20

Forecasting

How much will the economy grow over the next year? Where is the stock market headed? What about interest rates? How will consumer tastes be changing? What will be the hot new products?

Forecasters have answers to all these questions. Unfortunately, these answers will more than likely be wrong. Nobody can accurately predict the future every time.

Nevertheless, the future success of any business depends heavily on how savvy its management is in spotting trends and developing appropriate strategies. The leaders of the best companies often seem to have a sixth sense for when to change direction to stay a step ahead of the competition. These companies seldom get into trouble by badly misestimating what the demand will be for their products. Many other companies do. The ability to forecast well makes the difference.

The preceding chapter has presented a considerable number of models for the management of inventories. All these models are based on a forecast of future demand for a product, or at least a probability distribution for that demand. Therefore, the missing ingredient for successfully implementing these inventory models is an approach for forecasting demand.

Fortunately, when historical sales data are available, some proven **statistical forecasting methods** have been developed for using these data to forecast future demand. Such a method assumes that historical trends will continue, so management then needs to make any adjustments to reflect current changes in the marketplace.

Several **judgmental forecasting methods** that solely use expert judgment also are available. These methods are especially valuable when little or no historical sales data are available or when major changes in the marketplace make these data unreliable for forecasting purposes.

Forecasting product demand is just one important application of the various forecasting methods. A variety of applications are surveyed in the first section. The second section outlines the main judgmental forecasting methods. Section 20.3 then describes *time series,* which form the basis for the statistical forecasting methods presented in the subsequent five sections. Section 20.9 turns to another important type of statistical forecasting method, *regression analysis,* where the variable to be forecasted is expressed as a mathematical function of one or more other variables whose values will be known at the time of the forecast. The chapter then concludes by surveying forecasting practices in U.S. corporations.

20.1 SOME APPLICATIONS OF FORECASTING

We now will discuss some main areas in which forecasting is widely used. In each case, we will illustrate this use by mentioning one or more actual applications that have been described in published articles. A summary table at the end of the section will tell you where these articles can be found in case you want to read further.

Sales Forecasting

Any company engaged in selling goods needs to forecast the demand for those goods. Manufacturers need to know how much to produce. Wholesalers and retailers need to know how much to stock. Substantially underestimating demand is likely to lead to many lost sales, unhappy customers, and perhaps allowing the competition to gain the upper hand in the marketplace. On the other hand, significantly overestimating demand also is very costly due to (1) excessive inventory costs, (2) forced price reductions, (3) unneeded production or storage capacity, and (4) lost opportunities to market more profitable goods. Successful marketing and production managers understand very well the importance of obtaining good sales forecasts.

The *Merit Brass Company* is a family-owned company that supplies several thousand products to the pipe, valve, and fittings industry. In 1990, Merit Brass embarked on a modernization program that emphasized installing OR methodologies in statistical sales forecasting and finished-goods inventory management (two activities that go hand in glove). This program led to major improvements in customer service (as measured by product availability) while simultaneously achieving substantial cost reductions.

A major Spanish electric utility, *Hidroeléctrica Español,* has developed and implemented a hierarchy of OR models to assist in managing its system of reservoirs used for generating hydroelectric power. All these models are driven by forecasts of both energy demand (this company's sales) and reservoir inflows. A sophisticated statistical forecasting method is used to forecast energy demand on both a short-term and long-term basis. A hydrological forecasting model generates the forecasts of reservoir inflows.

Airline companies now depend heavily on the high fares paid by business people traveling on short notice while providing discount fares to others to help fill the seats. The decision on how to allocate seats to the different fare classes is a crucial one for maximizing revenue. *American Airlines,* for example, uses statistical forecasting of the demand at each fare to make this decision.

Forecasting the Need for Spare Parts

Although effective sales forecasting is a key for virtually any company, some organizations must rely on other types of forecasts as well. A prime example involves forecasts of the need for spare parts.

Many companies need to maintain an inventory of spare parts to enable them to quickly repair either their own equipment or their products sold or leased to customers. In some cases, this inventory is huge. For example, IBM's spare-parts inventory described in Sec. 19.8 is valued in the billions of dollars and includes many thousand different parts.

Just as for a finished-goods inventory ready for sale, effective management of a spare-parts inventory depends upon obtaining a reliable forecast of the demand for that inven-

tory. Although the types of costs incurred by misestimating demand are somewhat different, the consequences may be no less severe for spare parts. For example, the consequence for an airline not having a spare part available on location when needed to continue flying an airplane probably is at least one canceled flight.

To support its operation of several hundred aircraft, *American Airlines* maintains an extensive inventory of spare parts. Included are over 5,000 different types of rotatable parts (e.g., landing gear and wing flaps) with an average value of $5,000 per item. When a rotatable part on an airplane is found to be defective, it is immediately replaced by a corresponding part in inventory so the airplane can depart. However, the replaced part then is repaired and placed back into inventory for subsequent use as a replacement part.

American Airlines uses a PC-based forecasting system called the Rotatables Allocation and Planning System (RAPS) to forecast demand for the rotatable parts and to help allocate these parts to the various airports. The statistical forecast uses an 18-month history of parts usage and flying hours for the fleet, and then projects ahead based on planned flying hours.

Forecasting Production Yields

The yield of a production process refers to the percentage of the completed items that meet quality standards (perhaps after rework) and so do not need to be discarded. Particularly with high-technology products, the yield frequently is well under 100 percent.

If the forecast for the production yield is somewhat under 100 percent, the size of the production run probably should be somewhat larger than the order quantity to provide a good chance of fulfilling the order with acceptable items. (The difference between the run size and the order quantity is referred to as the *reject allowance*.) If an expensive setup is required for each production run, or if there is only time for one production run, the reject allowance may need to be quite large. However, an overly large value should be avoided to prevent excessive production costs.

Obtaining a reliable forecast of production yield is essential for choosing an appropriate value of the reject allowance.

This was the case for the *Albuquerque Microelectronics Operation,* a dedicated production source for radiation-hardened microchips. The first phase in the production of its microchips, the wafer fabrication process, was continuing to provide erratic production yields. For a given product, the yield typically would be quite small (0 to 40 percent) for the first several lots and then would gradually increase to a higher range (35 to 75 percent) for later lots. Therefore, a statistical forecasting method that considered this increasing trend was used to forecast the production yield.

Forecasting Economic Trends

With the possible exception of sales forecasting, the most extensive forecasting effort is devoted to forecasting economic trends on a regional, national, or even international level. How much will the nation's gross domestic product grow next quarter? Next year? What is the forecast for the rate of inflation? The unemployment rate? The balance of trade?

Statistical models to forecast economic trends (commonly called *econometric models*) have been developed in a number of governmental agencies, university research centers, large corporations, and consulting firms, both in the United States and elsewhere.

Using historical data to project ahead, these econometric models typically consider a very large number of factors that help drive the economy. Some models include hundreds of variables and equations. However, except for their size and scope, these models resemble some of the statistical forecasting methods used by businesses for sales forecasting, etc.

These econometric models can be very influential in determining governmental policies. For example, the forecasts provided by the U.S. Congressional Budget Office strongly guide Congress in developing the federal budgets. These forecasts also help businesses in assessing the general economic outlook.

As an example on a smaller scale, the *U.S. Department of Labor* contracted with a consulting firm to develop the *unemployment insurance econometric forecasting model* (UIEFM). The model is now in use by state employment security agencies around the nation. By projecting such fundamental economic factors as unemployment rates, wage levels, the size of the labor force covered by unemployment insurance, etc., UIEFM forecasts how much the state will need to pay in unemployment insurance. By projecting tax inflows into the state's unemployment insurance trust fund, UIEFM also forecasts trust fund balances over a 10-year period. Therefore, UIEFM has proved to be invaluable in managing state unemployment insurance systems and in guiding related legislative policies.

Forecasting Staffing Needs

One of the major trends in the American economy is a shifting emphasis from manufacturing to services. More and more of our manufactured goods are being produced outside the country (where labor is cheaper) and then imported. At the same time, an increasing number of American business firms are specializing in providing a service of some kind (e.g., travel, tourism, entertainment, legal aid, health services, financial, educational, design, maintenance, etc.). For such a company, forecasting "sales" becomes forecasting the demand for services, which then translates into forecasting staffing needs to provide those services.

For example, one of the fastest-growing service industries in the United States today is call centers. A call center receives telephone calls from the general public requesting a particular type of service. Depending on the center, the service might be providing technical assistance over the phone, or making a travel reservation, or filling a telephone order for goods, or booking services to be performed later, etc. There now are more than 350,000 call centers in the United States, with over $25 billion invested to date and an annual growth rate of 20 percent.

As with any service organization, an erroneous forecast of staffing requirements for a call center has serious consequences. Providing too few agents to answer the telephone leads to unhappy customers, lost calls, and perhaps lost business. Too many agents cause excessive personnel costs.

Section 3.5 described a major OR study that involved personnel scheduling at *United Airlines*. With over 4,000 reservations sales representatives and support personnel at its 11 reservations offices, and about 1,000 customer service agents at its 10 largest airports, a computerized planning system was developed to design the work schedules for these employees. Although several other OR techniques (including linear programming) were incorporated into this system, statistical forecasting of staffing requirements also was a key ingredient. This system provided annual savings of over $6 million as well as improved customer service and reduced support staff requirements.

TABLE 20.1 Some applications of statistical forecasting methods

Organization	Quantity Being Forecasted	Issue of *Interfaces*
Merit Brass Co.	Sales of finished goods	Jan.–Feb. 1993
Hidroeléctrica Español	Energy demand	Jan.–Feb. 1990
American Airlines	Demand for different fare classes	Jan.–Feb. 1992
American Airlines	Need for spare parts to repair airplanes	July–Aug. 1989
Albuquerque Microelectronics	Production yield in wafer fabrication	March–April 1994
U.S. Department of Labor	Unemployment insurance payments	March–April 1988
United Airlines	Demand at reservations offices and airports	Jan.–Feb. 1986
L.L. Bean	Staffing needs at call center	Nov.–Dec. 1995

L.L. Bean is a major retailer of high-quality outdoor goods and apparel. Over 70 percent of its total sales volume is generated through orders taken at the company's call center. Two 800 numbers are provided, one for placing orders and the second for making inquiries or reporting problems. Each of the company's agents is trained to answer just one of the 800 numbers. Therefore, separate statistical forecasting models were developed to forecast staffing requirements for the two 800 numbers on a weekly basis. The improved precision of these models is estimated to have saved L.L. Bean $300,000 annually through enhanced scheduling efficiency.

Other

Table 20.1 summarizes the actual applications of statistical forecasting methods presented in this section. The last column cites the issue of *Interfaces* which includes the article that describes each application in detail.

All five categories of forecasting applications discussed in this section use the types of forecasting methods presented in the subsequent sections. There also are other important categories (including forecasting weather, the stock market, and prospects for new products before market testing) that use specialized techniques that are not discussed here.

20.2 JUDGMENTAL FORECASTING METHODS

Judgmental forecasting methods are, by their very nature, subjective, and they may involve such qualities as intuition, expert opinion, and experience. They generally lead to forecasts that are based upon qualitative criteria.

These methods may be used when no data are available for employing a statistical forecasting method. However, even when good data are available, some decision makers prefer a judgmental method instead of a formal statistical method. In many other cases, a combination of the two may be used.

Here is a brief overview of the main judgmental forecasting methods.

1. **Manager's opinion:** This is the most informal of the methods, because it simply involves a single manager using his or her best judgment to make the forecast. In some cases, some data may be available to help make this judgment. In others, the manager may be drawing solely on experience and an intimate knowledge of the current conditions that drive the forecasted quantity.

2. **Jury of executive opinion:** This method is similar to the first one, except now it involves a small group of high-level managers who pool their best judgment to collectively make the forecast. This method may be used for more critical forecasts for which several executives share responsibility and can provide different types of expertise.
3. **Sales force composite:** This method is often used for sales forecasting when a company employs a sales force to help generate sales. It is a *bottom-up approach* whereby each salesperson provides an estimate of what sales will be in his or her region. These estimates then are sent up through the corporate chain of command, with managerial review at each level, to be aggregated into a corporate sales forecast.
4. **Consumer market survey:** This method goes even further than the preceding one in adopting a *grass-roots approach* to sales forecasting. It involves surveying customers and potential customers regarding their future purchasing plans and how they would respond to various new features in products. This input is particularly helpful for designing new products and then in developing the initial forecasts of their sales. It also is helpful for planning a marketing campaign.
5. **Delphi method:** This method employs a panel of experts in different locations who independently fill out a series of questionnaires. However, the results from each questionnaire are provided with the next one, so each expert then can evaluate this group information in adjusting his or her responses next time. The goal is to reach a relatively narrow spread of conclusions from most of the experts. The decision makers then assess this input from the panel of experts to develop the forecast. This involved process normally is used only at the highest levels of a corporation or government to develop long-range forecasts of broad trends.

The decision on whether to use one of these judgmental forecasting methods should be based on an assessment of whether the individuals who would execute the method have the background needed to make an informed judgment. Another factor is whether the expertise of these individuals or the availability of relevant historical data (or a combination of both) appears to provide a better basis for obtaining a reliable forecast.

The next seven sections discuss statistical forecasting methods based on relevant historical data.

20.3 TIME SERIES

Most statistical forecasting methods are based on using historical data from a *time series.*

> A **time series** is a series of observations over time of some quantity of interest (a random variable). Thus, if X_i is the random variable of interest at time i, and if observations are taken at times[1] $i = 1, 2, \ldots, t$, then the observed values $\{X_1 = x_1, X_2 = x_2, \ldots, X_t = x_t\}$ are a time series.

For example, the recent monthly sales figures for a product comprises a time series, as illustrated in Fig. 20.1.

Because a time series is a description of the past, a logical procedure for forecasting the future is to make use of these historical data. If the past data are indicative of what

---

[1]These times of observation sometimes are actually time periods (months, years, etc.), so we often will refer to the times as periods.
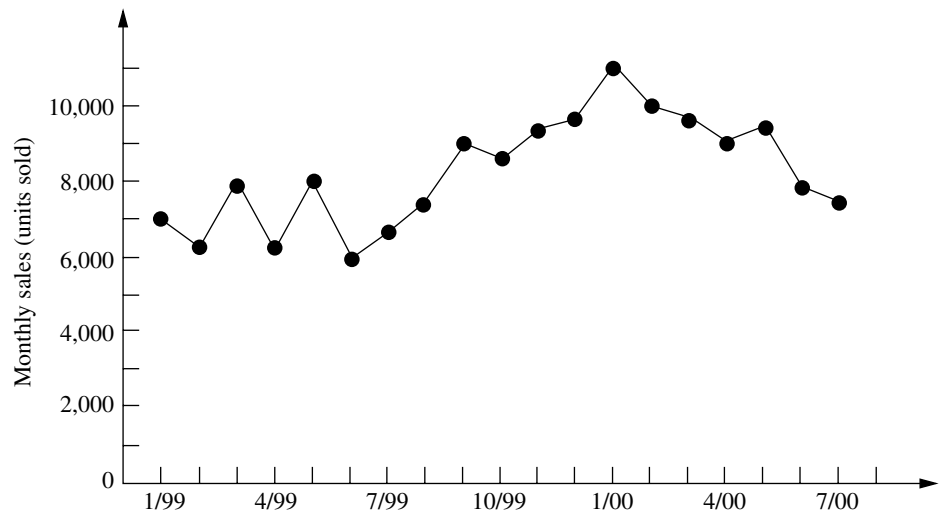
FIGURE 20.1
The evolution of the monthly sales of a product illustrates a time series.

we can expect in the future, we can postulate an underlying mathematical model that is representative of the process. The model can then be used to generate forecasts.

In most realistic situations, we do not have complete knowledge of the exact form of the model that generates the time series, so an approximate model must be chosen. Frequently, the choice is made by observing the pattern of the time series. Several typical time series patterns are shown in Fig. 20.2. Figure 20.2*a* displays a typical time series if the generating process were represented by a **constant level** superimposed with random fluctuations. Figure 20.2*b* displays a typical time series if the generating process were represented by a **linear trend** superimposed with random fluctuations. Finally, Fig. 20.2*c* shows a time series that might be observed if the generating process were represented by a constant level superimposed with a **seasonal effect** together with random fluctuations. There are many other plausible representations, but these three are particularly useful in practice and so are considered in this chapter.

FIGURE 20.2
Typical time series patterns, with random fluctuations around (*a*) a constant level, (*b*) a linear trend, and (*c*) a constant level plus seasonal effects.
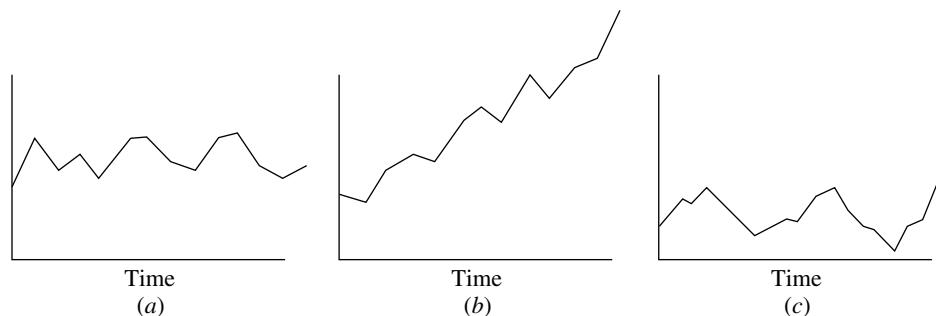
Time
(*a*)

Time
(*b*)

Time
(*c*)

Once the form of the model is chosen, a mathematical representation of the generating process of the time series can be given. For example, suppose that the generating process is identified as a **constant-level model** superimposed with random fluctuations, as illustrated in Fig. 20.2a. Such a representation can be given by

$$X_i = A + e_i, \qquad \text{for } i = 1, 2, \ldots,$$

where X_i is the random variable observed at time i, A is the constant level of the model, and e_i is the random error occurring at time i (assumed to have expected value equal to zero and constant variance). Let

F_{t+1} = forecast of the values of the time series at time $t + 1$, given the observed values, $X_1 = x_1, X_2 = x_2, \ldots, X_t = x_t$.

Because of the random error e_{t+1}, it is impossible for F_{t+1} to predict the value $X_{t+1} = x_{t+1}$ precisely, but the goal is to have F_{t+1} estimate the constant level $A = E(X_{t+1})$ as closely as possible. It is reasonable to expect that F_{t+1} will be a function of at least some of the observed values of the time series.

20.4   FORECASTING METHODS FOR A CONSTANT-LEVEL MODEL

We now present four alternative forecasting methods for the constant-level model introduced in the preceding paragraph. This model, like any other, is only intended to be an idealized representation of the actual situation. For the real time series, at least small shifts in the value of A may be occurring occasionally. Each of the following methods reflects a different assessment of how recently (if at all) a significant shift may have occurred.

Last-Value Forecasting Method

By interpreting t as the *current time,* the last-value forecasting procedure uses the value of the time series observed at time t (x_t) as the forecast at time $t + 1$. Therefore,

$$F_{t+1} = x_t.$$

For example, if x_t represents the sales of a particular product in the quarter just ended, this procedure uses these sales as the forecast of the sales for the next quarter.

This forecasting procedure has the disadvantage of being imprecise; i.e., its variance is large because it is based upon a sample of size 1. It is worth considering only if (1) the underlying assumption about the constant-level model is "shaky" and the process is changing so rapidly that anything before time t is almost irrelevant or misleading or (2) the assumption that the random error e_t has constant variance is unreasonable and the variance at time t actually is much smaller than at previous times.

The last-value forecasting method sometimes is called the **naive method,** because statisticians consider it naive to use just a *sample size of one* when additional relevant data are available. However, when conditions are changing rapidly, it may be that the last value is the only relevant data point for forecasting the next value under current conditions. Therefore, decision makers who are anything but naive do occasionally use this method under such circumstances.

Averaging Forecasting Method

This method goes to the other extreme. Rather than using just a sample size of one, this method uses *all* the data points in the time series and simply *averages* these points. Thus, the forecast of what the next data point will turn out to be is

$$F_{t+1} = \sum_{i=1}^{t} \frac{x_i}{t}.$$

This estimate is an excellent one if the process is entirely stable, i.e., if the assumptions about the underlying model are correct. However, frequently there exists skepticism about the persistence of the underlying model over an extended time. Conditions inevitably change eventually. Because of a natural reluctance to use very old data, this procedure generally is limited to young processes.

Moving-Average Forecasting Method

Rather than using very old data that may no longer be relevant, this method averages the data for only the last n periods as the forecast for the next period, i.e.,

$$F_{t+1} = \sum_{i=t-n+1}^{t} \frac{x_i}{n}.$$

Note that this forecast is easily updated from period to period. All that is needed each time is to lop off the first observation and add the last one.

The *moving-average* estimator combines the advantages of the *last value* and *averaging* estimators in that it uses only recent history *and* it uses multiple observations. A disadvantage of this method is that it places as much weight on x_{t-n+1} as on x_t. Intuitively, one would expect a good method to place more weight on the most recent observation than on older observations that may be less representative of current conditions. Our next method does just this.

Exponential Smoothing Forecasting Method

This method uses the formula

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

where α ($0 < \alpha < 1$) is called the **smoothing constant.** (The choice of α is discussed later.) Thus, the forecast is just a weighted sum of the last observation x_t and the preceding forecast F_t for the period just ended. Because of this recursive relationship between F_{t+1} and F_t, alternatively F_{t+1} can be expressed as

$$F_{t+1} = \alpha x_t + \alpha(1 - \alpha)x_{t-1} + \alpha(1 - \alpha)^2 x_{t-2} + \cdots.$$

In this form, it becomes evident that exponential smoothing gives the most weight to x_t and decreasing weights to earlier observations. Furthermore, the first form reveals that the forecast is simple to calculate because the data prior to period t need not be retained; all that is required is x_t and the previous forecast F_t.

Another alternative form for the exponential smoothing technique is given by

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

which gives a heuristic justification for this method. In particular, the forecast of the time series at time $t + 1$ is just the preceding forecast at time t plus the *product* of the forecasting error at time t and a discount factor α. This alternative form is often simpler to use.

A measure of effectiveness of exponential smoothing can be obtained under the assumption that the process is completely stable, so that $X_1, X_2, \ldots$ are independent, identically distributed random variables with variance σ^2. It then follows that (for large t)

$$\text{var}[F_{t+1}] \approx \frac{\alpha\sigma^2}{2 - \alpha} = \frac{\sigma^2}{(2 - \alpha)/\alpha},$$

so that the variance is statistically equivalent to a moving average with $(2 - \alpha)/\alpha$ observations. For example, if α is chosen equal to 0.1, then $(2 - \alpha)/\alpha = 19$. Thus, in terms of its variance, the exponential smoothing method with this value of α is *equivalent* to the moving-average method that uses 19 observations. However, if a change in the process does occur (e.g., if the mean starts increasing), exponential smoothing will react more quickly with better tracking of the change than the moving-average method.

An important drawback of exponential smoothing is that it lags behind a continuing trend; i.e., if the constant-level model is incorrect and the mean is increasing steadily, then the forecast will be several periods behind. However, the procedure can be easily adjusted for trend (and even seasonally adjusted).

Another disadvantage of exponential smoothing is that it is difficult to choose an appropriate smoothing constant α. Exponential smoothing can be viewed as a statistical filter that inputs raw data from a stochastic process and outputs smoothed estimates of a mean that varies with time. If α is chosen to be small, response to change is slow, with resultant smooth estimators. On the other hand, if α is chosen to be large, response to change is fast, with resultant large variability in the output. Hence, there is a need to compromise, depending upon the degree of stability of the process. It has been suggested that α should not exceed 0.3 and that a reasonable choice for α is approximately 0.1. This value can be increased temporarily if a change in the process is expected or when one is just starting the forecasting. At the start, a reasonable approach is to choose the forecast for period 2 according to

$$F_2 = \alpha x_1 + (1 - \alpha)(\text{initial estimate}),$$

where some initial estimate of the constant level A must be obtained. If past data are available, such an estimate may be the average of these data.

Your OR Courseware includes a pair of Excel templates for each of the four forecasting methods presented in this section. In each use, one template (*without seasonality*) applies the method just as described here. The second template (*with seasonality*) also incorporates into the method the seasonal factors discussed in the next section.

20.5   INCORPORATING SEASONAL EFFECTS INTO FORECASTING METHODS

It is fairly common for a time series to have a *seasonal pattern* with higher values at certain times of the year than others. For example, this occurs for the sales of a product that is a popular choice for Christmas gifts. Such a time series violates the basic assumption

of a *constant-level model,* so the forecasting methods presented in the preceding section should not be applied directly.

Fortunately, it is relatively straightforward to make *seasonal adjustments* in such a time series so that these forecasting methods based on a constant-level model can still be applied. We will illustrate the procedure with the following example.

Example.   The COMPUTER CLUB WAREHOUSE (commonly referred to as CCW) sells various computer products at bargain prices by taking telephone orders directly from customers at its call center. Figure 20.3 shows the average number of calls received per day in each of the four quarters of the past three years. Note how the call volume jumps up sharply in each Quarter 4 because of Christmas sales. There also is a tendency for the call volume to be a little higher in Quarter 3 than in Quarter 1 or 2 because of back-to-school sales.

To quantify these seasonal effects, the second column of Table 20.2 shows the average daily call volume for each quarter over the past three years. Underneath this column, the *overall average* over all four quarters is calculated to be 7,529. Dividing the average for each quarter by this overall average gives the *seasonal factor* shown in the third column.

> In general, the **seasonal factor** for any period of a year (a quarter, a month, etc.) measures how that period compares to the overall average for an entire year. Specifically, using historical data, the seasonal factor is calculated to be
>
> $$\text{Seasonal factor} = \frac{\text{average for the period}}{\text{overall average}}.$$

Your OR Courseware includes an Excel template for calculating these seasonal factors.

FIGURE 20.3
The average number of calls received per day at the CCW call center in each of the four quarters of the past three years.

	A	B	C	D	E	F	G	H	I
1		CCW's Average Daily Call Volume							
2									
3		Year	Quarter	Call Volume					
4		1	1	6,809					
5		1	2	6,465					
6		1	3	6,569					
7		1	4	8,266					
8		2	1	7,257					
9		2	2	7,064					
10		2	3	7,784					
11		2	4	8,724					
12		3	1	6,992					
13		3	2	6,822					
14		3	3	7,949					
15		3	4	9,650					
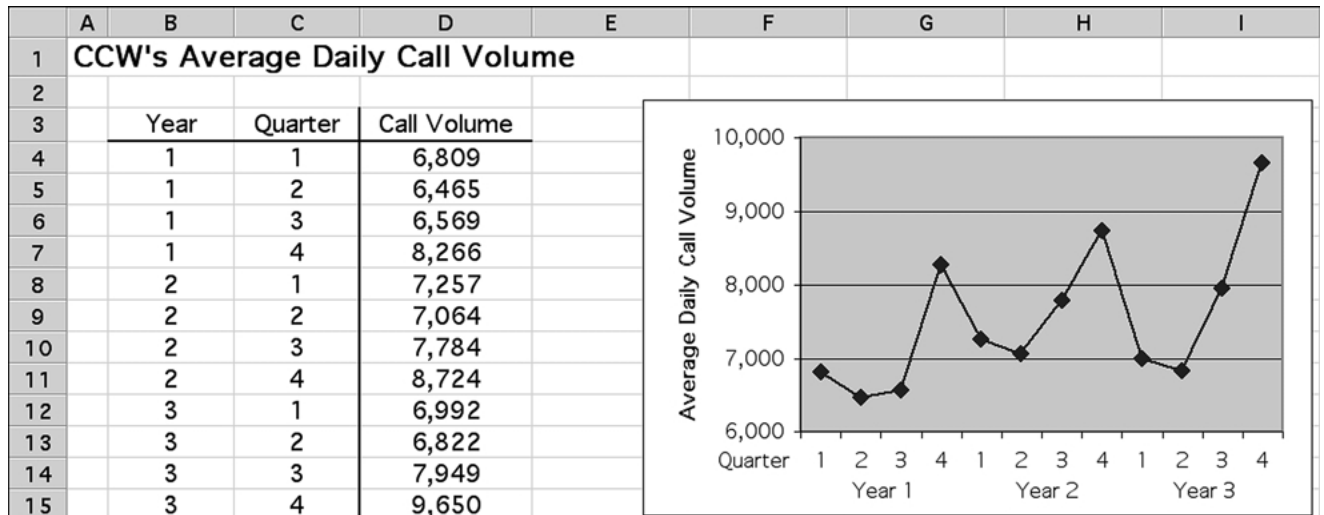
TABLE 20.2 Calculation of the seasonal factors for the CCW problem

Quarter	Three-Year Average	Seasonal Factor
1	7,019	$\dfrac{7,019}{7,529} = 0.93$
2	6,784	$\dfrac{6,784}{7,529} = 0.90$
3	7,434	$\dfrac{7,434}{7,529} = 0.99$
4	8,880	$\dfrac{8,880}{7,529} = 1.18$

$$\text{Total} = 30,117$$
$$\text{Average} = \frac{30,117}{4} = 7,529.$$

The Seasonally Adjusted Time Series

It is much easier to analyze a time series and detect new trends if the data are first adjusted to remove the effect of seasonal patterns. To remove the seasonal effects from the time series shown in Fig. 20.3, each of these average daily call volumes needs to be divided by the corresponding seasonal factor given in Table 20.2. Thus, the formula is

$$\text{Seasonally adjusted call volume} = \frac{\text{actual call volume}}{\text{seasonal factor}}.$$

Applying this formula to all 12 call volumes in Fig. 20.3 gives the seasonally adjusted call volumes shown in column F of the spreadsheet in Fig. 20.4.
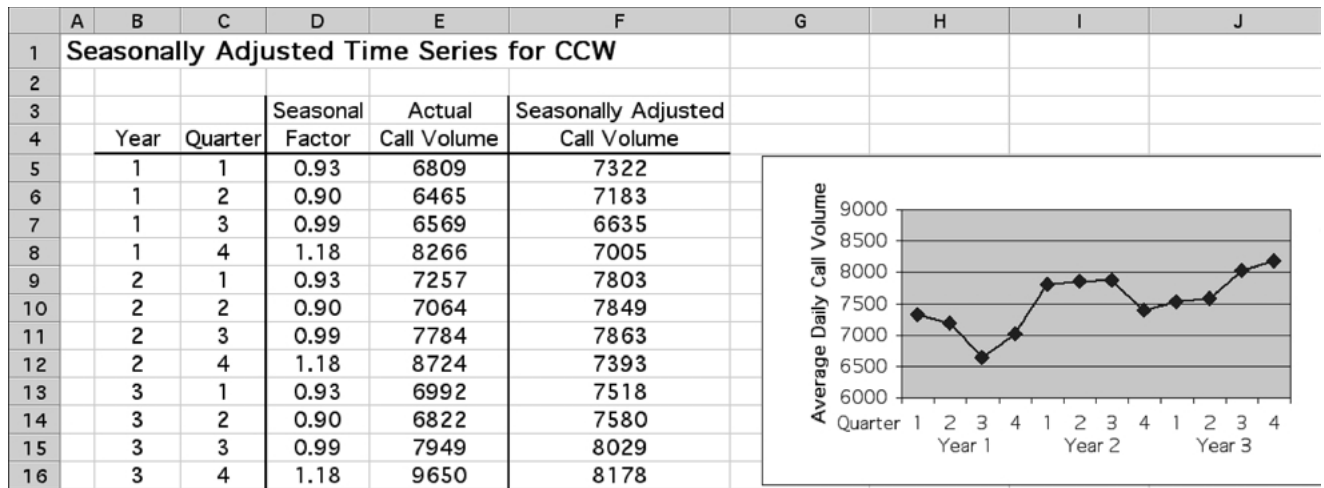
In effect, these seasonally adjusted call volumes show what the call volumes would have been if the calls that occur because of the time of the year (Christmas shopping, back-to-school shopping, etc.) had been spread evenly throughout the year instead. Compare the plots in Figs. 20.4 and 20.3. After considering the smaller vertical scale in Fig. 20.4, note how much less fluctuation this figure has than Fig. 20.3 because of removing seasonal effects. However, this figure still is far from completely flat because fluctuations in call volume occur for other reasons beside just seasonal effects. For example, hot new products attract a flurry of calls. A jump also occurs just after the mailing of a catalog. Some random fluctuations occur without any apparent explanation. Figure 20.4 enables seeing and analyzing these fluctuations in sales volumes that are not caused by seasonal effects.

The General Procedure

After seasonally adjusting a time series, any of the forecasting methods presented in the preceding section (or the next section) can then be applied. Here is an outline of the general procedure.

1. Use the following formula to seasonally adjust each value in the time series:

$$\text{Seasonally adjusted value} = \frac{\text{actual value}}{\text{seasonal factor}}.$$

	A	B	C	D	E	F	G	H	I	J
1		Seasonally Adjusted Time Series for CCW								
2										
3				Seasonal	Actual	Seasonally Adjusted				
4		Year	Quarter	Factor	Call Volume	Call Volume				
5		1	1	0.93	6809	7322				
6		1	2	0.90	6465	7183				
7		1	3	0.99	6569	6635				
8		1	4	1.18	8266	7005				
9		2	1	0.93	7257	7803				
10		2	2	0.90	7064	7849				
11		2	3	0.99	7784	7863				
12		2	4	1.18	8724	7393				
13		3	1	0.93	6992	7518				
14		3	2	0.90	6822	7580				
15		3	3	0.99	7949	8029				
16		3	4	1.18	9650	8178				

	F
5	=E5/D5
6	=E6/D6
7	=E7/D7
8	=E8/D8
9	:
10	:

FIGURE 20.4
The seasonally adjusted time series for the CCW problem obtained by dividing each actual average daily call volume in Fig. 20.3 by the corresponding seasonal factor obtained in Table 20.2.

2. Select a time series forecasting method.
3. Apply this method to the seasonally adjusted time series to obtain a forecast of the next *seasonally adjusted* value (or values).
4. Multiply this forecast by the corresponding seasonal factor to obtain a forecast of the next *actual* value (without seasonal adjustment).

As mentioned at the end of the preceding section, an Excel template that incorporates seasonal effects is available in your OR Courseware for each of the forecasting methods to assist you with combining the method with this procedure.

20.6 AN EXPONENTIAL SMOOTHING METHOD FOR A LINEAR TREND MODEL

Recall that the constant-level model introduced in Sec. 20.3 assumes that the sequence of random variables $\{X_1, X_2, \ldots, X_t\}$ generating the time series has a constant expected value denoted by A, where the goal of the forecast F_{t+1} is to estimate A as closely as possible. However, as was illustrated in Fig. 20.2b, some time series violate this assumption by having a continuing trend where the expected values of successive random variables

keep changing in the same direction. Therefore, a forecasting method based on the constant-level model (perhaps after adjusting for seasonal effects) would do a poor job of forecasting for such a time series because it would be continually lagging behind the trend. We now turn to another model that is designed for this kind of time series.

Suppose that the generating process of the observed time series can be represented by a *linear trend* superimposed with *random fluctuations,* as illustrated in Fig. 20.2*b*. Denote the slope of the linear trend by *B,* where the slope is called the **trend factor.** The model is represented by

$$X_i = A + Bi + e_i, \qquad \text{for } i = 1, 2, \ldots,$$

where X_i is the random variable that is observed at time i, A is a constant, B is the trend factor, and e_i is the random error occurring at time i (assumed to have expected value equal to zero and constant variance).

For a real time series represented by this model, the assumptions may not be completely satisfied. It is common to have at least small shifts in the values of A and B occasionally. It is important to detect these shifts relatively quickly and reflect them in the forecasts. Therefore, practitioners generally prefer a forecasting method that places substantial weight on recent observations and little if any weight on old observations. The exponential smoothing method presented next is designed to provide this kind of approach.

Adapting Exponential Smoothing to This Model

The exponential smoothing method introduced in Sec. 20.4 can be adapted to include the trend factor incorporated into this model. This is done by also using exponential smoothing to estimate this trend factor.

Let

T_{t+1} = exponential smoothing estimate of the trend factor B at time $t + 1$, given the observed values, $X_1 = x_1, X_2 = x_2, \ldots, X_t = x_t$.

Given T_{t+1}, the forecast of the value of the time series at time $t + 1$ (F_{t+1}) is obtained simply by adding T_{t+1} to the formula for F_{t+1} given in Sec. 20.4, so

$$F_{t+1} = \alpha x_t + (1 - \alpha)F_t + T_{t+1}.$$

To motivate the procedure for obtaining T_{t+1}, note that the model assumes that

$$B = E(X_{i+1}) - E(X_i), \qquad \text{for } i = 1, 2, \ldots.$$

Thus, the standard statistical estimator of B would be the *average* of the observed differences, $x_2 - x_1, x_3 - x_2, \ldots, x_t - x_{t-1}$. However, the exponential smoothing approach recognizes that the parameters of the stochastic process generating the time series (including A and B) may actually be gradually shifting over time so that the most recent observations are the most reliable ones for estimating the current parameters. Let

L_{t+1} = latest trend at time $t + 1$ based on the last two values (x_t and x_{t-1}) and the last two forecasts (F_t and F_{t-1}).

The exponential smoothing formula used for L_{t+1} is

$$L_{t+1} = \alpha(x_t - x_{t-1}) + (1 - \alpha)(F_t - F_{t-1}).$$

Then T_{t+1} is calculated as

$$T_{t+1} = \beta L_{t+1} + (1 - \beta)T_t,$$

where β is the **trend smoothing constant** which, like α, must be between 0 and 1. Calculating L_{t+1} and T_{t+1} in order then permits calculating F_{t+1} with the formula given in the preceding paragraph.

Getting started with this forecasting method requires making two initial estimates about the status of the time series just prior to beginning forecasting. These initial estimates are

x_0 = initial estimate of the *expected value* of the time series (A) if the conditions just prior to beginning forecasting were to remain unchanged without any trend;

T_1 = initial estimate of the *trend* of the time series (B) just prior to beginning forecasting.

The resulting forecasts for the first two periods are

$$F_1 = x_0 + T_1,$$
$$L_2 = \alpha(x_1 - x_0) + (1 - \alpha)(F_1 - x_0),$$
$$T_2 = \beta L_2 + (1 - \beta)T_1,$$
$$F_2 = \alpha x_1 + (1 - \alpha)F_1 + T_2.$$

The above formulas for L_{t+1}, T_{t+1}, and F_{t+1} then are used directly to obtain subsequent forecasts.

Since the calculations involved with this method are relatively involved, a computer commonly is used to implement the method. Your OR Courseware includes two Excel templates (one without seasonal adjustments and one with) for this method.
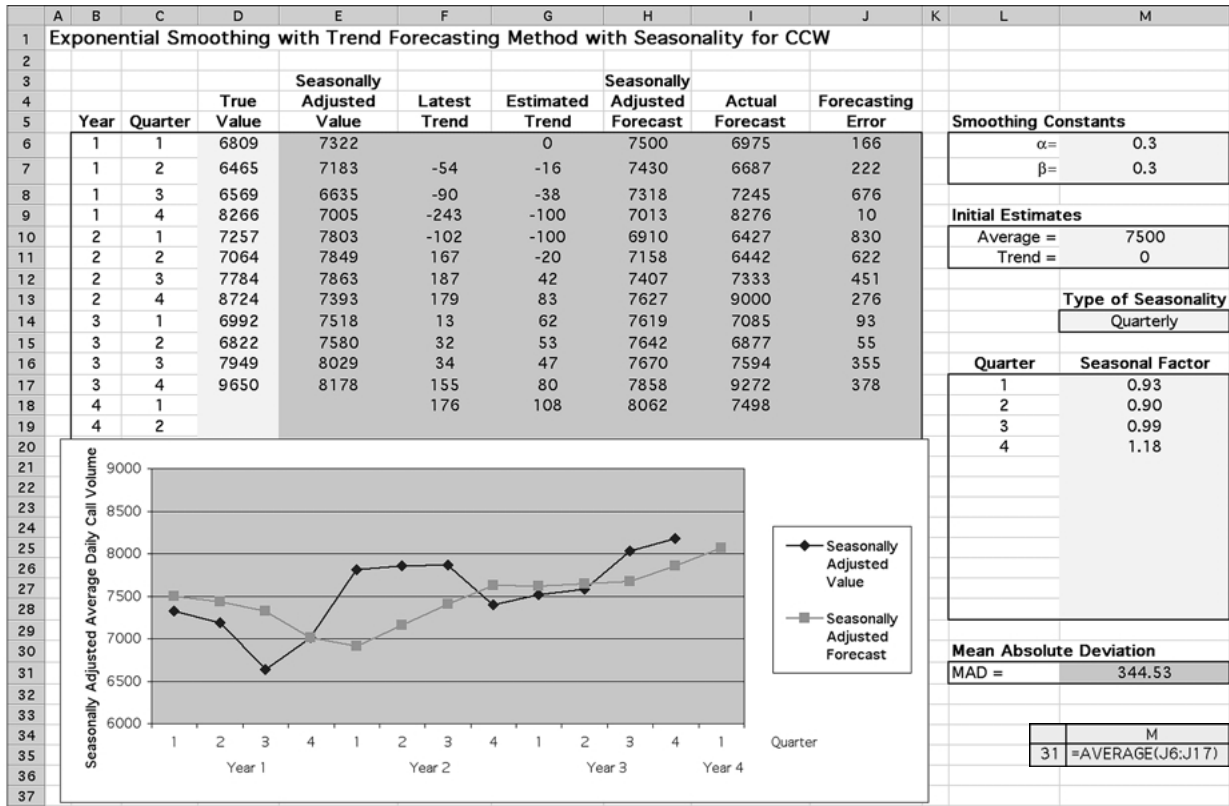
Application of the Method to the CCW Example

Reconsider the example involving the Computer Club Warehouse (CCW) that was introduced in the preceding section. Figure 20.3 shows the time series for this example (representing the average daily call volume quarterly for 3 years) and then Fig. 20.4 gives the seasonally adjusted time series based on the seasonal factors calculated in Table 20.2. We now will assume that these seasonal factors were determined *prior* to these three years of data and that the company then was using *exponential smoothing with trend* to forecast the average daily call volume quarter by quarter over the 3 years based on these data. CCW management has chosen the following initial estimates and smoothing constants:

$$x_0 = 7{,}500, \qquad T_1 = 0, \qquad \alpha = 0.3, \qquad \beta = 0.3.$$

Working with the seasonally adjusted call volumes given in Fig. 20.4, these initial estimates lead to the following seasonally adjusted forecasts.

Y1, Q1: $F_1 = 7{,}500 + 0 = 7{,}500.$
Y1, Q2: $L_2 = 0.3(7{,}322 - 7{,}500) + 0.7(7{,}500 - 7{,}500) = -53.4.$
     $T_2 = 0.3(-53.4) + 0.7(0) = -16.$
     $F_2 = 0.3(7{,}322) + 0.7(7{,}500) - 16 = 7{,}431.$

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Exponential Smoothing with Trend Forecasting Method with Seasonality for CCW											
2													
3					Seasonally				Seasonally				
4				True	Adjusted	Latest	Estimated	Adjusted	Actual	Forecasting			
5		Year	Quarter	Value	Value	Trend	Trend	Forecast	Forecast	Error		Smoothing Constants	
6		1	1	6809	7322		0	7500	6975	166		α=	0.3
7		1	2	6465	7183	-54	-16	7430	6687	222		β=	0.3
8		1	3	6569	6635	-90	-38	7318	7245	676			
9		1	4	8266	7005	-243	-100	7013	8276	10		Initial Estimates	
10		2	1	7257	7803	-102	-100	6910	6427	830		Average =	7500
11		2	2	7064	7849	167	-20	7158	6442	622		Trend =	0
12		2	3	7784	7863	187	42	7407	7333	451			
13		2	4	8724	7393	179	83	7627	9000	276		Type of Seasonality	
14		3	1	6992	7518	13	62	7619	7085	93		Quarterly	
15		3	2	6822	7580	32	53	7642	6877	55			
16		3	3	7949	8029	34	47	7670	7594	355		Quarter	Seasonal Factor
17		3	4	9650	8178	155	80	7858	9272	378		1	0.93
18		4	1			176	108	8062	7498			2	0.90
19		4	2									3	0.99
20												4	1.18

Mean Absolute Deviation
MAD = 344.53

	M
31	=AVERAGE(J6:J17)

	E	F	G	H	I	J
6	=D6/M17		=M11	=M10+M11	=H6*M17	=ABS(D6-I6)
7	=D7/M18	=$M$6*(E6-M10)+(1-$M$6)*(H6-M10)	=$M$7*F7+(1-$M$7)*G6	=$M$6*E6+(1-$M$6)*H6+G7	=H7*M18	=ABS(D7-I7)
8	=D8/M19	=$M$6*(E7-E6)+(1-$M$6)*(H7-H6)	=$M$7*F8+(1-$M$7)*G7	=$M$6*E7+(1-$M$6)*H7+G8	=H8*M19	=ABS(D8-I8)
9	=D9/M20	=$M$6*(E8-E7)+(1-$M$6)*(H8-H7)	=$M$7*F9+(1-$M$7)*G8	=$M$6*E8+(1-$M$6)*H8+G9	=H9*M20	=ABS(D9-I9)
10	=D10/M17	=$M$6*(E9-E8)+(1-$M$6)*(H9-H8)	=$M$7*F10+(1-$M$7)*G9	=$M$6*E9+(1-$M$6)*H9+G10	=H10*M17	=ABS(D10-I10)
11	=D11/M18	=$M$6*(E10-E9)+(1-$M$6)*(H10-H9)	=$M$7*F11+(1-$M$7)*G10	=$M$6*E10+(1-$M$6)*H10+G11	=H11*M18	=ABS(D11-I11)
12	:	:	:	:	:	:
13	:	:	:	:	:	:

FIGURE 20.5
The Excel template in your OR Courseware for the exponential smoothing with trend method with seasonal adjustments is applied here to the CCW problem.

Y1, Q3: $L_3 = 0.3(7{,}183 - 7{,}322) + 0.7(7{,}431 - 7{,}500) = -90.$
            $T_3 = 0.3(-90) + 0.7(-16) = -38.2.$
            $F_3 = 0.3(7{,}183) + 0.7(7{,}431) - 38.2 = 7{,}318.$
            $\vdots$

The Excel template in Fig. 20.5 shows the results from these calculations for all 12 quarters over the 3 years, as well as for the upcoming quarter. The middle of the figure shows the plots of all the seasonally adjusted call volumes and seasonally adjusted forecasts. Note how each trend up or down in the call volumes causes the forecasts to gradually trend in the same direction, but then the trend in the forecasts takes a couple of quarters to turn around when the trend in call volumes suddenly reverses direction. Each number in column I is calculated by multiplying the seasonally adjusted forecast in column H by the corresponding seasonal factor in column M to obtain the forecast of the actual value (not seasonally adjusted) for the average daily call volume. Column J then shows the resulting *forecasting errors* (the absolute value of the difference between columns D and I).

Forecasting More Than One Time Period Ahead

We have focused thus far on forecasting what will happen in the *next* time period (the next quarter in the case of CCW). However, decision makers sometimes need to forecast further into the future. How can the various forecasting methods be adapted to do this?

In the case of the methods for a constant-level model presented in Sec. 20.4, the forecast for the next period F_{t+1} also is the best available forecast for subsequent periods as well. However, when there is a *trend* in the data, as we are assuming in this section, it is important to take this trend into account for long-range forecasts. *Exponential smoothing with trend* provides a straightforward way of doing this. In particular, after determining the *estimated trend* T_{t+1}, this method's forecast for n time periods into the future is

$$F_{t+n} = \alpha x_t + (1 - \alpha)F_t + nT_{t+1}.$$

20.7 FORECASTING ERRORS

Several forecasting methods now have been presented. How does one choose the appropriate method for any particular application? Identifying the underlying model that best fits the time series (constant-level, linear trend, etc., perhaps in combination with seasonal effects) is an important first step. Assessing how *stable* the parameters of the model are, and so how much reliance can be placed on older data for forecasting, also helps to narrow down the selection of the method. However, the final choice between two or three methods may still not be clear. Some measure of performance is needed.

The goal is to generate forecasts that are as accurate as possible, so it is natural to base a measure of performance on the *forecasting errors.*

> The **forecasting error** (also called the *residual*) for any period t is the absolute value of the deviation of the forecast for period t (F_t) from what then turns out to be the observed value of the time series for period t (x_t). Thus, letting E_t denote this error,
>
> $$E_t = |x_t - F_t|.$$

For example, column J of the spreadsheet in Fig. 20.5 gives the forecasting errors when applying *exponential smoothing with trend* to the CCW example.

Given the forecasting errors for n time periods ($t = 1, 2, \ldots , n$), two popular measures of performance are available. One, called the **mean absolute deviation (MAD)** is simply the average of the errors, so

$$\text{MAD} = \frac{\sum\limits_{t=1}^{n} E_t}{n} .$$

This is the measure shown in cell M31 of Fig. 20.5. (Most of the Excel templates for this chapter use this measure.) The other measure, called the **mean square error (MSE),** instead averages the *square* of the forecasting errors, so

$$\text{MSE} = \frac{\sum\limits_{t=1}^{n} E_t^2}{n} .$$

The advantages of MAD are its ease of calculation and its straightforward interpretation. However, the advantage of MSE is that it imposes a relatively large penalty for a large forecasting error that can have serious consequences for the organization while almost ignoring inconsequentially small forecasting errors. In practice, managers often prefer to use MAD, whereas statisticians generally prefer MSE.

Either measure of performance might be used in two different ways. One is to compare alternative forecasting methods in order to choose one with which to begin forecasting. This is done by applying the methods *retrospectively* to the time series in the past (assuming such data exist). This is a very useful approach as long as the future behavior of the time series is expected to resemble its past behavior. Similarly, this retrospective testing can be used to help select the parameters for a particular forecasting method, e.g., the smoothing constant(s) for exponential smoothing. Second, after the real forecasting begins with some method, one of the measures of performance (or possibly both) normally would be calculated periodically to monitor how well the method is performing. If the performance is disappointing, the same measure of performance can be calculated for alternative forecasting methods to see if any of them would have performed better.

20.8   BOX-JENKINS METHOD

In practice, a forecasting method often is chosen without adequately checking whether the underlying model is an appropriate one for the application. The beauty of the **Box-Jenkins method** is that it carefully coordinates the model and the procedure. (Practitioners often use this name for the method because it was developed by G.E.P. Box and G.M. Jenkins. An alternative name is the *ARIMA method,* which is an acronym for autoregressive integrated moving average.) This method employs a systematic approach to identifying an appropriate model, chosen from a rich class of models. The historical data are used to test the validity of the model. The model also generates an appropriate forecasting procedure.

To accomplish all this, the Box-Jenkins method requires a great amount of past data (a minimum of 50 time periods), so it is used only for major applications. It also is a sophisticated and complex technique, so we will provide only a conceptual overview of the method. (See Selected References 2 and 6 for further details.)

The Box-Jenkins method is iterative in nature. First, a model is chosen. To choose this model, we must compute autocorrelations and partial autocorrelations and examine their patterns. An *autocorrelation* measures the correlation between time series values separated by a fixed number of periods. This fixed number of periods is called the *lag*. Therefore, the autocorrelation for a lag of two periods measures the correlation between every other observation; i.e., it is the correlation between the original time series and the same series moved forward two periods. The *partial autocorrelation* is a conditional autocorrelation between the original time series and the same series moved forward a fixed number of periods, holding the effect of the other lagged times fixed. Good estimates of both the autocorrelations and the partial autocorrelations for all lags can be obtained by using a computer to calculate the *sample* autocorrelations and the *sample* partial autocorrelations. (These are "good" estimates because we are assuming large amounts of data.)

From the autocorrelations and the partial autocorrelations, we can identify the functional form of one or more possible models because a rich class of models is characterized by these quantities. Next we must estimate the parameters associated with the model by using the historical data. Then we can compute the residuals (the forecasting errors when the forecasting is done retrospectively with the historical data) and examine their behavior. Similarly, we can examine the behavior of the estimated parameters. If both the residuals and the estimated parameters behave as expected under the presumed model, the model appears to be validated. If they do not, then the model should be modified and the procedure repeated until a model is validated. At this point, we can obtain an actual forecast for the next period.

For example, suppose that the sample autocorrelations and the sample partial autocorrelations have the patterns shown in Fig. 20.6. The sample autocorrelations appear to decrease exponentially as a function of the time lags, while the same partial autocorrelations have spikes at the first and second time lags followed by values that seem to be of negligible magnitude. This behavior is characteristic of the functional form

$$X_t = B_0 + B_1 X_{t-1} + B_2 X_{t-2} + e_t.$$

Assuming this functional form, we use the time series data to estimate B_0, B_1, and B_2. Denote these estimates by b_0, b_1, and b_2, respectively. Together with the time series data, we then obtain the residuals

$$x_t - (b_0 + b_1 x_{t-1} + b_2 x_{t-2}).$$

FIGURE 20.6
Plot of sample autocorrelation and partial autocorrelation versus time lags.
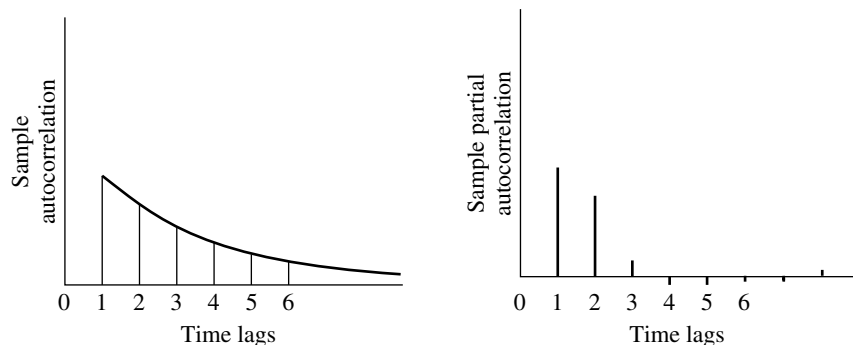
If the assumed functional form is adequate, the residuals and the estimated parameters should behave in a predictable manner. In particular, the sample residuals should behave approximately as independent, normally distributed random variables, each having mean 0 and variance σ^2 (assuming that e_t, the random error at time period t, has mean 0 and variance σ^2). The estimated parameters should be uncorrelated and significantly different from zero. Statistical tests are available for this diagnostic checking.

The Box-Jenkins procedure appears to be a complex one, and it is. Fortunately, computer software is available. The programs calculate the sample autocorrelations and the sample partial autocorrelations necessary for identifying the form of the model. They also estimate the parameters of the model and do the diagnostic checking. These programs, however, cannot accurately identify one or more models that are compatible with the autocorrelations and the partial autocorrelations. Expert human judgment is required. This expertise can be acquired, but it is beyond the scope of this text. Although the Box-Jenkins method is complicated, the resulting forecasts are extremely accurate and, when the time horizon is short, better than most other forecasting methods. Furthermore, the procedure produces a measure of the forecasting error.

20.9   CAUSAL FORECASTING WITH LINEAR REGRESSION

In the preceding six sections, we have focused on *time series forecasting methods,* i.e., methods that forecast the next value in a time series based on its previous values. We now turn to another type of approach to forecasting.

Causal Forecasting

In some cases, the variable to be forecasted has a rather direct relationship with one or more other variables whose values will be known at the time of the forecast. If so, it would make sense to base the forecast on this relationship. This kind of approach is called *causal forecasting.*

> **Causal forecasting** obtains a forecast of the quantity of interest (the *dependent variable*) by relating it directly to one or more other quantities (the *independent* variables) that drive the quantity of interest.

Table 20.3 shows some examples of the kinds of situations where causal forecasting sometimes is used. In each of the first three cases, the indicated dependent variable can be expected to go up or down rather directly with the independent variable(s) listed in the rightmost column. The last case also applies when some quantity of interest (e.g., sales

TABLE 20.3 Possible examples of causal forecasting

Type of Forecasting	Possible Dependent Variable	Possible Independent Variables
Sales	Sales of a product	Amount of advertising
Spare parts	Demand for spare parts	Usage of equipment
Economic trends	Gross domestic product	Various economic factors
Any quantity	This same quantity	Time

of a product) tends to follow a steady trend upward (or downward) with the passage of time (the independent variable that drives the quantity of interest).

As one specific example, Sec. 20.1 includes a description of American Airline's elaborate system for forecasting its need for expensive spare parts (its "rotatable" parts) to continue operating its fleet of several hundred airplanes. This system uses causal forecasting, where the demand for spare parts is the dependent variable and the number of flying hours is the independent variable. This makes sense because the demand for spare parts should be roughly proportional to the number of flying hours for the fleet.

Linear Regression

We will focus on the type of causal forecasting where the mathematical relationship between the dependent variable and the independent variable(s) is assumed to be a linear one (plus some random fluctuations). The analysis in this case is referred to as *linear regression.*

To illustrate the linear regression approach, suppose that a publisher of textbooks is concerned about the initial press run for her books. She sells books both through bookstores and through mail orders. This latter method uses an extensive advertising campaign on line, as well as through publishing media and direct mail. The advertising campaign is conducted prior to the publication of the book. The sales manager has noted that there is a rather interesting linear relationship between the number of mail orders and the number sold through bookstores during the first year. He suggests that this relationship be exploited to determine the initial press run for subsequent books.

Thus, if the number of mail order sales for a book is denoted by X and the number of bookstore sales by Y, then the random variables X and Y exhibit a *degree of association.* However there is *no functional relationship* between these two random variables; i.e., given the number of mail order sales, one does not expect to determine *exactly* the number of bookstore sales. For any given number of mail order sales, there is a range of possible bookstore sales, and vice versa.

What, then, is meant by the statement, "The sales manager has noted that there is a rather interesting linear relationship between the number of mail orders and the number sold through bookstores during the first year"? Such a statement implies that the *expected value* of the number of bookstore sales is linear with respect to the number of mail order sales, i.e.,

$$E[Y\,|\,X = x] = A + Bx.$$

Thus, if the number of mail order sales is x for many different books, the average number of corresponding bookstore sales would tend to be approximately $A + Bx$. This relationship between X and Y is referred to as a **degree of association model.**

As already suggested in Table 20.3, other examples of this degree of association model can easily be found. A college admissions officer may be interested in the relationship between a student's performance on the college entrance examination and subsequent performance in college. An engineer may be interested in the relationship between tensile strength and hardness of a material. An economist may wish to predict a measure of inflation as a function of the cost of living index, and so on.

The degree of association model is not the only model of interest. In some cases, there exists a **functional relationship** between two variables that may be linked linearly.

In a forecasting context, one of the two variables is time, while the other is the variable of interest. In Sec. 20.6, such an example was mentioned in the context of the generating process of the time series being represented by a linear trend superimposed with random fluctuations, i.e.,

$$X_t = A + Bt + e_t,$$

where A is a constant, B is the slope, and e_t is the random error, assumed to have expected value equal to zero and constant variance. (The symbol X_t can also be read as X given t or as $X|t$.) It follows that

$$E(X_t) = A + Bt.$$

Note that both the degree of association model and the *exact functional relationship* model lead to the same linear relationship, and their subsequent treatment is almost identical. Hence, the publishing example will be explored further to illustrate how to treat both kinds of models, although the special structure of the model

$$E(X_t) = A + Bt,$$

with t taking on integer values starting with 1, leads to certain simplified expressions. In the standard notation of regression analysis, X represents the **independent variable** and Y represents the **dependent variable** of interest. Consequently, the notational expression for this special time series model now becomes

$$Y_t = A + Bt + e_t.$$

Method of Least Squares

Suppose that bookstore sales and mail order sales are given for 15 books. These data appear in Table 20.4, and the resulting plot is given in Fig. 20.7.

TABLE 20.4 Data for the mail-order and bookstore sales example

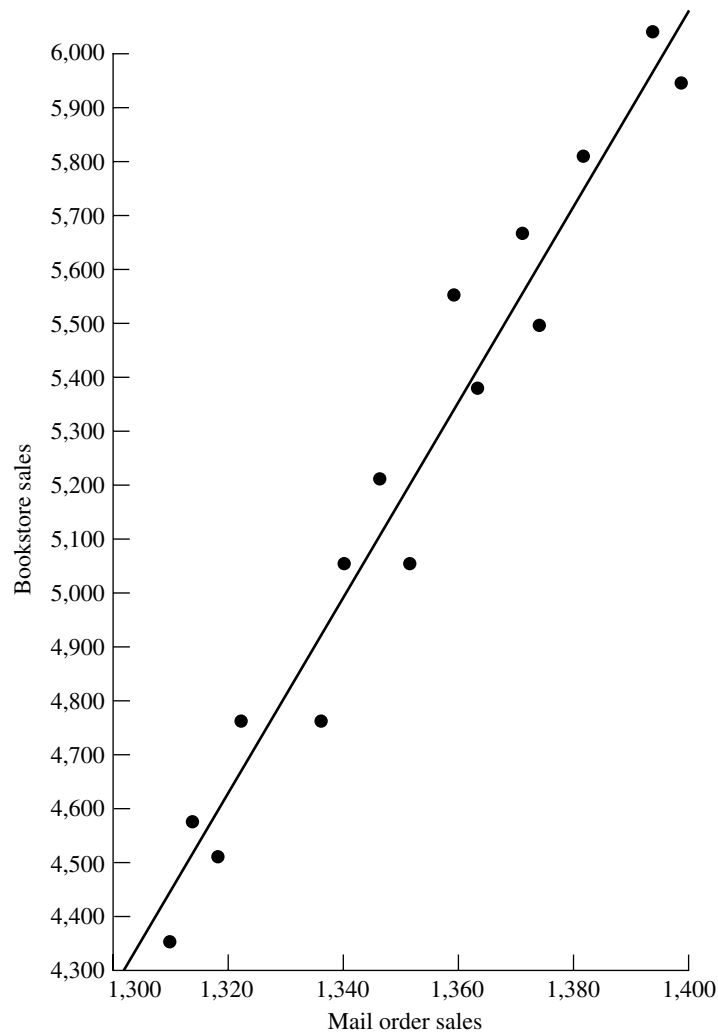| Mail-Order Sales	Bookstore Sales
1,310	4,360
1,313	4,590
1,320	4,520
1,322	4,770
1,338	4,760
1,340	5,070
1,347	5,230
1,355	5,080
1,360	5,550
1,364	5,390
1,373	5,670
1,376	5,490
1,384	5,810
1,395	6,060
1,400	5,940

FIGURE 20.7
Plot of mail order sales versus bookstore sales from Table 20.3.

It is evident that the points in Fig. 20.7 do not lie on a straight line. Hence, it is not clear where the line should be drawn to show the linear relationship. Suppose that an arbitrary line, given by the expression $\tilde{y} = a + bx$, is drawn through the data. A measure of how well this line fits the data can be obtained by computing the *sum of squares* of the vertical deviations of the actual points from the fitted line. Thus, let y_i represent the bookstore sales of the ith book and x_i the corresponding mail order sales. Denote by $\tilde{y}_i$ the point on the fitted line corresponding to the mail order sales of x_i. The proposed measure of fit is then given by

$$Q = (y_1 - \tilde{y}_1)^2 + (y_2 - \tilde{y}_2)^2 + \cdots + (y_{15} - \tilde{y}_{15})^2 = \sum_{i=1}^{15} (y_i - \tilde{y}_i)^2.$$

The usual method for identifying the "best" fitted line is the **method of least squares**. This method chooses that line $a + bx$ that makes Q a minimum. Thus, a and b are obtained simply by setting the partial derivatives of Q with respect to a and b equal to zero and solving the resulting equations. This method yields the solution

$$b = \frac{\displaystyle\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\displaystyle\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\displaystyle\sum_{i=1}^{n} x_i y_i - \left(\displaystyle\sum_{i=1}^{n} x_i \displaystyle\sum_{i=1}^{n} y_i\right)\Big/ n}{\displaystyle\sum_{i=1}^{n} x_i^2 - \left(\displaystyle\sum_{i=1}^{n} x_i\right)^2 \Big/ n}$$

and

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

where

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

and

$$\bar{y} = \sum_{i=1}^{n} \frac{y_i}{n}.$$

(Note that $\bar{y}$ is not the same as $\tilde{y} = a + bx$ discussed in the preceding paragraph.)

For the publishing example, the data in Table 20.4 and Fig. 20.7 yield

$$\bar{x} = 1{,}353.1,$$
$$\bar{y} = 5{,}219.3,$$
$$\sum_{i=1}^{15} (x_i - \bar{x})(y_i - \bar{y}) = 214{,}543.9,$$
$$\sum_{i=1}^{15} (x_i - \bar{x})^2 = 11{,}966,$$
$$a = -19{,}041.9,$$
$$b = 17.930.$$

Hence, the least-squares estimate of bookstore sales $\tilde{y}$ with mail order sales x is given by

$$\tilde{y} = -19{,}041.9 + 17.930x,$$

and this is the line drawn in Fig. 20.7. Such a line is referred to as a **regression line.**

An Excel template called Linear Regression is available in your OR Courseware for calculating a regression line in this way.

This regression line is useful for forecasting purposes. For a given value of x, the corresponding value of y represents the forecast.

The decision maker may be interested in some measure of uncertainty that is associated with this forecast. This measure is easily obtained provided that certain as-

sumptions can be made. Therefore, for the remainder of this section, it is assumed that

1. A random sample of n pairs $(x_1, Y_1), (x_2, Y_2), \ldots, (x_n, Y_n)$ is to be taken.
2. The Y_i are normally distributed with mean $A + Bx_i$ and variance σ^2 (independent of i).

The assumption that Y_i is normally distributed is not a critical assumption in determining the uncertainty in the forecast, but the assumption of constant variance is crucial. Furthermore, an estimate of this variance is required.

An unbiased estimate of σ^2 is given by $s_{y|x}^2$, where

$$s_{y|x}^2 = \sum_{i=1}^{n} \frac{(y_i - \tilde{y}_i)^2}{n - 2}.$$

Confidence Interval Estimation of $E(Y|x = x^*)$

A very important reason for obtaining the linear relationship between two variables is to use the line for future decision making. From the regression line, it is possible to estimate $E(Y|x)$ by a *point* estimate (the forecast) and a *confidence interval* estimate (a measure of forecast uncertainty).

For example, the publisher might want to use this approach to estimate the expected number of bookstore sales corresponding to mail order sales of, say, 1,400, by both a point estimate and a confidence interval estimate for forecasting purposes.

A point estimate of $E(Y|x = x^*)$ is given by

$$\tilde{y}^* = a + bx^*,$$

where x^* denotes the given value of the independent variable and $\tilde{y}^*$ is the corresponding point estimate.

The endpoints of a $(100)(1 - \alpha)$ percent confidence interval for $E(Y|x = x^*)$ are given by

$$a + bx^* - t_{\alpha/2;n-2}s_{y|x} \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

and

$$a + bx^* + t_{\alpha/2;n-2}s_{y|x} \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}},$$

where $s_{y|x}^2$ is the estimate of σ^2, and $t_{\alpha/2;n-2}$ is the $100\alpha/2$ percentage point of the t distribution with $n - 2$ degrees of freedom (see Table A5.2 of Appendix 5). Note that the interval is narrowest where $x^* = \bar{x}$, and it becomes wider as x^* departs from the mean.

In the publishing example with $x^* = 1,400$, $s_{y|x}^2$ is computed from the data in Table 20.4 to be 17,030, so $s_{y|x} = 130.5$. If a 95 percent confidence interval is required, Table A5.2 gives $t_{0.025;13} = 2.160$. The earlier calculation of a and b yields

$$a + bx^* = -19,041.9 + 17.930(1,400) = 6,060$$

as the point estimate of $E(Y \mid 1{,}400)$, that is, the forecast. Consequentially, the confidence limits corresponding to mail order sales of 1,400 are

$$\text{Lower confidence limit} = 6{,}060 - 2.160(130.5)\sqrt{\frac{1}{15} + \frac{46.9^2}{11{,}966}}$$

$$= 5{,}919,$$

$$\text{Upper confidence limit} = 6{,}060 + 2.160(130.5)\sqrt{\frac{1}{15} + \frac{46.9^2}{11{,}966}}$$

$$= 6{,}201.$$

The fact that the confidence interval was obtained at a data point ($x = 1{,}400$) is purely coincidental.

The Excel template for linear regression in your OR Courseware does most of the computational work involved in calculating these confidence limits. In addition to computing a and b (the regression line), it calculates $s_{y \mid x}^2$, $\bar{x}$, and $\sum_{i=1}^{n} (x_i - \bar{x})^2$.

Predictions

The confidence interval statement for the expected number of bookstore sales corresponding to mail order sales of 1,400 may be useful for budgeting purposes, but it is not too useful for making decisions about the *actual* press run. Instead of obtaining bounds on the *expected* number of bookstore sales, this kind of decision requires bounds on what the *actual* bookstore sales will be, i.e., a **prediction interval** on the value that the random variable (bookstore sales) takes on. This measure is a *different* measure of forecast uncertainty.

The two endpoints of a prediction interval are given by the expressions

$$a + bx_+ - t_{\alpha/2;n-2}s_{y \mid x}\sqrt{1 + \frac{1}{n} + \frac{(x_+ - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

and

$$a + bx_+ + t_{\alpha/2;n-2}s_{y \mid x}\sqrt{1 + \frac{1}{n} + \frac{(x_+ - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

For any given value of x (denoted here by x_+), the probability is $1 - \alpha$ that the value of the future Y_+ associated with x_+ will fall in this interval.

Thus, in the publishing example, if x_+ is 1,400, then the corresponding 95 percent prediction interval for the number of bookstore sales is given by $6{,}060 \pm 315$, which is naturally wider than the confidence interval for the expected number of bookstore sales, $6{,}060 \pm 141$.

This method of finding a prediction interval works fine if it is only being done once. However, it is not feasible to use the same data to find multiple prediction intervals with various values of x_+ in this way and then specify a probability that *all* these predictions will be correct. For example, suppose that the publisher wants prediction intervals for sev-

eral different books. For each individual book, she still is able to use these expressions to find the prediction interval and then make the prediction that the bookstore sales will be within this interval, where the probability is $1 - \alpha$ that the prediction will be correct. However, what she cannot do is specify a probability that *all* these predictions will be correct. The reason is that these predictions are all based upon the same statistical data, so the predictions are not statistically independent. *If* the predictions were independent and if k future bookstore sales were being predicted, with each prediction being made with probability $1 - \alpha$, then the probability would be $(1 - \alpha)^k$ that *all* k predictions of future bookstore sales will be correct. Unfortunately, the predictions are *not* independent, so the actual probability cannot be calculated, and $(1 - \alpha)^k$ does not even provide a reasonable approximation.

This difficulty can be overcome by using **simultaneous tolerance intervals.** Using this technique, the publisher can take the mail order sales of any book, find an interval (based on the previously determined linear regression line) that will contain the actual bookstore sales with probability at least $1 - \alpha$, and repeat this for any number of books having the same or different mail order sales. Furthermore, the probability is P that *all* these predictions will be correct. An alternative interpretation is as follows. If every publisher followed this procedure, each using his or her own linear regression line, then $100P$ percent of the publishers (on average) would find that at least $100(1 - \alpha)$ percent of their bookstore sales fell into the predicted intervals. The expression for the endpoints of each such tolerance interval is given by

$$a + bx_+ - c^{**}s_{y\,|\,x}\sqrt{\frac{1}{n} + \frac{(x_+ - \bar{x})^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}}$$

and

$$a + bx_+ + c^{**}s_{y\,|\,x}\sqrt{\frac{1}{n} + \frac{(x_+ - \bar{x})^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}},$$

where c^{**} is given in Table 20.5.

Thus, the publisher can predict that the bookstore sales corresponding to known mail order sales will fall in these tolerance intervals. Such statements can be made for as many books as the publisher desires. Furthermore, the probability is P that at least $100(1 - \alpha)$ percent of bookstore sales corresponding to mail order sales will fall in these intervals. If P is chosen as 0.90 and $\alpha = 0.05$, the appropriate value of c^{**} is 11.625. Hence, the number of bookstore sales corresponding to mail order sales of 1,400 books is predicted to fall in the interval 6,060 ± 759. If another book had mail order sales of 1,353, the bookstore sales are predicted to fall in the interval 5,258 ± 390, and so on. At least 95 percent of the bookstore sales will fall into their predicted intervals, and these statements are made with confidence 0.90.

To summarize, we now have described three *measures of forecast uncertainty*. The first (in the preceding subsection) is a *confidence interval* on the *expected value* of the random variable Y (for example, bookstore sales) given the observed value x of the independent variable X (for example, mail order sales). The second is a *prediction interval* on

TABLE 20.5 Values of c**

n	α = 0.50	α = 0.25	α = 0.10	α = 0.05	α = 0.01	α = 0.001
			P = 0.90			
4	7.471	10.160	13.069	14.953	18.663	23.003
6	5.380	7.453	9.698	11.150	14.014	17.363
8	5.037	7.082	9.292	10.722	13.543	16.837
10	4.983	7.093	9.366	10.836	13.733	17.118
12	5.023	7.221	9.586	11.112	14.121	17.634
14	5.101	7.394	9.857	11.447	14.577	18.232
16	5.197	7.586	10.150	11.803	15.057	18.856
18	5.300	7.786	10.449	12.165	15.542	19.484
20	5.408	7.987	10.747	12.526	16.023	20.140
			P = 0.95			
4	10.756	14.597	18.751	21.445	26.760	32.982
6	6.652	9.166	11.899	13.669	17.167	21.266
8	5.933	8.281	10.831	12.484	15.750	19.568
10	5.728	8.080	10.632	12.286	15.553	19.369
12	5.684	8.093	10.701	12.391	15.724	19.619
14	5.711	8.194	10.880	12.617	16.045	20.050
16	5.771	8.337	11.107	12.898	16.431	20.559
18	5.848	8.499	11.357	13.204	16.845	21.097
20	5.937	8.672	11.619	13.521	17.272	21.652
			P = 0.99			
4	24.466	33.019	42.398	48.620	60.500	74.642
6	10.444	14.285	18.483	21.215	26.606	32.920
8	8.290	11.453	14.918	17.166	21.652	26.860
10	7.567	10.539	13.796	15.911	20.097	24.997
12	7.258	10.182	13.383	15.479	19.579	24.403
14	7.127	10.063	13.267	15.355	19.485	24.316
16	7.079	10.055	13.306	15.410	19.582	24.467
18	7.074	10.111	13.404	15.552	19.794	24.746
20	7.108	10.198	13.566	15.745	20.065	25.122

Source: Reprinted by permission from G. J. Lieberman and R. G. Miller, "Simultaneous Tolerance Intervals in Regression," *Biometrika,* **50**(1 and 2): 164, 1963.

the *actual value* that Y will take on, given x. The third is *simultaneous tolerance intervals* on a *succession* of *actual values* that Y will take on given a succession of observed values of X.

20.10 FORECASTING IN PRACTICE

You now have seen the major forecasting methods used in practice. We conclude with a brief look at how widely the various methods are used.

To begin, consider the actual forecasting applications discussed in Sec. 20.1 and summarized in Table 20.1 there. Most of this table is repeated here in Table 20.6, but the right-

TABLE 20.6 The forecasting methods used in the actual applications presented in Section 20.1

Organization	Quantity Being Forecasted	Forecasting Model
Merit Brass Co.	Sales of finished goods	Exponential smoothing
Hidroeléctrica Español	Energy demand	ARIMA (Box-Jenkins), etc.
American Airlines	Demand for different fare classes	Exponential smoothing
American Airlines	Need for spare parts to repair airplanes	Linear regression
Albuquerque Microelectronics	Production yield in wafer fabrication	Exponential smoothing with a linear trend
U.S. Department of Labor	Unemployment insurance payments	Linear regression
United Airlines	Demand at reservation offices and airports	ARIMA (Box-Jenkins)
L.L. Bean	Staffing needs at call center	ARIMA (Box-Jenkins)

most column now identifies which forecasting method was used in each application. Not surprisingly, these major forecasting projects chose one of the more sophisticated statistical forecasting methods. (The most sophisticated is the ARIMA method, commonly called the Box-Jenkins method, that is described in Sec. 20.8.)

Every company needs to do at least some forecasting, but their methods often are not as sophisticated as with these major projects. Some insight into their general approach was provided by a survey conducted a few years ago[1] of sales forecasting practices at 500 U.S. corporations.

This survey indicates that, generally speaking, judgmental forecasting methods are somewhat more widely used than statistical methods. The main reasons given for using judgmental methods were accuracy and difficulty in obtaining the data required for statistical methods. Comments also were made that upper management is not familiar with quantitative techniques, that judgmental methods create a sense of ownership, and that these methods add a commonsense element to the forecast.

Among the judgmental methods, the most popular is a jury of executive opinion. This is especially true for companywide or industry sales forecasts but also holds true by a small margin over manager's opinion when forecasting sales of individual products or families of products.

Statistical forecasting methods also are fairly widely used, especially in companies with high sales. Compared to earlier surveys, familiarity with such methods is increasing. However, many survey respondents cited better data availability as the improvement they most wanted to see in their organizations. The availability of good data is crucial for the use of these methods.

The survey indicates that the moving-average method and linear regression are the most widely used statistical forecasting methods. The moving-average method is more popular for short- and medium-range forecasts (less than a year), as well as for forecasting sales of individual products and families of products. Linear regression is more popular for longer-range forecasts and for forecasting either companywide or industry sales. Both exponential smoothing and the last-value method also receive considerable use. However, the highest dissatisfaction is with the last-value method, and its popularity is decreasing compared to earlier surveys.

When statistical forecasting methods are used, it is fairly common to also use judgmental methods to adjust the forecasts.

[1]N. R. Sanders and K. B. Manrodt, "Forecasting Practices in U.S. Corporations: Survey Results," *Interfaces*, **24:** 92–100, March–April 1994.

As managers become more familiar with statistical methods, and more used to using the computer to compile data and implement OR techniques, we anticipate a continuing increase in the usage of statistical forecasting methods. However, there always will be an important role for judgmental methods, both alone and in combination with statistical methods.

20.11 CONCLUSIONS

The future success of any business depends heavily on the ability of its management to forecast well. Judgmental forecasting methods often play an important role in this process. However, the ability to forecast well is greatly enhanced if historical data are available to help guide the development of a statistical forecasting method. By studying these data, an appropriate model can be structured. A forecasting method that behaves well under the model should be selected. This method may require choosing one or more parameters—e.g., the smoothing constant α in exponential smoothing—and the historical data may prove useful in making this choice. After forecasting begins, the performance should be monitored carefully to assess whether modifications should be made in the method.

SELECTED REFERENCES

1. Armstrong, J. E. (ed.): *Handbook of Forecasting Principles,* Kluwer Academic Publishers, Boston, 2001.
2. Box, G. E. P., and G. M. Jenkins: *Time Series Analysis, Forecasting and Control,* Holden-Day, San Francisco, 1976.
3. Bunn, D., and G. Wright: "Interaction of Judgmental and Statistical Methods: Issues and Analysis," *Management Science,* **37:** 501–518, 1991.
4. Gardner, E. S., Jr.: "Exponential Smoothing: The State of the Art," *Journal of Forecasting,* **4:** 1–38, 1985.
5. Hillier, F. S., M. S. Hillier, and G. J. Lieberman: *Introduction to Management Science: A Modeling and Case Studies Approach with Spreadsheets,* Irwin/McGraw-Hill, Burr Ridge, IL, 2000, chap. 13.
6. Hoff, J. C.: *A Practical Guide to Box-Jenkins Forecasting,* Lifetime Learning Publications, Belmont, CA, 1983.
7. Marshall, K. T., and R. M. Oliver: *Decision Making and Forecasting,* McGraw-Hill, New York, 1995.
8. Sanders, N. R., and K. B. Manrodt: "Forecasting Practices in U.S. Corporations: Survey Results," *Interfaces,* **24**(2): 92–100, March–April 1994.
9. Rust, R. T., D. Simester, R. J. Brodie, and V. Nilikant: "Model Selection Criteria: An Investigation of Relative Accuracy, Posterior Probabilities, and Combinations of Criteria," *Management Science,* **41:** 322–333, 1995.
10. Yurkiewicz, J.: "Forecasting Software Survey," *OR/MS Today,* Dec. 1996, pp. 70–75.

LEARNING AIDS FOR THIS CHAPTER IN YOUR OR COURSEWARE

"Ch. 20—Forecasting" Excel File:

Template for *Seasonal Factors*
Templates for *Last-Value* Method (with and without Seasonality)