produce a forecast for period t, we do not have sufficient data to produce a naïve forecast for the first week of this time series.

A simple measure of forecast accuracy is the mean or average of the forecast errors. If we have n periods in our time series and k is the number of periods at the beginning of the time series for which we cannot produce a naïve forecast, the mean forecast error (MFE) is

$$\text{MFE} = \frac{\sum_{t=k+1}^{n} e_t}{n - k} \tag{15.2}$$

Table 15.7 shows that the sum of the forecast errors for the gasoline sales time series is 5; thus, the mean or average error is $5/11 = 0.45$. Because we do not have sufficient data to produce a naïve forecast for the first week of this time series, we must adjust our calculations in both the numerator and denominator accordingly. This is common in forecasting; we often use k past periods from the time series to produce forecasts, and so we frequently cannot produce forecasts for the first k periods. In those instances the summation in the numerator starts at the first value of t for which we have produced a forecast (so we begin the summation at $t = k + 1$), and the denominator (which is the number of periods in our time series for which we are able to produce a forecast) will also reflect these circumstances. In the gasoline example, although the time series consists of 12 values, to compute the mean error we divided the sum

TABLE 15.7 COMPUTING FORECASTS AND MEASURES OF FORECAST ACCURACY USING THE MOST RECENT VALUE AS THE FORECAST FOR THE NEXT PERIOD

Week	Time Series Value	Forecast	Forecast Error	Absolute Value of Forecast Error	Squared Forecast Error	Percentage Error	Absolute Value of Percentage Error
1	17						
2	21	17	4	4	16	19.05	19.05
3	19	21	−2	2	4	−10.53	10.53
4	23	19	4	4	16	17.39	17.39
5	18	23	−5	5	25	−27.78	27.78
6	16	18	−2	2	4	−12.50	12.50
7	20	16	4	4	16	20.00	20.00
8	18	20	−2	2	4	−11.11	11.11
9	22	18	4	4	16	18.18	18.18
10	20	22	−2	2	4	−10.00	10.00
11	15	20	−5	5	25	−33.33	33.33
12	22	15	7	7	49	31.82	31.82
		Total	5	41	179	1.19	211.69

of the forecast errors by 11 because there are only 11 forecast errors (we cannot generate forecast sales for the first week using this naïve forecasting method).

Also note that in the gasoline time series, the mean forecast error is positive, which implies that the method is generally underforecasting; in other words, the observed values tend to be greater than the forecasted values. Because positive and negative forecast errors tend to offset one another, the mean error is likely to be small; thus, the mean error is not a very useful measure of forecast accuracy.

The **mean absolute error**, denoted MAE, is a measure of forecast accuracy that avoids the problem of positive and negative forecast errors offsetting one another. As you might expect given its name, MAE is the average of the absolute values of the forecast errors:

$$\text{MAE} = \frac{\sum_{t=k+1}^{n} |e_t|}{n-k} \tag{15.3}$$

This is also referred to as the mean absolute deviation or MAD. Table 15.7 shows that the sum of the absolute values of the forecast errors is 41; thus

$$\text{MAE} = \text{average of the absolute value of forecast errors} = \frac{41}{11} = 3.73$$

Another measure that avoids the problem of positive and negative errors offsetting each other is obtained by computing the average of the squared forecast errors. This measure of forecast accuracy, referred to as the **mean squared error**, is denoted as MSE:

$$\text{MSE} = \frac{\sum_{t=k+1}^{n} e_t^2}{n-k} \tag{15.4}$$

From Table 15.7, the sum of the squared errors is 179; hence,

$$\text{MSE} = \text{average of the sum of squared forecast errors} = \frac{179}{11} = 16.27$$

The size of MAE and MSE depends upon the scale of the data. As a result, it is difficult to make comparisons for different time intervals (such as comparing a method of forecasting monthly gasoline sales to a method of forecasting weekly sales) or to make comparisons across different time series (such as monthly sales of gasoline and monthly sales of oil filters). To make comparisons such as these we need to work with relative or percentage error measures. The **mean absolute percentage error**, denoted as MAPE, is such a measure. To compute MAPE we must first compute the percentage error for each forecast:

$$\left(\frac{e_t}{Y_t}\right)100$$

For example, the percentage error corresponding to the forecast of 17 in Week 2 is computed by dividing the forecast error in Week 2 by the actual value in Week 2 and multiplying the result by 100. For Week 2 the percentage error is computed as follows:

$$\text{Percentage error for Week 2} = \left(\frac{e_2}{Y_2}\right)100 = \left(\frac{4}{21}\right)100 = 19.05\%$$

Thus, the forecast error for Week 2 is 19.05% of the observed value in Week 2. A complete summary of the percentage errors is shown in Table 15.7 in the column labeled Percentage Error. In the next column, we show the absolute value of the percentage error. Finally, we find the MAPE, which is calculated as

$$\text{MAPE} = \frac{\sum_{t=k+1}^{n} \left| \left(\frac{e_t}{Y_t}\right)100 \right|}{n - k} \tag{15.5}$$

Table 15.7 shows that the sum of the absolute values of the percentage errors is 211.69; thus

$$\begin{aligned}\text{MAPE} &= \text{average of the absolute value of percentage forecast errors}\\ &= \frac{211.69}{11} = 19.24\%\end{aligned}$$

In summary, using the naïve (most recent observation) forecasting method, we obtained the following measures of forecast accuracy:

$$\text{MAE} = 3.73$$
$$\text{MSE} = 16.27$$
$$\text{MAPE} = 19.24\%$$

Try Problem 1 for practice in computing measures of forecast accuracy.

These measures of forecast accuracy simply measure how well the forecasting method is able to forecast historical values of the time series. Now, suppose we want to forecast sales for a future time period, such as Week 13. In this case the forecast for Week 13 is 22, the actual value of the time series in Week 12. Is this an accurate estimate of sales for Week 13? Unfortunately there is no way to address the issue of accuracy associated with forecasts for future time periods. However, if we select a forecasting method that works well for the historical data, and we have reason to believe the historical pattern will continue into the future, we should obtain forecasts that will ultimately be shown to be accurate.

Before closing this section, let us consider another method for forecasting the gasoline sales time series in Table 15.1. Suppose we use the average of all the historical data available as the forecast for the next period. We begin by developing a forecast for Week 2. Since there is only one historical value available prior to Week 2, the forecast for Week 2 is

TABLE 15.8 COMPUTING FORECASTS AND MEASURES OF FORECAST ACCURACY USING THE AVERAGE OF ALL THE HISTORICAL DATA AS THE FORECAST FOR THE NEXT PERIOD

Week	Time Series Value	Forecast	Forecast Error	Absolute Value of Forecast Error	Squared Forecast Error	Percentage Error	Absolute Value of Percentage Error
1	17						
2	21	17.00	4.00	4.00	16.00	19.05	19.05
3	19	19.00	0.00	0.00	0.00	0.00	0.00
4	23	19.00	4.00	4.00	16.00	17.39	17.39
5	18	20.00	−2.00	2.00	4.00	−11.11	11.11
6	16	19.60	−3.60	3.60	12.96	−22.50	22.50
7	20	19.00	1.00	1.00	1.00	5.00	5.00
8	18	19.14	−1.14	1.14	1.31	−6.35	6.35
9	22	19.00	3.00	3.00	9.00	13.64	13.64
10	20	19.33	0.67	0.67	0.44	3.33	3.33
11	15	19.40	−4.40	4.40	19.36	−29.33	29.33
12	22	19.00	3.00	3.00	9.00	13.64	13.64
		Total	4.52	26.81	89.07	2.75	141.34

just the time series value in Week 1; thus, the forecast for Week 2 is 17,000 gallons of gasoline. To compute the forecast for Week 3, we take the average of the sales values in Weeks 1 and 2. Thus,

$$\hat{Y}_3 = \frac{17 + 21}{2} = 19$$

Similarly, the forecast for Week 4 is

$$\hat{Y}_4 = \frac{17 + 21 + 19}{3} = 19$$

The forecasts obtained using this method for the gasoline time series are shown in Table 15.8 in the column labeled Forecast. Using the results shown in Table 15.8, we obtained the following values of MAE, MSE, and MAPE:

$$\text{MAE} = \frac{26.81}{11} = 2.44$$

$$\text{MSE} = \frac{89.07}{11} = 8.10$$

$$\text{MAPE} = \frac{141.34}{11} = 12.85\%$$

We can now compare the accuracy of the two forecasting methods we have considered in this section by comparing the values of MAE, MSE, and MAPE for each method.

	Naïve Method	Average of Past Values
MAE	3.73	2.44
MSE	16.27	8.10
MAPE	19.24%	12.85%

For each of these measures, the average of past values provides more accurate forecasts than using the most recent observation as the forecast for the next period. In general, if the underlying time series is stationary, the average of all the historical data will provide the most accurate forecasts.

Evaluating different forecasts based on historical accuracy is only helpful if historical patterns continue in to the future. As we note in Section 15.1, the 12 observations of Table 15.1 comprise a stationary time series. In Section 15.1 we mentioned that changes in business conditions often result in a time series that is not stationary. We discussed a situation in which the gasoline distributor signed a contract with the Vermont State Police to provide gasoline for state police cars located in southern Vermont. Table 15.2 shows the number of gallons of gasoline sold for the original time series and the 10 weeks after signing the new contract, and Figure 15.2 shows the corresponding time series plot. Note the change in level in Week 13 for the resulting time series. When a shift to a new level such as this occurs, it takes several periods for the forecasting method that uses the average of all the historical data to adjust to the new level of the time series. However, in this case the simple naïve method adjusts very rapidly to the change in level because it uses only the most recent observation available as the forecast.

Measures of forecast accuracy are important factors in comparing different forecasting methods, but we have to be careful to not rely too heavily upon them. Good judgment and knowledge about business conditions that might affect the value of the variable to be forecast also have to be considered carefully when selecting a method. Historical forecast accuracy is not the sole consideration, especially if the pattern exhibited by the time series is likely to change in the future.

In the next section we will introduce more sophisticated methods for developing forecasts for a time series that exhibits a horizontal pattern. Using the measures of forecast accuracy developed here, we will be able to assess whether such methods provide more accurate forecasts than we obtained using the simple approaches illustrated in this section. The methods that we will introduce also have the advantage that they adapt well to situations in which the time series changes to a new level. The ability of a forecasting method to adapt quickly to changes in level is an important consideration, especially in short-term forecasting situations.

15.3 MOVING AVERAGES AND EXPONENTIAL SMOOTHING

In this section we discuss three forecasting methods that are appropriate for a time series with a horizontal pattern: moving averages, weighted moving averages, and exponential smoothing. These methods are also capable of adapting well to changes in the level of a horizontal pattern such as what we saw with the extended gasoline sales time series (Table 15.2 and Figure 15.2). However, without modification they are not appropriate when considerable trend, cyclical, or seasonal effects are present. Because the objective of each of these methods is to "smooth out" random fluctuations in the time series, they are referred to as smoothing methods. These methods are easy to use and generally provide a high level of accuracy for short-range forecasts, such as a forecast for the next time period.

Moving Averages

The moving averages method uses the average of the most recent k data values in the time series as the forecast for the next period. Mathematically, a **moving average** forecast of order k is as follows:

$$\hat{Y}_{t+1} = \frac{\sum(\text{most recent } k \text{ data values})}{k} = \frac{\sum_{i=t-k+1}^{t} Y_i}{k}$$
$$= \frac{Y_{t-k+1} + \cdots + Y_{t-1} + Y_t}{k} \qquad (15.6)$$

where

$$\hat{Y}_{t+1} = \text{forecast of the time series for period } t + 1$$
$$Y_i = \text{actual value of the time series in period } i$$
$$k = \text{number of periods of time series data used to generate the forecast}$$

The term *moving* is used because every time a new observation becomes available for the time series, it replaces the oldest observation in the equation and a new average is computed. Thus, the periods over which the average is calculated change, or move, with each ensuing period.

To illustrate the moving averages method, let us return to the original 12 weeks of gasoline sales data in Table 15.1 and Figure 15.1. The time series plot in Figure 15.1 indicates that the gasoline sales time series has a horizontal pattern. Thus, the smoothing methods of this section are applicable.

To use moving averages to forecast a time series, we must first select the order k, or number of time series values to be included in the moving average. If only the most recent values of the time series are considered relevant, a small value of k is preferred. If a greater number of past values are considered relevant, then we generally opt for a larger value of k. As mentioned earlier, a time series with a horizontal pattern can shift to a new level over time. A moving average will adapt to the new level of the series and resume providing good forecasts in k periods. Thus a smaller value of k will track shifts in a time series more quickly (the naïve approach discussed earlier is actually a moving average for $k = 1$). On the other hand, larger values of k will be more effective in smoothing out random fluctuations. Thus, managerial judgment based on an understanding of the behavior of a time series is helpful in choosing an appropriate value of k.

To illustrate how moving averages can be used to forecast gasoline sales, we will use a three-week moving average ($k = 3$). We begin by computing the forecast of sales in Week 4 using the average of the time series values in Weeks 1 to 3.

$$\hat{Y}_4 = \text{average of Weeks 1 to 3} = \frac{17 + 21 + 19}{3} = 19$$

Thus, the moving average forecast of sales in Week 4 is 19 or 19,000 gallons of gasoline. Because the actual value observed in Week 4 is 23, the forecast error in Week 4 is $e_4 = 23 - 19 = 4$.

We next compute the forecast of sales in Week 5 by averaging the time series values in Weeks 2–4.

$$\hat{Y}_5 = \text{average of Weeks 2 to 4} = \frac{21 + 19 + 23}{3} = 21$$

Hence, the forecast of sales in Week 5 is 21 and the error associated with this forecast is $e_5 = 18 - 21 = -3$. A complete summary of the three-week moving average forecasts for the gasoline sales time series is provided in Table 15.9. Figure 15.7 shows the original time series plot and the three-week moving average forecasts. Note how the graph of the moving average forecasts has tended to smooth out the random fluctuations in the time series.

Can you now use moving averages to develop forecasts? Try Problem 7.

To forecast sales in Week 13, the next time period in the future, we simply compute the average of the time series values in Weeks 10, 11, and 12.

$$\hat{Y}_{13} = \text{average of Weeks 10 to 12} = \frac{20 + 15 + 22}{3} = 19$$

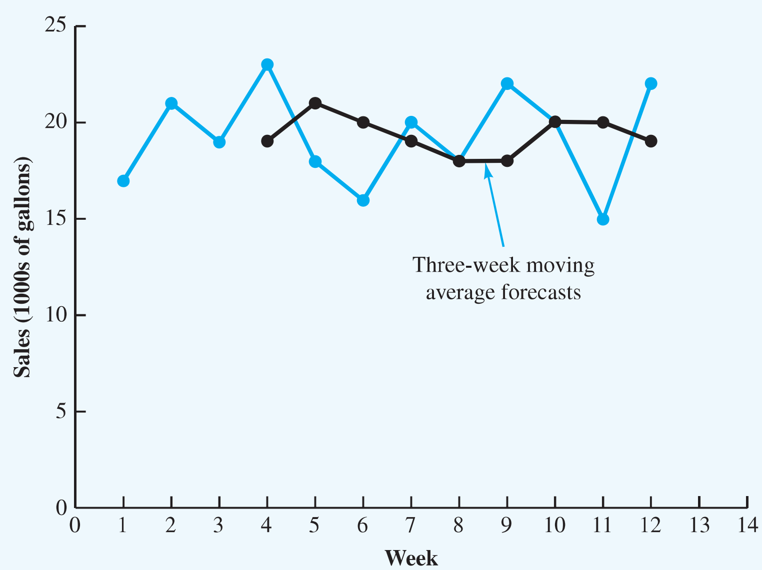
Thus, the forecast for Week 13 is 19 or 19,000 gallons of gasoline.

Forecast Accuracy In Section 15.2 we discussed three measures of forecast accuracy: mean absolute error (MAE); mean squared error (MSE); and mean absolute percentage error

TABLE 15.9 SUMMARY OF THREE-WEEK MOVING AVERAGE CALCULATIONS

Week	Time Series Value	Forecast	Forecast Error	Absolute Value of Forecast Error	Squared Forecast Error	Percentage Error	Absolute Value of Percentage Error
1	17						
2	21						
3	19						
4	23	19	4	4	16	17.39	17.39
5	18	21	−3	3	9	−16.67	16.67
6	16	20	−4	4	16	−25.00	25.00
7	20	19	1	1	1	5.00	5.00
8	18	18	0	0	0	0.00	0.00
9	22	18	4	4	16	18.18	18.18
10	20	20	0	0	0	0.00	0.00
11	15	20	−5	5	25	−33.33	33.33
12	22	19	3	3	9	13.64	13.64
		Total	0	24	92	−20.79	129.21

FIGURE 15.7 GASOLINE SALES TIME SERIES PLOT AND THREE-WEEK MOVING AERAGE FORECASTS



(MAPE). Using the three-week moving average calculations in Table 15.9, the values for these three measures of forecast accuracy are

$$\text{MAE} = \frac{\sum_{t=4}^{12} |e_t|}{12 - 3} = \frac{24}{9} = 2.67$$

$$\text{MSE} = \frac{\sum_{t=4}^{12} e_t^2}{12 - 3} = \frac{92}{9} = 10.22$$

$$\text{MAPE} = \frac{\sum_{t=4}^{12} \left| \left(\frac{e_t}{Y_t} \right) 100 \right|}{12 - 3} = \frac{129.21}{9} = 14.36\%$$

In situations where you need to compare forecasting methods for different time periods, such as comparing a forecast of weekly sales to a forecast of monthly sales, relative measures such as MAPE are preferred.

In Section 15.2 we showed that using the most recent observation as the forecast for the next week (a moving average of order $k = 1$) resulted in values of MAE = 3.73, MSE = 16.27, and MAPE = 19.24%. Thus, in each case the three-week moving average approach has provided more accurate forecasts than simply using the most recent observation as the forecast. Also note how the formulas for the MAE, MSE, and MAPE reflect that our use of a three-week moving average leaves us with insufficient data to generate forecasts for the first three weeks of our time series.

To determine if a moving average with a different order k can provide more accurate forecasts, we recommend using trial and error to determine the value of k that minimizes the MSE. For the gasoline sales time series, it can be shown that the minimum value of MSE corresponds to a moving average of order $k = 6$ with MSE = 6.79. If we are willing to assume that the order of the moving average that is best for the historical data will also be best for future values of the time series, the most accurate moving average forecasts of gasoline sales can be obtained using a moving average of order $k = 6$.

Weighted Moving Averages

A moving average forecast of order k is just a special case of the weighted moving averages method in which each weight is equal to 1/k; for example, a moving average forecast of order k = 3 is just a special case of the weighted moving averages method in which each weight is equal to $\frac{1}{3}$.

In the moving averages method, each observation in the moving average calculation receives equal weight. One variation, known as **weighted moving averages**, involves selecting a different weight for each data value in the moving average and then computing a weighted average of the most recent k values as the forecast.

$$\hat{Y}_{t+1} = w_t Y_t + w_{t-1} Y_{t-1} + \cdots + w_{t-k+1} Y_{t-k+1} \qquad (15.7)$$

where

$\hat{Y}_{t+1}$ = forecast of the time series for period $t + 1$

Y_t = actual value of the time series in period t

w_t = weight applied to the actual time series value for period t

k = number of periods of time series data used to generate the forecast

Generally the most recent observation receives the largest weight, and the weight decreases with the relative age of the data values. Let us use the gasoline sales time series in Table 15.1 to illustrate the computation of a weighted three-week moving average. We will assign a weight of $w_t = \frac{3}{6}$ to the most recent observation, a weight of $w_{t-1} = \frac{2}{6}$ to the second most recent observation, and a weight of $w_{t-2} = \frac{1}{6}$ to the third most recent observation. Using this weighted average, our forecast for Week 4 is computed as follows:

$$\text{Forecast for Week 4} = \frac{1}{6}(17) + \frac{2}{6}(21) = \frac{3}{6}(19) = 19.33$$

Use Problem 8 to practice using weighted moving averages to produce forecasts.

Note that the sum of the weights is equal to 1 for the weighted moving average method.

Forecast Accuracy To use the weighted moving averages method, we must first select the number of data values to be included in the weighted moving average and then choose weights for each of these data values. In general, if we believe that the recent past is a better predictor of the future than the distant past, larger weights should be given to the more recent observations. However, when the time series is highly variable, selecting approximately equal weights for the data values may be preferable. The only requirements in selecting the weights are that they be nonnegative and that their sum must equal 1. To determine whether one particular combination of number of data values and weights provides a more accurate forecast than another combination, we recommend using MSE as the measure of forecast accuracy. That is, if we assume that the combination that is best for the past will also be best for the future, we would use the combination of number of data values and weights that minimized MSE for the historical time series to forecast the next value in the time series.

Exponential Smoothing

Exponential smoothing also uses a weighted average of past time series values as a forecast; it is a special case of the weighted moving averages method in which we select only one weight—the weight for the most recent observation. The weights for the other data values are computed automatically and become smaller as the observations move farther into the past. The exponential smoothing model follows.

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha)\hat{Y}_t \tag{15.8}$$

where

$$\begin{aligned} \hat{Y}_{t+1} &= \text{forecast of the time series for period } t + 1 \\ Y_t &= \text{actual value of the time series in period } t \\ \hat{Y}_t &= \text{forecast of the time series for period } t \\ \alpha &= \text{smoothing constant } (0 \leq \alpha \leq 1) \end{aligned}$$

There are several exponential smoothing procedures. Because it has a single smoothing constant α, the method presented here is often referred to as single exponential smoothing.

Equation (15.8) shows that the forecast for period $t + 1$ is a weighted average of the actual value in period t and the forecast for period t. The weight given to the actual value in period t is the **smoothing constant** α and the weight given to the forecast in period t is $1 - \alpha$. It turns out that the exponential smoothing forecast for any period is actually a weighted average of *all the previous actual values* of the time series. Let us illustrate by working with a time series involving only three periods of data: Y_1, Y_2, and Y_3.

To initiate the calculations, we let $\hat{Y}_1$ equal the actual value of the time series in period 1; that is, $\hat{Y}_1 = Y_1$. Hence, the forecast for period 2 is

$$\begin{aligned} \hat{Y}_2 &= \alpha Y_1 + (1 - \alpha)\hat{Y}_1 \\ &= \alpha Y_1 + (1 - \alpha) Y_1 \\ &= Y_1 \end{aligned}$$

We see that the exponential smoothing forecast for period 2 is equal to the actual value of the time series in period 1.

The forecast for period 3 is

$$\hat{Y}_3 = \alpha Y_2 + (1 - \alpha)\hat{Y}_2 = \alpha Y_2 + (1 - \alpha) Y_1$$

Finally, substituting this expression for $\hat{Y}_3$ into the expression for $\hat{Y}_4$, we obtain

$$\begin{aligned} \hat{Y}_4 &= \alpha Y_3 + (1 - \alpha)\hat{Y}_3 \\ &= \alpha Y_3 + (1 - \alpha)[\alpha Y_2 + (1 - \alpha) Y_1] \\ &= \alpha Y_3 + \alpha(1 - \alpha) Y_2 + (1 - \alpha)^2 Y_1 \end{aligned}$$

The term exponential smoothing comes from the exponential nature of the weighting scheme for the historical values.

We now see that $\hat{Y}_4$ is a weighted average of the first three time series values. The sum of the coefficients, or weights, for Y_1, Y_2, and Y_3 equals 1. A similar argument can be made to show that, in general, any forecast $\hat{Y}_{t+1}$ is a weighted average of all the t previous time series values.

Despite the fact that exponential smoothing provides a forecast that is a weighted average of all past observations, all past data do not need to be retained to compute the forecast for the next period. In fact, equation (15.8) shows that once the value for the smoothing constant α is selected, only two pieces of information are needed to compute the forecast for period $t + 1$: Y_t, the actual value of the time series in period t; and $\hat{Y}_t$, the forecast for period t.

To illustrate the exponential smoothing approach to forecasting, let us again consider the gasoline sales time series in Table 15.1 and Figure 15.1. As indicated previously, to initialize the calculations we set the exponential smoothing forecast for period 2 equal to the actual value of the time series in period 1. Thus, with $Y_1 = 17$, we set $\hat{Y}_2 = 17$ to initiate the computations. Referring to the time series data in Table 15.1, we find an actual time series value in period 2 of $Y_2 = 21$. Thus, in period 2 we have a forecast error of $e_2 = 21 - 17 = 4$.

Continuing with the exponential smoothing computations using a smoothing constant of $\alpha = 0.2$, we obtain the following forecast for period 3:

$$\hat{Y}_3 = 0.2Y_2 + 0.8\hat{Y}_2 = 0.2(21) + 0.8(17) = 17.8$$

Once the actual time series value in period 3, $Y_3 = 19$, is known, we can generate a forecast for period 4 as follows:

$$\hat{Y}_4 = 0.2Y_3 + 0.8\hat{Y}_3 = 0.2(19) + 0.8(17.8) = 18.04$$

Continuing the exponential smoothing calculations, we obtain the weekly forecast values shown in Table 15.10. Note that we have not shown an exponential smoothing forecast or a forecast error for Week 1 because no forecast was made (we used actual sales for Week 1 as the forecasted sales for Week 2 to initialize the exponential smoothing process). For Week 12, we have $Y_{12} = 22$ and $\hat{Y}_{12} = 18.48$. We can we use this information to generate a forecast for Week 13.

$$\hat{Y}_{13} = 0.2Y_{12} + 0.8\hat{Y}_{12} = 0.2(22) + 0.8(18.48) = 19.18$$

Try Problem 9 for practice using exponential smoothing to produce forecasts.

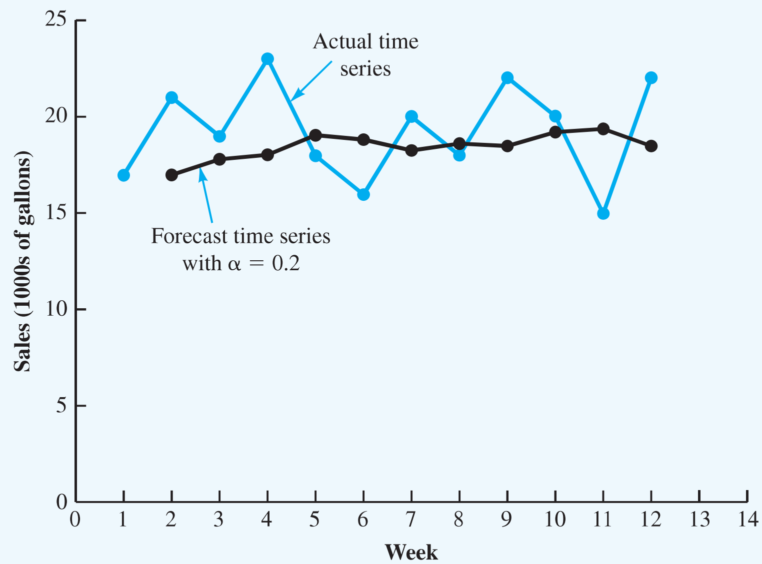
Thus, the exponential smoothing forecast of the amount sold in Week 13 is 19.18, or 19,180 gallons of gasoline. With this forecast, the firm can make plans and decisions accordingly.

Figure 15.8 shows the time series plot of the actual and forecast time series values. Note in particular how the forecasts "smooth out" the irregular or random fluctuations in the time series.

TABLE 15.10 SUMMARY OF THE EXPONENTIAL SMOOTHING FORECASTS AND FORECAST ERRORS FOR THE GASOLINE SALES TIME SERIES WITH SMOOTHING CONSTANT $\alpha = 0.2$

Week	Time Series Value	Forecast	Forecast Error	Squared Forecast Error
1	17			
2	21	17.00	4.00	16.00
3	19	17.80	1.20	1.44
4	23	18.04	4.96	24.60
5	18	19.03	−1.03	1.06
6	16	18.83	−2.83	8.01
7	20	18.26	1.74	3.03
8	18	18.61	−0.61	0.37
9	22	18.49	3.51	12.32
10	20	19.19	0.81	0.66
11	15	19.35	−4.35	18.92
12	22	18.48	3.52	12.39
		Total	10.92	98.80

FIGURE 15.8 ACTUAL AND FORECAST GASOLINE TIME SERIES WITH SMOOTHING CONSTANT $\alpha = 0.2$



Forecast Accuracy In the preceding exponential smoothing calculations, we used a smoothing constant of $\alpha = 0.2$. Although any value of α between 0 and 1 is acceptable, some values will yield more accurate forecasts than others. Insight into choosing a good value for α can be obtained by rewriting the basic exponential smoothing model as follows:

$$
\begin{aligned}
\hat{Y}_{t+1} &= \alpha Y_t + (1 - \alpha)\hat{Y}_t \\
\hat{Y}_{t+1} &= \alpha Y_t + \hat{Y}_t - \alpha\hat{Y}_t \\
\hat{Y}_{t+1} &= \hat{Y}_t + \alpha(Y_t - \hat{Y}_t) = \hat{Y}_t + \alpha e_t
\end{aligned}
\tag{15.9}
$$

Thus, the new forecast $\hat{Y}_{t+1}$ is equal to the previous forecast $\hat{Y}_t$ plus an adjustment, which is the smoothing constant α times the most recent forecast error, $e_t = Y_t - \hat{Y}_t$. That is, the forecast in period $t + 1$ is obtained by adjusting the forecast in period t by a fraction of the forecast error from period t. If the time series contains substantial random variability, a small value of the smoothing constant is preferred. The reason for this choice is that if much of the forecast error is due to random variability, we do not want to overreact and adjust the forecasts too quickly. For a time series with relatively little random variability, a forecast error is more likely to represent a real change in the level of the series. Thus, larger values of the smoothing constant provide the advantage of quickly adjusting the forecasts to changes in the time series; this allows the forecasts to react more quickly to changing conditions.

The criterion we will use to determine a desirable value for the smoothing constant α is the same as the criterion we proposed for determining the order or number of periods of data to include in the moving averages calculation. That is, we choose the value of α that minimizes the MSE. A summary of the MSE calculations for the exponential smoothing forecast of gasoline sales with $\alpha = 0.2$ is shown in Table 15.10. Note that there is one less squared error term than the number of time periods; this is because we had no past values with which to make a forecast for period 1. The value of the sum of squared forecast errors is 98.80; hence MSE $= 98.80/11 = 8.98$. Would a different value of α provide better results in terms of a lower MSE value? Trial and error is often used to determine if a different smoothing constant α can provide more accurate forecasts, but we can avoid trial and error and determine the value of α that minimizes MSE through the use of nonlinear optimization as discussed in Chapter 8 (see Problem 8.12).

NOTES AND COMMENTS

1. Spreadsheet packages are effective tools for implementing exponential smoothing. With the time series data and the forecasting formulas in a spreadsheet as shown in Table 15.10, you can use the MAE, MSE, and MAPE to evaluate different values of the smoothing constant α.
2. We presented the moving average, weighted moving average, and exponential smoothing methods in the context of a stationary time series. These methods can also be used to forecast a nonstationary time series that shifts in level but exhibits no trend or seasonality. Moving averages with small values of k adapt more quickly than moving averages with larger values of k. Weighted moving averages that place relatively large weights on the most recent values adapt more quickly than weighted moving averages that place relatively equal weights on the k time series values used in calculating the forecast. Exponential smoothing models with smoothing constants closer to 1 adapt more quickly than models with smaller values of the smoothing constant.

15.4 LINEAR TREND PROJECTION

In this section we present forecasting methods that are appropriate for time series exhibiting trend patterns. Here we show how **regression analysis** may be used to forecast a time series with a linear trend. In Section 15.1 we used the bicycle sales time series in Table 15.3 and Figure 15.3 to illustrate a time series with a trend pattern. Let us now use this time series to illustrate how regression analysis can be used to forecast a time series with a linear trend. The data for the bicycle time series are repeated in Table 15.11 and Figure 15.9.

Although the time series plot in Figure 15.9 shows some up and down movement over the past 10 years, we might agree that the linear trend line shown in Figure 15.10 provides a reasonable approximation of the long-run movement in the series. We can use regression analysis to develop such a linear trend line for the bicycle sales time series.

In regression analysis we use known values of variables to estimate the relationship between one variable (called the **dependent variable**) and one or more other related variables (called **independent variables**). This relationship is usually found in a manner that minimizes the sum of squared errors (and so also minimizes the MSE). With this relationship we can then use values of the independent variables to estimate the associated value of the dependent variable. When we estimate a linear relationship between the dependent variable (which is usually denoted as y) and a single independent variable (which is usually denoted as x), this is referred to as **simple linear regression**. Estimating the relationship between the dependent variable and a single independent

DATA *file*
Bicycle

TABLE 15.11 BICYCLE SALES TIME SERIES

Year	Sales (1000s)
1	21.6
2	22.9
3	25.5
4	21.9
5	23.9
6	27.5
7	31.5
8	29.7
9	28.6
10	31.4

FIGURE 15.9 BICYCLE SALES TIME SERIES PLOT

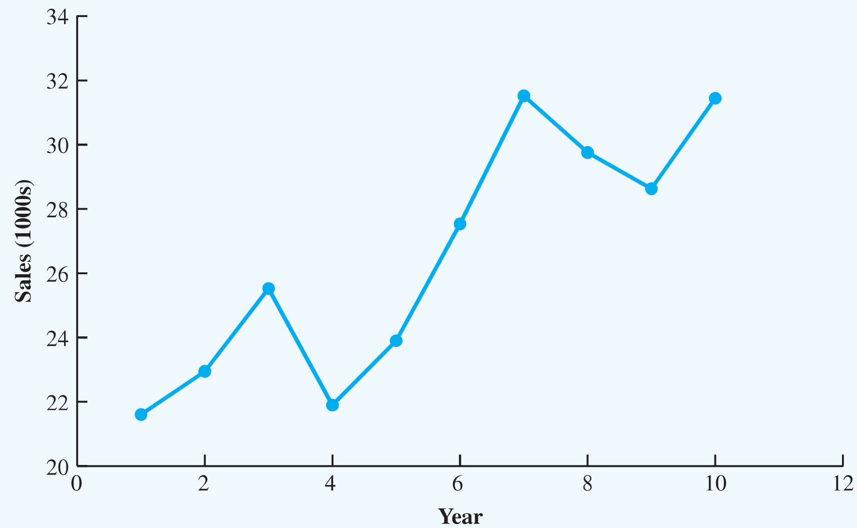
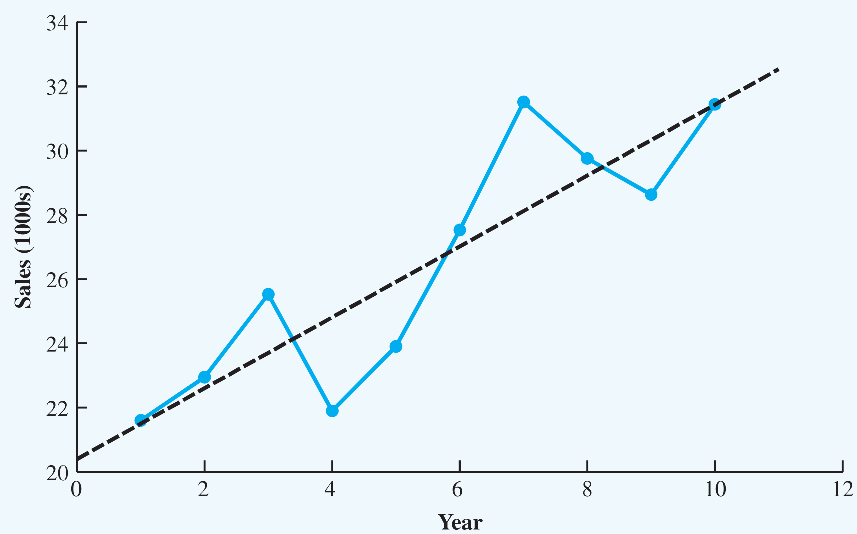


FIGURE 15.10 TREND REPRESENTED BY A LINEAR FUNCTION FOR THE BICYCLE SALES TIME SERIES



variable requires that we find the values of parameters b_0 and b_1 for the straight line $y = b_0 + b_1 x$.

Because our use of simple linear regression analysis yields the linear relationship between the independent variable and the dependent variable that minimizes the MSE, we can use this approach to find a best-fitting line to a set of data that exhibits a linear trend. In finding a linear trend, the variable to be forecasted (Y_t, the actual value of the time series in period t) is the dependent variable and the trend variable (time period t) is the independent variable. We will use the following notation for our linear trendline:

$$\hat{Y}_t = b_0 + b_1 t \tag{15.10}$$

where

t = the time period
$\hat{Y}_t$ = linear trend forecast in period t (i.e., the estimated value of Y_t in period t)
b_0 = the Y-intercept of the linear trendline
b_1 = the slope of the linear trendline

In equation (15.10) the time variable begins at $t = 1$ corresponding to the first time series observation (Year 1 for the bicycle sales time series) and continues until $t = n$ corresponding to the most recent time series observation (Year 10 for the bicycle sales time series). Thus, for the bicycle sales time series $t = 1$ corresponds to the oldest time series value and $t = 10$ corresponds to the most recent year. Calculus may be used to show that the equations given below for b_0 and b_1 yield the line that minimizes the MSE. The equations for computing the values of b_0 and b_1 are

$$b_1 = \frac{\sum_{t=1}^{n} tY_t - \sum_{t=1}^{n} t \sum_{t=1}^{n} Y_t \Big/ n}{\sum_{t=1}^{n} t^2 - \left(\sum_{t=1}^{n} t\right)^2 \Big/ n} \tag{15.11}$$

$$b_0 = \overline{Y} - b_1 \bar{t} \tag{15.12}$$

where

t = the time period
Y_t = actual value of the time series in period t
n = number of periods in the time series
$\overline{Y}$ = average value of the time series; that is, $\overline{Y} = \sum_{t=1}^{n} Y_t \Big/ n$
$\bar{t}$ = mean value of t; that is, $\bar{t} = \sum_{t=1}^{n} t \Big/ n$

Let us calculate b_0 and b_1 for the bicycle data in Table 15.11; the intermediate summary calculations necessary for computing the values of b_0 and b_1 are

	t	Y_t	tY_t	t^2
	1	21.6	21.6	1
	2	22.9	45.8	4
	3	25.5	76.5	9
	4	21.9	87.6	16
	5	23.9	119.5	25
	6	27.5	165.0	36
	7	31.5	220.5	49
	8	29.7	237.6	64
	9	28.6	257.4	81
	10	31.4	314.0	100
Total	55	264.5	1545.5	385

And the final calculations of the values of b_0 and b_1 are

$$\bar{t} = \frac{55}{10} = 5.5$$
$$\overline{Y} = \frac{264.5}{10} = 26.45$$
$$b_1 = \frac{1545.5 - (55)(264.5)/10}{385 - 55^2/10} = 1.10$$
$$b_0 = 26.45 - 1.10(5.5) = 20.40$$

Problem 20 provides additional practice in using regression analysis to estimate the linear trend in a time series data set.

Therefore,

$$\hat{Y}_t = 20.4 + 1.1t \qquad (15.13)$$

is the regression equation for the linear trend component for the bicycle sales time series.

The slope of 1.1 in this trend equation indicates that over the past 10 years, the firm has experienced an average growth in sales of about 1100 units per year. If we assume that the past 10-year trend in sales is a good indicator for the future, we can use equation (15.13) to project the trend component of the time series. For example, substituting $t = 11$ into equation (15.13) yields next year's trend projection, $\hat{Y}_{11}$:

$$\hat{Y}_{11} = 20.4 + 1.1(11) = 32.5$$

Thus, the linear trend model yields a sales forecast of 32,500 bicycles for the next year.

Table 15.12 shows the computation of the minimized sum of squared errors for the bicycle sales time series. As previously noted, minimizing sum of squared errors also minimizes the commonly used measure of accuracy, MSE. For the bicycle sales time series,

$$\text{MSE} = \frac{\sum_{t=1}^{n} e_t^2}{n} = \frac{30.7}{10} = 3.07$$

Note that in this example we are not using past values of the time series to produce forecasts, and so $k = 0$; that is, we can produce a forecast for each period of the time series and so do not have to adjust our calculations of the MAE, MSE, or MAPE for k.

TABLE 15.12 SUMMARY OF THE LINEAR TREND FORECASTS AND FORECAST ERRORS FOR THE BICYCLE SALES TIME SERIES

Week	Sales (1000s) Y_t	Forecast $\hat{Y}_t$	Forecast Error	Squared Forecast Error
1	21.6	21.5	0.1	0.01
2	22.9	22.6	0.3	0.09
3	25.5	23.7	1.8	3.24
4	21.9	24.8	−2.9	8.41
5	23.9	25.9	−2.0	4.00
6	27.5	27.0	0.5	0.25
7	31.5	28.1	3.4	11.56
8	29.7	29.2	0.5	0.25
9	28.6	30.3	−1.7	2.89
10	31.4	31.4	0.0	0.00
			Total	30.70

We can also use the trendline to forecast sales farther into the future. For instance, using equation (15.13), we develop annual forecasts for two and three years into the future as follows:

$$\hat{Y}_{12} = 20.4 + 1.1(12) = 33.6$$
$$\hat{Y}_{13} = 20.4 + 1.1(13) = 34.7$$

Note that the forecasted value increases by 1100 bicycles in each year.

NOTES AND COMMENTS

1. Statistical packages such as Minitab and SAS, as well as Excel, have routines to perform regression analysis. Regression analysis minimizes the sum of squared error and under certain assumptions it also allows the analyst to make statistical statements about the parameters and the forecasts.
2. While the use of a linear function to model the trend is common, some time series exhibit a curvilinear (nonlinear) trend. More advanced texts discuss how to develop nonlinear models such as quadratic models and exponential models for these more complex relationships.

In this section we used simple linear regression to estimate the relationship between the dependent variable (Y_t, the actual value of the time series in period t) and a single independent variable (the trend variable t). However, some regression models include several independent variables. When we estimate a linear relationship between the dependent variable and more than one independent variable, this is referred to as multiple linear regression. In the next section we will apply multiple linear regression to time series that include seasonal effects and to time series that include both seasonal effects and a linear trend.

15.5 SEASONALITY

In this section we show how to develop forecasts for a time series that has a seasonal pattern. To the extent that seasonality exists, we need to incorporate it into our forecasting models to ensure accurate forecasts. We begin the section by considering a seasonal time series with no trend and then discuss how to model seasonality with a linear trend.

Seasonality Without Trend

Let us consider again the data from Table 15.5, the number of umbrellas sold at a clothing store over the past five years. We repeat the data here in Table 15.13, and Figure 15.11 again shows the corresponding time series plot. The time series plot does not indicate any long-term trend in sales. In fact, unless you look carefully at the data, you might conclude that the data follow a horizontal pattern with random fluctuation and that single exponential smoothing could be used to forecast sales. However, closer inspection of the time series plot reveals a pattern in the fluctuations. That is, the first and third quarters have moderate sales, the second quarter the highest sales, and the fourth quarter tends to be the lowest quarter in terms of sales volume. Thus, we conclude that a quarterly seasonal pattern is present.

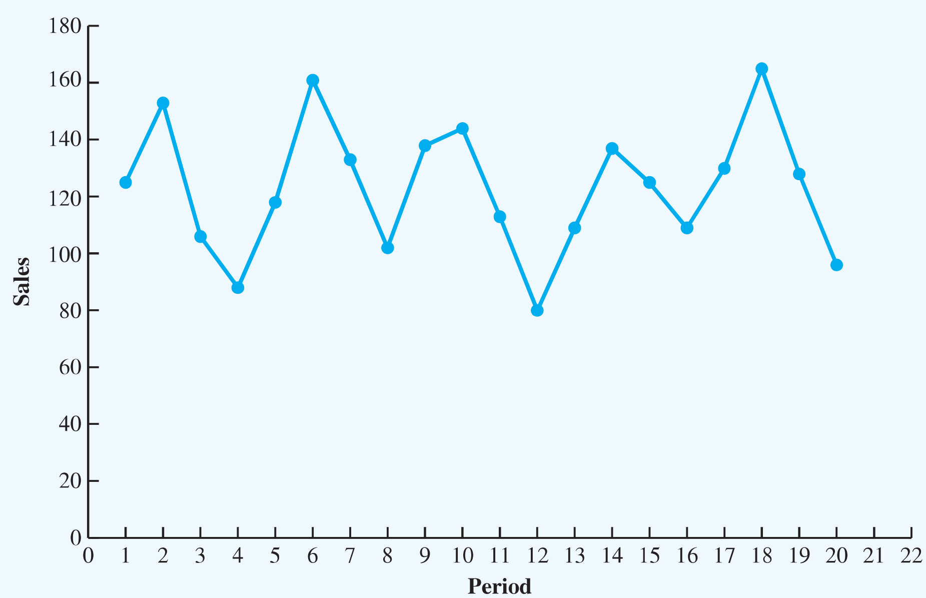
We can model a time series with a seasonal pattern by treating the season as a categorical variable. **Categorical variables** are data used to categorize observations of data. When a categorical variable has k levels, $k - 1$ dummy variables (sometimes called 0-1 variables) are required. So if there are four seasons, we need three dummy variables. For instance, in the umbrella sales time series, the quarter to which each observation corresponds is treated as a season; it is a categorical variable with four levels: Quarter 1, Quarter 2, Quarter 3, and

TABLE 15.13 UMBRELLA SALES TIME SERIES

Year	Quarter	Sales
1	1	125
	2	153
	3	106
	4	88
2	1	118
	2	161
	3	133
	4	102
3	1	138
	2	144
	3	113
	4	80
4	1	109
	2	137
	3	125
	4	109
5	1	130
	2	165
	3	128
	4	96

DATA file
Umbrella

FIGURE 15.11 UMBRELLA SALES TIME SERIES PLOT



Quarter 4. Thus, to model the seasonal effects in the umbrella time series we need $4 - 1 = 3$ dummy variables. The three dummy variables can be coded as follows:

$$\text{Qtr1}_t = \begin{cases} 1 & \text{if period } t \text{ is a Quarter 1} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Qtr2}_t = \begin{cases} 1 & \text{if period } t \text{ is a Quarter 2} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Qtr3}_t = \begin{cases} 1 & \text{if period } t \text{ is a Quarter 3} \\ 0 & \text{otherwise} \end{cases}$$

Using $\hat{Y}_t$ to denote the forecasted value of sales for period t, the general form of the equation relating the number of umbrellas sold to the quarter the sales take place is as follows.

$$\hat{Y}_t = b_0 + b_1 \text{ Qtr1}_t + b_2 \text{ Qtr2}_t + b_3 \text{ Qtr3}_t \tag{15.14}$$

Note that we have numbered the observations in Table 15.14 as periods 1 to 20. For example, Year 3, quarter 3 is observation 11.

Note that the fourth quarter will be denoted by a setting of all three dummy variables to 0. Table 15.14 shows the umbrella sales time series with the coded values of the dummy variables shown. We can use a multiple linear regression model to find the values of b_0, b_1, b_2, and b_3 that minimize the sum of squared errors. For this regression model Y_t is the dependent variable and the quarterly dummy variables Qtr1_t, Qtr2_t, and Qtr3_t are the independent variables.

Using the data in Table 15.14 and regression analysis, we obtain the following equation:

$$\hat{Y}_t = 95.0 + 29.0 \text{ Qtr1}_t + 57.0 \text{ Qtr2}_t + 26.0 \text{ Qtr3}_t \tag{15.15}$$

we can use Equation (15.15) to forecast quarterly sales for next year.

For practice using categorical variables to estimate seasonal effects, try Problem 24.

Quarter 1: Sales = 95.0 + 29.0(1) + 57.0(0) + 26.0(0) = 124
Quarter 2: Sales = 95.0 + 29.0(0) + 57.0(1) + 26.0(0) = 152
Quarter 3: Sales = 95.0 + 29.0(0) + 57.0(0) + 26.0(1) = 121
Quarter 4: Sales = 95.0 + 29.0(0) + 57.0(0) + 26.0(0) = 95

TABLE 15.14 UMBRELLA SALES TIME SERIES WITH DUMMY VARIABLES

Period	Year	Quarter	Qtr1	Qtr2	Qtr3	Sales
1	1	1	1	0	0	125
2		2	0	1	0	153
3		3	0	0	1	106
4		4	0	0	0	88
5	2	1	1	0	0	118
6		2	0	1	0	161
7		3	0	0	1	133
8		4	0	0	0	102
9	3	1	1	0	0	138
10		2	0	1	0	144
11		3	0	0	1	113
12		4	0	0	0	80
13	4	1	1	0	0	109
14		2	0	1	0	137
15		3	0	0	1	125
16		4	0	0	0	109
17	5	1	1	0	0	130
18		2	0	1	0	165
19		3	0	0	1	128
20		4	0	0	0	96

It is interesting to note that we could have obtained the quarterly forecasts for next year by simply computing the average number of umbrellas sold in each quarter, as shown in the following table:

Year	Quarter 1	Quarter 2	Quarter 3	Quarter 4
1	125	153	106	88
2	118	161	133	102
3	138	144	113	80
4	109	137	125	109
5	130	165	128	96
Average	124	152	121	95

Nonetheless, for more complex problem situations, such as dealing with a time series that has both trend and seasonal effects, this simple averaging approach will not work.

Seasonality with Trend

We now consider situations for which the time series contains both a seasonal effect and a linear trend by showing how to forecast the quarterly smartphone sales time series introduced in Section 15.1. The data for the smartphone time series are shown in Table 15.15. The time series plot in Figure 15.12 indicates that sales are lowest in the second quarter of each year and increase in quarters 3 and 4. Thus, we conclude that a seasonal pattern exists for smartphone sales. However, the time series also has an upward linear trend that will need to be accounted for in order to develop accurate forecasts of quarterly sales. This is easily done by combining the dummy variable approach for handling seasonality with the approach we discussed in Section 15.4 for handling a linear trend.

The general form of the regression equation for modeling both the quarterly seasonal effects and the linear trend in the smartphone time series is:

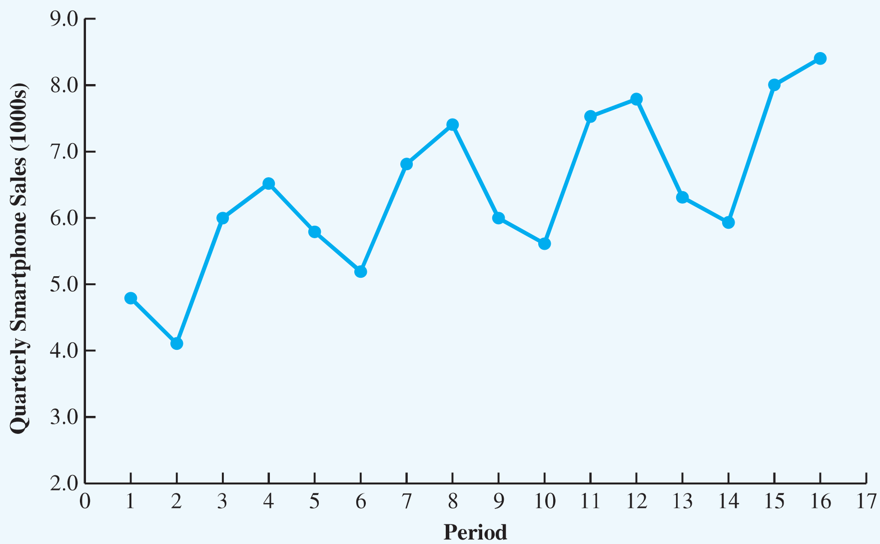
$$\hat{Y}_t = b_0 + b_1\text{Qtr1}_t + b_2\text{Qtr2}_t + b_3\text{Qtr3}_t + b_4 t \tag{15.16}$$

TABLE 15.15 SMARTPHONE SALES TIME SERIES

DATA *file*
SmartPhoneSales

Year	Quarter	Sales (1000s)
1	1	4.8
	2	4.1
	3	6.0
	4	6.5
2	1	5.8
	2	5.2
	3	6.8
	4	7.4
3	1	6.0
	2	5.6
	3	7.5
	4	7.8
4	1	6.3
	2	5.9
	3	8.0
	4	8.4

FIGURE 15.12 SMARTPHONE SALES TIME SERIES PLOT



where

$\hat{Y}_t$ = forecast of sales in period t

Qtr1_t = 1 if time period t corresponds to the first quarter of the year; 0, otherwise

Qtr2_t = 1 if time period t corresponds to the second quarter of the year; 0, otherwise

Qtr3_t = 1 if time period t corresponds to the third quarter of the year; 0, otherwise

t = time period

For this regression model Y_t is the dependent variable and the quarterly dummy variables Qtr1_t, Qtr2_t, and Qtr3_t and the time period t are the independent variables.

Table 15.16 shows the revised smartphone sales time series that includes the coded values of the dummy variables and the time period t. Using the data in Table 15.16 with the regression model that includes both the seasonal and trend components, we obtain the following equation that minimizes our sum of squared errors:

$$\hat{Y}_t = 6.07 - 1.36\ \text{Qtr1}_t - 2.03\ \text{Qtr2}_t - 0.304\ \text{Qtr3}_t + 0.146t \qquad (15.17)$$

We can now use equation (15.17) to forecast quarterly sales for next year. Next year is Year 5 for the smartphone sales time series; that is, time periods 17, 18, 19, and 20.

Forecast for Time Period 17 (Quarter 1 in Year 5)

$$\hat{Y}_{17} = 6.07 - 1.36(1) - 2.03(0) - 0.304(0) + 0.146(17) = 7.19$$

Forecast for Time Period 18 (Quarter 2 in Year 5)

$$\hat{Y}_{18} = 6.07 - 1.36(0) - 2.03(1) - 0.304(0) + 0.146(18) = 6.67$$

Forecast for Time Period 19 (Quarter 3 in Year 5)

$$\hat{Y}_{19} = 6.07 - 1.36(0) - 2.03(0) - 0.304(1) + 0.146(19) = 8.54$$

Forecast for Time Period 20 (Quarter 4 in Year 5)

$$\hat{Y}_{20} = 6.07 - 1.36(0) - 2.03(0) - 0.304(0) + 0.146(20) = 8.99$$

Thus, accounting for the seasonal effects and the linear trend in smartphone sales, the estimates of quarterly sales in Year 5 are 7190, 6670, 8540, and 8990.

TABLE 15.16 SMARTPHONE SALES TIME SERIES WITH DUMMY VARIABLES AND TIME PERIOD

Period	Year	Quarter	Qtr1	Qtr2	Qtr3	Sales (1000s)
1	1	1	1	0	0	4.8
2		2	0	1	0	4.1
3		3	0	0	1	6.0
4		4	0	0	0	6.5
5	2	1	1	0	0	5.8
6		2	0	1	0	5.2
7		3	0	0	1	6.8
8		4	0	0	0	7.4
9	3	1	1	0	0	6.0
10		2	0	1	0	5.6
11		3	0	0	1	7.5
12		4	0	0	0	7.8
13	4	1	1	0	0	6.3
14		2	0	1	0	5.9
15		3	0	0	1	8.0
16		4	0	0	0	8.4

The dummy variables in the equation actually provide four equations, one for each quarter. For instance, if time period t corresponds to quarter 1, the estimate of quarterly sales is

$$\text{Quarter 1: Sales} = 6.07 - 1.36(1) - 2.03(0) - 0.304(0) + 0.146t = 4.71 + 0.146t$$

Similarly, if time period t corresponds to quarters 2, 3, and 4, the estimates of quarterly sales are

$$\text{Quarter 2: Sales} = 6.07 - 1.36(0) - 2.03(1) - 0.304(0) + 0.146t = 4.04 + 0.146t$$
$$\text{Quarter 3: Sales} = 6.07 - 1.36(0) - 2.03(0) - 0.304(1) + 0.146t = 5.77 + 0.146t$$
$$\text{Quarter 4: Sales} = 6.07 - 1.36(0) - 2.03(0) - 0.304(0) + 0.146t = 6.07 + 0.146t$$

Problem 28 provides another example of using regression analysis to forecast time series data with both trend and seasonal effects.

The slope of the trend line for each quarterly forecast equation is 0.146, indicating a consistent growth in sales of about 146 sets per quarter. The only difference in the four equations is that they have different intercepts.

Models Based on Monthly Data

Whenever a categorical variable such as season has k levels, k − 1 dummy variables are required.

In the preceding smartphone sales example, we showed how dummy variables can be used to account for the quarterly seasonal effects in the time series. Because there were four levels for the categorical variable season, three dummy variables were required. However, many businesses use monthly rather than quarterly forecasts. For monthly data, season is a categorical variable with 12 levels, and thus $12 - 1 = 11$ dummy variables are required. For example, the 11 dummy variables could be coded as follows:

$$\text{Month1} = \begin{cases} 1 & \text{if January} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Month2} = \begin{cases} 1 & \text{if February} \\ 0 & \text{otherwise} \end{cases}$$

$$\vdots$$

$$\text{Month11} = \begin{cases} 1 & \text{if November} \\ 0 & \text{otherwise} \end{cases}$$

Other than this change, the approach for handling seasonality remains the same.

SUMMARY

This chapter provided an introduction to basic methods of time series analysis and forecasting. We first showed that the underlying pattern in the time series can often be identified by constructing a time series plot. Several types of data patterns can be distinguished, including a horizontal pattern, a trend pattern, and a seasonal pattern. The forecasting methods we have discussed are based on which of these patterns are present in the time series.

We also discussed that the accuracy of the method is an important factor in determining which forecasting method to use. We considered three measures of forecast accuracy: mean absolute error (MAE), mean squared error (MSE), and mean absolute percentage error (MAPE). Each of these measures is designed to determine how well a particular forecasting method is able to reproduce the time series data that are already available. By selecting the method that is most accurate for the data already known, we hope to increase the likelihood that we will obtain more accurate forecasts for future time periods.

For a time series with a horizontal pattern, we showed how moving averages, weighted moving averages, and exponential smoothing can be used to develop a forecast. The moving averages method consists of computing an average of past data values and then using that average as the forecast for the next period. In the weighted moving average and exponential smoothing methods, weighted averages of past time series values are used to compute forecasts. These methods also adapt well to a horizontal pattern that shifts to a different level and then resumes a horizontal pattern.

For time series that have only a long-term linear trend, we showed how regression analysis can be used to make trend projections. For a time series with a seasonal pattern, we showed how dummy variables and regression analysis can be used to develop an equation with seasonal effects. We then extended the approach to include situations where the time series contains both a seasonal and a linear trend effect by showing how to combine the dummy variable approach for handling seasonality with the approach for handling a linear trend.

GLOSSARY

Categorical (dummy) variable A variable used to categorize observations of data. Used when modeling a time series with a seasonal pattern.

Cyclical pattern A cyclical pattern exists if the time series plot shows an alternating sequence of points below and above the trend line lasting more than one year.

Dependent variable The variable that is being predicted or explained in a regression analysis.

Exponential smoothing A forecasting method that uses a weighted average of past time series values as the forecast; it is a special case of the weighted moving averages method in which we select only one weight—the weight for the most recent observation.

Forecast error The difference between the actual time series value and the forecast.

Independent variable A variable used to predict or explain values of the dependent variable in regression analysis.

Mean absolute error (MAE) The average of the absolute values of the forecast errors.

Mean absolute percentage error (MAPE) The average of the absolute values of the percentage forecast errors.

Mean squared error (MSE) The average of the sum of squared forecast errors.

Moving averages A forecasting method that uses the average of the k most recent data values in the time series as the forecast for the next period.

Regression analysis A procedure for estimating values of a dependent variable given the values of one or more independent variables in a manner that minimizes the sum of the squared errors.

Seasonal pattern A seasonal pattern exists if the time series plot exhibits a repeating pattern over successive periods.

Smoothing constant A parameter of the exponential smoothing model that provides the weight given to the most recent time series value in the calculation of the forecast value.

Stationary time series A time series whose statistical properties are indepepndent of time. For a stationary time series, the process generating the data has a constant mean and the variability of the time series is constant over time.

Time series A sequence of observations on a variable measured at successive points in time or over successive periods of time.

Time series plot A graphical presentation of the relationship between time and the time series variable. Time is shown on the horizontal axis and the time series values are shown on the verical axis.

Trend pattern A trend pattern exists if the time series plot shows gradual shifts or movements to relatively higher or lower values over a longer period of time.

Weighted moving averages A forecasting method that involves selecting a different weight for the k most recent data values in the time series and then computing a weighted average of the of the values. The sum of the weights must equal one.

PROBLEMS

SELF*test*

1. Consider the following time series data:

Week	1	2	3	4	5	6
Value	18	13	16	11	17	14

Using the naïve method (most recent value) as the forecast for the next week, compute the following measures of forecast accuracy:

a. Mean absolute error
b. Mean squared error
c. Mean absolute percentage error
d. What is the forecast for Week 7?

SELF*test*

2. Refer to the time series data in Exercise 1. Using the average of all the historical data as a forecast for the next period, compute the following measures of forecast accuracy:

a. Mean absolute error
b. Mean squared error
c. Mean absolute percentage error
d. What is the forecast for Week 7?

SELF*test*

3. Exercises 1 and 2 used different forecasting methods. Which method appears to provide the more accurate forecasts for the historical data? Explain.

4. Consider the following time series data:

Month	1	2	3	4	5	6	7
Value	24	13	20	12	19	23	15

a. Compute MSE using the most recent value as the forecast for the next period. What is the forecast for Month 8?
b. Compute MSE using the average of all the data available as the forecast for the next period. What is the forecast for Month 8?
c. Which method appears to provide the better forecast?

SELF*test*

5. Consider the following time series data:

Week	1	2	3	4	5	6
Value	18	13	16	11	17	14

a. Construct a time series plot. What type of pattern exists in the data?
b. Develop a three-week moving average for this time series. Compute MSE and a forecast for Week 7.
c. Use $\alpha = 0.2$ to compute the exponential smoothing values for the time series. Compute MSE and a forecast for Week 7.
d. Compare the three-week moving average forecast with the exponential smoothing forecast using $\alpha = 0.2$. Which appears to provide the better forecast based on MSE? Explain.
e. Use trial and error to find a value of the exponential smoothing coefficient α that results in a smaller MSE than what you calculated for $\alpha = 0.2$.

6. Consider the following time series data:

Month	1	2	3	4	5	6	7
Value	24	13	20	12	19	23	15

a. Construct a time series plot. What type of pattern exists in the data?
b. Develop a three-week moving average for this time series. Compute MSE and a forecast for Month 8.
c. Use $\alpha = 0.2$ to compute the exponential smoothing values for the time series. Compute MSE and a forecast for Month 8.
d. Compare the three-week moving average forecast with the exponential smoothing forecast using $\alpha = 0.2$. Which appears to provide the better forecast based on MSE?
e. Use trial and error to find a value of the exponential smoothing coefficient α that results in a smaller MSE than what you calculated for $\alpha = 0.2$.

DATA *file*
Gasoline

7. Refer to the gasoline sales time series data in Table 15.1.
a. Compute four-week and five-week moving averages for the time series.
b. Compute the MSE for the four-week and five-week moving average forecasts.
c. What appears to be the best number of weeks of past data (three, four, or five) to use in the moving average computation? Recall that MSE for the three-week moving average is 10.22.

SELF*test*

8. Refer again to the gasoline sales time series data in Table 15.1.
a. Using a weight of 1/2 for the most recent observation, 1/3 for the second most recent, and 1/6 for third most recent, compute a three-week weighted moving average for the time series.
b. Compute the MSE for the weighted moving average in part (a). Do you prefer this weighted moving average to the unweighted moving average? Remember that the MSE for the unweighted moving average is 10.22.
c. Suppose you are allowed to choose any weights as long as they sum to 1. Could you always find a set of weights that would make the MSE smaller for a weighted moving average than for an unweighted moving average? Why or why not?

SELF*test*

9. With the gasoline time series data from Table 15.1, show the exponential smoothing forecasts using $\alpha = 0.1$.
a. Applying the MSE measure of forecast accuracy, would you prefer a smoothing constant of $\alpha = 0.1$ or $\alpha = 0.2$ for the gasoline sales time series?
b. Are the results the same if you apply MAE as the measure of accuracy?
c. What are the results if MAPE is used?

10. With a smoothing constant of $\alpha = 0.2$, equation (15.8) shows that the forecast for Week 13 of the gasoline sales data from Table 15.1 is given by $\hat{Y}_{13} = 0.2Y_{12} + 0.8\,\hat{Y}_{12}$. However, the forecast for Week 12 is given by $\hat{Y}_{12} = 0.2Y_{11} + 0.8\hat{Y}_{11}$. Thus, we could combine these two results to show that the forecast for Week 13 can be written as

$$\hat{Y}_{13} = 0.2Y_{12} + 0.8(0.2Y_{11} + 0.8\hat{Y}_{11}) = 0.2Y_{12} + 0.16Y_{11} + 0.64\hat{Y}_{11}$$

a. Making use of the fact that $\hat{Y}_{11} = 0.2Y_{10} + 0.8\hat{Y}_{10}$ (and similarly for $\hat{Y}_{10}$ and $\hat{Y}_9$), continue to expand the expression for $\hat{Y}_{13}$ until it is written in terms of the past data values Y_{12}, Y_{11}, Y_{10}, Y_9, Y_8, and the forecast for Week 8.

b. Refer to the coefficients or weights for the past values Y_{12}, Y_{11}, Y_{10}, Y_9, and Y_8. What observation can you make about how exponential smoothing weights past data values in arriving at new forecasts? Compare this weighting pattern with the weighting pattern of the moving averages method.

11. For the Hawkins Company, the monthly percentages of all shipments received on time over the past 12 months are 80, 82, 84, 83, 83, 84, 85, 84, 82, 83, 84, and 83.
a. Construct a time series plot. What type of pattern exists in the data?
b. Compare a three-month moving average forecast with an exponential smoothing forecast for $\alpha = 0.2$. Which provides the better forecasts using MSE as the measure of model accuracy?
c. What is the forecast for next month?

12. Corporate triple A bond interest rates for 12 consecutive months follow.

9.5 9.3 9.4 9.6 9.8 9.7 9.8 10.5 9.9 9.7 9.6 9.6

a. Construct a time series plot. What type of pattern exists in the data?
b. Develop three-month and four-month moving averages for this time series. Does the three-month or four-month moving average provide the better forecasts based on MSE? Explain.
c. What is the moving average forecast for the next month?

SELF*test*

13. The values of Alabama building contracts (in millions of dollars) for a 12-month period follow.

240 350 230 260 280 320 220 310 240 310 240 230

a. Construct a time series plot. What type of pattern exists in the data?
b. Compare a three-month moving average forecast with an exponential smoothing forecast. Use $\alpha = 0.2$. Which provides the better forecasts based on MSE?
c. What is the forecast for the next month?

14. The following time series shows the sales of a particular product over the past 12 months:

Month	Sales	Month	Sales
1	105	7	145
2	135	8	140
3	120	9	100
4	105	10	80
5	90	11	100
6	120	12	110

a. Construct a time series plot. What type of pattern exists in the data?
b. Use $\alpha = 0.3$ to compute the exponential smoothing values for the time series.
c. Use trial and error to find a value of the exponential smoothing coefficient α that results in a relatively small MSE.

15. Ten weeks of data on the Commodity Futures Index are 7.35, 7.40, 7.55, 7.56, 7.60, 7.52, 7.52, 7.70, 7.62, and 7.55.
a. Construct a time series plot. What type of pattern exists in the data?
b. Use trial and error to find a value of the exponential smoothing coefficient α that results in a relatively small MSE.

DATA *file*
SuperBowlRatings

16. Since its inception in 1967, the Super Bowl is one of the most watched events on television in the United States every year. The number of U.S. households that tuned in for each Super Bowl, reported by Nielson.com, is provided in the data set SuperBowlRatings.
a. Construct a time series plot for the data. What type of pattern exists in the data? Discuss some of the patterns that may have resulted in the pattern exhibited in the time series plot of the data.
b. Given the pattern of the time series plot developed in part (a), do you think the forecasting methods discussed in this chapter are appropriate to develop forecasts for this time series? Explain.
c. Use simple linear regression analysis to find the parameters for the line that minimizes MSE for this time series.

SELF*test*

17. Consider the following time series:

t	1	2	3	4	5
Y_t	6	11	9	14	15

a. Construct a time series plot. What type of pattern exists in the data?
b. Use simple linear regression analysis to find the parameters for the line that minimizes MSE for this time series.
c. What is the forecast for $t = 6$?

18. The following table reports the percentage of stocks in a portfolio for nine quarters from 2012 to 2014:

Quarter	Stock%
1st—2012	29.8
2nd—2012	31.0
3rd—2012	29.9
4th—2012	30.1
1st—2013	32.2
2nd—2013	31.5
3rd—2013	32.0
4th—2013	31.9
1st—2014	30.0

a. Construct a time series plot. What type of pattern exists in the data?
b. Use trial and error to find a value of the exponential smoothing coefficient α that results in a relatively small MSE.
c. Using the exponential smoothing model you developed in part (b), what is the forecast of the percentage of stocks in a typical portfolio for the second quarter of 2014?

19. Consider the following time series:

t	1	2	3	4	5	6	7
Y_t	120	110	100	96	94	92	88

a. Construct a time series plot. What type of pattern exists in the data?
b. Use simple linear regression analysis to find the parameters for the line that minimizes MSE for this time series.
c. What is the forecast for $t = 8$?

SELF*test*

20. Because of high tuition costs at state and private universities, enrollments at community colleges have increased dramatically in recent years. The following data show the enrollment (in thousands) for Jefferson Community College for the nine most recent years:

Year	Enrollment (1000s)
1	6.5
2	8.1
3	8.4
4	10.2
5	12.5
6	13.3
7	13.7
8	17.2
9	18.1

a. Construct a time series plot. What type of pattern exists in the data?
b. Use simple linear regression analysis to find the parameters for the line that minimizes MSE for this time series.
c. What is the forecast for Year 10?

DATA file
AdultSmokers

21. The Centers for Disease Control and Prevention Office on Smoking and Health (OSH) is the lead federal agency responsible for comprehensive tobacco prevention and control. OSH was established in 1965 to reduce the death and disease caused by tobacco use and exposure to second-hand smoke. One of the many responsibilities of the OSH is to collect data on tobacco use. The following data show the percentage of adults in the United States who were users of tobacco from 2001 through 2011 (http://www.cdc.gov/tobacco/data_statistics/tables/trends/cig_smoking/index.htm):

Year	Percentage of Adults Who Smoke
2001	22.8
2002	22.5
2003	21.6
2004	20.9
2005	20.9
2006	20.8
2007	19.8
2008	20.6
2009	20.6
2010	19.3
2011	18.9

a. Construct a time series plot. What type of pattern exists in the data?
b. Use simple linear regression to find the parameters for the line that minimizes MSE for this time series.
c. One of OSH's *Healthy People 2020 Goals* is to cut the percentage of adults in the United States who were users of tobacco to 12% or less by the year 2020. Does your regression model from part (b) suggest that the OSH is on target to meet this goal? If not, use your model from part (b) to estimate the year in which the OSH will achieve this goal.

22. The president of a small manufacturing firm is concerned about the continual increase in manufacturing costs over the past several years. The following figures provide a time series of the cost per unit for the firm's leading product over the past eight years:

Year	Cost/Unit ($)	Year	Cost/Unit ($)
1	20.00	5	26.60
2	24.50	6	30.00
3	28.20	7	31.00
4	27.50	8	36.00

a. Construct a time series plot. What type of pattern exists in the data?
b. Use simple linear regression analysis to find the parameters for the line that minimizes MSE for this time series.
c. What is the average cost increase that the firm has been realizing per year?
d. Compute an estimate of the cost/unit for next year.

DATA file
Exercise

23. The medical community unanimously agrees on the health benefits of regular exercise, but are adults listening? During each of the past 15 years, a polling organization has surveyed Americans about their exercise habits. In the most recent of these polls, slightly over half of all American adults reported that they exercise for 30 or more minutes at least three times per week. The following data show the percentages of adults who reported that they exercise for 30 or more minutes at least three times per week during each of the 15 years of this study:

Year	Percentage of Adults Who Reported That They Exercise for 30 or More Minutes At Least Three Times per Week
1	41.0
2	44.9
3	47.1
4	45.7
5	46.6
6	44.5
7	47.6
8	49.8
9	48.1
10	48.9
11	49.9
12	52.1
13	50.6
14	54.6
15	52.4

a. Construct a time series plot. Does a linear trend appear to be present?
b. Use simple linear regression to find the parameters for the line that minimizes MSE for this time series.
c. Use the trend equation from part (b) to forecast the percentage of adults next year (Year 16 of the study) who will report that they exercise for 30 or more minutes at least three times per week.
d. Would you feel comfortable using the trend equation from part (b) to forecast the percentage of adults three years from now (Year 18 of the study) who will report that they exercise for 30 or more minutes at least three times per week?

SELF*test*

24. Consider the following time series:

Quarter	Year 1	Year 2	Year 3
1	71	68	62
2	49	41	51
3	58	60	53
4	78	81	72

a. Construct a time series plot. What type of pattern exists in the data?
b. Use a multiple linear regression model with dummy variables as follows to develop an equation to account for seasonal effects in the data. Qtr1 $=$ 1 if Quarter 1, 0 otherwise; Qtr2 $=$ 1 if Quarter 2, 0 otherwise; Qtr3 $=$ 1 if Quarter 3, 0 otherwise.
c. Compute the quarterly forecasts for next year.

25. Consider the following time series data:

Quarter	Year 1	Year 2	Year 3
1	4	6	7
2	2	3	6
3	3	5	6
4	5	7	8

a. Construct a time series plot. What type of pattern exists in the data?
b. Use a multiple regression model with dummy variables as follows to develop an equation to account for seasonal effects in the data. Qtr1 $=$ 1 if Quarter 1, 0 otherwise; Qtr2 $=$ 1 if Quarter 2, 0 otherwise; Qtr3 $=$ 1 if Quarter 3, 0 otherwise.
c. Compute the quarterly forecasts for next year.

26. The quarterly sales data (number of copies sold) for a college textbook over the past three years follow.

Quarter	Year 1	Year 2	Year 3
1	1690	1800	1850
2	940	900	1100
3	2625	2900	2930
4	2500	2360	2615

a. Construct a time series plot. What type of pattern exists in the data?
b. Use a regression model with dummy variables as follows to develop an equation to account for seasonal effects in the data. Qtr1 = 1 if Quarter 1, 0 otherwise; Qtr2 = 1 if Quarter 2, 0 otherwise; Qtr3 = 1 if Quarter 3, 0 otherwise.
c. Compute the quarterly forecasts for next year.
d. Let $t = 1$ to refer to the observation in Quarter 1 of Year 1; $t = 2$ to refer to the observation in Quarter 2 of Year 1; . . . ; and $t = 12$ to refer to the observation in Quarter 4 of Year 3. Using the dummy variables defined in part (b) and also using t, develop an equation to account for seasonal effects and any linear trend in the time series. Based upon the seasonal effects in the data and linear trend, compute the quarterly forecasts for next year.

DATA *file*
Pollution

27. Air pollution control specialists in southern California monitor the amount of ozone, carbon dioxide, and nitrogen dioxide in the air on an hourly basis. The hourly time series data exhibit seasonality, with the levels of pollutants showing patterns that vary over the hours in the day. On July 15, 16, and 17, the following levels of nitrogen dioxide were observed for the 12 hours from 6:00 A.M. to 6:00 P.M.:

July 15:	25	28	35	50	60	60	40	35	30	25	25	20
July 16:	28	30	35	48	60	65	50	40	35	25	20	20
July 17:	35	42	45	70	72	75	60	45	40	25	25	25

a. Construct a time series plot. What type of pattern exists in the data?
b. Use a multiple linear regression model with dummy variables as follows to develop an equation to account for seasonal effects in the data:

Hour1 = 1 if the reading was made between 6:00 A.M. and 7:00 A.M.; 0 otherwise

Hour2 = 1 if the reading was made between 7:00 A.M. and 8:00 A.M.; 0 otherwise

⋮

Hour11 = 1 if the reading was made between 4:00 P.M. and 5:00 P.M.; 0 otherwise

Note that when the values of the 11 dummy variables are equal to 0, the observation corresponds to the 5:00 P.M. to 6:00 P.M. hour.
c. Using the equation developed in part (b), compute estimates of the levels of nitrogen dioxide for July 18.
d. Let $t = 1$ to refer to the observation in Hour 1 on July 15; $t = 2$ to refer to the observation in Hour 2 of July 15; . . . ; and $t = 36$ to refer to the observation in Hour 12 of July 17. Using the dummy variables defined in part (b) and t, develop an equation to account for seasonal effects and any linear trend in the time series. Based upon the seasonal effects in the data and linear trend, compute estimates of the levels of nitrogen dioxide for July 18.

DATA *file*
SouthShore

28. South Shore Construction builds permanent docks and seawalls along the southern shore of Long Island, New York. Although the firm has been in business for only five years, revenue has increased from \$308,000 in the first year of operation to \$1,084,000 in the most recent year. The following data show the quarterly sales revenue in thousands of dollars:

Quarter	Year 1	Year 2	Year 3	Year 4	Year 5
1	20	37	75	92	176
2	100	136	155	202	282
3	175	245	326	384	445
4	13	26	48	82	181

a. Construct a time series plot. What type of pattern exists in the data?
b. Use a multiple regression model with dummy variables as follows to develop an equation to account for seasonal effects in the data. Qtr1 = 1 if Quarter 1, 0 otherwise; Qtr2 = 1 if Quarter 2, 0 otherwise; Qtr3 = 1 if Quarter 3, 0 otherwise.
c. Let Period = 1 to refer to the observation in Quarter 1 of Year 1; Period = 2 to refer to the observation in Quarter 2 of Year 1; . . . and Period = 20 to refer to the observation in Quarter 4 of Year 5. Using the dummy variables defined in part (b) and Period, develop an equation to account for seasonal effects and any linear trend in the time series. Based upon the seasonal effects in the data and linear trend, compute estimates of quarterly sales for Year 6.

Case Problem 1 FORECASTING FOOD AND BEVERAGE SALES

The Vintage Restaurant, on Captiva Island near Fort Myers, Florida, is owned and operated by Karen Payne. The restaurant just completed its third year of operation. During that time, Karen sought to establish a reputation for the restaurant as a high-quality dining establishment that specializes in fresh seafood. Through the efforts of Karen and her staff, her restaurant has become one of the best and fastest-growing restaurants on the island.

To better plan for future growth of the restaurant, Karen needs to develop a system that will enable her to forecast food and beverage sales by month for up to one year in advance. Table 15.17 shows the value of food and beverage sales ($1000s) for the first three years of operation.

Managerial Report

Perform an analysis of the sales data for the Vintage Restaurant. Prepare a report for Karen that summarizes your findings, forecasts, and recommendations. Include the following:

1. A time series plot. Comment on the underlying pattern in the time series.
2. Using the dummy variable approach, forecast sales for January through December of the fourth year.

DATA file
Vintage

TABLE 15.17 FOOD AND BEVERAGE SALES FOR THE VINTAGE RESTAURANT ($1000s)

Month	First Year	Second Year	Third Year
January	242	263	282
February	235	238	255
March	232	247	265
April	178	193	205
May	184	193	210
June	140	149	160
July	145	157	166
August	152	161	174
September	110	122	126
October	130	130	148
November	152	167	173
December	206	230	235

Assume that January sales for the fourth year turn out to be $295,000. What was your forecast error? If this error is large, Karen may be puzzled about the difference between your forecast and the actual sales value. What can you do to resolve her uncertainty in the forecasting procedure?

Case Problem 2 FORECASTING LOST SALES

The Carlson Department Store suffered heavy damage when a hurricane struck on August 31. The store was closed for four months (September through December), and Carlson is now involved in a dispute with its insurance company about the amount of lost sales during the time the store was closed. Two key issues must be resolved: (1) the amount of sales Carlson would have made if the hurricane had not struck and (2) whether Carlson is entitled to any compensation for excess sales due to increased business activity after the storm. More than $8 billion in federal disaster relief and insurance money came into the county, resulting in increased sales at department stores and numerous other businesses.

Table 15.18 gives Carlson's sales data for the 48 months preceding the storm. Table 15.19 reports total sales for the 48 months preceding the storm for all department stores in the county, as well as the total sales in the county for the four months the Carlson Department Store was closed. Carlson's managers asked you to analyze these data and develop estimates of the lost sales at the Carlson Department Store for the months of September through December. They also asked you to determine whether a case can be made for excess storm-related sales during the same period. If such a case can be made, Carlson is entitled to compensation for excess sales it would have earned in addition to ordinary sales.

Managerial Report

Prepare a report for the managers of the Carlson Department Store that summarizes your findings, forecasts, and recommendations. Include the following:

1. An estimate of sales for Carlson Department Store had there been no hurricane
2. An estimate of countywide department store sales had there been no hurricane
3. An estimate of lost sales for the Carlson Department Store for September through December

In addition, use the countywide actual department stores sales for September through December and the estimate in part (2) to make a case for or against excess storm-related sales.

DATA file
CarlsonSales

TABLE 15.18 SALES FOR CARLSON DEPARTMENT STORE ($ MILLIONS)

Month	Year 1	Year 2	Year 3	Year 4	Year 5
January		1.45	2.31	2.31	2.56
February		1.80	1.89	1.99	2.28
March		2.03	2.02	2.42	2.69
April		1.99	2.23	2.45	2.48
May		2.32	2.39	2.57	2.73
June		2.20	2.14	2.42	2.37
July		2.13	2.27	2.40	2.31
August		2.43	2.21	2.50	2.23
September	1.71	1.90	1.89	2.09	
October	1.90	2.13	2.29	2.54	
November	2.74	2.56	2.83	2.97	
December	4.20	4.16	4.04	4.35	

TABLE 15.19 DEPARTMENT STORE SALES FOR THE COUNTY ($ MILLIONS)

DATA *file*
CountySales

Month	Year 1	Year 2	Year 3	Year 4	Year 5
January		46.80	46.80	43.80	48.00
February		48.00	48.60	45.60	51.60
March		60.00	59.40	57.60	57.60
April		57.60	58.20	53.40	58.20
May		61.80	60.60	56.40	60.00
June		58.20	55.20	52.80	57.00
July		56.40	51.00	54.00	57.60
August		63.00	58.80	60.60	61.80
September	55.80	57.60	49.80	47.40	69.00
October	56.40	53.40	54.60	54.60	75.00
November	71.40	71.40	65.40	67.80	85.20
December	117.60	114.00	102.00	100.20	121.80

Appendix 15.1 FORECASTING WITH EXCEL DATA ANALYSIS TOOLS

In this appendix we show how Excel can be used to develop forecasts using three forecasting methods: moving averages, exponential smoothing, and trend projection. We also show how to use Excel Solver for least-squares fitting of models to data.

Moving Averages

*If the **Data Analysis** option does not appear in the **Analyze** group, you will have to include the Add-In in Excel. To do so, click on the **File** tab, then click **Options**, and then **Add-Ins**. Click **Go** next to the **Excel Add-Ins** drop-down box. Click the box next to **Analysis ToolPak** and click **OK**.*

To show how Excel can be used to develop forecasts using the moving averages method, we develop a forecast for the gasoline sales time series in Table 15.1 and Figure 15.1. We assume that the user has entered the week in rows 2 through 13 of column A and the sales data for the 12 weeks into worksheet rows 2 through 13 of column B (as in Figure 15.13).

The following steps can be used to produce a three-week moving average:

Step 1. Select the **Data** tab
Step 2. From the **Analyze** group select the **Data Analysis** option
Step 3. When the **Data Analysis** dialog box appears, choose **Moving Average** and click **OK**
Step 4. When the **Moving Average** dialog box appears:
Enter B2:B13 in the **Input Range** box
Enter 3 in the **Interval** box
Enter C2 in the **Output Range** box
Click **OK**

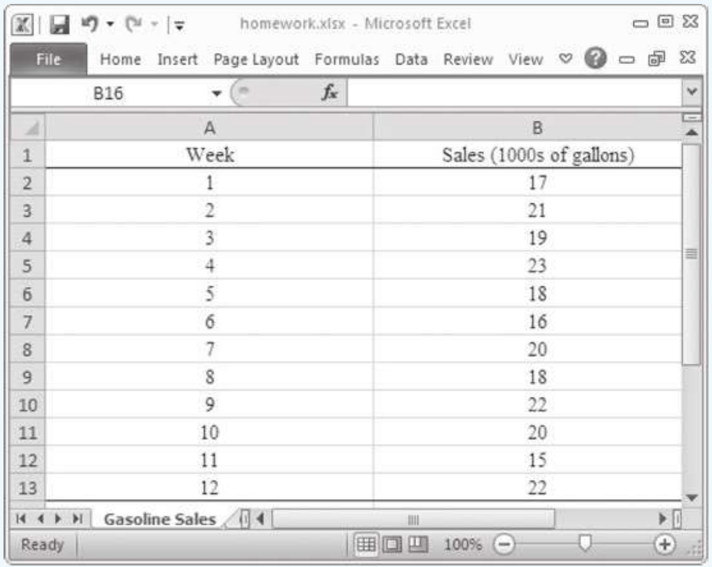
Once you have completed this step (as shown in Figure 15.14), the three-week moving average forecasts will appear in column C of the worksheet as in Figure 15.15. Note that forecasts for periods of other lengths can be computed easily by entering a different value in the **Interval** box.

Exponential Smoothing

To show how Excel can be used for exponential smoothing, we again develop a forecast for the gasoline sales time series in Table 15.1 and Figure 15.1. We assume that the user has entered the week in rows 2 through 13 of column A and the sales data for the 12 weeks into worksheet rows 2 through 13 of column B (as in Figure 15.13), and that the smoothing constant is $\alpha = 0.2$. The following steps can be used to produce a forecast:

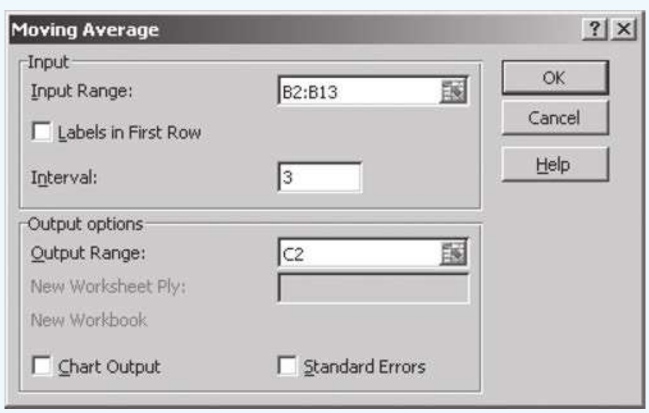
Step 1. Select the **Data** tab
Step 2. From the **Analyze** group select the **Data Analysis** option

FIGURE 15.13 GASOLINE SALES DATA IN EXCEL ARRANGED TO USE THE MOVING AVERAGES FUNCTION TO DEVELOP FORECASTS



	A	B
1	Week	Sales (1000s of gallons)
2	1	17
3	2	21
4	3	19
5	4	23
6	5	18
7	6	16
8	7	20
9	8	18
10	9	22
11	10	20
12	11	15
13	12	22

FIGURE 15.14 EXCEL MOVING AVERAGE DIALOGUE BOX FOR A 3-PERIOD MOVING AVERAGE



Step 3. When the **Data Analysis** dialog box appears, choose **Exponential Smoothing** and click **OK**

Step 4. When the **Exponential Smoothing** dialog box appears:
Enter B2:B13 in the **Input Range** box
Enter 0.8 in the **Damping factor** box
Enter C2 in the **Output Range** box
Click **OK**

Once you have completed this step (as shown in Figure 15.16), the exponential smoothing forecasts will appear in column C of the worksheet (as in Figure 15.17). Note that the value we entered in the **Damping factor** box is $1 - \alpha$; forecasts for other smoothing constants can be computed easily by entering a different value for $1 - \alpha$ in the **Damping factor** box.

FIGURE 15.15 GASOLINE SALES DATA AND OUTPUT OF MOVING AVERAGES FUNCTION IN EXCEL

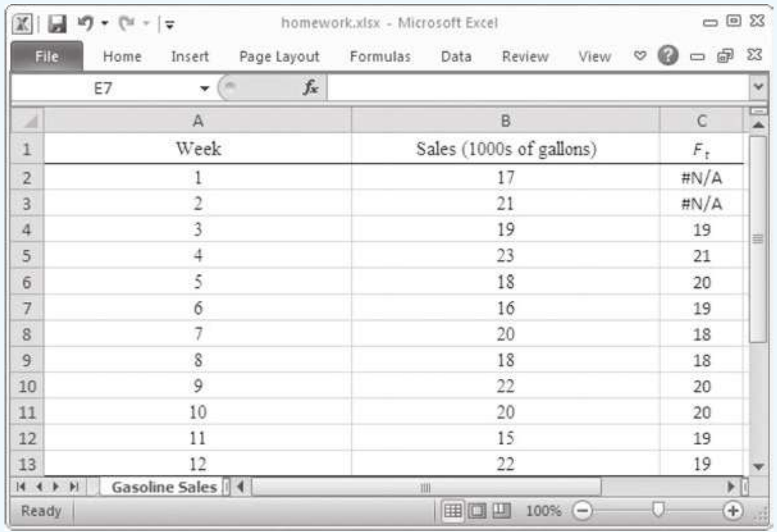
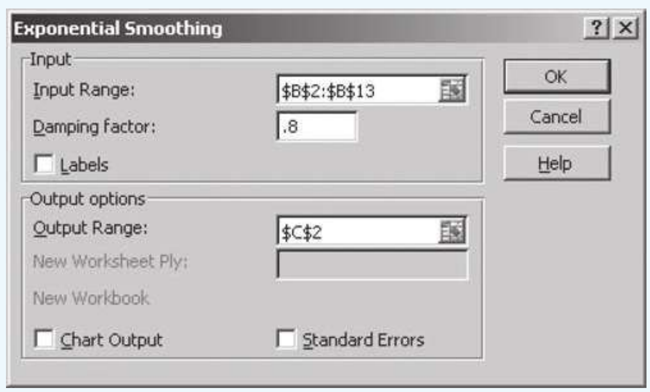


FIGURE 15.16 EXCEL EXPONENTIAL SMOOTHING DIALOGUE BOX FOR $\alpha = 0.20$

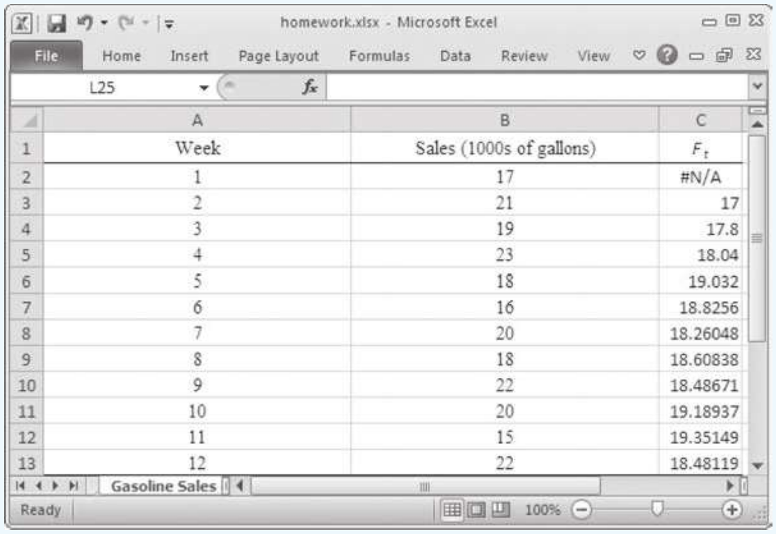


Trend Projection

To show how Excel can be used for trend projection, we develop a forecast for the bicycle sales time series in Table 15.3 and Figure 15.3. We assume that the user has entered the year (1–10) for each observation into worksheet rows 2 through 11 of column A and the sales values into worksheet rows 2 through 11 of column B as shown in Figure 15.18. The following steps can be used to produce a forecast for Year 11 by trend projection:

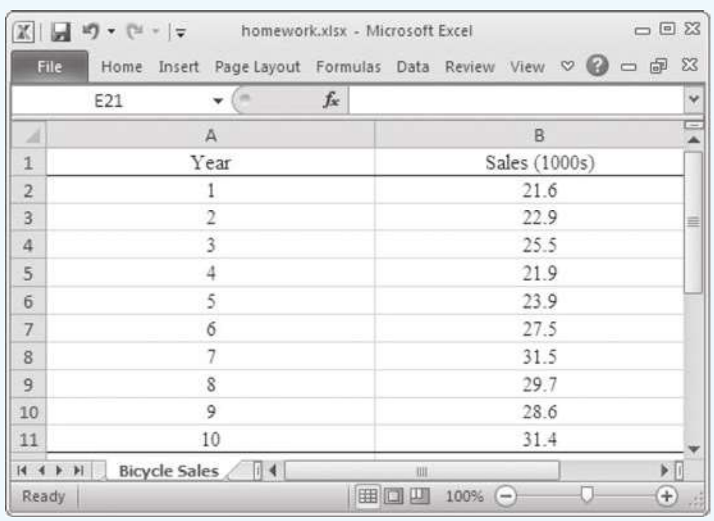
Step 1. Select the **Formulas** tab
Step 2. Select two cells in the row where you want the regression coefficients b_1 and b_0 to appear (for this example, choose D1 and E1)
Step 3. Click on the **Insert Function** key
Step 4. When the **Insert Function** dialog box appears:
Choose **Statistical** in the **Or select a category** box
Choose **Linest** in the **Select a function** box
Click **OK**

FIGURE 15.17 GASOLINE SALES DATA AND OUTPUT OF EXPONENTIAL SMOOTHING FUNCTION IN EXCEL



	A	B	C
1	Week	Sales (1000s of gallons)	F_t
2	1	17	#N/A
3	2	21	17
4	3	19	17.8
5	4	23	18.04
6	5	18	19.032
7	6	16	18.8256
8	7	20	18.26048
9	8	18	18.60838
10	9	22	18.48671
11	10	20	19.18937
12	11	15	19.35149
13	12	22	18.48119

FIGURE 15.18 BICYCLE SALES DATA IN EXCEL ARRANGED TO USE THE LINEST FUNCTION TO FIND THE LINEAR TREND



	A	B
1	Year	Sales (1000s)
2	1	21.6
3	2	22.9
4	3	25.5
5	4	21.9
6	5	23.9
7	6	27.5
8	7	31.5
9	8	29.7
10	9	28.6
11	10	31.4

See Figure 15.19 for an example of this step.

Step 5. When the **Function Arguments** dialog box appears:
Enter B2:B11 in the **Known_y's** box
Enter A2:A11 in the **Known_x's** box
Click **OK**

FIGURE 15.19 EXCEL INSERT FUNCTION DIALOGUE BOX FOR FINDING THE TREND LINE USING THE LINEST FUNCTION IN EXCEL

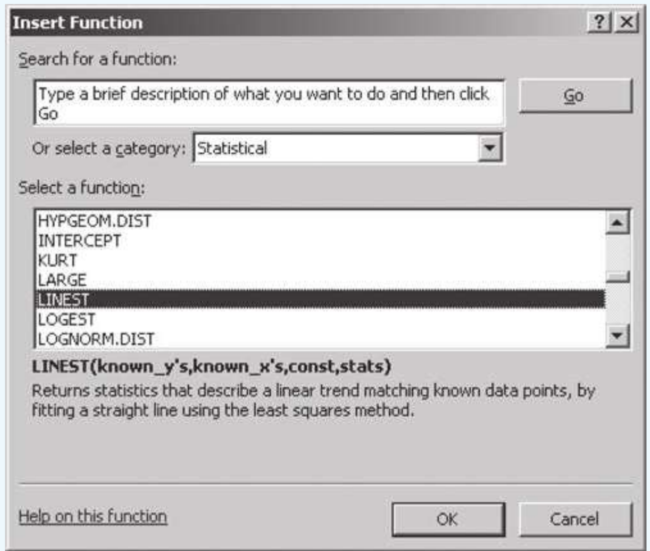
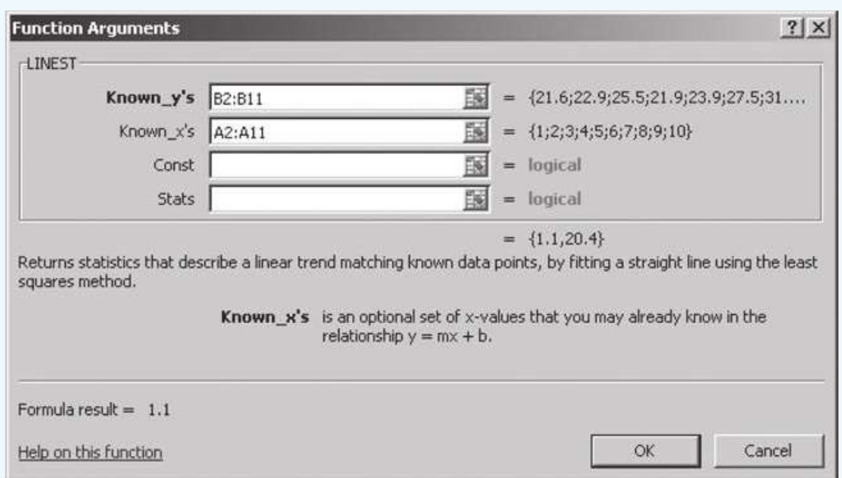


FIGURE 15.20 EXCEL FUNCTION ARGUMENTS DIALOGUE BOX FOR FINDING THE TREND LINE USING THE LINEST FUNCTION IN EXCEL



See Figure 15.20 for an example of this step.

Step 6. Hit the F2 key and then simultaneously hit the Shift, Control, and Enter keys (Shift + Control + Enter) to create an array that contains the values of the regression coefficients b_1 and b_0

At this point you have generated the regression coefficients b_1 and b_0 in the two cells you originally selected in step 1. It is important to note that cell D1 contains b_1 and cell E1 contains b_0.

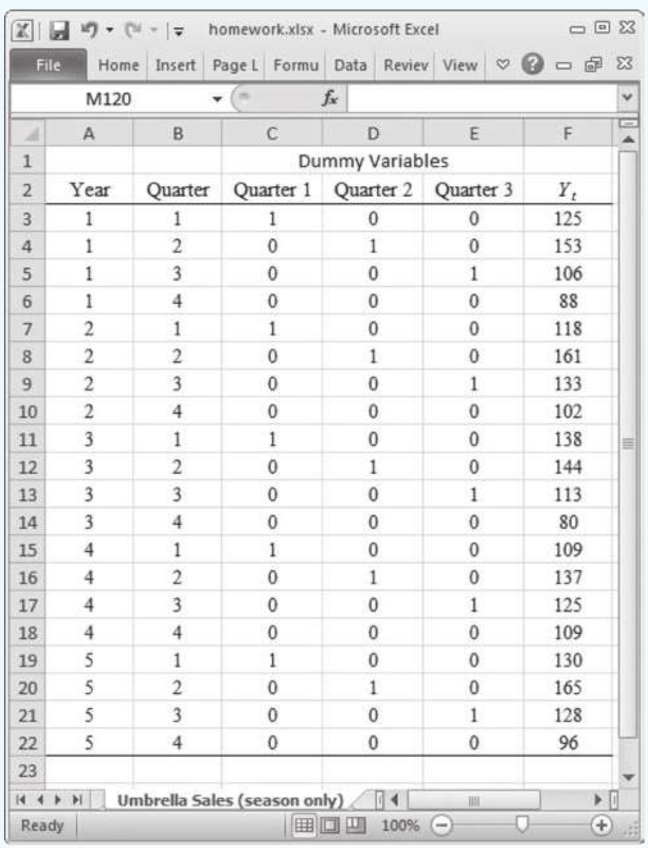
To generate a forecast, in a blank cell, multiply the value of the independent variable t by b_1 and add the value of b_0 to this product. For example, if you wish to use this linear trend model to generate a forecast for Year 11 and the value of b_1 is in cell D1 and the value of b_0 is in cell E1, then enter =11*D1+E1 in a blank cell. The forecast for Year 11, in this case 32.5, will appear in the blank cell in which you enter this formula.

Models with Seasonality and No Trend

To show how Excel can be used to fit models with seasonality, we develop a forecast for the umbrella sales time series in Table 15.13 and Figure 15.11. We assume that the user has entered the year (1–5) for each observation into worksheet rows 3 through 22 of column A; the values for the quarter in worksheet rows 3 through 22 of column B; the values for the quarterly dummy variables $Qtr1_t$, $Qtr2_t$, and $Qtr3_t$ in worksheet rows 3 through 22 of columns C, D, and E, respectively; and the sales values into worksheet rows 3 through 22 of column F. The following steps can be used to produce a forecast for Year 11 by trend projection as shown in Figure 15.21.

Step 1. Select the **Formulas** tab
Step 2. Select four cells in the row where you want the regression coefficients b_3, b_2, b_1, and b_0 to appear (for this example, choose G1:J1)
Step 3. Click on the **Insert Function** key
Step 4. When the **Insert Function** dialog box appears:
Choose **Statistical** in the **Or select a category** box

FIGURE 15.21 UMBRELLA SALES DATA IN EXCEL ARRANGED TO USE THE LINEST FUNCTION TO FIND THE SEASONAL COMPONENTS



	A	B	C	D	E	F
1				Dummy Variables		
2	Year	Quarter	Quarter 1	Quarter 2	Quarter 3	Y_t
3	1	1	1	0	0	125
4	1	2	0	1	0	153
5	1	3	0	0	1	106
6	1	4	0	0	0	88
7	2	1	1	0	0	118
8	2	2	0	1	0	161
9	2	3	0	0	1	133
10	2	4	0	0	0	102
11	3	1	1	0	0	138
12	3	2	0	1	0	144
13	3	3	0	0	1	113
14	3	4	0	0	0	80
15	4	1	1	0	0	109
16	4	2	0	1	0	137
17	4	3	0	0	1	125
18	4	4	0	0	0	109
19	5	1	1	0	0	130
20	5	2	0	1	0	165
21	5	3	0	0	1	128
22	5	4	0	0	0	96
23						

Choose **LINEST** in the **Select a function** box
Click **OK**

Step 5. When the **Function Arguments** dialog box appears:
Enter F3:F22 in the **Known_y's** box
Enter C3:E22 in the **Known_x's** box
Click **OK**
See Figure 15.22 for an example of this step.

Step 6. Hit the F2 key and then simultaneously hit the Shift, Control, and Enter keys (Shift + Control + Enter) to create an array that contains the values of the regression coefficients b_3, b_2, b_1, and b_0

At this point you have generated the regression coefficients b_3, b_2, b_1, and b_0 in cells G1:J1 selected in step 1. It is important to note that the first cell you selected contains b_3, the second cell you selected contains b_2, the third cell you selected contains b_1, and the fourth cell you selected contains b_0 (i.e., if you selected cells G1:J1 in step 1, the value of b_1 will be in cell G1, the value of b_2 will be in H1, the value of b_1 will be in I1, and the value of b_0 will be in cell J1).

To generate a forecast, in a blank cell, add together b_0 and the product of b_1 and Qtr1$_t$, the product of b_2 and Qtr2$_t$, and the product of b_3 and Qtr3$_t$. For example, if you wish to use this linear trend model to generate a forecast for the first quarter of next year and the value of b_3 is in cell G1, the value of b_2 is in cell H1, the value of b_1 is in cell I1, and the value of b_0 is in cell J1, then enter =1*G1+0*H1+0*I1+J1 in a blank cell. The forecast for the first quarter of next year, in this case 124.0, will appear in the blank cell in which you enter this formula.

Models with Seasonality and Linear Trend

To show how Excel can be used to fit models with seasonality and a linear trend, we develop a forecast for the umbrella set time series in Table 15.13 and Figure 15.11. We assume that the user has entered the year (1–5) for each observation into worksheet rows 3 through 22 of column A; the values for the quarter in worksheet rows 3 through 22 of column B; the values for the quarterly dummy variables Qtr1$_t$, Qtr2$_t$, and Qtr3$_t$ into worksheet rows 3 through 22 of columns C, D, and E, respectively; the values of period t into worksheet rows 3 through 22 of column F; and the sales values into worksheet rows 3 through 22 of column G. The following steps can be used to produce a forecast for Year 11 by trend projection as shown in Figure 15.23.

FIGURE 15.22 EXCEL FUNCTION ARGUMENTS DIALOGUE BOX FOR FINDING THE SEASONAL COMPONENTS USING THE LINEST FUNCTION IN EXCEL

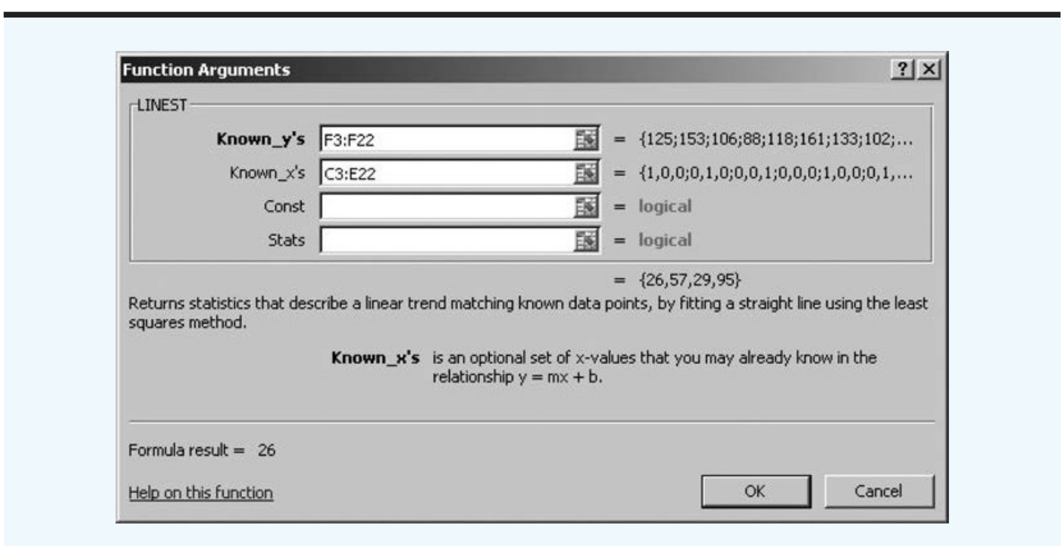


FIGURE 15.23 UMBRELLA TIME SERIES DATA IN EXCEL ARRANGED TO USE THE LINEST FUNCTION TO FIND BOTH THE SEASONAL COMPONENTS AND TREND COMPONENT

	A	B	C	D	E	F	G
1				Dummy Variables			
2	Year	Quarter	Quarter 1	Quarter 2	Quarter 3	t	Y_t
3	1	1	1	0	0	1	125
4	1	2	0	1	0	2	153
5	1	3	0	0	1	3	106
6	1	4	0	0	0	4	88
7	2	1	1	0	0	5	118
8	2	2	0	1	0	6	161
9	2	3	0	0	1	7	133
10	2	4	0	0	0	8	102
11	3	1	1	0	0	9	138
12	3	2	0	1	0	10	144
13	3	3	0	0	1	11	113
14	3	4	0	0	0	12	80
15	4	1	1	0	0	13	109
16	4	2	0	1	0	14	137
17	4	3	0	0	1	15	125
18	4	4	0	0	0	16	109
19	5	1	1	0	0	17	130
20	5	2	0	1	0	18	165
21	5	3	0	0	1	19	128
22	5	4	0	0	0	20	96
23							

Step 1. Select the **Formulas** tab

Step 2. Select five cells in the row where you want the regression coefficients b_4, b_3, b_2, b_1, and b_0 to appear for this example; choose H1:L1

Step 3. Click on the **Insert Function** key

Step 4. When the **Insert Function** dialog box appears:
Choose **Statistical** in the **Or select a category** box
Choose **LINEST** in the **Select a function** box
Click **OK**

Step 5. When the **Function Arguments** dialog box appears:
Enter G3:G22 in the **Known_y's** box
Enter C3:F22 in the **Known_x's** box
Click **OK**

Step 6. Hit the F2 key and then simultaneously hit the Shift, Control, and Enter keys (Shift + Control + Enter) to create an array that contains the values of the regression coefficients b_4, b_3, b_2, b_1, and b_0

At this point you have generated the regression coefficients b_4, b_3, b_2, b_1, and b_0 in cells H1:L1 selected in step 1. It is important to note that the first cell you selected contains b_4, the second cell you selected contains b_3, the third cell you selected contains b_2, the fourth cell you selected contains b_1, and the fifth cell you selected contains b_0 (i.e., if you selected cells H1:L1 in step 1, the value of b_4 will be in cell H1, the value of b_1 will be in cell I1, the value of b_2 will be in J1, the value of b_1 will be in K1, and the value of b_0 will be in cell L1).

To generate a forecast, in a blank cell, add together b_0 and the product of b_1 and Qtr1_t, the product of b_2 and Qtr2_t, the product of b_3 and Qtr3_t, and the product of b_4 and t. For example, if you wish to use this linear trend model to generate a forecast for the first quarter of Year 5 and the value of b_4 is in cell H1, the value of b_3 is in cell I1, the value of b_2 is in cell J1, the value of b_1 is in cell K1, and the value of b_0 is in cell L1, then enter = 17*H1+1*I1+0*J1+0*K1+L1 in a blank cell. The forecast for the first quarter of next year, in this case 7.19, will appear in the blank cell in which you enter this formula.

Appendix 15.2 USING THE EXCEL FORECAST SHEET

Excel 2016 features a new tool called Forecast Sheet. This interface automatically produces forecasts using the Holt–Winters additive seasonal smoothing model, which is an exponential smoothing approach to estimating additive linear trend and seasonal effects. It also generates a variety of other outputs that are useful in assessing the accuracy of the forecast model it produces.

Excel refers to the forecasting approach used by Forecast: Sheet as the AAA exponential smoothing (ETS) algorithm, where AAA stands for additive error, additive trend, and additive seasonality.

We will demonstrate Forecast Sheet on the four years of quarterly smartphone sales that are provided in Table 15.6. A review of the time series plot of these data in Figure 15.6 provides clear evidence of an increasing linear trend and a seasonal pattern (sales are consistently lowest in the second quarter of each year and highest in quarters 3 and 4). We concluded in Section 15.1 that we need to use a forecasting method that is capable of dealing with both trend and seasonality when developing a forecasting model for this time series, and so it is appropriate to use Forecast Sheet to produce forecasts for these data.

We begin by putting the data into the format required by Forecast Sheet. The time series data must be collected on a consistent interval (i.e., annually, quarterly, monthly, etc.), and the spreadsheet must include two data series in contiguous columns or rows that include:

- a series with the dates or periods in the time series
- a series with corresponding time series values

DATA *file*
SmartPhoneSales

First, open the file *SmartPhoneSales*, then insert a column between column B (Quarter) and Column C (Sales (1000s)). Enter *Period* into cell C1; this will be the heading for the column of values that will represent the periods in our data. Next enter *1* in cell C2, 2 in cell C3, *3* in cell C4, and so on, ending with *16* in Cell C17 as shown in Figure 15.24.

Now that the data are properly formatted for Forecast Sheet, the following steps can be used to produce forecasts for the next four quarters (periods 17 through 20) with Forecast Sheet:

Step 1. Highlight cells C1:D17 (the data in column C of this highlighted section is what Forecast Sheet refers to as the **Timeline Range** and the data in column D is the **Values Range**).
Step 2. Click the **Data** tab in the Ribbon
Step 3. Click Forecast Sheet in the Forecast group
Step 4. When the **Create Forecast Worksheet** dialog box appears (Figure 15.25):
Select **20** for **Forecast End**
Click **Options** to expand the **Create Forecast Worksheet** dialog box and show the options (Figure 15.25)
Select **16** for **Forecast Start**
Select **95%** for **Confidence Interval**
Under **Seasonality**, click on **Set Manually** and select **4**
Select the checkbox for **Include forecast statistics**
Click **Create**

Forecast Sheet requires that the period selected for Forecast Start is one of the periods of the original time series.

The results of Forecast Sheet will be output to a new worksheet as shown in Figure 15.26. The output of Forecast Sheet includes the following.

- The period for each of the 16 time series observations and the forecasted time periods in column A
- The actual time series data for periods 1 to 16 in column B

FIGURE 15.24 SMARTPHONE DATA REFORMATTED FOR FORECAST SHEET

	A	B	C	D
1	**Year**	**Quarter**	**Period**	**Sales (1000s)**
2	1	1	1	4.8
3	1	2	2	4.1
4	1	3	3	6.0
5	1	4	4	6.5
6	2	1	5	5.8
7	2	2	6	5.2
8	2	3	7	6.8
9	2	4	8	7.4
10	3	1	9	6.0
11	3	2	10	5.6
12	3	3	11	7.5
13	3	4	12	7.8
14	4	1	13	6.3
15	4	2	14	5.9
16	4	3	15	8.0
17	4	4	16	8.4

- The forecasts for periods 16 to 20 in column C
- The lower confidence bounds for the forecasts for periods 16 to 20 in column D
- The upper confidence bounds for the forecasts for periods 16 to 20 in column E
- A line graph of the time series, forecast values, and forecast interval
- The values of the three parameters (alpha, beta, and gamma) used in the Holt–Winters additive seasonal smoothing model in cells H2:H4 (these values are determined by an algorithm in Forecast Sheet)
- Measures of forecast accuracy in cells H5:H8, including:
  - the MASE, or mean absolute scaled error, in cell H5; MASE, which was not discussed in this chapter, is defined as:

$$\text{MASE} = \frac{1}{n}\sum_{t=1}^{n}\frac{|e_t|}{\frac{1}{n-1}\sum_{t=1}^{n}|y_t - y_{t-1}|}$$

    MASE compares the forecast error, e_t to a naïve forecast error given by $|y_t - y_{t-1}|$. If MASE > 1, then the forecast is considered inferior to a naïve forecast; if MASE < 1 the forecast is considered superior to a naïve forecast.

  - the SMAPE, or symmetric mean absolute percentage error, in cell H6; SMAPE, which was not discussed in this chapter, is defined as:

$$\text{SMAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{|e_t|}{(|y_t| + |\hat{y}_t)/2}$$

    SMAPE is similar to mean absolute percentage error (MAPE), discussed in Section 8.2; both SMAPE and MAPE measure forecast error relative to actual values.

FIGURE 15.25 CREATE FORECAST WORKSHEET DIALOG BOX WITH OPTIONS OPEN FOR QUARTERLY SMARTPHONE SALES DATA

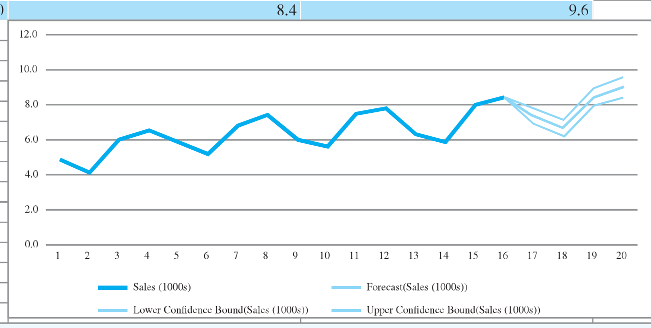
- the MAE, or mean absolute error, (as defined in equation 8.3) in cell H7
- the RMSE, or root mean squared error, (which is the square root of the MSE, defined in equation 8.4) in cell H8

Forecast Sheet also features an algorithm for finding the number of time periods over which the seasonal pattern recurs. To use this algorithm, select the option for **Detect Automatically** under Seasonality in the **Create Forecast Worksheet** dialog box before clicking Create. We suggest using this feature only to confirm a suspected seasonal pattern (Forecast Sheet actually does successfully detect a four-period seasonal pattern in the quarterly smartphone sales data). Using this feature to find a seasonal effect may lead to identification of a spurious pattern that does not actually reflect seasonality and cannot be expected to persist in future periods. This would result in a model that is overfit on the observed time series data and would likely produce very inaccurate forecasts. A forecast model with seasonality should only be fit when the modeler has reason to suspect a specific seasonal pattern.

Forecast Sheet is actually an interface that implements several functions that are new to Excel 2016. We can recreate the output from Forecast Sheet using these Excel functions. For example, after reformatting the data in the same manner as we did in preparation

FIGURE 15.26 FORECAST SHEET RESULTS FOR QUARTERLY SMARTPHONE SALES DATA

Period	Sales (1000s)	Forecast(Sales (1000s))	Lower Confidence Bound(Sales (1000s))	Upper Confidence Bound(Sales (1000s))		Statistic	Value
1	4.8					Alpha	0.50
2	4.1					Beta	0.00
3	6.0					Gamma	0.00
4	6.5					MASE	0.22
5	5.8					SMAPE	0.03
6	5.2					MAE	0.20
7	6.8					RMSE	0.27
8	7.4						
9	6.0						
10	5.6						
11	7.5						
12	7.8						
13	6.3						
14	5.9						
15	8.0						
16	8.4	8.4	8.4	8.4			
17		7.3	6.9	7.8			
18		6.6	6.2	7.1			
19		8.4	7.9	9.0			
20		9.0	8.4	9.6			



for using Forecast Sheet, we enter the values 17, 18, 19, and 20 into cells C18 through C21, respectively, to denote the periods for which we will be generating forecasts. We then enter the column titles *Forecast, Lower Confidence Interval, Upper Confidence Interval, Statistic,* and *Value* in cells E1 through I1, respectively. Next we enter the statistic labels *Alpha, Beta, Gamma, MASE, SMAPE, MAE*, and *RMSE* in cells H2 through H8, respectively. Finally, we enter the label Seasonality in cell H12. This updated worksheet is shown in Figure 15.27.

We can now recreate the Forecast Sheet results in the updated worksheet shown in Figure 15.27 as follows:

- The forecast generated by Forecast Sheet for period 17 for the smartphone quarterly sales data can be found by using the formula:

  =FORECAST.ETS(C18,D2:D17,C2:C17,TRUE).

  The arguments for this function are the forecast period, the time series values, the timeline associated with the time series values, and a seasonality indicator that is TRUE if Excel is to automatically detect a seasonal pattern for the forecast and FALSE otherwise.

- The margin of error for the confidence bounds generated by Forecast Sheet for period 17 for the smartphone quarterly sales data can be found by using the formula:

  =FORECAST.ETS.CONFINT(C18,D2:D17,C2:C17,0.95,TRUE)

  The confidence bounds generated by Forecast Sheet for period 17 for the smartphone quarterly sales data can be found by using the formulas:

  =E18-FORECAST.ETS.CONFINT(C18,D2:D17,C2:C17,0.95,TRUE)

FIGURE 15.27 SMARTPHONE DATA REFORMATTED FOR USE WITH EXCEL FORECAST FUNCTIONS

	A	B	C	D	E	F	G	H	I
1	Year	Quarter	Period	Sales (1000s)	Forecast	Lower Confidence Bound	Upper Confidence Bound	Statistic	Value
2	1	1	1	4.8				Alpha	
3	1	2	2	4.1				Beta	
4	1	3	3	6.0				Gamma	
5	1	4	4	6.5				MASE	
6	2	1	5	5.8				SMAPE	
7	2	2	6	5.2				MAE	
8	2	3	7	6.8				RMSE	
9	2	4	8	7.4					
10	3	1	9	6.0					
11	3	2	10	5.6					
12	3	3	11	7.5				Seasonality	
13	3	4	12	7.8					
14	4	1	13	6.3					
15	4	2	14	5.9					
16	4	3	15	8.0					
17	4	4	16	8.4					
18			17						
19			18						
20			19						
21			20						

and

=E18+FORECAST.ETS.CONFINT(C18,D2:D17,C2:C17,0.95,TRUE).

The arguments for this function are identical to the arguments for the FORECAST.ETS function.

The statistics generated by Forecast Sheet for the smartphone quarterly sales data can be found by using the formulas:

- Alpha

  =FORECAST.ETS.STAT(D2:D17,C2:C17,1,TRUE)

- Beta

  =FORECAST.ETS.STAT(D2:D17,C2:C17,2,TRUE)

- Gamma

  =FORECAST.ETS.STAT(D2:D17,C2:C17,3,TRUE)

- MASE

  =FORECAST.ETS.STAT(D2:D17,C2:C17,4,TRUE)

- SMAPE

  =FORECAST.ETS.STAT(D2:D17,C2:C17,5,TRUE)

- MAE

  =FORECAST.ETS.STAT(D2:D17,C2:C17,6,TRUE)

- RMSE

  = FORECAST.ETS.STAT(D2:D17,C2:C17,7,TRUE)

The arguments for this function are the time series values, the timeline associated with the time series values, the statistic type, and a seasonality indicator that is TRUE if Excel is to automatically detect a seasonal pattern for the forecast and FALSE otherwise. The statistic-type argument indicates which statistic will be produced by this function. Values for the statistic-type argument include the following:

- Statistic type = 1: requests the alpha parameter used in the Holt–Winters additive seasonal smoothing model
- Statistic type = 2: requests the beta parameter used in the Holt–Winters additive seasonal smoothing model
- Statistic type = 3: requests the gamma parameter used in the Holt–Winters additive seasonal smoothing model
- Statistic type = 4: requests the MASE that results when the Holt–Winters additive seasonal smoothing model is applied to the original time series data
- Statistic type = 5: requests the SMAPE that results when the Holt–Winters additive seasonal smoothing model is applied to the original time series data
- Statistic type = 6: requests the MAE that results when the Holt–Winters additive seasonal smoothing model is applied to the original time series data
- Statistic type = 7: requests the RMSE that results when the Holt–Winters additive seasonal smoothing model is applied to the original time series data

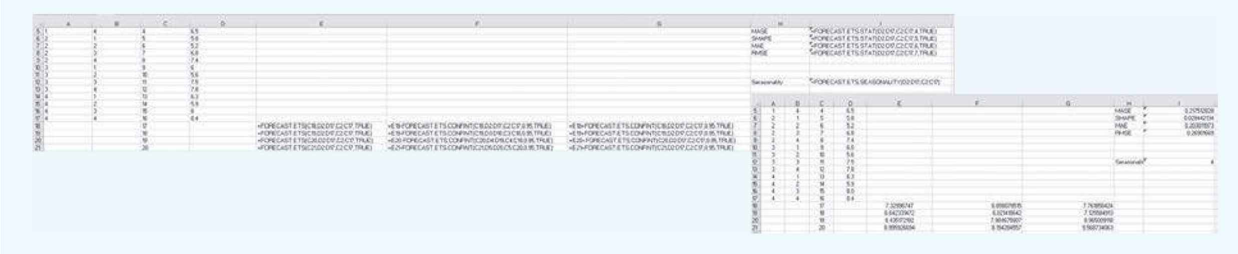
We can also use the formula FORECAST.ETS.SEASONALITY to determine the number of periods in the seasonal pattern detected by the FORECAST.ETS formula in the smartphone quarterly sales data by entering the following formula:

=FORECAST.ETS.SEASONALITY(D2:D17,C2:C17)

We now illustrate the use of these Excel functions on the smartphone quarterly sales data. We use the:

- FORECAST.ETS function to generate forecasts for periods 17 to 20 in cells E18:E21
- FORECAST.ETS.CONFINT function to generate confidence bounds for these four forecasts in cells F18:G21 (we subtract these values from the corresponding forecast values in cells E18:E21 to create lower confidence bounds, and we add these values to the corresponding forecast values in cells E18:E21 to create upper confidence bounds)
- FORECAST.ETS.STAT function with the appropriate values for the statistics type argument to generate the parameters of our Holt–Winters additive seasonal smoothing model and measures of forecast accuracy in cells I2:I8
- FORECAST.ETS.SEASONALITY function in cell I12 to determine the number of periods in the seasonal pattern detected by the = FORECAST.ETS function.

FIGURE 15.28 FORECAST RESULTS FOR QUARTERLY SMARTPHONE SALES DATA USING EXCEL FUNCTIONS



These results are provided in Figure 15.28.

Finally, note that:

- **Forecast Start** in the **Create Forecast Worksheet** dialog box controls both the first period to be forecasted and the last period to be used to generate the forecast model. If we had selected 15 for Forecast Start, we would have generated a forecast model for the smartphone monthly sales data based on only the first 15 periods of data in the original time series.
- Forecast Sheet can accommodate multiple observations for a single period of the time series. The **Aggregate Duplicates Using** option in the **Create Forecast Worksheet** dialog box allows the user to select from several ways to deal with this issue.
- Forecast Sheet allows for up to 30% of the values for the time series variable to be missing. In the smartphone quarterly sales data, the value of sales for up to 30% of the 16 periods (or 4 periods) could be missing and Forecast Sheet will still produce forecasts. The **Fill Missing Points Using** option in the **Create Forecast Worksheet** dialog box allows the user to select whether the missing values will be replaced with zero or with the result of linearly interpolating existing values in the time series.