CHAPTER 12

Forecasting

Anton_Ivanov / Shutterstock

LEARNING OBJECTIVES

After reading this chapter, you will be able to:

- Discuss the strategic role of forecasting in supply chain management.
- Describe the forecasting process and identify the components of forecasting demand.
- Forecast demand using various time series models, including exponential smoothing, and trend and seasonal adjustments.
- Discuss and calculate various methods for evaluating forecast accuracy.
- Use Excel to create various forecast models.
- Develop forecasting models with linear and multiple regression analysis.

Forecasting the Global Smartphone Market

Since the introduction of both the first iPhone by Apple and the Android platform in 2007, the global smartphone market has experienced extraordinary growth with steadily increasing sales. In 2008, 140 million smartphones were sold worldwide, and by 2015 sales with a market value over $300 billion had increased to over 1.4 billion units, a 900% increase, and the forecast is for this trend to continue with sales of smartphones estimated to approach 2 billion by 2019. As smartphone sales have increased, the sale of mobile non-smartphones have correspondingly decreased from over a billion in 2008 to 620 million in 2015. Further, various forecasts estimate that by 2018 mobile devices (e.g., smartphones and tablets) will constitute almost 90% of the market for devices that connect to the Internet compared to a little over 10% for PCs and laptops, and this trend will also likely continue into the future.

In the immediate (five-year) future, smartphone sales are forecast to continue to grow, but likely at a much slower rate. In the first part of this decade emerging markets in China and India dominated new user sales as the smartphone market in Europe, Japan, Australia, and North America increasingly became more dependent on the replacement phone market as the number of new users dwindled. In 2015, a little over 420 million smartphones were sold in China, an increase of almost 450% from the 77 million sold in 2011. However, there is evidence that the market for new (first-time) users in China has slowed, and with a 90% market penetration, China has now become a replacement market as well, and will continue to be so in the future. As the Chinese smartphone market emerged many forecasters predicted Apple would not be able to take advantage of the world's largest smartphone market with over 520 million possible users because of its high-end pricing strategy. However, almost the opposite occurred and by 2015 China became Apple's largest market, superseding the United States, primarily because so many consumers in China joined the middle class.

By 2019 it is estimated that new mobile phone users will account for only around 7% of the total global market. As the global smartphone market growth slows and new players enter the market, especially in China, the selling price of a smartphone is expected to decline from an average of about $300 in 2014 to about $240 by 2019, reducing revenues. Forecasters expect Apple and Samsung to continue to dominate the smartphone market, but perhaps not to the extent as in the past as competition increases among traditional and new competitors, especially in China, that are expected to introduce smartphones priced below $150 for consumers who currently own less-sophisticated mobile phones and will switch to low-priced devices.

Although Apple's global market share is expected to decline, its product and pricing strategy aimed at users at the high end of the market should enable it to continue to retain the largest portion of market revenues and profits, while Android-platform smartphones (e.g., Samsung) are likely to continue to make deep penetrations into the market. In the future in China and other mature markets, convincing existing users to upgrade to new smartphones will be the key to further growth. For competitive smartphone companies this will likely depend heavily on creating new features that users do not currently know they want or need (something that Apple has been very good at), increasingly large expenditures on marketing and advertising (something that Apple has the financial assets to do), trying to expand sales through companies' own-brand retail shops and direct-to-customer online sales (something Apple already does well), and introducing smartphones aimed at the high-end user to compete directly with Apple.

In this chapter we will learn about the important role forecasting plays in supply chain management, and some of the quantitative models, techniques, and technologies that companies in the supply chain use to forecast product demand.

A forecast is a prediction of what will occur in the future. Meteorologists forecast the weather, sportscasters and gamblers predict the winners of football games, and companies attempt to predict how much of their product will be sold in the future. A forecast of product demand is the basis for most important planning decisions. Planning decisions regarding scheduling, inventory, production, facility layout and design, workforce, distribution, purchasing, and so on, are functions of customer demand. Long-range strategic plans by top management are based on forecasts of the type of products consumers will demand in the future and the size and location of product markets.

Forecasting is an uncertain process. It is not possible to predict consistently what the future will be, even with the help of a crystal ball or a deck of tarot cards. Management generally hopes to forecast demand with as much accuracy as possible, which is becoming increasingly difficult to do. In the current international business environment, consumers have more product choices and more information on which to base choices. They also demand and receive greater product diversity, made possible by rapid technological advances. This makes forecasting products and product demand more difficult. Consumers and markets have never been stationary targets, but they are moving more rapidly now than they ever have before.

Companies sometimes use **qualitative forecast methods** based on judgment, opinion, past experience, or best guesses, to make forecasts. A number of **quantitative forecasting methods** are also available to aid management in making planning decisions. In this chapter, we discuss two of the traditional types of mathematical forecasting methods, time series analysis and regression, as well as several nonmathematical, qualitative approaches to forecasting. Although no technique will result in a totally accurate forecast, these methods can provide reliable guidelines in making decisions.

Qualitative forecast methods Subjective methods.

Quantitative forecast methods Are based on mathematical formulas.

The Strategic Role of Forecasting in Supply Chain Management

In today's global business environment, strategic planning and design tend to focus on supply chain management and quality management.

Supply Chain Management

A company's supply chain encompasses all of the facilities, functions, and activities involved in producing a product or service from suppliers (and their suppliers) to customers (and their customers). Supply chain functions include purchasing, inventory, production, scheduling, facility location, transportation, and distribution. All these functions are affected in the short run by product demand and in the long run by new products and processes, technology advances, and changing markets.

Forecasts of product demand determine how much inventory is needed, how much product to make, and how much material to purchase from suppliers to meet forecasted customer needs. This in turn determines the kind of transportation that will be needed and where plants, warehouses, and distribution centers should be located so that products and services can be delivered on time. Without accurate forecasts, large stocks of costly inventory must be kept at each stage of the supply chain to compensate for the uncertainties of customer demand. If there are insufficient inventories, customer service suffers because of late deliveries and stockouts. This is especially hurtful in today's competitive global business environment, where customer service and on-time delivery are critical factors. **Figure 12.1** illustrates the effects of bad forecasting on the supply chain.

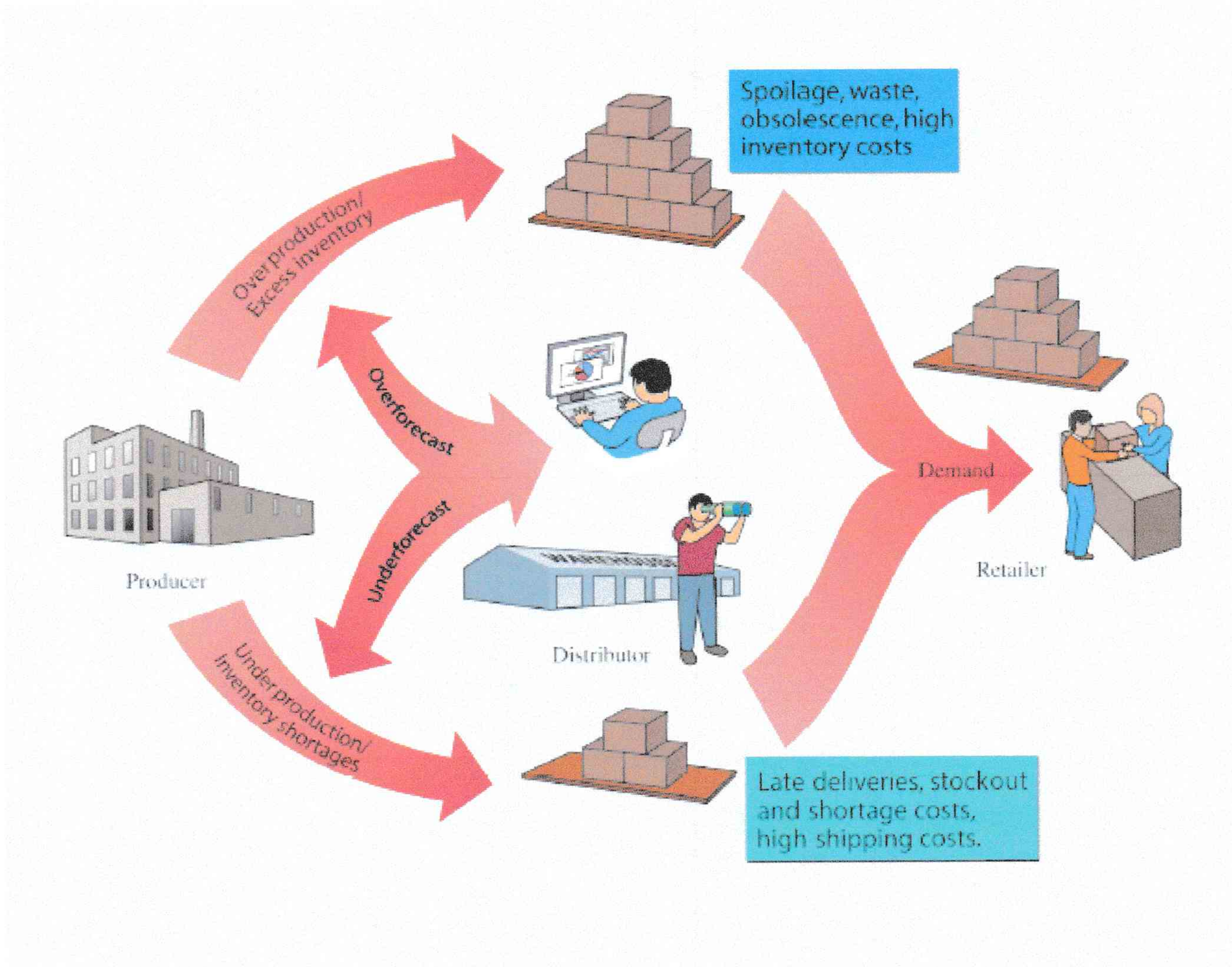


FIGURE 12.1 **The Effect of Inaccurate Forecasting on the Supply Chain**

While accurate forecasts are necessary, completely accurate forecasts are never possible. Hopefully, the forecast will reduce uncertainty about the future as much as possible, but it will never eliminate uncertainty. Thus, all of the supply chain processes need to be flexible to respond to some degree of uncertainty.

In Chapter 10 on supply chain management, we talked about the "bullwhip effect" and its negative impact on the supply chain. The bullwhip effect is the distortion of information about product demand (including forecasts) as it is transmitted back through the supply chain toward suppliers. As demand moves further away from the ultimate end-use consumer, the variation in demand becomes greater and demand forecasts become less reliable. This increased variation can result in excessive, costly safety stock inventories at each stage in the supply chain and poorer customer service.

The bullwhip effect is caused when slight demand variability is magnified as information moves back upstream in the supply chain (see Figure 10.4). It is created when supply chain members make ordering decisions with an eye to their own self-interest and/or they do not have accurate demand forecasts from adjacent supply chain members. If each supply chain member is uncertain and not confident about what the actual demand is for the succeeding member it supplies, and it's making its own demand forecast, then it will stockpile extra inventory to compensate for the uncertainty; that is, the member creates a security blanket of inventory. One way to cope with the bullwhip effect is to develop demand forecasts that will reduce uncertainty and for supply chain members to share these forecasts with one another. Ideally, a single forecast of demand for the final customer in the supply chain would drive the development of subsequent forecasts for each supply chain member back up through the supply chain.

One trend in supply chain design is *continuous replenishment,* wherein continuous updating of data is shared between suppliers and customers. In this system, customers are continuously being replenished, daily or even more often, by their suppliers based on actual sales. Continuous replenishment, typically managed by the supplier, reduces inventory for the company and speeds customer delivery. Variations of continuous replenishment include quick response, just-in-time (JIT), VMI (vendor-managed inventory), SMI (supplier managed inventory), and stockless inventory. Such systems rely heavily on accurate short-term forecasts, usually on a weekly basis, of end-use sales to the ultimate customer. The supplier at one end of a company's supply chain must forecast the company's customer demand at the other end of the supply chain in order to maintain continuous replenishment. The forecast also has to be able to respond to sudden, quick changes in demand. Longer forecasts based on historical sales data for 6 to 12 months into the future are also generally required to help make weekly forecasts and suggest trend changes.

Quality Management

Forecasting is also crucial in a quality management environment. More and more, customers perceive good-quality service to mean having a product when they demand it. This holds true for manufacturing and service companies. When customers walk into a McDonald's to order a meal, they do not expect to wait long to place and receive orders. They expect McDonald's to have the item they want, and they expect to receive their orders within a short period of time. A good forecast of customer traffic flow and product demand enables McDonald's to schedule enough servers, to stock enough food, and to schedule food production to provide high-quality service. An inaccurate forecast causes service to break down, resulting in poor quality. For manufacturing operations, especially for suppliers, customers expect parts to be provided when demanded. Accurately forecasting customer demand is a crucial part of providing high-quality service.

Strategic Planning

There can be no strategic planning without forecasting. The ultimate objective of strategic planning is to determine what the company should be in the future—what markets to compete in, with what products, to be successful and grow. To answer these questions, the company needs to know what new products its customers will want, how much of these products customers will want, and the level of quality and other features that will be expected in these products. Forecasting answers these questions and is a key to a company's long-term competitiveness and success. The determination of future new products and their design subsequently determines process design, the kinds of new equipment and technologies that will be needed, and the design of the supply chain, including the facilities, transportation, and distribution systems that will be required. These elements are ultimately based on the company's forecast of the long-run future.

Along the Supply Chain

Forecasting Denim Jeans Trends

Denim jeans consumers may think that new apparel trends just happen out of the "blue," however, that's rarely the case. In most cases, jeans looks, styles, and colors can be traced back months and often years, and are the result of vast amounts of data and information, sophisticated forecasting methodologies, and expert professional analysis. In the denim jeans industry trends evolve over time, as distinct from fads, which are short term and are driven by extraneous factors like instant celebrity or a music video. Most large-volume denim jeans brands like Levi's, Gap, Lee, and Wrangler, and more fashionable brands like H&M, Calvin Klein, Tommy Hilfiger, and Zara, follow a similar methodology for trend forecasting. They first identify basic fundamental facts about past trends and forecasts such as the colors and styles that have had the greatest demand, then they determine what factors caused changes in the past and compare differences between past forecasts and what actually occurred. Next they determine what factors are most likely to affect future trends, which can be economic and technological changes as well as fashion changes. Factors that go into denim jeans trend forecasting could include new cotton fiber innovations, the price and availability of cotton, advances in manufacturing processes and machinery, shifts in global manufacturing locations, shipping changes, shifting global markets (for example into less economically developed countries), and sustainability issues, in addition to fashion factors like design, style, color, media, blogs, celebrity, and apparel trade shows. The resulting information and data are used in forecasting tools and techniques to develop trend forecasts, with particular attention paid to accuracy and reliability. Once the trend forecast is developed it is closely monitored to determine reasons for significant deviations from what is actually happening, and the forecast is revised as needed. Ultimately it is this trend forecast combined with shorter term and seasonal forecasts that drives a jeans company's global supply chain; and it is sudden changes in fashion trends and customer tastes that alter the forecast and make the denim jeans supply chain so complex and difficult to manage.

Discuss how forecasting trends for an apparel item like denim jeans is different from forecasting trends for an electronic device like a smartphone or tablet.

Source: Based on Randi Gollin, "Trend Forecasters Talk Fall/Winter 2012/2013," Apparel Insiders, www.apparelinsiders.com.

Components of Forecasting Demand

The type of forecasting method to use depends on several factors, including the **time frame** of the forecast (i.e., how far in the future is being forecasted), the *behavior* of demand, and the possible existence of patterns (trends, seasonality, and so on), and the *causes* of demand behavior.

Time frame Indicates how far into the future is forecast.

Time Frame

Forecasts are either short- to mid-range, or long-range. **Short-range (to mid-range) forecasts** are typically for daily, weekly, or monthly sales demand for up to approximately two years into the future, depending on the company and the type of industry. They are primarily used to determine production and delivery schedules and to establish inventory levels. At Hewlett-Packard monthly forecasts for printers are constructed from 12 to 18 months into the future, while at Levi Strauss weekly forecasts for jeans are prepared for five years into the future.

Short- to mid-range forecast Typically encompasses the immediate future—daily up to two years.

A **long-range forecast** is usually for a period longer than two years into the future. A long-range forecast is normally used for strategic planning—to establish long-term goals, plan new products for changing markets, enter new markets, develop new facilities, develop technology, design the supply chain, and implement strategic programs. At Unisys, long-range strategic forecasts project three years into the future; Hewlett-Packard's long-term forecasts are developed for years 2 through 6; and at Fiat, the Italian automaker, strategic plans for new and continuing products go 10 years into the future.

Long-range forecast Usually encompasses a period of time longer than two years.

These classifications are generalizations. The line between short- and long-range forecasts is not always distinct. For some companies a short-range forecast can be several years, and for other firms a long-range forecast can be in terms of months. The length of a forecast depends a lot on how rapidly the product market changes and how susceptible the market is to technological changes.

Demand Behavior

Demand sometimes behaves in a random, irregular way. At other times it exhibits predictable behavior, with trends or repetitive patterns, which the forecast may reflect. The three types of demand behavior are *trends, cycles*, and *seasonal patterns*.

A **trend** is a gradual, long-term up or down movement of demand. For example, the demand for houses has followed an upward trend during the past few decades, without any sustained downward movement in the market. Trends are often the starting points for developing forecasts. Figure 12.2(*a*) illustrates a demand trend in which there is a general upward movement, or increase. Notice that Figure 12.2(*a*) also includes several random movements up and down. **Random variations** are movements that are not predictable and follow no pattern (and thus are virtually unpredictable). They are routine variations that have no "assignable" cause.

Trend A gradual, long-term up or down movement of demand.

Cycle An up-and-down repetitive movement in demand.

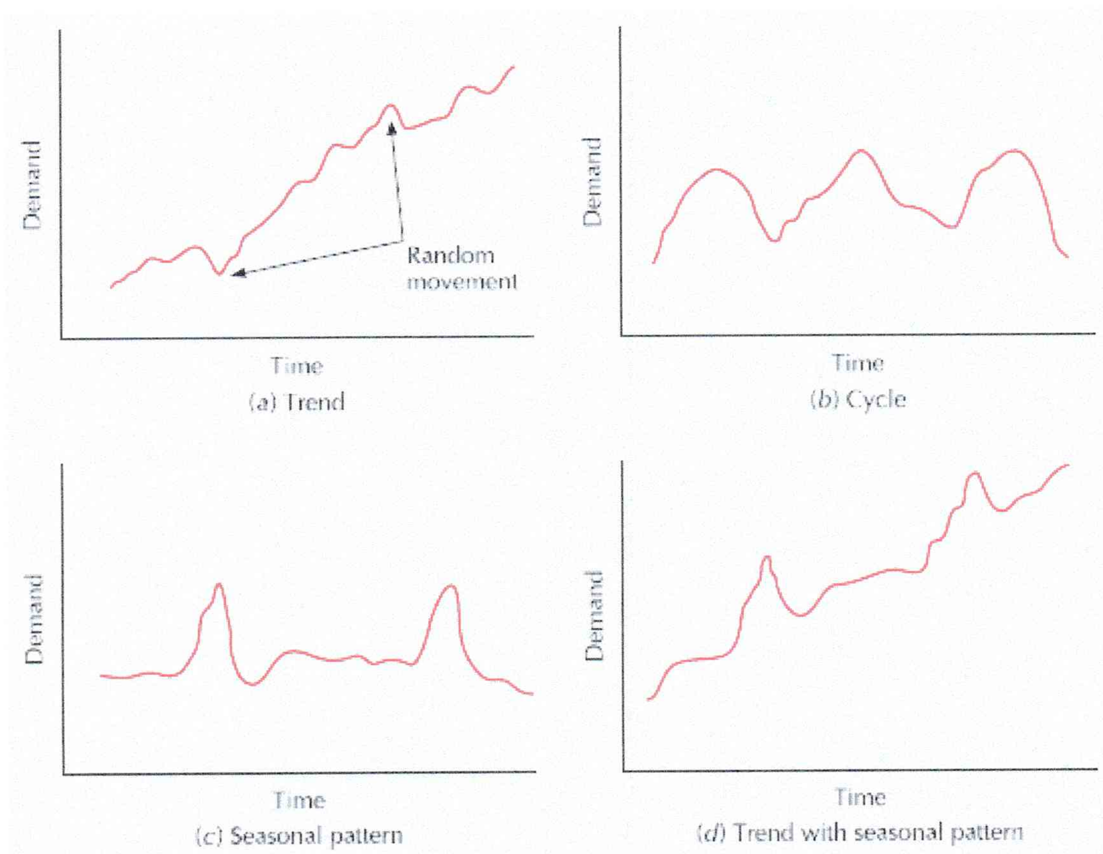


FIGURE 12.2 **Forms of Forecast Movement**

A **cycle** is an up-and-down movement in demand that repeats itself over a lengthy time span (i.e., more than a year). For example, new housing starts and, thus, construction-related products tend to follow cycles in the economy. Automobile sales also tend to follow cycles. The demand for winter sports equipment increases every four years before and after the Winter Olympics. Figure 12.2(*b*) shows the behavior of a demand cycle.

A **seasonal pattern** is an oscillating movement in demand that occurs periodically (in the short run) and is repetitive. Seasonality is often weather-related. For example, every winter the demand for snowblowers and skis increases, and retail sales in general increase during the holiday season. However, a seasonal pattern can occur on a daily or weekly basis. For example, some restaurants are busier at lunch than at dinner, and shopping mall stores and theaters tend to have higher demand on weekends. At FedEx seasonalities include the month of the year, day of the week, and day of the month, as well as various holidays. Figure 12.2(*c*) illustrates a seasonal pattern in which the same demand behavior is repeated each year at the same time.

Seasonal pattern An up-and-down repetitive movement in demand occurring periodically.

Demand behavior frequently displays several of these characteristics simultaneously. Although housing starts display cyclical behavior, there has been an upward trend in new house construction over the years. Demand for skis is seasonal; however, there has been an upward trend in the demand for winter sports equipment during the past two decades. Figure 12.2(*d*) displays the combination of two demand patterns, a trend with a seasonal pattern.

Instances when demand behavior exhibits no pattern are referred to as *irregular movements,* or variations. For example, a local flood might cause a momentary increase in carpet demand, or a competitor's promotional campaign might cause a company's product demand to drop for a time. Although this behavior has a cause and, thus, is not totally random, it still does not follow a pattern that can be reflected in a forecast.

Forecasting Methods

The factors discussed previously in this section determine to a certain extent the type of forecasting method that can or should be used. In this chapter we are going to discuss three basic types of forecasting: *time series methods*, *regression methods*, and *qualitative methods*.

Time series methods are statistical techniques that use historical demand data to predict future demand. **Regression** (or causal) **forecasting methods** attempt to develop a mathematical relationship (in the form of a regression model) between demand and factors that cause it to behave the way it does. Most of the remainder of this chapter will be about time series and regression forecasting methods. In this section we will focus our discussion on qualitative forecasting.

Regression forecasting methods Relate demand to other factors that cause demand behavior.

Qualitative (or judgmental) methods use management judgment, expertise, and opinion to make forecasts. Often called "the jury of executive opinion," they are the most common type of forecasting method for the long-term strategic planning process. There are normally individuals or groups within an organization whose judgments and opinions regarding the future are as valid or more valid than those of outside experts or other structured approaches. Top managers are the key group involved in the development of forecasts for strategic plans. They are generally most familiar with their firms' own capabilities and resources and the markets for their products. The sales force of a company represents a direct point of contact with the consumer. This contact provides an awareness of consumer expectations in the future that others may not possess. Engineering personnel have an innate understanding of the technological aspects of the type of products that might be feasible and likely in the future.

Consumer (or market) *research* is an organized approach using surveys and other research techniques to determine what products and services customers want and will purchase, and to identify new markets and sources of customers. Consumer and market research is normally conducted by the marketing department within an organization, by industry organizations and groups, and by private marketing or consulting firms. Although market research can provide accurate and useful forecasts of product demand, it must be skillfully and correctly conducted, and it can be expensive.

The **Delphi method** is a procedure for acquiring informed judgments and opinions from knowledgeable individuals using a series of questionnaires to develop a consensus forecast about what will occur in the future. It was developed at the Rand Corporation shortly after World War II to forecast the impact of a hypothetical nuclear attack on the United States. Although the Delphi method has been used for a variety of applications, forecasting has been one of its primary uses. It has been especially useful for forecasting technological change and advances.

Delphi method Involves soliciting forecasts about technological advances from experts.

Technological forecasting has become increasingly crucial to compete in the modern international business environment. New enhanced computer technology, new production methods, and advanced machinery and equipment are constantly being made available to companies. These advances enable them to introduce more new products into the marketplace faster than ever before. The companies that succeed manage to get a "technological" jump on their competitors by accurately predicting what technology will be available in the future and how it can be exploited. What new products and services will be technologically feasible, when they can be introduced, and what their demand will be are questions about the future for which answers cannot be predicted from historical data. Instead, the informed opinion and judgment of experts are necessary to make these types of single, long-term forecasts.

Data mining is a recent addition to the library of methods and techniques companies have available for forecasting, brought about by the evolution in information technology. It is a process and set of tools for analyzing large amounts of data in order to identify patterns, trends, and relationships among and between groups of customers, markets, and products. Data mining is made possible by the vast amounts of data that companies now have available to them from various electronic transactions throughout their supply chains, and the ability to store this amount of data inexpensively.

Data mining Process and set of tools for analyzing large amounts of data.

Forecasting Process

Forecasting is not simply identifying and using a method to compute a numerical estimate of what demand will be in the future. It is a continuing process that requires constant monitoring and adjustment illustrated by the steps in **Figure 12.3**.

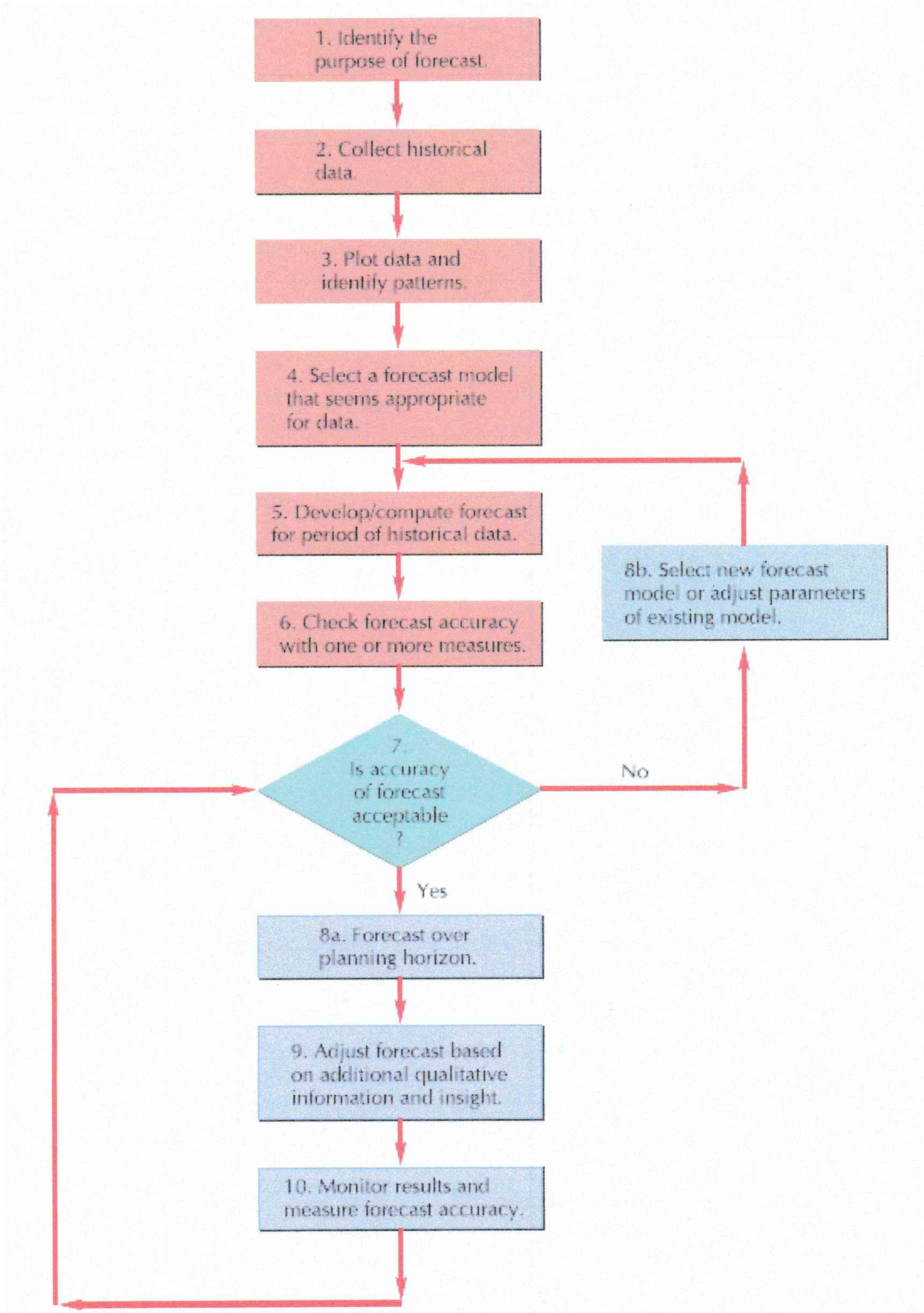


FIGURE 12.3 **Steps of the Forecasting Process**

Along the Supply Chain

Global Supply Chain Forecasting at Hershey's

Global companies such as Hershey's have numerous opportunities to use forecasting along various points in its supply chain. Downstream in its supply chain Hershey's attempts to forecast product demand, which is subject to uncertainties resulting from new and current products from competitors, competitors' promotional events and price changes, and changing consumer tastes for its own current and new products. Issues related to product quality and safety, ingredients, or packaging could adversely affect demand. Negative publicity related to product recalls due to contamination or product tampering, whether valid or not, might also negatively impact product demand. All of these factors affect the forecasting process. Upstream in its supply chain, Hershey's sources many different commodities including cocoa products, sugar, dairy products, peanuts, almonds, corn sweeteners, natural gas, and fuel oil. Commodities are subject to price volatility and changes in supply caused by numerous factors, including commodity market fluctuations and speculative influences; currency exchange rates; the effect of weather on crop yield and distribution channels; trade agreements among producing and consuming nations; political unrest in producing countries; and changes in governmental agricultural programs and energy policies. Other factors that can create uncertainties in its global supply chain and operations that make forecasting difficult include global economic and environmental changes that can result in interruptions in supply and decreased demand for its products overseas; changes in tariff and trade agreements; political instability; nationalization of Hershey's properties; and disruptions in shipping or reduced availability of freight transportation.

Given the various uncertainties inherent in the forecasting process at Hershey's, what kind of forecasting models might you suggest for the company?

Source: Hershey's website at www.thehersheycompany.com

In the next few sections we present several different forecasting methods applicable for different patterns of demand behavior. Thus, one of the first steps in the forecasting process is to plot the available historical demand data and, by looking at them, to attempt to determine the forecasting method that best seems to fit the patterns the data exhibit. Historical demand is usually past sales or orders data. There are several measures for comparing historical demand with the forecast to see how accurate the forecast is. Following our discussion of the forecasting methods, we present several measures of forecast accuracy. If the forecast does not seem to be accurate, another method can be tried until an accurate forecast method is identified. After the forecast is made over the desired planning horizon, it may be possible to use judgment, experience, knowledge of the market, or even intuition to adjust the forecast to enhance its accuracy. Finally, as demand actually occurs over the planning period, it must be monitored and compared with the forecast in order to assess the performance of the forecast method. If the forecast is accurate, then it is appropriate to continue using the forecast method. If it is not accurate, a new model or adjusting the existing one should be considered.

Moving Average

Time series methods are statistical techniques that make use of historical data accumulated over a period of time. Time series methods assume that what has occurred in the past will continue to occur in the future. As the name *time series* suggests, these methods relate the forecast to only one factor—time. These methods assume that identifiable historical patterns or trends for demand over time will repeat themselves. They include the moving average, exponential smoothing, and linear trend line, and they are among the most popular methods for short-range forecasting among service and manufacturing companies. Time series is the most popular forecasting method by far. It's likely that the majority of businesses have used time series to some extent. One of the reasons time series models are so popular is that they are relatively easy to understand and use. The survey also showed that the most popular time series models are moving averages and exponential smoothing.

Time series methods Use historical demand data over a period of time to predict future demand.

A time series forecast can be as simple as using demand in the current period to predict demand in the next period. This is sometimes called a *naive* or *intuitive* forecast. For example, if demand is 100 units this week, the forecast for next week's demand is 100 units; if demand turns out to be 90 units instead, then the following week's demand is 90 units, and so on. This type of forecasting method does not take into account historical demand *behavior*; it relies only on demand in the current period. It reacts directly to the normal, random movements in demand.

The simple **moving average** method uses several demand values during the recent past to develop a forecast. This tends to *dampen*, or *smooth out,* the random increases and decreases of a forecast that uses only one period. The simple moving average is useful for forecasting demand that is stable and does not display any pronounced demand behavior, such as a trend or seasonal pattern.

Moving average Method uses average demand for a fixed sequence of periods.

Moving averages are computed for specific periods, such as three months or five months, depending on how much the forecaster desires to "smooth" the demand data. The longer the moving average period, the smoother it will be. (Alternatively, a shorter moving average is more susceptible to simple random variation.) The formula for computing the simple moving average is

$$MA_n \frac{\sum_{i=1}^{n} D_i}{n}$$

where

n = number of periods in the moving average

D_i = demand in period i

The disadvantage of the moving average method is that it does not react to variations that occur for a reason, such as cycles and seasonal effects. Factors that cause changes are generally ignored. It is basically a "mechanical" method that reflects historical data in a consistent way. However, the moving average method does have the advantage of being easy to use, quick, and relatively inexpensive. In general, this method can provide a good forecast for the short run, but it should not be pushed too far into the future.

EXAMPLE 12.1 | Computing a Simple Moving Average

The Heartland Produce Company sells and delivers food produce to restaurants and catering services within a 100-mile radius of its warehouse. The food supply business is competitive, and the ability to deliver orders promptly is a factor in getting new customers and keeping old ones. The manager of the company wants to be certain enough drivers and vehicles are available to deliver orders promptly and they have adequate inventory in stock. Therefore, the manager wants to be able to forecast the number of orders that will occur during the next month (i.e., to forecast the demand for deliveries).

From records of delivery orders, management has accumulated the following data for the past 10 months, from which it wants to compute three- and five-month moving averages.

MONTH	ORDERS
January	120
February	90
March	100
April	75
May	110
June	50

July	75
August	130
September	110
October	90

Solution: Let us assume that it is the end of October. The forecast resulting from either the three- or five-month moving average is typically for the next month in the sequence, which in this case is November. The moving average is computed from the demand for orders for the prior three months in the sequence according to the following formula:

$$\begin{aligned} MA_3 &= \frac{\sum_{i=1}^{3} D_i}{3} \\ &= \frac{90+110+130}{3} \\ &= 110 \text{ orders for November} \end{aligned}$$

The five-month moving average is computed from the prior five months of demand data as follows:

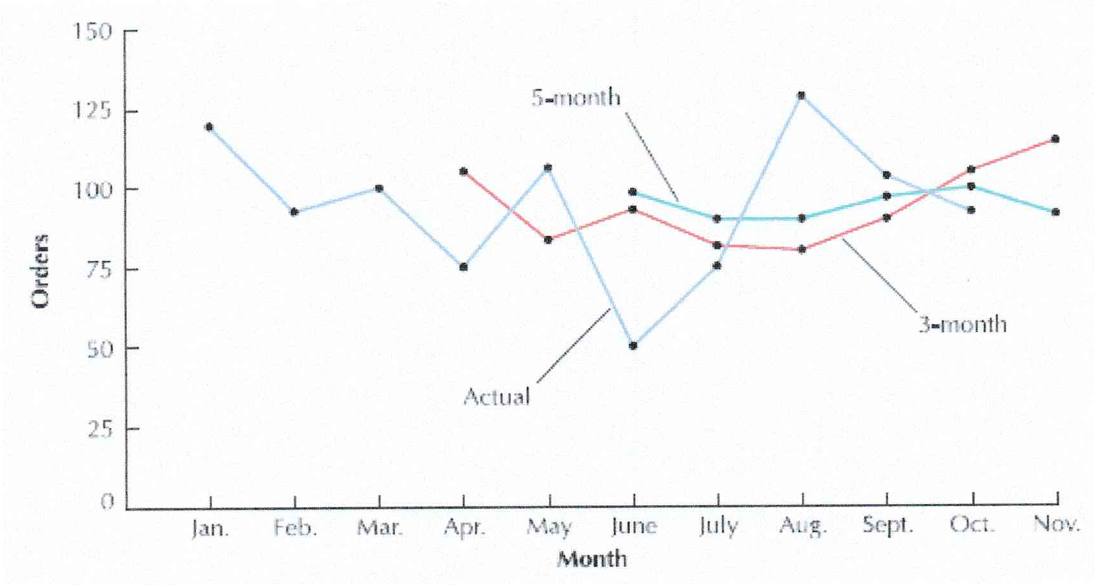
$$\begin{aligned} MA_5 &= \frac{\sum_{i=1}^{5} D_i}{5} \\ &= \frac{90+110+130+75+50}{5} \\ &= 91 \text{ orders for November} \end{aligned}$$

The three- and five-month moving average forecasts for all the months of demand data are shown in the following table. Actually, the manager would use only the forecast for November based on the most recent monthly demand. However, the earlier forecasts for prior months allow us to compare the forecast with actual demand to see how accurate the forecasting method is—that is, how well it does.

Three- and Five-Month Averages

MONTH	ORDERS PER MONTH	THREE-MONTH MOVING AVERAGE	FIVE-MONTH MOVING AVERAGE
January	120	—	—
February	90	—	—
March	100	—	—
April	75	103.3	—
May	110	88.3	—
June	50	95.0	99.0
July	75	78.3	85.0
August	130	78.3	82.0
September	110	85.0	88.0
October	90	105.0	95.0
November	—	110.0	91.0

Both moving average forecasts in the preceding table tend to smooth out the variability occurring in the actual data. This smoothing effect can be observed in the following figure in which the three-month and five-month averages have been superimposed on a graph of the original data:



The five-month moving average in the previous figure smooths out fluctuations to a greater extent than the three-month moving average. However, the three-month average more closely reflects the most recent data available to the produce company manager. In general, forecasts using the longer-period moving average are slower to react to recent changes in demand than those made using shorter-period moving averages. The extra periods of data dampen the speed with which the forecast responds. Establishing the appropriate number of periods to use in a moving average forecast often requires some amount of trial-and-error experimentation.

Weighted Moving Average

The moving average method can be adjusted to more closely reflect fluctuations in the data. In the **weighted moving average** method, weights are assigned to the most recent data according to the following formula:

Weighted moving average Weights are assigned to the most recent data.

$$WMA_n = \sum_{i=1}^{n} W_i D_i$$

where

$$W_i = \text{the weight for period } i\text{, between 0 and 100 percent}$$
$$\sum w_i = 1.00$$

Determining the precise weights to use for each period of data usually requires some trial-and-error experimentation, as does determining the number of periods to include in the moving average. If the most recent periods are weighted too heavily, the forecast might overreact to a random fluctuation in demand. If they are weighted too lightly, the forecast might underreact to actual changes in demand behavior.

EXAMPLE 12.2 | Computing a Weighted Moving Average

The Heartland Produce Company in Example 12.1 wants to compute a three-month weighted moving average with a weight of 50% for the October data, a weight of 33% for the September data, and a weight of 17% for the August data. These weights reflect the company's desire to have the most recent data influence the forecast most strongly.

Solution: The weighted moving average is computed as

$$\begin{aligned} WMA_3 &= \sum_{i=1}^{3} W_i D_i \\ &= (0.50)(90) + (0.33)(110) + (0.17)(130) \\ &= 103.4 \text{ orders} \end{aligned}$$

Notice that the forecast includes a fractional part, 0.4. In general, the fractional parts need to be included in the computation to achieve mathematical accuracy, but when the final forecast is achieved, it must be rounded up or down.

This forecast is slightly lower than our previously computed three-month average forecast of 110 orders, reflecting the lower number of orders in October (the most recent month in the sequence).

Along the Supply Chain

Demand Forecasts for Solar Energy Development at GE

General Electric (GE) Energy is a global leader in energy technology involved in all areas of the energy industry, including conventional energy sources, coal, oil, natural gas, nuclear energy, and renewable sources such as water, wind, solar, biogas, and other alternative fuels. GE has received support from the U.S. Department of Energy to produce electricity generated from solar (photoelectric) cells that convert sunlight into electricity, as shown in photo, and make its generation and distribution cost competitive with conventional sources of electricity. GE Energy expects its emerging solar business to grow to over $1 billion in revenues by mid-decade and for it to continue to grow rapidly in the future. However, the solar industry is in an embryonic stage and thus requires GE to not only engage in R&D to develop solar cell technology, but to also establish partnerships with companies to develop a manufacturing and distribution infrastructure to install solar energy systems for commercial and utility consumers. This requires a significant capital investment for manufacturing solar power equipment for high-volume consumer usage, which requires estimating costs and forecasting solar energy demand in order to make good investment decisions in the face of uncertainties in technology, costs, demand, and energy policy.

A key aspect of this development process for GE is being able to forecast solar energy demand by year, region of the country, and market type over a 10- to 15-year period, as well as research and development costs. GE's forecasting model for solar energy demand first forecasts the potential in megawatts that can physically be installed on rooftops by region of the country, using existing data for available floor space. GE then estimates potential revenues generated by an average rooftop installation, which when compared with installation costs enables GE to determine the payback period for its investment. These forecasts are then used in combination with previously developed solar energy market penetration functions to determine market penetration in megawatts as a function of payback years.

Solar energy is an emerging technology; discuss how GE might use "subjective" expert opinions as a forecasting tool.

Source: Based on Bex G. Thomas and Srinivas Bollapragada, "General Electric Uses an Integrated Framework for Product Costing, Demand Forecasting, and Capacity Planning of New Photovoltaic Technology Products," Interfaces, 40 (5) (September–October 2010), pp. 353–367.

Solar parks are becoming increasingly popular, especially in Europe. Although, solar power accounts for less than 1 percent of all electricity generation worldwide, it is growing at an annual rate above 50 percent, which makes forecasting solar demand for energy companies critically important.

Exponential Smoothing

Exponential smoothing is also an averaging method that weights the most recent data more strongly. As such, the forecast will react more to recent changes in demand. This is useful if the recent changes in the data are significant and unpredictable instead of just random fluctuations (for which a simple moving average forecast would suffice).

Exponential smoothing An averaging method that reacts more strongly to recent changes in demand.

Exponential smoothing is one of the more popular and frequently used forecasting techniques, for a variety of reasons. Exponential smoothing requires minimal data. Only the forecast for the current period, the actual demand for the current period, and a weighting factor called a smoothing constant are necessary. The mathematics of the technique are easy for management to understand. Virtually all forecasting computer software packages include modules for exponential smoothing. Most importantly, exponential smoothing has a good track record of success. It has been employed over the years by many companies that have found it to be an accurate method of forecasting.

The exponential smoothing forecast is computed using the formula

$$F_{t+1} = \alpha D_t + (1 - \alpha) F_t$$

where

F_{t+1} = the forecast for the next period
D_t = actual demand in the present period
F_t = the previously determined forecast for the present period
α = a weighting factor referred to as the **smoothing constant**

Smoothing constant The weighting factor given to the most recent data in exponential smoothing forecasts.

The smoothing constant, α, is between 0.0 and 1.0. It reflects the weight given to the most recent demand data. For example, if α = 20,

$$F_{t+1} = 0.20D_t + 0.80F_t$$

which means that our forecast for the next period is based on 20% of recent demand (D_t) and 80% of past demand (in the form of forecast F_t, since F_t is derived from previous demands and forecasts). If we go to one extreme and let α= 0.0, then

$$\begin{aligned} F_{t+1} &= 0D_t + 1F_t \\ &= F_t \end{aligned}$$

and the forecast for the next period is the same as the forecast for this period. In other words, *the forecast does not reflect the most recent demand at all.*

On the other hand, if α = 10, then

$$\begin{aligned} F_{t+1} &= 1D_t + 0F_t \\ &= 1D_t \end{aligned}$$

and we have considered only the most recent data (demand in the present period) and nothing else. Thus, the higher α is, the more sensitive the forecast will be to changes in recent demand, and the smoothing will be less. The closer α is to zero, the greater will be the dampening, or smoothing, effect. As α approaches zero, the forecast will react and adjust more slowly to differences between the actual demand and the forecasted demand. The most commonly used values of α are in the range of .01 to .50. However, the determination of α is usually judgmental and subjective and is often based on trial-and-error experimentation. An inaccurate estimate of α can limit the usefulness of this forecasting technique. (As α approaches 1.0, the forecast is the same as the naive result.)

EXAMPLE 12.3 | Computing an Exponentially Smoothed Forecast

HiTek Computer Services repairs and services personal computers at its store, and it makes local service calls. It primarily uses part-time State University students as technicians. The company has had steady growth since it started. It purchases generic computer parts in volume at a discount from a variety of sources whenever it sees a good deal. Thus, they need a good forecast of demand for repairs so that they will know how many computer component parts to purchase and stock, and how many technicians to hire.

The company has accumulated the demand data shown in the accompanying table for repair and service calls for the past 12 months, from which it wants to consider exponential smoothing forecasts using smoothing constants (α) equal to .30 and .50.

Demand for Repair and Service Calls

PERIOD	MONTH	DEMAND	PERIOD	MONTH	DEMAND
1	January	37	7	July	43
2	February	40	8	August	47
3	March	41	9	September	56
4	April	37	10	October	52
5	May	45	11	November	55
6	June	50	12	December	54

Solution: To develop the series of forecasts for the data in this table, we will start with period 1 (January) and compute the forecast for period 2 (February) using $\alpha = .30$. The formula for exponential smoothing also requires a forecast for period 1, which we do not have, so we will use the demand for period 1 as both *demand* and *forecast* for period 1. (Other ways to determine a starting forecast include averaging the first three or four periods or making a subjective estimate.) Thus, the forecast for February is

$$\begin{aligned} F_2 &= \alpha D_1 + (1-\alpha)\,F_1 \\ &= (0.30)\,(37) + (0.70)\,(37) \\ &= 37 \text{ service calls} \end{aligned}$$

The forecast for period 3 is computed similarly:

$$\begin{aligned} F_3 &= \alpha D_2 + (1-\alpha)\,F_2 \\ &= (0.30)\,(40) + (0.70)\,(37) \\ &= 37.9 \text{ service calls} \end{aligned}$$

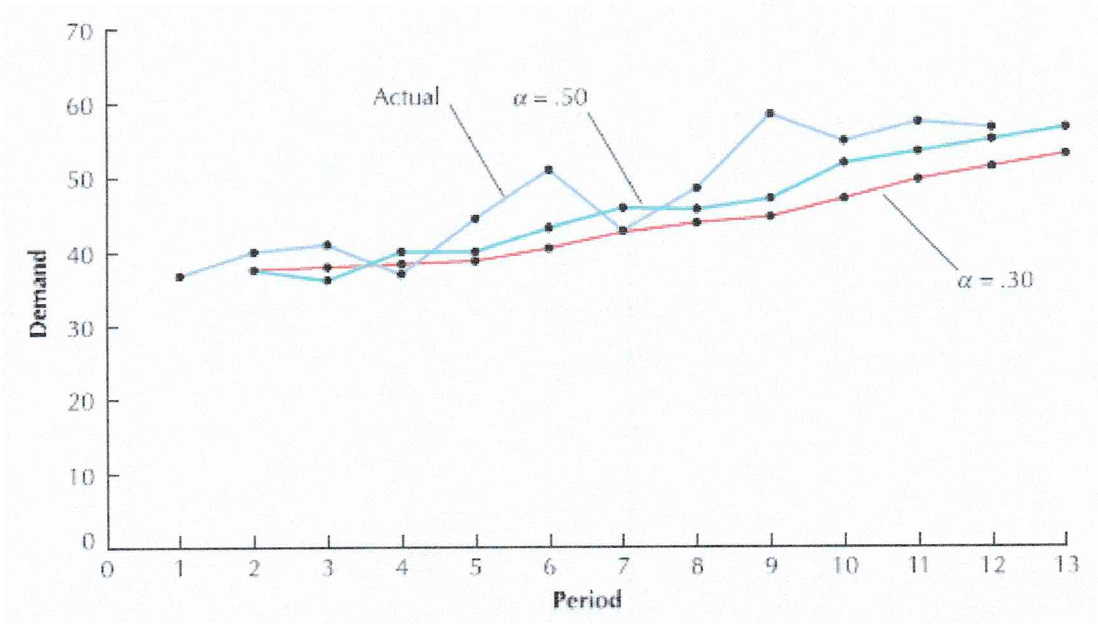
The remainder of the monthly forecasts are shown in the following table. The final forecast is for period 13, January, and is the forecast of interest to HiTek:

$$\begin{aligned} F_{13} &= \alpha D_{12} + (1-\alpha) F_{12} \\ &= (0.30)(54) + (0.70)(50.84) \\ &= 51.79 \text{ service calls} \end{aligned}$$

Exponential Smoothing Forecasts, α = .30 and α = .50

			FORECAST, F_{t+1}	
PERIOD	MONTH	DEMAND	$\alpha = .30$	$\alpha = .50$
1	January	37	—	—
2	February	40	37.00	37.00
3	March	41	37.90	38.50
4	April	37	38.83	39.75
5	May	45	38.28	38.37
6	June	50	40.29	41.68
7	July	43	43.20	45.84
8	August	47	43.14	44.42
9	September	56	44.30	45.71
10	October	52	47.81	50.85
11	November	55	49.06	51.42
12	December	54	50.84	53.21
13	January	—	51.79	53.61

This table also includes the forecast values using α = .50. Both exponential smoothing forecasts are shown in the following figure together with the actual data.



In Example 12.3, the forecast using the higher smoothing constant, α = .51, seems to react more strongly to changes in demand than does the forecast with α = .30, although both smooth out the random fluctuations in the forecast. Notice that both forecasts lag behind the actual demand. For example, a pronounced downward change in demand in July is not reflected in the forecast until August. If these changes mark a change in trend (i.e., a long-term upward or downward movement) rather than just a random fluctuation, then the forecast will always lag behind this trend. We can see a general upward trend in service calls throughout the year. Both forecasts tend to be consistently lower than the actual demand; that is, the forecasts lag the trend.

Based on simple observation of the two forecasts in Example 12.3, $\alpha = .50$ seems to be the more accurate of the two in the sense that it seems to follow the actual data more closely. (Later in this chapter we discuss several quantitative methods for determining forecast accuracy.) When demand is relatively stable without any trend, a small value for α is more appropriate to simply smooth out the forecast. When actual demand displays an increasing (or decreasing) trend, as is the case in the figure, a larger value of α is better. It will react more quickly to more recent upward or downward movements in the actual data. In some approaches to exponential smoothing, the accuracy of the forecast is monitored in terms of the difference between the actual values and the forecasted values. If these differences become larger, then α is changed (higher or lower) in an attempt to adapt the forecast to the actual data. However, the exponential smoothing forecast can also be adjusted for the effects of a trend.

In Example 12.3, the final forecast computed was for one month, January. A forecast for two or three months could have been computed by grouping the demand data into the required number of periods and then using these values in the exponential smoothing computations. For example, if a three-month forecast were needed, demand for January, February, and March could be summed and used to compute the average forecast for the next three-month period, and so on, until a final three-month forecast results. Alternatively, if a trend is present, the final period forecast can be used for an extended forecast by adjusting it by a trend factor.

Adjusted Exponential Smoothing

The **adjusted exponential smoothing forecast** consists of the exponential smoothing forecast with a trend adjustment factor added to it:

Adjusted exponential smoothing forecast An exponential smoothing forecast with an adjustment for a trend added to it.

$$AF_{t+1} = F_{t+1} + T_{t+1}$$

where

T = an exponentially smoothed trend factor

The trend factor is computed much the same as the exponentially smoothed forecast. It is, in effect, a forecast model for trend:

$$T_{t+1} = \beta\left(F_{t+1} - F_t\right) + (1 - \beta)\,T_t$$

where

$$\begin{aligned} T_1 &= \text{the last period's trend factor} \\ \beta &= \text{a smoothing constant for trend} \end{aligned}$$

β is a value between 0.0 and 1.0. It reflects the weight given to the most recent trend data. β is usually determined subjectively based on the judgment of the forecaster. A high β reflects trend changes more than a low β. It is not uncommon for β to equal α in this method.

Notice that this formula for the trend factor reflects a weighted measure of the increase (or decrease) between the next period forecast, F_{t+1}, and the current forecast, F_t.

EXAMPLE 12.4 | Computing an Adjusted Exponentially Smoothed Forecast

HiTek Computer Services now wants to develop an adjusted exponentially smoothed forecast using the same 12 months of demand shown in the table for Example 12.3. It will use the exponentially smoothed forecast with $\alpha = .5$ computed in Example 12.3 with a smoothing constant for trend, β, of .30.

Solution: The formula for the adjusted exponential smoothing forecast requires an initial value for T_t to start the computational process. This initial trend factor is often an estimate determined subjectively or based on past data by the forecaster. In this case, since we have a long sequence of demand data (i.e., 12 months), we will start with the trend T_t equal to zero. By the time the forecast value of interest F_{13} is computed, we should have a relatively good value for the trend factor.

The adjusted forecast for February, AF_2, is the same as the exponentially smoothed forecast, since the trend computing factor will be zero (i.e., F_1 and F_2 are the same and $T_2 = 0$). Thus, we compute the adjusted forecast for March, AF_3, as follows, starting with the determination of the trend factor, T_3:

$$\begin{aligned} T_3 &= \beta(F_3 - F_2) + (1-\beta)T_2 \\ &= (0.30)(38.5 - 37.0) + (0.70)(0) \\ &= 0.45 \end{aligned}$$

and

$$\begin{aligned} AF_3 &= F_3 + T_3 \\ &= 38.5 + 0.45 \\ &= 38.95 \text{ service cells} \end{aligned}$$

This adjusted forecast value for period 3 is shown in the accompanying table, with all other adjusted forecast values for the 12-month period plus the forecast for period 13, computed as follows:

$$\begin{aligned} T_{13} &= \beta(F_{13} - F_{12}) + (1-\beta)T_{12} \\ &= (0.30)(53.61 - 53.21) + (0.70)(1.77) \\ &= 1.36 \end{aligned}$$

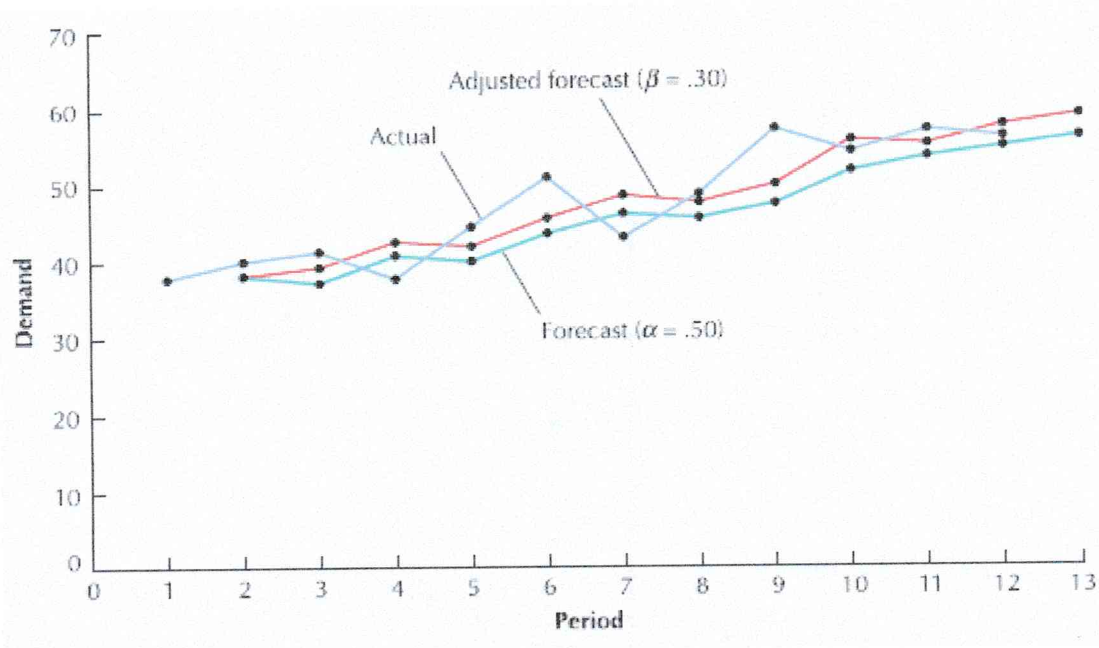
and

$$\begin{aligned} AF_{13} &= F_{13} + T_{13} \\ &= 53.61 + 1.36 \\ &= 54.97 \text{ service calls} \end{aligned}$$

Adjusted Exponential Smoothing Forecast Values

PERIOD	MONTH	DEMAND	FORECAST F_{t+1}	TREND T_{t+1}	ADJUSTED FORECAST AF_{t+1}
1	January	37	37.00	—	—
2	February	40	37.00	0.00	37.00
3	March	41	38.50	0.45	38.95
4	April	37	39.75	0.69	40.44
5	May	45	38.37	0.07	38.44
6	June	50	41.68	1.04	42.73
7	July	43	45.84	1.97	47.82
8	August	47	44.42	0.95	45.37
9	September	56	45.71	1.05	46.76
10	October	52	50.85	2.28	53.13
11	November	55	51.42	1.76	53.19
12	December	54	53.21	1.77	54.98
13	January	—	53.61	1.36	54.97

The adjusted exponentially smoothed forecast values shown in the above table are compared with the exponentially smoothed forecast values and the actual data in the figure. Notice that the adjusted forecast is consistently higher than the exponentially smoothed forecast and is thus more reflective of the generally increasing trend of the actual data. However, in general, the pattern, or degree of smoothing, is very similar for both forecasts.


Adjusted forecast (β = .30)
Actual
Forecast (α = .50)
Demand
Period
70
60
50
40
30
20
10
0
0
1
2
3
4
5
6
7
8
9
10
11
12
13

Linear Trend Line

Linear regression is a method of forecasting in which a mathematical relationship is developed between demand and some other factor that causes demand behavior. However, when demand displays an obvious trend over time, a least squares regression line, or **linear trend line**, that relates demand to time can be used to forecast demand.

Linear trend line A linear regression model relating demand to time.

A linear trend line relates a dependent variable, which for our purposes is demand, to one independent variable, time, in the form of a linear equation:

$$y = a + bx$$

where

$$\begin{aligned} a &= \text{intercept (at period 0)} \\ b &= \text{slope of the line} \\ x &= \text{the time period} \\ y &= \text{forecast for demand for period } x \end{aligned}$$

These parameters of the linear trend line can be calculated using the least squares formulas for linear regression:

$$\begin{aligned} b &= \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2} \\ a &= \bar{y} - b\bar{x} \end{aligned}$$

where

$$\begin{aligned} n &= \text{number of periods} \\ \bar{x} &= \frac{\sum x}{n} = \text{the mean of the } x \text{ values} \\ \bar{y} &= \frac{\sum y}{n} = \text{the mean of the } y \text{ values} \end{aligned}$$

EXAMPLE 12.5 | Computing a Linear Trend Line

The data for HiTek Computer Services (shown in the table for Example 12.3) appears to follow an increasing linear trend. The company wants to compute a linear trend line to see if it is more accurate than the exponential smoothing and adjusted exponential smoothing forecasts developed in Example 12.3 and 12.4.

Solution: The values required for the least squares calculations are as follows:

Least Squares Calculations

x (PERIOD)	y (DEMAND)	xy	x^2
1	37	37	1
2	40	80	4
3	41	123	9
4	37	148	16
5	45	225	25
6	50	300	36
7	43	301	49
8	47	376	64
9	56	504	81
10	52	520	100
11	55	605	121
12	54	648	144
78	557	3867	650

Using these values, we can compute the parameters for the linear trend line as follows:

$$
\begin{aligned}
\bar{x} &= \frac{78}{12} = 6.5 \\
\bar{y} &= \frac{557}{12} = 46.42 \\
b &= \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2} \\
&= \frac{3867 - (12)(6.5)(46.42)}{650 - 12(6.5)^2} \\
&= 1.72 \\
a &= \bar{y} - b\bar{x} \\
&= 46.42 - (1.72)(6.5) \\
&= 35.2
\end{aligned}
$$

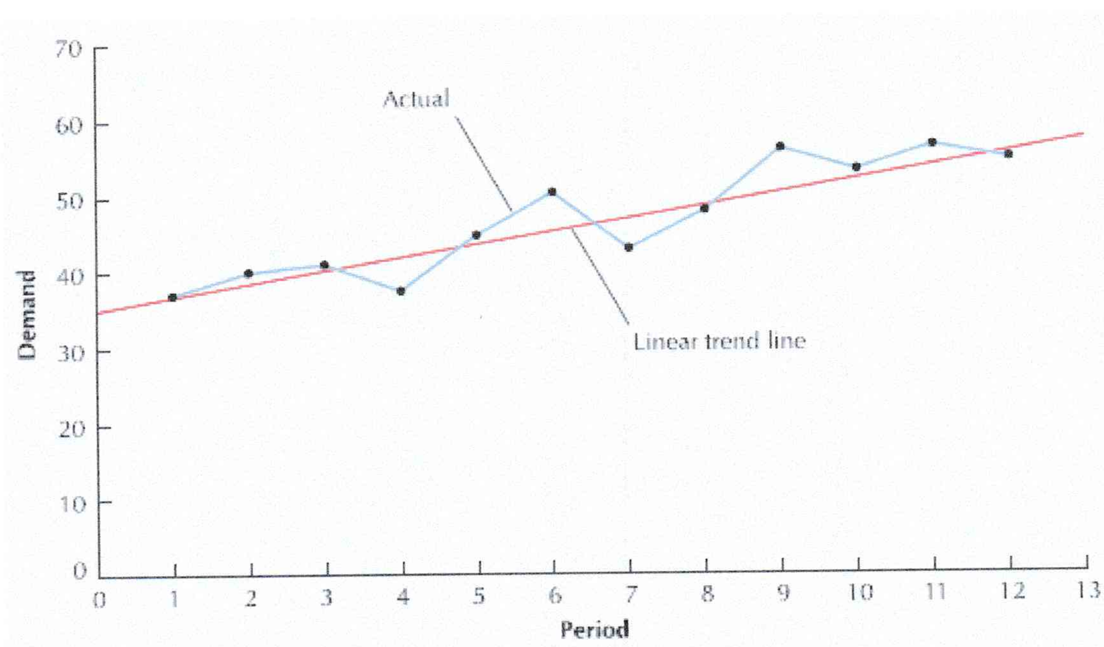
Therefore, the linear trend line equation is

$$y = 35.2 + 1.72x$$

To calculate a forecast for period 13, let $x = 13$ in the linear trend line:

$$
\begin{aligned}
y &= 35.2 + 1.72\,(13) \\
&= 57.56 \text{ service calls}
\end{aligned}
$$

The graph shows the linear trend line compared with the actual data. The trend line appears to reflect closely the actual data—that is, to be a good fit—and would thus be a good forecast model for this problem. However, a disadvantage of the linear trend line is that it will not adjust to a change in the trend, as the exponential smoothing forecast methods will; that is, it is assumed that all future forecasts will follow a straight line. This limits the use of this method to a shorter time frame in which you can be relatively certain that the trend will not change.



Seasonal Adjustments

A seasonal pattern is a repetitive increase and decrease in demand. Many demand items exhibit seasonal behavior. Clothing sales follow annual seasonal patterns, with demand for warm clothes increasing in the fall and winter and declining in the spring and summer as the demand for cooler clothing increases. Another example is the snow skiing industry as shown in photo. Demand for many retail items, including toys, sports equipment, clothing, electronic appliances, hams, turkeys, wine, and fruit, increases during the holiday season. Greeting card demand increases in conjunction with special days such as Valentine's Day and Mother's Day. Seasonal patterns can also occur on a monthly, weekly, or even daily basis. Some restaurants have higher demand in the evening than at lunch or on weekends as opposed to weekdays. Traffic—hence sales—at shopping malls picks up on Friday and Saturday.

PhotoDisc/Getty Images, Inc.

Snow skiing is an industry that exhibits several different patterns of demand behavior. It is primarily a seasonal (i.e., winter) industry and over a long period of time has exhibited a generally increasing growth trend. Random factors can cause variations, or abrupt peaks and valleys, in demand. For example, demand for skiing products always shows a pronounced increase after the Winter Olympics.

There are several methods for reflecting seasonal patterns in a time series forecast. We will describe one of the simpler methods using a *seasonal factor*. A **seasonal factor** is a numerical value that is multiplied by the normal forecast to get a seasonally adjusted forecast.

Seasonal factor Adjust for seasonality by multiplying the normal forecast by a seasonal factor.

One method for developing a demand for seasonal factors is to divide the demand for each seasonal period by total annual demand, according to the following formula:

$$S_i = \frac{D_i}{\sum D}$$

The resulting seasonal factors between 0 and 1.0 are, in effect, the portion of total annual demand assigned to each season. These seasonal factors are multiplied by the annual forecasted demand to yield adjusted forecasts for each season.

EXAMPLE 12.6 | COMPUTING A FORECAST WITH SEASONAL ADJUSTMENTS

Wishbone Farms grows turkeys to sell to a meat-processing company throughout the year, as seen in photo. However, its peak season is obviously during the fourth quarter of the year, from October to December. Wishbone Farms has experienced the demand for turkeys for the past three years shown in the following table:

Demand for Turkeys at Wishbone Farms

Virtual Tours

	DEMAND (1000S) PER QUARTER				
YEAR	1	2	3	4	TOTAL
2014	12.6	8.6	6.3	17.5	45.0
2015	14.1	10.3	7.5	18.2	50.1
2016	15.3	10.6	8.1	19.6	53.6
Total	42.0	29.5	21.9	55.3	148.7

Solution: Because we have three years of demand data, we can compute the seasonal factors by dividing total quarterly demand for the three years by total demand across all three years:

$$S_1 = \frac{D_1}{\sum D} = \frac{42.0}{148.7} = 0.28$$
$$S_2 = \frac{D_2}{\sum D} = \frac{29.5}{148.7} = 0.20$$
$$S_3 = \frac{D_3}{\sum D} = \frac{21.9}{148.7} = 0.15$$
$$S_4 = \frac{D_4}{\sum D} = \frac{55.3}{148.7} = 0.37$$

Next, we want to multiply the forecasted demand for the next year, 2014, by each of the seasonal factors to get the forecasted demand for each quarter. To accomplish this, we need a demand forecast for 2017. In this case, since the demand data in the table seem to exhibit a generally increasing trend, we compute a linear trend line for the three years of data in the table to get a rough forecast estimate:

$$\begin{aligned} y &= 40.97 + 4.30x \\ &= 40.97 + 4.30\,(4) \\ &= 58.17 \end{aligned}$$

Thus, the forecast for 2017 is 58.17, or 58,170 turkeys.

Using this annual forecast of demand, we find that the seasonally adjusted forecasts, SF_i, for 2017 are

$$\begin{aligned} SF_1 &= (S_1)\,(F_5) = (0.28)\,(58.17) = 16.28 \\ SF_2 &= (S_2)\,(F_5) = (0.20)\,(58.17) = 11.63 \\ SF_3 &= (S_3)\,(F_5) = (0.15)\,(58.17) = 8.73 \\ SF_4 &= (S_4)\,(F_5) = (0.37)\,(58.17) = 21.53 \end{aligned}$$

Comparing these quarterly forecasts with the actual demand values in the table, we see that they would seem to be relatively good forecast estimates, reflecting both the seasonal variations in the data and the general upward trend.

Along the Supply Chain

Demand Forecasting for Global Distribution at Zara

Zara is a subsidiary of the Inditex Group, one of the world's largest clothing retailers, that includes over 100 companies involved in textile design, manufacturing, and distribution. Zara is one of the world's most recognized and successful apparel retailers with more than 2000 stores in 88 countries and annual revenues over 11.5 billion euros. Its supply chain model involves frequent and rapid store changes with trendy fashion items provided at competitive prices. Within Zara's supply chain a continual flow of information from stores to designers conveys customer's changing tastes and instigates rapid orders for new designs from manufacturing suppliers.

Zara's supply chain includes four primary logistics platforms (warehouses and distribution centers) in Spain that receive finished clothing shipments from suppliers around the world and then ship items directly to stores around the world twice a week. The time between when orders are received at distribution centers to store delivery averages 24 hours in Europe and is less than 48 hours to U.S. and Asian stores. Individual store managers differentiate between major (S, M, and L) sizes and minor sizes (XXS, XXL, etc.), and when a store runs out of a major size for a clothing article it pulls all the item's inventory from the shelves and replaces it with a new article, so that customers won't be frustrated by wanting to buy an item that's not available in their size. The removed article may be returned to the store shelves if missing sizes can be shipped from warehouses, or the removed inventory may be transferred to another store where it is consolidated.

The challenge for Zara is to determine the exact number of units of each of eight sizes, of each of up to 3,000 items, to include in a shipment to each of its 1,500 stores. These order decisions, which across Zara's supply chain can reach several million each week, must be determined in a few hours after the store manager receives the relevant information including store inventory and the previous day's sales history. Any further delays can delay store replenishment by a day because of warehouse processing times and transportation schedules. Zara uses previous store shipment orders and past sales data to develop demand forecasts, which are combined with the current store and warehouse inventory to determine optimal shipment quantities that will maximize global sales. The demand forecasting model uses standard regression analysis to predict the upcoming weekly demand for each size of each article in each of Zara's stores.

As an apparel retailer, what are some of the unique supply chain forecasting factors that might affect Zara?

Sources: Based on Felipe Caro, Jeremie Gallien, Miguel Diaz, Javier Garcia, Jose Corredoira, Marcos Montes, Jose A. Ramos, and Juan Correa, "Zara Uses Operations Research to Reengineer Its Global Distribution Process," Interfaces 40 (1) (January–February 2010), pp. 71–84; and Zara website at www.zara.com.

Alamy Images

Turkeys are an example of a product with a long-term trend for increasing demand with a seasonal pattern. Turkey sales show a distinct seasonal pattern by increasing markedly during the Thanksgiving holiday season. For example, turkey sales are lowest from January to May, they begin to rise in June and July and peak in August when distributors begin to build up their inventory of frozen turkeys for increased sales in November. Sales remain high for September, October, and November and then begin to decline in December and January.

Forecast Accuracy

A forecast is never completely accurate; forecasts will always deviate from the actual demand. This difference between the forecast and the actual is the **forecast error**. Although forecast error is inevitable, the objective of forecasting is that it be as slight as possible. A large degree of error may indicate that either the forecasting technique is the wrong one or it needs to be adjusted by changing its parameters (for example, α in the exponential smoothing forecast).

Forecast error The difference between the forecast and actual demand.

There are different measures of forecast error. We will discuss several of the more popular ones: mean absolute deviation (MAD), mean absolute percent deviation (MAPD), cumulative error, and average error or bias $(\bar{E})$.

Mean Absolute Deviation

The **mean absolute deviation**, or **MAD**, is one of the most popular and simplest to use measures of forecast error. MAD is an average of the difference between the forecast and actual demand, as computed by the following formula:

Mean absolute deviation (MAD) The average absolute difference between the forecast and demand.

$$\text{MAD} = \frac{\sum |D_t - F_t|}{n}$$

where

$$\begin{aligned} t &= \text{the period number} \\ D_t &= \text{demand in period } t \\ F_t &= \text{the forecast for period } t \\ n &= \text{the total number of periods} \\ || &= \text{absolute value} \end{aligned}$$

EXAMPLE 12.7 | Measuring Forecasting Accuracy with MAD

In Examples Example 12.3, 12.4, and 12.5, forecasts were developed using exponential smoothing (α = .30 and α = .50), adjusted exponential smoothing (α = .30 and β = .30), and a linear trend line, respectively, for the demand data for HiTek Computer Services. The company wants to compare the accuracy of these different forecasts using MAD.

Solution: We will compute MAD for all four forecasts; however, we will present the computational detail for the exponential smoothing forecast with $\alpha = .30$ only. The following table shows the values necessary to compute M for the exponential smoothing forecast:

Computational Values for MAD

PERIOD	DEMAND, D_t	FORECAST F_t ($\alpha_t = .30$)	ERROR (e_t) ($D_t - F_t$)	$\lvert D_t - F_t \rvert$
1	37	37.00	—	—
2	40	37.00	3.00	3.00
3	41	37.90	3.10	3.10
4	37	38.83	−1.83	1.83
5	45	38.28	6.72	6.72
6	50	40.29	9.69	9.69
7	43	43.20	−0.20	0.20
8	47	43.14	3.86	3.86
9	56	44.30	11.70	11.70
10	52	47.81	4.19	4.19
11	55	49.06	5.94	5.94
12	54	50.84	3.15	3.15
	557		49.32	53.38

[a] The computation of MAD will be based on 11 periods, periods 2 through 12, excluding the initial demand and forecast values for period 1 since they both equal 37.

Using the data in the table, MAD is computed as

$$\begin{aligned} \text{MAD} &= \frac{\sum |D_t - F_t|}{n} \\ &= \frac{53.39}{11} \\ &= 4.85 \end{aligned}$$

The smaller the value of MAD, the more accurate the forecast, although viewed alone, MAD is difficult to assess. In this example, the data values were relatively small, and the MAD value of 4.85 should be judged accordingly. Overall, it would seem to be a "low" value; that is, the forecast appears to be relatively accurate. However, if the magnitude of the data values were in the thousands or millions, then a MAD value of a similar magnitude might not be bad, either. The point is, you cannot compare a MAD value of 4.85 with a MAD value of 485 and say the former is good and the latter is bad; they depend to a certain extent on the relative magnitude of the data.

One benefit of MAD is to compare the accuracy of several different forecasting techniques, as we are doing in this example. The MAD values for the remaining forecasts are as follows:

$$\begin{aligned} \text{Exponential smoothing } (\alpha = 0.50) : \text{MAD} &= 4.04 \\ \text{Adjusted exponential smoothing } (\alpha = 50, \beta = 0.30) : \text{MAD} &= 3.81 \\ \text{Linear trend line} : \text{MAD} &= 2.29 \end{aligned}$$

Since the linear trend line has the lowest MAD value of 2.29, it would seem to be the most accurate, although it does not appear to be significantly better than the adjusted exponential smoothing forecast. Furthermore, we can deduce from these MAD values that increasing α from .30 to .50 enhanced the accuracy of the exponentially smoothed forecast. The adjusted forecast is even more accurate.

The **mean absolute percent deviation (MAPD)** measures the absolute error as a percentage of demand rather than per period. As a result, it eliminates the problem of interpreting the measure of accuracy relative to the magnitude of the demand and forecast values, as MAD does. The mean absolute percent deviation is computed according to the following formula:

mean absolute percent deviation (MAPD) The absolute error as a percentage of demand.

$$\text{MAPD} = \frac{\sum |D_t - F_t|}{\sum D_t}$$

Using the data from the table in Example 12.7 for the exponential smoothing forecast (α = .30) for HiTek Computer Services, we find

$$\begin{aligned}\text{MAPD} &= \frac{53.39}{557}\\ &= 0.096 \text{ or } 9.6\%\end{aligned}$$

A lower percent deviation implies a more accurate forecast. The MAPD values for our other three forecasts are

$$\begin{aligned}\text{Exponential smoothing } (\alpha = 0.50) &: \text{MAPD} = 7.9\%\\ \text{Adjusted exponential smoothing } (\alpha = 0.50, \beta = 0.30) &: \text{MAPD} = 7.5\%\\ \text{Linear trend line} &: \text{MAPD} = 4.9\%\end{aligned}$$

Cumulative Error

Cumulative error is computed simply by summing the forecast errors, as shown in the following formula.

Cumulative error The sum of the forecast errors.

$$E = \sum e_t$$

A large positive value indicates that the forecast is probably consistently lower than the actual demand, or is biased low. A large negative value implies that the forecast is consistently higher than actual demand, or is biased high. Also, when the errors for each period are scrutinized, a preponderance of positive values shows the forecast is consistently less than the actual value and vice versa.

The cumulative error for the exponential smoothing forecast (α = .30) for HiTek Computer Services can be read directly from the table in Example 12.7; it is simply the sum of the values in the "Error" column:

$$\begin{aligned}E &= \sum e_t\\ &= 49.31\end{aligned}$$

This large positive error for cumulative error, plus the fact that the individual errors for all but two of the periods in the table are positive, indicates that this forecast is consistently below the actual demand. A quick glance back at the plot of the exponential smoothing ($\alpha = .30$) forecast in Example 12.3 visually verifies this result.

The cumulative error for the other forecasts are

$$\text{Exponential smoothing } (\alpha = 0.50) : E = 33.21$$
$$\text{Adjusted exponential smoothing } (\alpha = 0.50, \beta = 0.30) : E = 21.14$$

We did not show the cumulative error for the linear trend line. E will always be near zero for the linear trend line.

A measure closely related to cumulative error is the **average error**, or *bias*. It is computed by averaging the cumulative error over the number of time periods:

Average error The per-period average of cumulative error.

$$\overline{E} = \frac{\sum e_t}{n}$$

For example, the average error for the exponential smoothing forecast ($\alpha = .30$) is computed as follows. (Notice a value of 11 was used for n, since we used actual demand for the first-period forecast, resulting in no error, that is, $D_1 = F_1 = 37$.)

$$\overline{E} = \frac{49.32}{11} = 4.48$$

The average error is interpreted similarly to the cumulative error. A positive value indicates low bias, and a negative value indicates high bias. A value close to zero implies a lack of bias.

Table 12.1 summarizes the measures of forecast accuracy we have discussed in this section for the four example forecasts we developed in Example 12.3, 12.4, and 12.5 for HiTek Computer Services. The results are consistent for all four forecasts, indicating that for the HiTek Computer Services example data, a larger value of α is preferable for the exponential smoothing forecast. The adjusted forecast is more accurate than the exponential smoothing forecasts, and the linear trend is more accurate than all the others. Although these results are for specific examples, they indicate how the different forecast measures for accuracy can be used to adjust a forecasting method or select the best method.

TABLE 12.1 Comparison of Forecasts for HiTek Computer Services

FORECAST	MAD	MAPD	E	$\bar{E}$
Exponential smoothing ($\alpha = .30$)	4.85	9.6%	49.31	4.48
Exponential smoothing ($\alpha = .50$)	4.04	8.5%	33.21	3.02
Adjusted exponential smoothing ($\alpha = .50, \beta = .30$)	3.81	7.5%	21.14	1.92
Linear trend line	2.29	4.9%	—	—

Forecast Control

There are several ways to monitor forecast error over time to make sure that the forecast is performing correctly—that is, the forecast is in control. Forecasts can go "out of control" and start providing inaccurate forecasts for several reasons, including a change in trend, the unanticipated appearance of a cycle, or an irregular variation such as unseasonable weather, a promotional campaign, new competition, or a political event that distracts consumers.

A **tracking signal** indicates if the forecast is consistently biased high or low. It is computed by dividing the cumulative error by MAD, according to the formula

Tracking signal Monitors the forecast to see if it is biased high or low.

$$\text{Tracking signal} = \frac{\sum (D_t - F_t)}{\text{MAD}} = \frac{E}{\text{MAD}}$$

The tracking signal is recomputed each period, with updated, "running" values of cumulative error and MAD. The movement of the tracking signal is compared to *control limits*: as long as the tracking signal is within these limits, the forecast is in control.

Typically, forecast errors are normally distributed, which results in the following relationship between MAD and the standard deviation of the distribution of error, σ:

$$1 \text{ MAD } \cong 0.8\sigma$$

This enables us to establish statistical control limits for the tracking signal that correspond to the more familiar normal distribution. For example, statistical control limits of ±3 standard deviations, corresponding to 99.7% of the errors, would translate to ±3.75 MADs; that is, $3\sigma \div 0.8 = 3.75$ MADs. Control limits of ±2 to ±5 are used most frequently.

EXAMPLE 12.8 | Developing a Tracking Signal

In Example 12.7, the mean absolute deviation was computed for the exponential smoothing forecast ($\alpha = .30$) for HiTek Computer Services. Using a tracking signal, monitor the forecast accuracy using control limits of ±3 MADs.

Solution: To use the tracking signal, we must recompute MAD each period as the cumulative error is computed.

Using MAD = 3.00, we find that the tracking signal for period 2 is

$$TS_2 = \frac{E}{\text{MAD}} = \frac{3.00}{3.00} = 1.00$$

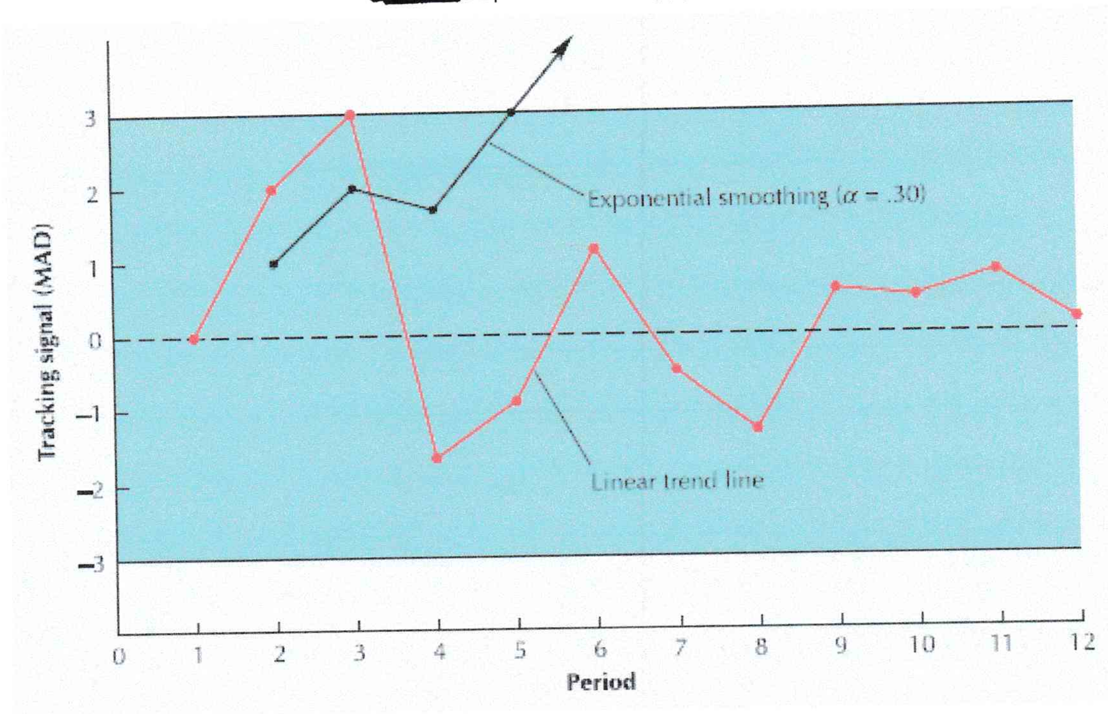
The tracking signal for period 3 is

$$TS_3 = \frac{6.10}{3.05} = 2.00$$

The remaining tracking signal values are shown in the following table:

Tracking Signal Values

PERIOD	DEMAND, D_t	FORECAST, F_t	ERROR, $D_t - F_t$	$E = \Sigma(D_t - F_t)$	MAD	TRACKING SIGNAL
1	37	37.00	—	—	—	—
2	40	37.00	3.00	3.00	3.00	1.00
3	41	37.90	3.10	6.10	3.05	2.00
4	37	38.83	−1.83	4.27	2.64	1.62
5	45	38.28	6.72	10.99	3.66	3.00
6	50	40.29	9.69	20.68	4.87	4.25
7	43	43.20	−0.20	20.48	4.09	5.01
8	47	43.14	3.86	24.34	4.06	6.00
9	56	44.30	11.70	36.04	5.01	7.19
10	52	47.81	4.19	40.23	4.92	8.18
11	55	49.06	5.94	46.17	5.02	9.20
12	54	50.84	3.15	49.32	4.85	10.17



The tracking signal values in the table above move outside ±3 MAD control limits (i.e., ±3.00) in period 5 *and* continue increasing. This suggests that the forecast is not performing accurately or, more precisely, is consistently biased low (i.e., actual demand consistently exceeds the forecast). This is illustrated in the graph. Notice that the tracking signal moves beyond the upper limit of 3 following period 5 and continues to rise. For the sake of comparison, the tracking signal for the linear trend line forecast computed in Example 12.5 is also plotted on this graph. Notice that it remains within the limits (touching the upper limit in period 3), indicating a lack of consistent bias.

Another method for monitoring forecast error is statistical control charts. For example, ±3σ control limits would reflect 99.7% of the forecast errors (assuming they are normally distributed). The sample standard deviation, σ, is computed as

$$\sigma = \sqrt{\frac{\sum (D_t - F_t)^2}{n-1}}$$

This formula without the square root is known as the **mean squared error (MSE)**, and it is sometimes used as a measure of forecast error. It reacts to forecast error much as MAD does. (For our Example 12.8, MSE = 37.57.)

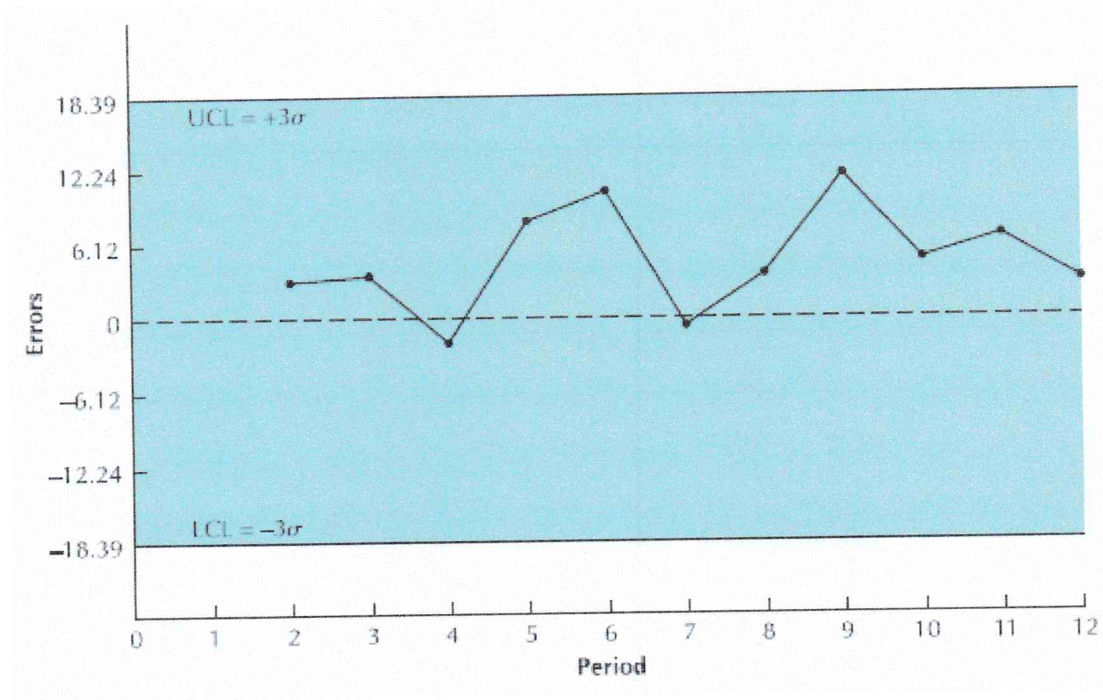
Mean squared error (MSE) The average of the squared forecast errors.

EXAMPLE 12.9 | Forecast Error with Statistical Control Charts

Using the same example for the exponential smoothing forecast (α = .30) for HiTek Computer Services as in Example 12.8, we compute the standard deviation as

$$\sigma = \sqrt{\frac{375.61}{10}} = 6.13$$

Using this value of σ we can compute statistical control limits for forecast errors for our exponential smoothing forecast (α = .30) example for HiTek Computer Services. Plus or minus 3σ control limits, reflecting 99.7% of the forecast errors, gives ±3(6.13), or ±18.39. Although it can be observed from the table in Example 12.8 that all the error values are within the control limits, we can still detect that most of the errors are positive, indicating a low bias in the forecast estimates. This is illustrated in the following graph of the control chart with the errors plotted on it.



Time Series Forecasting Using Excel

Excel can be used to develop forecasts using the moving average, exponential smoothing, adjusted exponential smoothing, and linear trend line techniques. Various surveys of companies across different industries that use forecasting, show that over half use Excel spreadsheets for forecasting, and most use a variety of different forecasting software packages.

First we will demonstrate how to determine exponentially smoothed and adjusted exponentially smoothed forecasts using Excel, as shown in **Exhibit 12.1**. We will demonstrate Excel using Example12.3 and 12.4 for forecasting demand at HiTek Computer Services, including the Excel spreadsheets showing the exponentially smoothed forecast with $\alpha = .5$ and the adjusted exponentially smoothed forecast with $\beta = .3$. We have also computed the values for MAD, MAPD, and *E*.

EXHIBIT 12.1

Excel File

Example 12.4: Exponentially Smoothed and Adjusted Exponentially Smoothed Forecasts

Alpha =	0.5
Beta =	0.3

Month	*Demand*	*Forecast*	*Trend*	*Adjusted Forecast*	*Error*	*Absolute Error*
January	37	37.00				
February	40	37.00	0.00	37.00	3.00	3.00
March	41	38.50	0.45	38.95	2.05	2.05
April	37	39.75	0.69	40.44	-3.44	3.44
May	45	38.38	0.07	38.45	6.55	6.55
June	50	41.69	1.04	42.73	7.27	7.27
July	43	45.84	1.98	47.82	-4.82	4.82
August	47	44.42	0.96	45.38	1.62	1.62
September	56	45.71	1.06	46.77	9.23	9.23
October	52	50.86	2.28	53.14	-1.14	1.14
November	55	51.43	1.77	53.20	1.80	1.80
December	54	53.21	1.77	54.99	-0.99	0.97
January		53.61	1.36	54.97		
					21.14	41.90

MAD =	3.81	
MAPD =	8.1	percent
E =	21.14	

=B5*(C11 – C10)+ (1 – B5)*D10

=C10 + D10

=ABS(B10 – E10)

=G22/11

=SUM(F10:F20)

Along the Supply Chain

Forecasting Empty Shipping Containers

Compana Sud Americana de Vapores (CSAV), headquartered in Chile, is one of the world's largest shipping companies. It ships cargo using containers transported by over 180 ships operated by CSAV or other companies. Its fleet consists of about 70,000 containers of different sizes and types, about 5% of which CSAV owns with the rest leased on long-term contracts. Each week the company makes hundreds of thousands of container logistics decisions over its vast shipping network, involving different nationalities, culture, and time zones, which makes the decision-making process very difficult. Typically an empty container is first loaded on a truck at a container depot and transported to a shipper by truck, filled and then shipped by truck to a port where it is loaded on a vessel after meeting customs requirements. At its destination port the container is unloaded from the vessel, and transported by truck, train, or feeder vessel to the customer, who unloads the container and returns it to the shipping company. Managing this fleet of empty containers is a complex process partly as a result of the imbalance in demand among regions; some regions are net exporters whereas some regions are net importers of containers. Another challenge is forecasting demand for empty containers of different types and sizes at each location for specific dates, given the uncertainty in the demand for empty containers, which depends on factors like market conditions, travel times, delays in returning empty containers, and the availability of vessel capacity to transport empty containers. To forecast demand CSAV uses a combination of several forecasting approaches. It forecasts returned containers using a moving average of the past *n* days, and a trended seasonal moving average using past demand from the same season with a yearly trend computed from previous years. These time series forecasts are complemented by a sales forecast based on the demand expectations of their sales agents around the world. Over a million and a half demand forecasts are generated based on updated information and revised settings. Logistics planners at various locations then use the forecast model that best fits their own experience and forecast accuracy at their location. These forecasts are subsequently used in a network flow optimization model to plan the movement of empty containers in CSAV's system. This solution approach resulted in savings of $81 million in the first year it was implemented.

Using the Internet, identify and discuss forecasting demand of another part or process (other than containers) related to distribution and transportation in the supply chain.

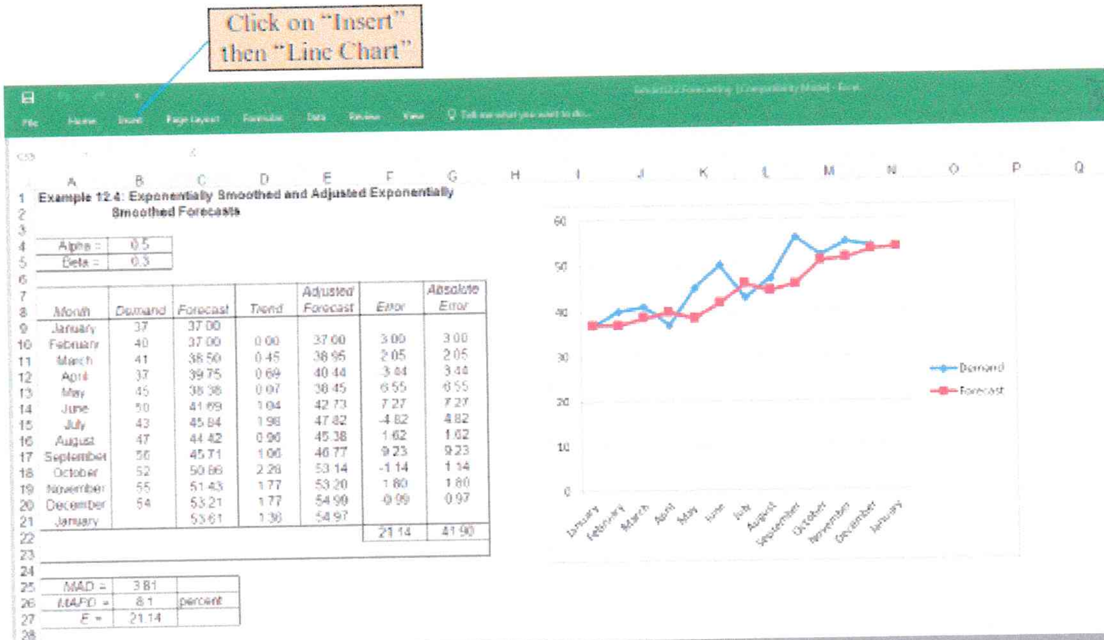
Sources: Based on R. Epstein, A. Neely, A. Weintraub, F. Valenzuela, S. Hurtado, G. Gonzalez, A. Beiza, M. Naveas, F. Infante, F. Alarcon, G. Angulo, C. Berner, J. Catalan, C. Gonzalez and D. Yung, "A Strategic Empty Container Logistics Optimization in A Major Shipping Company," Interfaces, vol. 42, no. 1 (January–February 2012): 5–16.

Notice that the formula in Exhibit 12.1 for computing the exponentially smoothed forecast for March is embedded in cell C11 and shown on the formula bar at the top of the screen. The same formula is used to compute all the other forecast values in column C. The formula for computing the trend value for March is = B5*(C11 – C10) + (1 – B5)*D10. The formula for the adjusted forecast in column *E* is computed by typing the formula = C10 + D10 in cell E10 and copying it to cells E11:E21 (using the copy and paste options from the right mouse key). The error is computed for the adjusted forecast, and the formula for computing the error for March is = B11 – E11, while the formula for absolute error for March is = ABS (F11).

A graph of the forecast can also be developed with Excel. To plot the exponentially smoothed forecast in column C and demand in column B, cover all cells from A8 to C21 with the mouse and click on "Insert" on the toolbar at the top of the worksheet. Next click on "Line" on the "Chart" toolbar. The resulting graph for demand and the exponentially smoothed forecast for our example are shown in **Exhibit 12.2**.

EXHIBIT 12.2

Excel File



Example 12.4: Exponentially Smoothed and Adjusted Exponentially Smoothed Forecasts

Alpha =	0.5
Beta =	0.3

Month	Demand	Forecast	Trend	Adjusted Forecast	Error	Absolute Error
January	37	37.00				
February	40	37.00	0.00	37.00	3.00	3.00
March	41	38.50	0.45	38.95	2.05	2.05
April	37	39.75	0.69	40.44	-3.44	3.44
May	45	38.38	0.07	38.45	6.55	6.55
June	50	41.69	1.04	42.73	7.27	7.27
July	43	45.84	1.98	47.82	-4.82	4.82
August	47	44.42	0.96	45.38	1.62	1.62
September	56	45.71	1.06	46.77	9.23	9.23
October	52	50.86	2.28	53.14	-1.14	1.14
November	55	51.43	1.77	53.20	1.80	1.80
December	54	53.21	1.77	54.98	-0.98	0.97
January		53.61	1.36	54.97		
					21.14	41.90

MAD =	3.81	
MAPD =	8.1	percent
E =	21.14	

The exponential smoothing forecast can also be developed directly from Excel without "customizing" a spreadsheet and entering formulas as we did in Exhibit 12.1. From the Tools menu at the top of the spreadsheet select the "Data" option and then "Data Analysis" option. **Exhibit 12.3** shows the "Data Analysis" window and the "Exponential Smoothing" menu item, which should be selected by clicking on "OK." The resulting "Exponential Smoothing" window is shown in **Exhibit 12.4**. The "input range" includes the demand values in column C in Exhibit 12.1, the damping factor is α, which in this case is .5, and the output should be placed in column C in Exhibit 12.1. Clicking on "OK" will result in the same forecast values in column C of Exhibit 12.1 as we computed using our own exponential smoothing formula. Note that the Data Analysis group of analysis tools does not have an adjusted exponential smoothing selection; that is one reason we developed our own customized spreadsheet in Exhibit 12.1. The "Data Analysis" tools also have a moving average menu item that you can use to compute a moving average forecast.

EXHIBIT 12.3

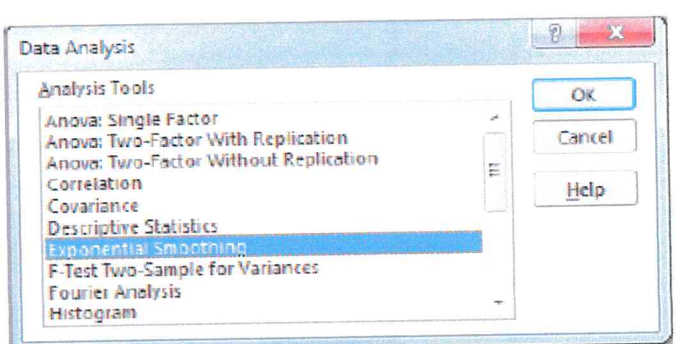
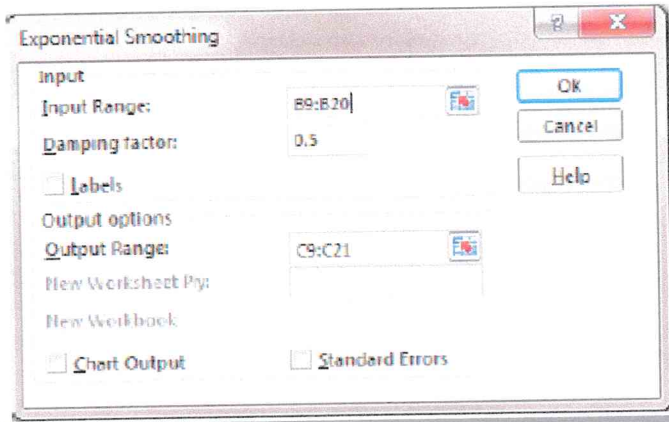


EXHIBIT 12.4



Excel can also be used to develop more customized forecast models, like seasonal forecasts. **Exhibit 12.5** shows an Excel screen for the seasonal forecast model developed in Example 12.6. Notice that the computation of the seasonal forecast for the first quarter (SF1) in cell B12 uses the formula shown on the formula bar at the top of the screen. The forecast value for SF1 is slightly different from the value in Example 12.6 because of rounding.

EXHIBIT 12.5

Excel File

Example 12.6: Computing a Forecast with Seasonal Adjustments

Year	Demand (1,000s) per Quarter 1	2	3	4	Total
2014	12.6	8.6	6.3	17.5	45.0
2015	14.1	10.3	7.5	18.2	50.1
2016	15.3	10.6	8.1	19.6	53.6
Total	42.0	29.5	21.9	55.3	148.7

Linear trend line forecast for 2017 = 58.17

SF1 =	16.43
SF2 =	11.54
SF3 =	8.57
SF4 =	21.63

Forecasting with OM Tools

OM Tools has modules for all of the forecasting methods presented in this chapter. As an example, **Exhibit 12.6** shows the OM Tools spreadsheet for the exponential smoothing model ($\alpha = .30$) in Example 12.3.

EXHIBIT 12.6

OM Tools

Exponentially Smoothed Forecasts OM Student - Example 12.3

Input:

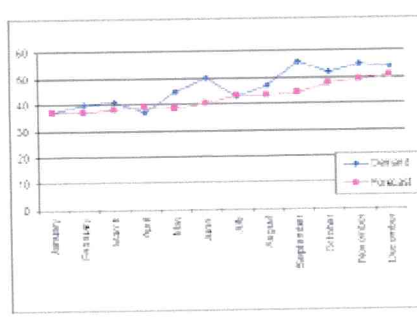
No. of demand periods	12
Alpha	0.30

Label periods, input demand data and smoothing constant, alpha. Scroll down for output values.

Exponential Smoothing

$$F_{t+1} = \alpha D_t + (1 - \alpha)F_t$$

Period	Demand	Forecast	Error	Absolute Error	Squared Error
January	37	37.00			
February	40	37.00	3.00	3.00	9.00
March	41	37.90	3.10	3.10	9.61
April	37	38.83	-1.83	1.83	3.35
May	45	38.28	6.72	6.72	45.14
June	50	40.30	9.70	9.70	94.15
July	43	43.21	-0.21	0.21	0.04
August	47	43.15	3.85	3.85	14.85
September	56	44.30	11.70	11.70	136.85
October	52	47.81	4.19	4.19	17.55
November	55	49.07	5.93	5.93	35.19
December	54	50.85	3.15	3.15	9.94
Total	567.00		49.31	53.39	375.68



Output:

MAD	4.85
MAPD	0.10
E	49.31
$\bar{E}$	4.48
MSE	37.57

$$MAD = \frac{\Sigma|D_t - F_t|}{n}$$

$$MAPD = \frac{\Sigma|D_t - F_t|}{D_t}$$

$$\bar{E} = \frac{\Sigma e_t}{n}$$

$$MSE = \frac{\Sigma(D_t - F_t)^2}{n - 1}$$

$$\bar{x} = \frac{\sum x}{n} = \text{mean of the } x \text{ data}$$

$$\bar{y} = \frac{\sum y}{n} = \text{mean of the } y \text{ data}$$

EXAMPLE 12.10 | Developing a Linear Regression Forecast

The State University athletic department wants to develop its budget for the coming year using a forecast for football attendance. Football attendance accounts for the largest portion of its revenues, and the athletic director believes attendance is directly related to the number of wins by the team. The business manager has accumulated total annual average attendance figures for the past eight years.

WINS	ATTENDANCE	WINS	ATTENDANCE
4	36,300	6	44,000
6	40,100	7	45,600
6	41,200	5	39,000
8	53,000	7	47,500

Regression Methods

The second most popular forecasting technique among various industrial firms is regression. Regression is used for forecasting by establishing a mathematical relationship between two or more variables. We are interested in identifying relationships between variables and demand. If we know that something has caused demand to behave in a certain way in the past, we would like to identify that relationship so if the same thing happens again in the future, we can predict what demand will be. For example, there is a relationship between increased demand in new housing and lower interest rates. Correspondingly, a whole myriad of building products and services display increased demand if new housing starts increase.

The simplest form of regression is linear regression, which we used previously to develop a linear trend line for forecasting. Now we will show how to develop a regression model for variables related to demand other than time.

Linear Regression

Linear regression is a mathematical technique that relates one variable, called an *independent variable*, to another, the *dependent variable*, in the form of an equation for a straight line. A linear equation has the following general form:

Linear regression A mathematical technique that relates a dependent variable to an independent variable in the form of a linear equation.

$$y = a + bx$$

where

$$\begin{aligned} y &= \text{the dependent variable} \\ a &= \text{the intercept} \\ b &= \text{the slope of the line} \\ x &= \text{the independent variable} \end{aligned}$$

Because we want to use linear regression as a forecasting model for demand, the dependent variable, y, represents demand, and x is an independent variable that causes demand to behave in a linear manner.

To develop the linear equation, the slope, b, and the intercept, a, must first be computed using the following least squares formulas:

$$a = \bar{y} - b\bar{x}$$

$$b = \frac{\sum xy - n\,\overline{xy}}{\sum x^2 - n\,\bar{x}^2}$$

where

Given the number of returning starters and the strength of the schedule, the athletic director believes the team will win at least seven games next year. Develop a simple regression equation for this data to forecast attendance for this level of success.

Solution: The computations necessary to compute a and b using the least squares formulas are summarized in the accompanying table. (Note that y is given in 1000s to make manual computation easier.)

Least Squares Computations

x (WINS)	y (ATTENDANCE, 1000s)	xy	x^2
4	36.3	145.2	16
6	40.1	240.6	36
6	41.2	247.2	36
8	53.0	424.0	64
6	44.0	264.0	36
7	45.6	319.2	49
5	39.0	195.0	25
7	47.5	332.5	49
49	346.9	2167.7	311

$$\begin{aligned}
\bar{x} &= \frac{49}{8} = 6.125 \\
\bar{y} &= \frac{346.9}{8} = 43.36 \\
b &= \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2} \\
&= \frac{(2167.7) - (8)(6.125)(43.36)}{\left(311\right) - \left(8\right)(6.125)^2} \\
&= 4.06 \\
a &= \bar{y} - b\bar{x} \\
&= 43.36 - (4.06)\,(6.125) \\
&= 18.46
\end{aligned}$$

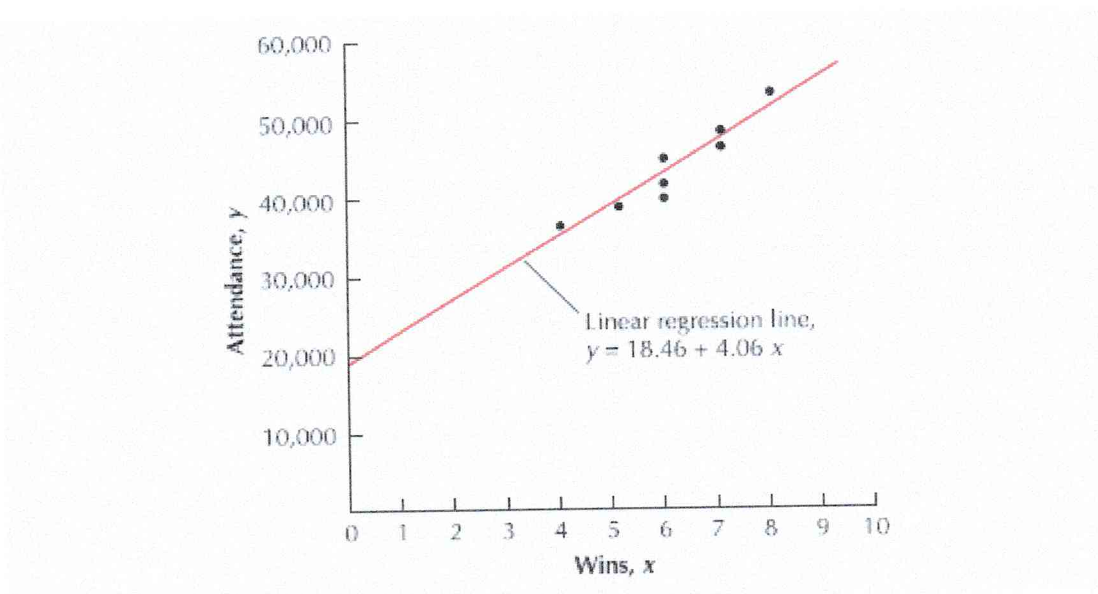
Substituting these values for a and b into the linear equation line, we have

$$y = 18.46 + 4.06x$$

Thus, for $x = 7$ (wins), the forecast for attendance is

$$\begin{aligned}
y &= 18.46 + 4.06\,(7) \\
&= 46.88\text{, or } 46{,}880
\end{aligned}$$

The data points with the regression line are shown in the following figure. Observing the regression line relative to the data points, it would appear that the data follow a distinct upward linear trend, which would indicate that the forecast should be relatively accurate. In fact, the MAD value for this forecasting model is 1.41, which suggests an accurate forecast.



Correlation

Correlation in a linear regression equation is a measure of the strength of the relationship between the independent and dependent variables. The formula for the correlation coefficient is

Correlation A measure of the strength of the relationship between independent and dependent variables.

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right]\left[n\sum y^2 - \left(\sum y\right)^2\right]}}$$

The value of r varies between −1.00 and −1.00, with a value of +1.00 indicating a strong linear relationship between the variables. If r = 1.00, then an increase in the independent variable will result in a corresponding linear increase in the dependent variable. If r = −1.00, an increase in the dependent variable will result in a linear decrease in the dependent variable. A value of r near zero implies that there is little or no linear relationship between variables.

We can determine the correlation coefficient for the linear regression equation determined in Example 12.9 by substituting most of the terms calculated for the least squares formula (except for $\sum y^2$) into the formula for r:

$$\begin{aligned} r &= \frac{(8)(2167.7)-(49)(346.9)}{\sqrt{\left[(8)(311)-(49)^2\right]\left[(8)(15{,}224.7)-(346.9)^2\right]}} \\ &= 0.947 \end{aligned}$$

This value for the correlation coefficient is very close to 1.00, indicating a strong linear relationship between the number of wins and home attendance.

Another measure of the strength of the relationship between the variables in a linear regression equation is the **coefficient of determination**. It is computed by squaring the value of r. It indicates the percentage of the variation in the dependent variable that is a result of the behavior of the independent variable. For our example, r = 0.947; thus, the coefficient of determination is

Coefficient of determination The percentage of the variation in the dependent variable that results from the independent variable.

$$\begin{aligned} r^2 &= (0.947)^2 \\ &= 0.897 \end{aligned}$$

This value for the coefficient of determination means that 89.7% of the amount of variation in attendance can be attributed to the number of wins by the team (with the remaining 10.3% due to other unexplained factors, such as weather, a good or poor start, or publicity). A value of 1.00

(or 100%) would indicate that attendance depends totally on wins. However, since 10.3% of the variation is a result of other factors, some amount of forecast error can be expected.

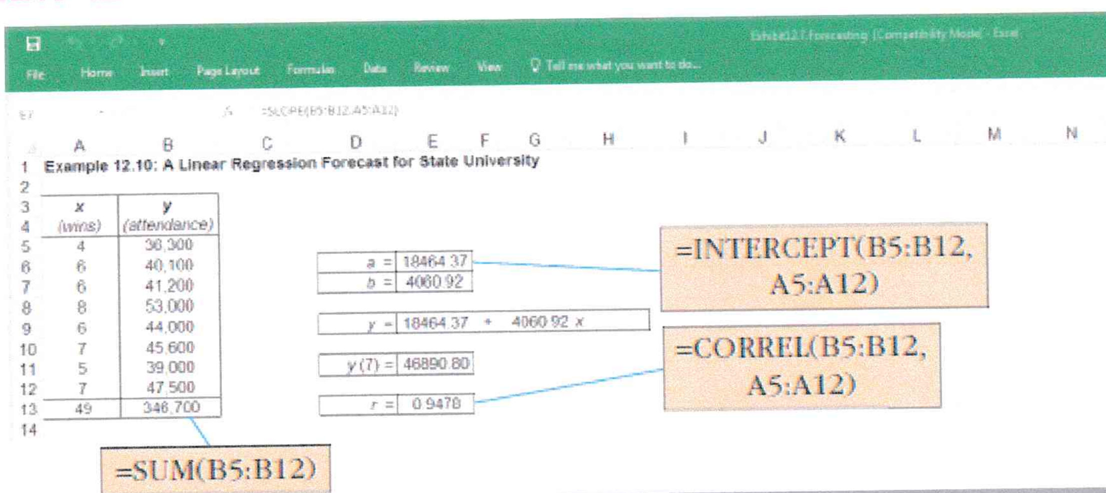
Regression Analysis with Excel

The development of the simple linear regression equation and the correlation coefficient for our example was not too difficult because the amount of data was relatively small. However, manual computation of the components of simple linear regression equations can become very time-consuming and cumbersome as the amount of data increases. Excel has the capability of performing linear regression.

Exhibit 12.7 shows a spreadsheet set up to develop the linear regression forecast for Example 12.10 for the State University Athletic Department. Notice that Excel computes the slope directly with the formula "=SLOPE(B5:B12,A5:A12)" entered in cell E7 and shown on the formula bar at the top of the spreadsheet. The formula for the intercept in cell E6 is "=INTERCEPT(B5:B12,A5:A12)." The values for the slope and intercept are subsequently entered into cells E9 and G9 to form the linear regression equation. The correlation coefficient in cell E13 is computed using the formula "=CORREL(B5:B12, A5:A12)." Although it is not shown on the spreadsheet, the coefficient of determination (r^2) could be computed using the formula "=RSQ(B5:B12,A5:A12)."

EXHIBIT 12.7

Excel File



A linear regression forecast can also be developed directly with Excel using the "Data Analysis" option from the Tools menu we accessed previously to develop an exponentially smoothed forecast. **Exhibit 12.8** shows the selection of "Regression" from the Data Analysis menu, and **Exhibit 12.9** shows the Regression window. We first enter the cells from Exhibit 12.7 that include the y values (for attendance), B5:B12. Next enter the x value cells, A5:A12. The output range is the location on the spreadsheet where you want to put the output results. This range needs to be large (18 cells by 9 cells) and not overlap with anything else on the spreadsheet. Clicking on "OK" will result in the spreadsheet shown in **Exhibit 12.10**. (Note that the "Summary Output" has been slightly moved around so that all the results could be included on the screen in Exhibit 12.9.)

EXHIBIT 12.8

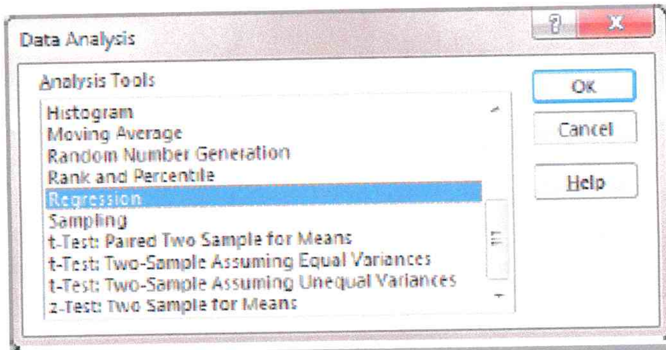
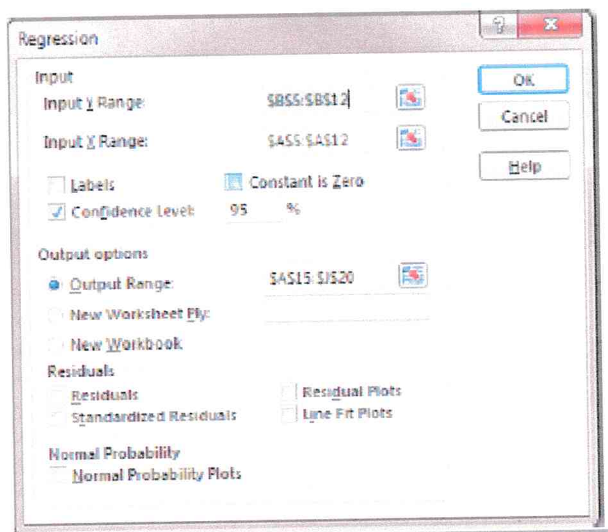


EXHIBIT 12.9



Excel File

EXHIBIT 12.10

Example 12.10: State University Athletic Department

x (wins)	y (attendance)
4	36,300
6	40,100
6	41,200
8	53,000
6	44,000
7	45,600
5	39,000
7	47,500
49	346,700

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.948
R Square	0.898
Adjusted R Square	0.881
Standard Error	1839.311
Observations	8

ANOVA

	df	SS	MS	F	Significance F
Regression	1	179340359.2	179340359.2	53.01120496	0.00034182
Residual	6	20298390.8	3383065.134		
Total	7	199638750			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	18464.37	3477.57	5.31	0.00	9955.06	26973.68	9955.06	26973.68
X Variable 1	4060.92	557.75	7.28	0.00	2696.15	5425.69	2696.15	5425.69

The "Summary Output" in Exhibit 12.10 provides a large amount of statistical information, the explanation and use of which are beyond the scope of this book. The essential items that we are interested in are the intercept and slope (labeled "X Variable 1") in the "Coefficients" column at the bottom of the spreadsheet, and the "Multiple R" (or correlation coefficient) value shown under "Regression Statistics."

Multiple Regression with Excel

Another causal method of forecasting is **multiple regression**, a more powerful extension of linear regression. Linear regression relates demand to one other independent variable, whereas multiple regression reflects the relationship between a dependent variable and two or more independent variables. A multiple regression model has the following general form:

Multiple regression A relationship of demand to two or more independent variables.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

where

$$\begin{aligned} \beta_0 &= \text{the intercept} \\ \beta_1, \ldots, \beta_k &= \text{parameters representing the contribution of the independent variables} \\ x_1, \ldots, x_k &= \text{independent variables} \end{aligned}$$

For example, the demand for new housing (y) in a region might be a function of several independent variables, including interest rates, population, housing prices, and personal income. Development and computation of the multiple regression equation, including the compilation of data, is more complex than linear regression. The only means for forecasting using multiple regression is with a computer.

EXAMPLE 12.11 | Developing a Multiple Regression Forecast with Excel

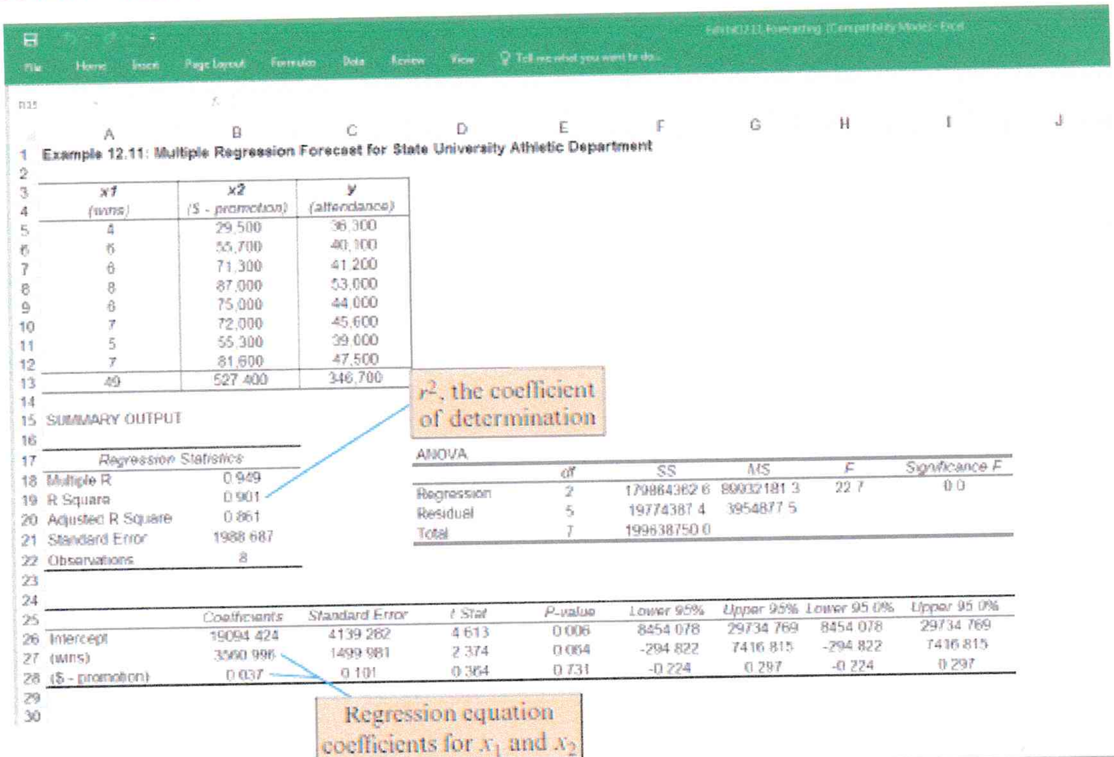
To demonstrate the capability to solve multiple regression problems with Excel spreadsheets, we will expand our State University athletic department Example 12.10 for forecasting attendance at football games that we used to demonstrate linear regression. Instead of attempting to predict attendance based on only one variable, wins, we will include a second variable for advertising and promotional expenditures as follows:

WINS	PROMOTION ($)	ATTENDANCE
4	29,500	36,300
6	55,700	40,100
6	71,300	41,200
8	87,000	53,000
6	75,000	44,000
7	72,000	45,600
5	55,300	39,000
7	81,600	47,500

We will use the "Data Analysis" option (add-in) from the Tools menu at the top of the spreadsheet that we used in the previous section to develop our linear regression equation, and then the "Regression" option from the "Data Analysis" menu. The resulting spreadsheet with the multiple regression statistics is shown in **Exhibit 12.11**.

EXHIBIT 12.11

Excel File



Example 12.11: Multiple Regression Forecast for State University Athletic Department

x1 (wins)	x2 ($ - promotion)	y (attendance)
4	29,500	36,300
6	55,700	40,100
6	71,300	41,200
8	87,000	53,000
6	75,000	44,000
7	72,000	45,600
5	55,300	39,000
7	81,600	47,500
49	527,400	346,700

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.949
R Square	0.901
Adjusted R Square	0.861
Standard Error	1988.687
Observations	8

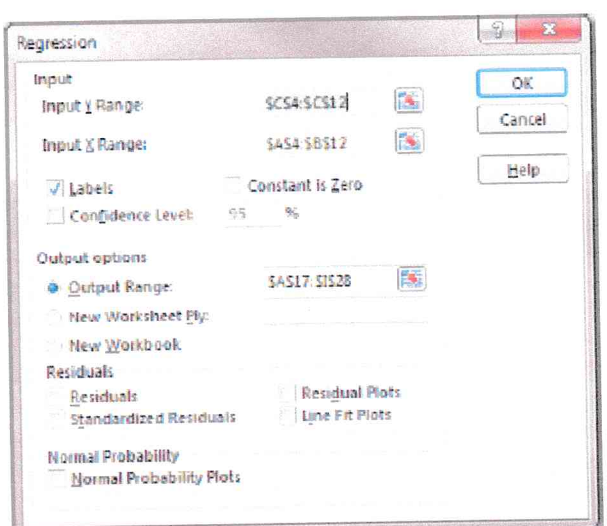
ANOVA

	df	SS	MS	F	Significance F
Regression	2	179864362.6	89932181.3	22.7	0.0
Residual	5	19774387.4	3954877.5		
Total	7	199638750.0			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	19094.424	4139.282	4.613	0.006	8454.078	29734.769	8454.078	29734.769
(wins)	3560.996	1499.981	2.374	0.064	-294.822	7416.815	-294.822	7416.815
($ - promotion)	0.037	0.101	0.364	0.731	-0.224	0.297	-0.224	0.297

Solution: Note that the data must be set up on the spreadsheet so that the two x variables are in adjacent columns (in this case A and B). Then we enter the "Input X Range" as **A4:B12** as shown in Exhibit 12.12.

EXHIBIT 12.12



The regression coefficients for our x variables, wins and promotion, are shown in cells B27 and B28. Thus, the multiple regression equation is formulated as

$$y = 19{,}094.42 + 3560.99x_1 + 0.0368x_2$$

This equation can now be used to forecast attendance based on both projected football wins and promotional expenditure. For example, if the athletic department expects the team to win seven games and plans to spend $60,000 on promotion and advertising, the forecasted attendance is

$$\begin{aligned} y &= 19{,}094.42 + 3560.99\,(7) + 0.0368\,(60{,}000) \\ &= 46{,}229.35 \end{aligned}$$

If the promotional expenditure is held constant, every win will increase attendance by 3560.99, whereas if the wins are held constant, every $1000 of advertising spent will increase attendance by 36.8 fans. This would seem to suggest that the number of wins has a more significant impact on attendance than promotional expenditures.

r^2, the coefficient of determination shown in cell B19, is 0.900, which suggests that 90% of the amount of variation in attendance can be attributed to the number of wins and the promotional expenditures. However, as we have already noted, the number of wins would probably appear to account for a larger part of the variation in attendance.

A problem often encountered in multiple regression is *multicollinearity*, or the amount of "overlapping" information about the dependent variable that is provided by several independent variables. This problem usually occurs when the independent variables are highly correlated, as in this example, in which wins and promotional expenditures are both positively correlated (i.e., more wins coincide with higher promotional expenditures and vice versa). (Possibly the athletic department increased promotional expenditures when it thought it would have a better team that would achieve more wins.) The topic of multicollinearity and how to cope with it is beyond the scope of this book and this brief section on multiple regression; however, most statistics textbooks discuss this topic in detail.

Data Mining

At one time, business forecasters typically did not have enough data to develop accurate forecasts; however, that's no longer a problem. In fact, the situation is just the opposite—businesses are overwhelmed with data. Data are collected wherever and whenever a customer swipes a credit card or accesses a website, or a company employee scans a box or a pallet in a warehouse or goods on a store shelf. It is estimated that all the data in databases in the world double in less than every two years. In addition, large, inexpensive storage devises make it possible for companies to store this explosion of data. Thus, the problem for forecasters is no longer having enough data, it's how to use the huge amount of data they now have for forecasting purposes. Data mining helps forecasters use this wealth of data to develop useful forecasts.

Uses for Data Mining

Data mining is a process of exploring, categorizing, and modeling large amounts of data in order to identify meaningful, logical patterns and relationships among key variables. It is used to discover trends, predict future events, and assess possible courses of action. For example, police can use data mining to predict when and where crimes might occur; weather analysts use data mining to discover weather patterns and make forecasts; utility companies use data mining to forecast energy consumption under different weather conditions in different geographic regions; hotels can use data mining to determine returning customers' preferences. Data mining allows companies to search for useful information such as patterns or trends that might be useful for forecasting—for example, to determine if certain products are sold together, if events trigger the sale of certain products, or, if more of a product is sold in a certain geographic region. Data mining can help companies like Amazon or Barnes & Noble determine what types of books or music various groups of customers categorized by such things as age, gender, income, or location might prefer; or help denim jeans companies identify fashion trends.

Data mining for business forecasting purposes is different from time series and regression methods. In these traditional forecasting methods forecasters first attempt to determine if there is a pattern in the demand data such as a trend or seasonality, and then find the forecasting model (i.e., exponential smoothing, weighted average, regression, etc.) that best fits the data pattern. When they find a model that reflects the data they use it to forecast future demand. The opposite happens with data mining; forecasters don't attempt to know what patterns may fit the data; many times, forecasters don't even know what kind of pattern they're looking for and or that may be found. Instead, data mining is a way of letting the data identify patterns and then using that information for forecasting.

Tools of Data Mining

Data mining uses and analyzes data that are stored in *databases*, *data warehouses*, and *data marts*. A database is a collection of related data organized to help a company analyze some relevant activity—for example, historical sales data for a specific product. A data warehouse is typically a company's repository for its own current and historical data, including information from its own operations (i.e., marketing, sales, ERP systems, supply chain, and external data) for such things as markets, demographics, customers, materials, competitors, suppliers, and financial data. Data warehouses can be subdivided into data marts, which store subsets of data that hold special information, and data that is grouped to help the company make decisions. Data mining employs several tools or tasks that are used to extract patterns and relationships from databases, data warehouses, and data marts. All of these categories of tools employ various software packages, such as SAS, SPSS, and Excel products, to mine data—to import, process, and analyze data, and help identify patterns and relationships.

Association rule learning is a data mining tool that searches for relationships between variables, such as a retail store determining which apparel products are frequently purchased together by certain categories of customers, such as a shirt with denim jeans; or common features across a variety of product variations that a group of customers prefer. *Clustering analysis* is a tool that uses software packages to identify groups in the data that naturally fall together, such as suppliers who use a type of packaging or method of transportation, or an age or income group that purchases a type of jeans. *Classification* tools attempt to distinguish different classes of objects or actions. For example, an email may be classified as "legitimate" or "spam," or a company or product (like jeans) may be considered "green" or not. *Prediction* tools are adopted from traditional forecasting methods, such as regression, for predicting a variable's value, like monthly demand. *Summarization* provides a compact representation of the data, including a report.

Summary

Forecasts of product demand are a necessity for almost all aspects of operational planning. Short-range demand forecasts determine the daily resource requirements needed for production, including labor and material, as well as for developing work schedules and shipping dates and controlling inventory levels. Long-range forecasts are needed to plan new products for development and changes in existing products and to acquire the plant, equipment, personnel, resources, and supply chain necessary for future operations.

We have presented several methods of forecasting useful for different time frames. These quantitative forecasting techniques are easy to understand, simple to use, and not especially costly unless the data requirements are substantial. They also have exhibited a good track record of performance for many companies that have used them. For these reasons, regression methods, and especially times series, are popular.

When managers and students are first introduced to forecasting methods, they are sometimes surprised and disappointed at the lack of exactness of the forecasts. However, they soon learn that forecasting is not easy, and exactness is not possible. Nonetheless, companies that have the skill and experience to obtain more accurate forecasts than their competitors' will gain a competitive edge.

Key Terms

adjusted exponential smoothing An exponential smoothing forecast adjusted for trend.

average error The cumulative error averaged over the number of time periods.

coefficient of determination The correlation coefficient squared; it measures the portion of the variation in the dependent variable that can be attributed to the independent variable.

correlation A measure of the strength of the causal relationship between the independent and dependent variables in a linear regression equation.

cumulative error A sum of the forecast errors; also known as bias.

cycle An up-and-down movement in demand over time.

data mining Is a process for analyzing large amounts of data to identify patterns, trends and relationships in groups.

Delphi method A procedure for acquiring informed judgments and opinions from knowledgeable individuals to use as a subjective forecast.

exponential smoothing An averaging method that weights the most recent data more strongly than more distant data.

forecast error The difference between actual and forecasted demand.

linear regression A mathematical technique that relates a dependent variable to an independent variable in the form of a linear equation.

linear trend line A forecast using the linear regression equation to relate demand to time.

long-range forecast A forecast encompassing a period longer than two years into the future.

mean absolute deviation (MAD) The per-period average of the absolute difference between actual and forecasted demand.

mean absolute percent deviation (MAPD) The absolute forecast error measured as a percentage of demand.

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

moving average Average demand for a fixed sequence of periods including the most recent period.

multiple regression A mathematical relationship that relates a dependent variable to two or more independent variables.

qualitative forecast methods Nonquantitative, subjective forecasts based on judgment, opinion, experience, and expert opinion.

quantitative forecast methods Forecasts derived from a mathematical formula.

random variations Movements in demand that are not predictable and follow no pattern.

regression forecasting methods A class of mathematical techniques that relate demand to factors that cause demand behavior.

seasonal factor A numerical value that is multiplied by the normal forecast to get a seasonally adjusted forecast.

seasonal pattern An oscillating movement in demand that occurs periodically in the short run and is repetitive.

short- to mid-range forecast A forecast encompassing the immediate future, usually days or weeks, but up to two years.

smoothing constant The weighting factor given to the most recent data in exponential smoothing forecasts.

time frame How far into the future something is forecasted.

time series methods A class of statistical methods that uses historical demand data over a period of time to predict future demand.

tracking signal A measure computed by dividing the cumulative error by MAD; used for monitoring bias in a forecast.

trend A gradual, long-term up or down movement of demand.

weighted moving average A moving average with more recent demand values adjusted with weights.