CHAPTER 15

Time-Series Forecasting and Index Numbers

LEARNING OBJECTIVES

This chapter discusses the general use of forecasting in business, several tools that are available for making business forecasts, the nature of time-series data, and the role of index numbers in business, thereby enabling you to:

1. Differentiate among various measurements of forecasting error, including mean absolute deviation and mean square error, in order to assess which forecasting method to use
2. Describe smoothing techniques for forecasting models, including na

Tim Hurst/Masterfile

Forecasting Air Pollution

For the past two decades, there has been a heightened awareness of and increased concern over pollution in various forms in the United States. One of the main areas of environmental concern is air pollution, and the U.S. Environmental Protection Agency (EPA) regularly monitors the quality of air around the country. Some of the air pollutants monitored include carbon monoxide emissions, nitrogen oxide emissions, volatile organic compounds, sulfur dioxide emissions, particulate matter, fugitive dust, and lead emissions. Shown below are emission data for two of these air pollution variables, carbon monoxide and nitrogen oxides, over a 19-year period reported by the EPA in millions short-tons.

Year	Carbon Monoxide	Nitrogen Oxides
1985	176.84	25.76
1986	173.67	25.42
1987	172.97	25.58
1988	174.42	26.07
1989	160.52	25.38
1990	154.19	25.53
1991	147.13	25.18
1992	140.90	25.26
1993	135.90	25.36
1994	133.56	25.35
1995	126.78	24.96
1996	128.86	24.79
1997	117.91	24.71
1998	115.38	24.35
1999	114.54	22.84
2000	114.47	22.60
2001	106.30	21.55
2002	111.06	21.14
2003	106.24	20.33
2004	101.43	19.52
2005	96.62	18.71
2006	92.13	17.69
2007	88.25	17.03

Managerial and Statistical Questions

1. Is it possible to forecast the emissions of carbon monoxide or nitrogen oxides for the year 2011, 2015, or even 2025 using these data?
2. What techniques best forecast the emissions of carbon monoxide or nitrogen oxides for future years from these data?

Source: Adapted from statistics published as National Transportation Statistics by the Bureau of Transportation Statistics (U.S. government) at: http://www.bts.gov/publications/national_transportation_statistics/html/table_04_40.html; http://www.bts.gov/publications/national_transportation_statistics/html/table_04_41.html

Every day, **forecasting**—*the art or science of predicting the future*—is used in the decision-making process to help business people reach conclusions about buying, selling, producing, hiring, and many other actions. As an example, consider the following items:

- Market watchers predict a resurgence of stock values next year.
- City planners forecast a water crisis in Southern California.
- Future brightens for solar power.
- Energy secretary sees rising demand for oil.
- CEO says difficult times won't be ending soon for U.S. airline industry.
- Life insurance outlook fades.
- Increased competition from overseas businesses will result in significant layoffs in the U.S. computer chip industry.

How are these and other conclusions reached? What forecasting techniques are used? Are the forecasts accurate? In this chapter we discuss several forecasting techniques, how to measure the error of a forecast, and some of the problems that can occur in forecasting. In addition, this chapter will focus only on data that occur over time, time-series data.

Time-series data are *data gathered on a given characteristic over a period of time at regular intervals.* Time-series forecasting techniques attempt to account for changes over time by examining patterns, cycles, or trends, or using information about previous time periods to predict the outcome for a future time period. Time-series methods include naïve methods, averaging, smoothing, regression trend analysis, and the decomposition of the possible time-series factors, all of which are discussed in subsequent sections.

15.1 INTRODUCTION TO FORECASTING

Virtually all areas of business, including production, sales, employment, transportation, distribution, and inventory, produce and maintain time-series data. Table 15.1 provides an example of time-series data released by the Office of Market Finance, U.S. Department of the Treasury. The table contains the bond yield rates of three-month Treasury Bills for a 17-year period.

Why does the average yield differ from year to year? Is it possible to use these time series data to predict average yields for year 18 or ensuing years? Figure 15.1 is a graph of these data over time. Often graphical depiction of time-series data can give a clue about any trends, cycles, or relationships that might exist there. Does the graph in Figure 15.1 show that bond yields are decreasing? Will next year's yield rate be lower or is a cycle occurring in these data that will result in an increase? To answer such questions, it is sometimes helpful to determine which of the four components of time-series data exist in the data being studied.

TABLE 15.1

Bond Yields of Three-Month Treasury Bills

Year	Average Yield
1	14.03%
2	10.69
3	8.63
4	9.58
5	7.48
6	5.98
7	5.82
8	6.69
9	8.12
10	7.51
11	5.42
12	3.45
13	3.02
14	4.29
15	5.51
16	5.02
17	5.07

Time-Series Components

It is generally believed that time-series data are composed of four elements: trend, cyclicality, seasonality, and irregularity. Not all time-series data have all these elements. Consider Figure 15.2, which shows the effects of these time-series elements on data over a period of 13 years.

The long-term general direction of data is referred to as **trend**. Notice that even though the data depicted in Figure 15.2 move through upward and downward periods, the general direction or trend is increasing (denoted in Figure 15.2 by the line). **Cycles** are *patterns of highs and lows through which data move over time periods usually of more than a year.* Notice that the data in Figure 15.2 seemingly move through two periods or cycles of highs and lows over a 13-year period. Time-series data that do not extend over a long period of time may not have enough "history" to show **cyclical effects**. **Seasonal effects**, on the other hand, are *shorter cycles, which usually occur in time periods of less than one year.* Often seasonal effects are measured by the month, but they may occur by quarter, or may be measured in as small a time frame as a week or even a day. Note the seasonal effects shown in Figure 15.2 as up and down cycles, many of which occur during a 1-year period. **Irregular fluctuations** are *rapid changes or "bleeps" in the data, which occur in even shorter time frames than seasonal effects.* Irregular fluctuations can happen as often as day to day. They are subject to momentary change and are often unexplained. Note the irregular fluctuations in the data of Figure 15.2.

FIGURE 15.1

Excel Graph of Bond Yield Time-Series Data (new)

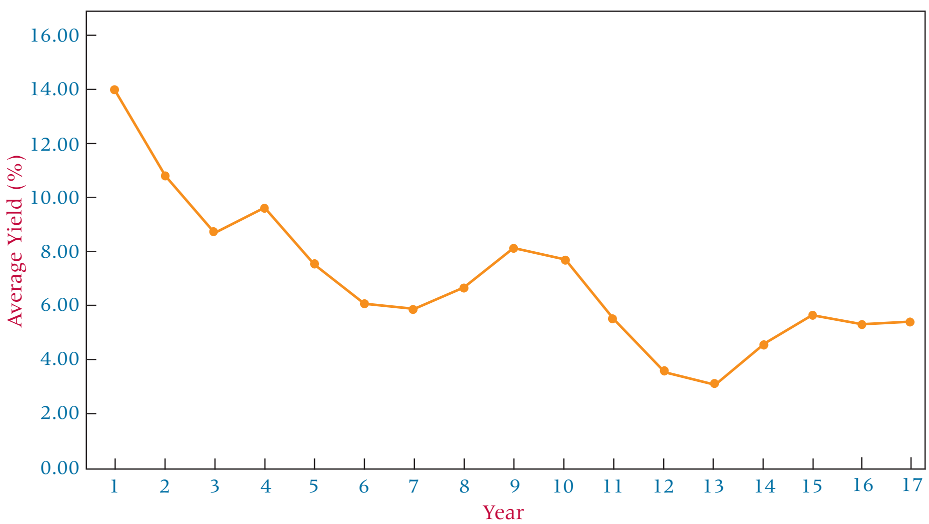
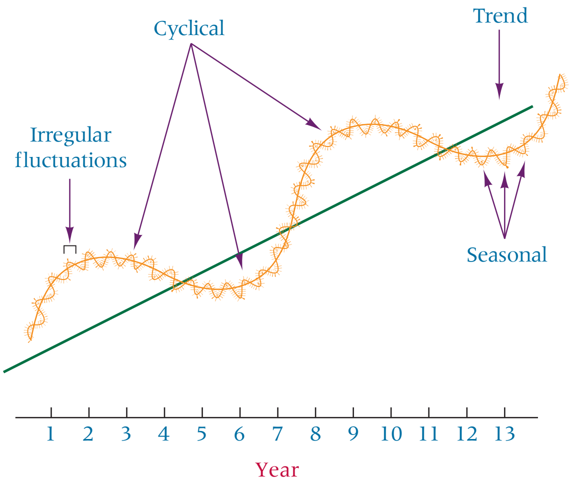


FIGURE 15.2
Time-Series Effects



Observe again the bond yield data depicted in Figure 15.1. The general trend seems to move downward and contain two cycles. Each of the cycles traverses approximately 5 to 8 years. It is possible, although not displayed here, that seasonal periods of highs and lows within each year result in seasonal bond yields. In addition, irregular daily fluctuations of bond yield rates may occur but are unexplainable.

Time-series data that contain no trend, cyclical, or seasonal effects are said to be **stationary**. Techniques used to forecast stationary data analyze only the irregular fluctuation effects.

The Measurement of Forecasting Error

In this chapter, several forecasting techniques will be introduced that typically produce different forecasts. How does a decision maker know which forecasting technique is doing the best job in predicting the future? One way is to compare forecast values with actual values and determine the amount of **forecasting error** a technique produces. An examination of individual errors gives some insight into the accuracy of the forecasts. However, this process can be tedious, especially for large data sets, and often a single measurement of overall forecasting error is needed for the entire set of data under consideration. Any of several methods can be used to compute error in forecasting. The choice depends on the forecaster's objective, the forecaster's familiarity with the technique, and the method of error measurement used by the computer forecasting software. Several techniques can be used to measure overall error, including mean error (ME), mean absolute deviation (MAD), mean square error (MSE), mean percentage error (MPE), and mean absolute percentage error (MAPE). Here we will consider the mean absolute deviation (MAD) and the mean square error (MSE).

Error

The **error of an individual forecast** is *the difference between the actual value and the forecast of that value.*

ERROR OF AN INDIVIDUAL FORECAST

$$e_t = x_t - F_t$$

where

e_t = the error of the forecast
x_t = the actual value
F_t = the forecast value

Mean Absolute Deviation (MAD)

One measure of overall error in forecasting is the mean absolute deviation, MAD. The **mean absolute deviation (MAD)** is *the mean, or average, of the absolute values of the errors.*

TABLE 15.2
Nonfarm Partnership Tax Returns

Year	Actual	Forecast	Error
1	1,402	—	—
2	1,458	1,402	56.0
3	1,553	1,441.2	111.8
4	1,613	1,519.5	93.5
5	1,676	1,585.0	91.0
6	1,755	1,648.7	106.3
7	1,807	1,723.1	83.9
8	1,824	1,781.8	42.2
9	1,826	1,811.3	14.7
10	1,780	1,821.6	−41.6
11	1,759	1,792.5	−33.5

Table 15.2 presents the nonfarm partnership tax returns in the United States over an 11-year period along with the forecast for each year and the error of the forecast. An examination of these data reveals that some of the forecast errors are positive and some are negative. In summing these errors in an attempt to compute an overall measure of error, the negative and positive values offset each other resulting in an underestimation of the total error. The mean absolute deviation overcomes this problem by taking the absolute value of the error measurement, thereby analyzing the magnitude of the forecast errors without regard to direction.

MEAN ABSOLUTE DEVIATION

$$\text{MAD} = \frac{\Sigma |e_i|}{\text{Number of Forecasts}}$$

The mean absolute error can be computed for the forecast errors in Table 15.2 as follows.

$$\text{MAD} = \frac{|56.0| + |111.8| + |93.5| + |91.0| + |106.3| + |83.9| + |42.2| + |14.7| + |-41.6| + |-33.5|}{10} = 67.45$$

Mean Square Error (MSE)

The **mean square error (MSE)** is another way to circumvent the problem of the canceling effects of positive and negative forecast errors. The MSE is *computed by squaring each error (thus creating a positive number) and averaging the squared errors.* The following formula states it more formally.

MEAN SQUARE ERROR

$$\text{MSE} = \frac{\Sigma e_i^2}{\text{Number of Forecasts}}$$

The mean square error can be computed for the errors shown in Table 15.2 as follows.

$$\text{MSE} = \frac{(56.0)^2 + (111.8)^2 + (93.5)^2 + (91.0)^2 + (106.3)^2 + (83.9)^2 + (42.2)^2 + (14.7)^2 + (-41.6)^2 + (-33.5)^2}{10} = 5{,}584.7$$

Selection of a particular mechanism for computing error is up to the forecaster. It is important to understand that different error techniques will yield different information. The business researcher should be informed enough about the various error measurement techniques to make an educated evaluation of the forecasting results.

15.1 PROBLEMS

15.1 Use the forecast errors given here to compute MAD and MSE. Discuss the information yielded by each type of error measurement.

Period	*e*
1	2.3
2	1.6
3	−1.4
4	1.1
5	.3
6	−.9
7	−1.9
8	−2.1
9	.7

15.2 Determine the error for each of the following forecasts. Compute MAD and MSE.

Period	Value	Forecast	Error
1	202	—	—
2	191	202	
3	173	192	
4	169	181	
5	171	174	
6	175	172	
7	182	174	
8	196	179	
9	204	189	
10	219	198	
11	227	211	

15.3 Using the following data, determine the values of MAD and MSE. Which of these measurements of error seems to yield the best information about the forecasts? Why?

Period	Value	Forecast
1	19.4	16.6
2	23.6	19.1
3	24.0	22.0
4	26.8	24.8
5	29.2	25.9
6	35.5	28.6

15.4 Figures for acres of tomatoes harvested in the United States from an 11-year period follow. The data are published by the U.S. Department of Agriculture. With these data, forecasts have been made by using techniques presented later in this chapter. Compute MAD and MSE on these forecasts. Comment on the errors.

Year	Number of Acres	Forecast
1	140,000	—
2	141,730	140,000
3	134,590	141,038
4	131,710	137,169
5	131,910	133,894
6	134,250	132,704
7	135,220	133,632
8	131,020	134,585
9	120,640	132,446
10	115,190	125,362
11	114,510	119,259

Several techniques are available to forecast time-series data that are stationary or that include no significant trend, cyclical, or seasonal effects. These techniques are often referred to as **smoothing techniques** because they *produce forecasts based on "smoothing out" the irregular fluctuation effects in the time-series data.* Three general categories of smoothing techniques are presented here: (1) naïve forecasting models, (2) averaging models, and (3) exponential smoothing.

Na

FIGURE 15.3

Excel Graph of Shipments of Bell Peppers over a 12-Month Period

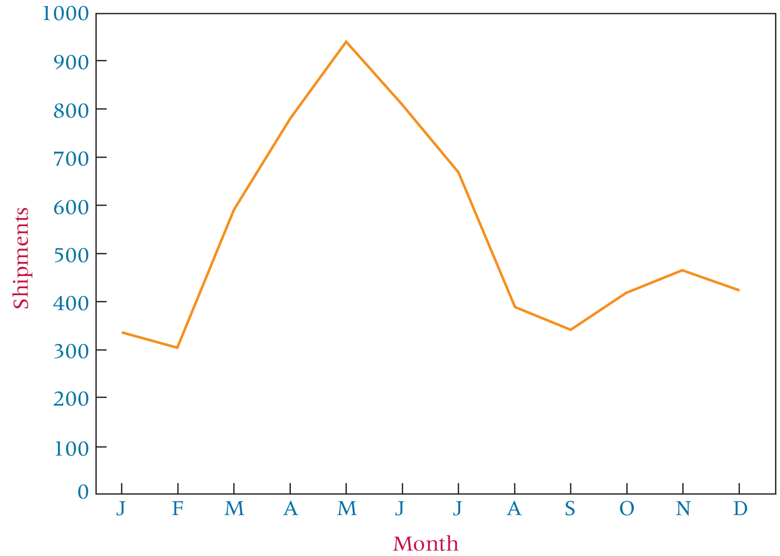


TABLE 15.4

Cost of Residential Heating Oil (cents per gallon)

Time Frame	Cost of Heating Oil
January (year 1)	66.1
February	66.1
March	66.4
April	64.3
May	63.2
June	61.6
July	59.3
August	58.1
September	58.9
October	60.9
November	60.7
December	59.4
January (year 2)	61.3
February	63.3
March	62.1
April	59.8
May	58.4
June	57.6
July	55.7
August	55.1
September	55.7
October	56.7
November	57.2
December	58.0
January (year 3)	58.2
February	58.3
March	57.7
April	56.7
May	56.8
June	55.5
July	53.8
August	52.8

Averaging Models

Many naïve model forecasts are based on the value of one time period. Often such forecasts become a function of irregular fluctuations of the data; as a result, the forecasts are "oversteered." Using averaging models, a forecaster enters information from several time periods into the forecast and "smoothes" the data. **Averaging models** are computed by *averaging data from several time periods and using the average as the forecast for the next time period.*

Simple Averages

The most elementary of the averaging models is the **simple average model**. With this model, *the forecast for time period t is the average of the values for a given number of previous time periods*, as shown in the following equation.

$$F_t = \frac{X_{t-1} + X_{t-2} + X_{t-3} + \cdots + X_{t-n}}{n}$$

The data in Table 15.4 provide the costs of residential heating oil in the United States for 3 years. Figure 15.4 displays a Minitab graph of these data.

A simple 12-month average could be used to forecast the cost of residential heating oil for September of year 3 from the data in Table 15.4 by averaging the values for September of year 2 through August of year 3 (the preceding 12 months).

$$F_{\text{Sept, year 3}} = \frac{55.7 + 56.7 + 57.2 + 58.0 + 58.2 + 58.3 + 57.7 + 56.7 + 56.8 + 55.5 + 53.8 + 52.8}{12} = 56.45$$

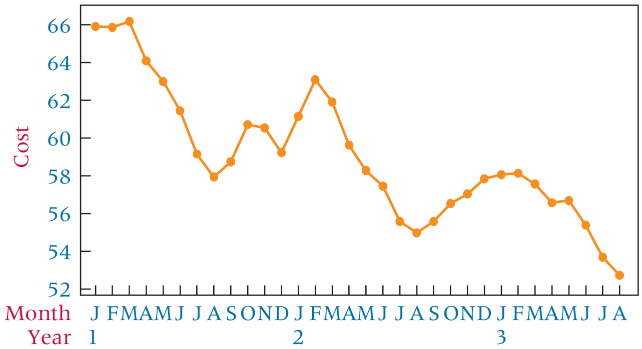
With this **simple average**, the forecast for year 3 September heating oil cost is 56.45 cents. Note that none of the previous 12-month figures equal this value and that this average is not necessarily more closely related to values early in the period than to those late in the period. The use of the simple average over 12 months tends to smooth the variations, or fluctuations, that occur during this time.

Moving Averages

Suppose we were to attempt to forecast the heating oil cost for October of year 3 by using averages as the forecasting method. Would we still use the simple average for September of year 2 through August of year 3 as we did to forecast for September of year 3? Instead of using the same 12 months' average used to forecast September of year 3, it would seem to

FIGURE 15.4

Minitab Graph of Heating Oil Cost Data



make sense to use the 12 months prior to October of year 3 (October of year 2 through September of year 3) to average for the new forecast. Suppose in September of year 3 the cost of heating oil is 53.3 cents. We could forecast October of year 3 with a new average that includes the same months used to forecast September of year 3, but without the value for September of year 2 and with the value of September of year 3 added.

$$F_{\text{Sept, year 3}} = \frac{56.7 + 57.2 + 58.0 + 58.2 + 58.3 + 57.7 + 56.7 + 56.8 + 55.5 + 53.8 + 52.8 + 53.3}{12} = 56.25$$

Computing an average of the values from October of year 2 through September of year 3 produces a moving average, which can be used to forecast the cost of heating oil for October of year 3. In computing this moving average, the earliest of the previous 12 values, September of year 2, is dropped and the most recent value, September of year 3, is included.

A **moving average** is *an average that is updated or recomputed for every new time period being considered.* The most recent information is utilized in each new moving average. This advantage is offset by the disadvantages that (1) it is difficult to choose the optimal length of time for which to compute the moving average, and (2) moving averages do not usually adjust for such time-series effects as trend, cycles, or seasonality. To determine the more optimal lengths for which to compute the moving averages, we would need to forecast with several different average lengths and compare the errors produced by them.

DEMONSTRATION PROBLEM 15.1

Shown here are shipments (in millions of dollars) for electric lighting and wiring equipment over a 12-month period. Use these data to compute a 4-month moving average for all available months.

Month	Shipments
January	1056
February	1345
March	1381
April	1191
May	1259
June	1361
July	1110
August	1334
September	1416
October	1282
November	1341
December	1382

Solution

The first moving average is

$$\text{4-Month Moving Average} = \frac{1056 + 1345 + 1381 + 1191}{4} = 1243.25$$

This first 4-month moving average can be used to forecast the shipments in May. Because 1259 shipments were actually made in May, the error of the forecast is

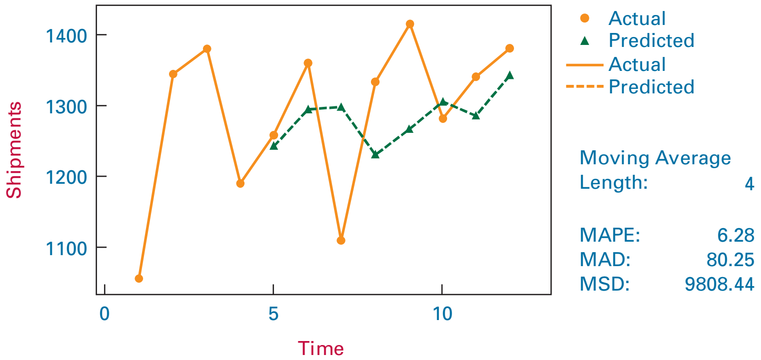
$$\text{Error}_{\text{May}} = 1259 - 1243.25 = 15.75$$

Shown next, along with the monthly shipments, are the 4-month moving averages and the errors of forecast when using the 4-month moving averages to predict the next month's shipments. The first moving average is displayed beside the month of May because it is computed by using January, February, March, and April and because it is being used to forecast the shipments for May. The rest of the 4-month moving averages and errors of forecast are as shown.

4-Month Moving Forecast

Month	Shipments	Average	Error
January	1056	—	—
February	1345	—	—
March	1381	—	—
April	1191	—	—
May	1259	1243.25	15.75
June	1361	1294.00	67.00
July	1110	1298.00	−188.00
August	1334	1230.25	103.75
September	1416	1266.00	150.00
October	1282	1305.25	−23.25
November	1341	1285.50	55.50
December	1382	1343.25	38.75

The following Minitab graph shows the actual shipment values and the forecast shipment values based on the 4-month moving averages. Notice that the moving averages are "smoothed" in comparison with the individual data values. They appear to be less volatile and seem to be attempting to follow the general trend of the data.



Weighted Moving Averages

A forecaster may want to place more weight on certain periods of time than on others. For example, a forecaster might believe that the previous month's value is three times as important in forecasting as other months. *A moving average in which some time periods are weighted differently than others* is called a **weighted moving average.**

As an example, suppose a 3-month weighted average is computed by weighting last month's value by 3, the value for the previous month by 2, and the value for the month before that by 1. This weighted average is computed as

$$\bar{x}_{weighted} = \frac{3(M_{t-1}) + 2(M_{t-2}) + 1(M_{t-3})}{6}$$

where

M_{t-1} = last month's value
M_{t-2} = value for the previous month
M_{t-3} = value for the month before the previous month

Notice that the divisor is 6. With a weighted average, the divisor always equals the total number of weights. In this example, the value of M_{t-1} counts three times as much as the value for M_{t-3}.

DEMONSTRATION PROBLEM 15.2

Compute a 4-month weighted moving average for the electric lighting and wiring data from Demonstration Problem 15.1, using weights of 4 for last month's value, 2 for the previous month's value, and 1 for each of the values from the 2 months prior to that.

Solution

The first weighted average is

$$\frac{4(1191) + 2(1381) + 1(1345) + 1(1056)}{8} = 1240.875$$

This moving average is recomputed for each ensuing month. Displayed next are the monthly values, the weighted moving averages, and the forecast error for the data.

Month	Shipments	4-Month Weighted Moving Average Forecast	Error
January	1056	—	—
February	1345	—	—
March	1381	—	—
April	1191	—	—
May	1259	1240.9	18.1
June	1361	1268.0	93.0
July	1110	1316.8	−206.8
August	1334	1201.5	132.5
September	1416	1272.0	144.0
October	1282	1350.4	−68.4
November	1341	1300.5	40.5
December	1382	1334.8	47.2

Note that in this problem the errors obtained by using the 4-month weighted moving average were greater than most of the errors obtained by using an unweighted 4-month moving average, as shown here.

Forecast Error, Unweighted 4-Month Moving Average	Forecast Error, Weighted 4-Month Moving Average
—	—
—	—
—	—
—	—
15.8	18.1
67.0	93.0
−188.0	−206.8
103.8	132.5
150.0	144.0
−23.3	−68.4
55.5	40.5
38.8	47.2

Larger errors with weighted moving averages are not always the case. The forecaster can experiment with different weights in using the weighted moving average as a technique. Many possible weighting schemes can be used.

Exponential Smoothing

Another forecasting technique, **exponential smoothing**, is *used to weight data from previous time periods with exponentially decreasing importance in the forecast.* Exponential smoothing is accomplished by multiplying the actual value for the present time period, X_t, by a value between 0 and 1 (the exponential smoothing constant) referred to as α (not the same α used for a Type I error) and adding that result to the product of the present time period's forecast, F_t and $(1 - \alpha)$. The following is a more formalized version.

EXPONENTIAL SMOOTHING

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

where

F_{t+1} = the forecast for the next time period $(t + 1)$
F_t = the forecast for the present time period (t)
X_t = the actual value for the present time period
α = a value between 0 and 1 referred to as the exponential smoothing constant.

The value of α is determined by the forecaster. The essence of this procedure is that the new forecast is a combination of the present forecast and the present actual value. If α is chosen to be less than .5, less weight is placed on the actual value than on the forecast of that value. If α is chosen to be greater than .5, more weight is being put on the actual value than on the forecast value.

As an example, suppose the prime interest rate for a time period is 5% and the forecast of the prime interest rate for this time period was 6%. If the forecast of the prime interest rate for the next period is determined by exponential smoothing with $\alpha = .3$, the forecast is

$$F_{t+1} = (.3)(5\%) + (1.0 - .3)(6\%) = 5.7\%$$

Notice that the forecast value of 5.7% for the next period is weighted more toward the previous forecast of 6% than toward the actual value of 5% because α is .3. Suppose we use $\alpha = .7$ as the exponential smoothing constant. Then,

$$F_{t+1} = (.7)(5\%) + (1.0 - .7)(6\%) = 5.3\%$$

This value is closer to the actual value of 5% than the previous forecast of 6% because the exponential smoothing constant, α, is greater than .5.

To see why this procedure is called exponential smoothing, examine the formula for exponential smoothing again.

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

If exponential smoothing has been used over a period of time, the forecast for F_t will have been obtained by

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

Substituting this forecast value, F_t, into the preceding equation for F_{t+1} produces

$$\begin{aligned} F_{t+1} &= \alpha \cdot X_t + (1 - \alpha)[\alpha \cdot X_{t-1} + (1 - \alpha) \cdot F_{t-1}] \\ &= \alpha \cdot X_t + \alpha(1 - \alpha) \cdot X_{t-1} + (1 - \alpha)^2 F_{t-1} \end{aligned}$$

but

$$F_{t-1} = \alpha \cdot X_{t-2} + (1 - \alpha)F_{t-2}$$

Substituting this value of F_{t-1} into the preceding equation for F_{t+1} produces

$$\begin{aligned} F_{t+1} &= \alpha \cdot X_t + \alpha(1 - \alpha) \cdot X_{t-1} + (1 - \alpha)^2 F_{t-1} \\ &= \alpha \cdot X_t + \alpha(1 - \alpha) \cdot X_{t-1} + (1 - \alpha)^2[\alpha \cdot X_{t-2} + (1 - \alpha)F_{t-2}] \\ &= \alpha \cdot X_t + \alpha(1 - \alpha) \cdot X_{t-1} + \alpha(1 - \alpha)^2 \cdot X_{t-2} + (1 - \alpha)^3 F_{t-2} \end{aligned}$$

Continuing this process shows that the weights on previous-period values and forecasts include $(1-\alpha)^n$ (exponential values). The following chart shows the values of α, $(1-\alpha)$, $(1-\alpha)^2$, and $(1-\alpha)^3$ for three different values of α. Included is the value of $\alpha(1-\alpha)^3$, which is the weight of the actual value for three time periods back. Notice the rapidly decreasing emphasis on values for earlier time periods. The impact of exponential smoothing on time-series data is to place much more emphasis on recent time periods. The choice of α determines the amount of emphasis.

α	$1-\alpha$	$(1-\alpha)^2$	$(1-\alpha)^3$	$\alpha(1-\alpha)^3$
.2	.8	.64	.512	.1024
.5	.5	.25	.125	.0625
.8	.2	.04	.008	.0064

Some forecasters use the computer to analyze time-series data for various values of α. By setting up criteria with which to judge the forecasting errors, forecasters can select the value of α that best fits the data.

The exponential smoothing formula

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

can be rearranged algebraically as

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

This form of the equation shows that the new forecast, F_{t+1}, equals the old forecast, F_t, plus an adjustment based on α times the error of the old forecast $(X_t - F_t)$. The smaller α is, the less impact the error has on the new forecast and the more the new forecast is like the old. It demonstrates the dampening effect of α on the forecasts.

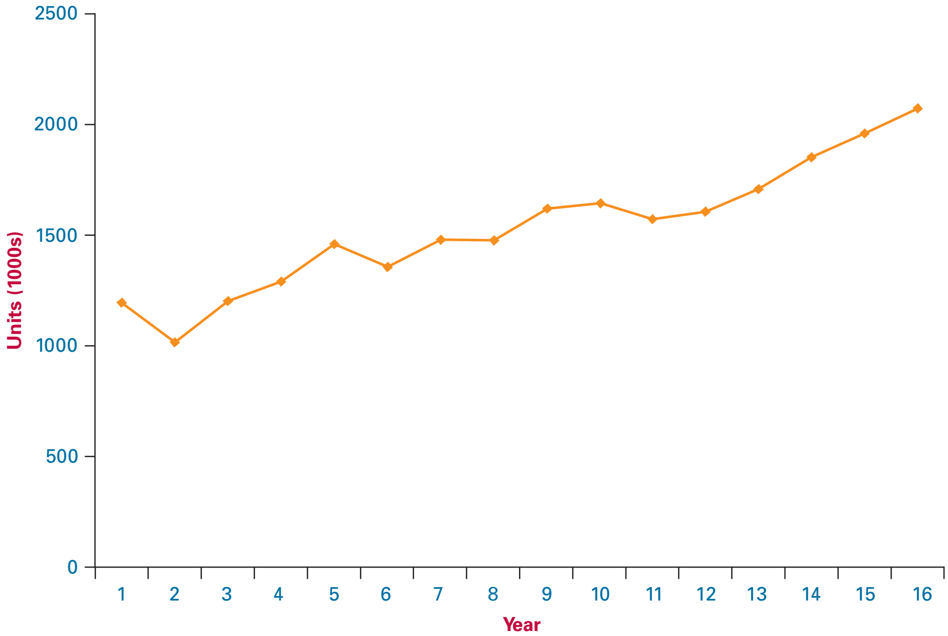
DEMONSTRATION PROBLEM 15.3

The U.S. Census Bureau reports the total units of new privately owned housing started over a 16-year recent period in the United States are given here. Use exponential smoothing to forecast the values for each ensuing time period. Work the problem using $\alpha = .2$, .5, and .8.

Year	Total Units (1000)
1	1193
2	1014
3	1200
4	1288
5	1457
6	1354
7	1477
8	1474
9	1617
10	1641
11	1569
12	1603
13	1705
14	1848
15	1956
16	2068

Solution

An Excel graph of these data is shown here.



The following table provides the forecasts with each of the three values of alpha. Note that because no forecast is given for the first time period, we cannot compute a forecast based on exponential smoothing for the second period. Instead, we use the actual value for the first period as the forecast for the second period to get started. As examples, the forecasts for the third, fourth, and fifth periods are computed for $\alpha = .2$ as follows.

$$F_3 = .2(1014) + .8(1193) = 1157.2$$
$$F_4 = .2(1200) + .8(1157.2) = 1165.8$$
$$F_5 = .2(1288) + .8(1165.8) = 1190.2$$

		$\alpha = .2$		$\alpha = .5$		$\alpha = .8$	
Year	Total Units (1000)	*F*	*e*	*F*	*e*	*F*	*e*
1	1193	—	—	—	—	—	—
2	1014	1193.0	−179.0	1193.0	−179.0	1193.0	−179.0
3	1200	1157.2	42.8	1103.5	96.5	1049.8	150.2
4	1288	1165.8	122.2	1151.8	136.2	1170.0	118.0
5	1457	1190.2	266.8	1219.9	237.1	1264.4	192.6
6	1354	1243.6	110.4	1338.4	15.6	1418.5	−64.5
7	1477	1265.7	211.3	1346.2	130.8	1366.9	110.1
8	1474	1307.9	166.1	1411.6	62.4	1455.0	19.0
9	1617	1341.1	275.9	1442.8	174.2	1470.2	146.8
10	1641	1396.3	244.7	1529.9	111.1	1587.6	53.4
11	1569	1445.2	123.8	1585.5	−16.5	1630.3	−61.3
12	1603	1470.0	133.0	1577.2	25.8	1581.3	21.7
13	1705	1496.6	208.4	1590.1	114.9	1598.7	106.3
14	1848	1538.3	309.7	1647.6	200.4	1683.7	164.3
15	1956	1600.2	355.8	1747.8	208.2	1815.1	140.9
16	2068	1671.4	396.6	1851.9	216.1	1927.8	140.2

	$\alpha = .2$	$\alpha = .5$	$\alpha = .8$
MAD:	209.8	128.3	111.2
MSE:	53,110.5	21,628.6	15,245.4

Which value of alpha works best on the data? At the bottom of the preceding analysis are the values of two different measurements of error for each of the three different values of alpha. With each measurement of error, $\alpha = .8$ produces the smallest measurement of error. Observe from the Excel graph of the original data that the data vary up and down considerably. In exponential smoothing, the value of alpha is multiplied by the actual value and $1 - \alpha$ is multiplied by the forecast value to get the next forecast. Because the actual values are varying considerably, the exponential smoothing value with the largest alpha seems to be forecasting the best. By placing the greatest weight on the actual values, the new forecast seems to predict the new value better.

STATISTICS IN BUSINESS TODAY

Forecasting the Economy by Scrap Metal Prices?

Economists are constantly on the lookout for valid indicators of a country's economy. Forecasters have sifted through oil indicators, the price of gold on the world markets, the Dow Jones averages, government-published indexes, and practically anything else that might seem related in some way to the state of the economy.

Would you believe that the price of scrap metal is a popular indicator of economic activity in the United States? Several well-known and experienced economic forecasters, including Federal Reserve chairman Alan Greenspan and the chief market analyst for Chase Manhattan, Donald Fine, believe that the price of scrap metal is a good indicator of the industrial economy.

Scrap metal is leftover copper, steel, aluminum, and other metals. Scrap metal is a good indicator of industrial activity because as manufacturing increases, the demand for scrap metals increases, as does the price of scrap metal. Donald Fine says that "scrap metal is the beginning of the production chain"; hence, an increasing demand for it is an indicator of increasing manufacturing production. Mr. Fine goes on to say that scrap metal is sometimes a better indicator of the future direction of the economy than many governmental statistics. In some cases, scrap metal correctly predicted no economic recovery when some government measures indicated that a recovery was underway.

Source: Anita Raghavan and David Wessel, "In Scraping Together Economic Data, Forecasters Turn to Scrap-Metal Prices," *The Wall Street Journal* (April 27, 1992), C1.

15.2 PROBLEMS

15.5 Use the following time-series data to answer the given questions.

Time Period	Value	Time Period	Value
1	27	6	66
2	31	7	71
3	58	8	86
4	63	9	101
5	59	10	97

a. Develop forecasts for periods 5 through 10 using 4-month moving averages.

b. Develop forecasts for periods 5 through 10 using 4-month weighted moving averages. Weight the most recent month by a factor of 4, the previous month by 2, and the other months by 1.

c. Compute the errors of the forecasts in parts (a) and (b) and observe the differences in the errors forecast by the two different techniques.

15.6 Following are time-series data for eight different periods. Use exponential smoothing to forecast the values for periods 3 through 8. Use the value for the first period as the forecast for the second period. Compute forecasts using two different values of alpha,

$\alpha = .1$ and $\alpha = .8$. Compute the errors for each forecast and compare the errors produced by using the two different exponential smoothing constants.

Time Period	Value	Time Period	Value
1	211	5	242
2	228	6	227
3	236	7	217
4	241	8	203

15.7 Following are time-series data for nine time periods. Use exponential smoothing with constants of .3 and .7 to forecast time periods 3 through 9. Let the value for time period 1 be the forecast for time period 2. Compute additional forecasts for time periods 4 through 9 using a 3-month moving average. Compute the errors for the forecasts and discuss the size of errors under each method.

Time Period	Value	Time Period	Value
1	9.4	6	11.0
2	8.2	7	10.3
3	7.9	8	9.5
4	9.0	9	9.1
5	9.8		

15.8 The U.S. Census Bureau publishes data on factory orders for all manufacturing, durable goods, and nondurable goods industries. Shown here are factory orders in the United States over a 13-year period ($ billion).

a. Use these data to develop forecasts for the years 6 through 13 using a 5-year moving average.

b. Use these data to develop forecasts for the years 6 through 13 using a 5-year weighted moving average. Weight the most recent year by 6, the previous year by 4, the year before that by 2, and the other years by 1.

c. Compute the errors of the forecasts in parts (a) and (b) and observe the differences in the errors of the forecasts.

Year	Factory Orders ($ billion)
1	2,512.7
2	2,739.2
3	2,874.9
4	2,934.1
5	2,865.7
6	2,978.5
7	3,092.4
8	3,356.8
9	3,607.6
10	3,749.3
11	3,952.0
12	3,949.0
13	4,137.0

15.9 The following data show the number of issues from initial public offerings (IPOs) for a 13-year period released by the Securities Data Company. Use these data to develop forecasts for the years 3 through 13 using exponential smoothing techniques with alpha values of .2 and .9. Let the forecast for year 2 be the value for year 1. Compare the results by examining the errors of the forecasts.

Year	Number of Issues
1	332
2	694
3	518
4	222
5	209
6	172
7	366
8	512
9	667
10	571
11	575
12	865
13	609

15.3 TREND ANALYSIS

There are several ways to determine trend in time-series data and one of the more prominent is regression analysis. In Section 12.9, we explored the use of simple regression analysis in determining the equation of a trend line. In time-series regression trend analysis, the response variable, Y, is the variable being forecast, and the independent variable, X, represents time.

Many possible trend fits can be explored with time-series data. In this section we examine only the linear model and the quadratic model because they are the easiest to understand and simplest to compute. Because seasonal effects can confound trend analysis, it is assumed here that no seasonal effects occur in the data or they were removed prior to determining the trend.

Linear Regression Trend Analysis

The data in Table 15.5 represent 35 years of data on the average length of the workweek in Canada for manufacturing workers. A regression line can be fit to these data by using the

TABLE 15.5

Average Hours per Week in Manufacturing by Canadian Workers

Time Period	Hours	Time Period	Hours
1	37.2	19	36.0
2	37.0	20	35.7
3	37.4	21	35.6
4	37.5	22	35.2
5	37.7	23	34.8
6	37.7	24	35.3
7	37.4	25	35.6
8	37.2	26	35.6
9	37.3	27	35.6
10	37.2	28	35.9
11	36.9	29	36.0
12	36.7	30	35.7
13	36.7	31	35.7
14	36.5	32	35.5
15	36.3	33	35.6
16	35.9	34	36.3
17	35.8	35	36.5
18	35.9		

Source: Data prepared by the U.S. Bureau of Labor Statistics, Office of Productivity and Technology.

FIGURE 15.5

Excel Regression Output for Hours Worked Using Linear Trend

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.782
R Square	0.611
Adjusted R Square	0.600
Standard Error	0.5090
Observations	35

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	13.4467	13.4467	51.91	0.000000029
Residual	33	8.5487	0.2591		
Total	34	21.9954			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	37.4161	0.1758	212.81	0.000000000
Year	–0.0614	0.0085	–7.20	0.000000029

time periods as the independent variable and length of workweek as the dependent variable. Because the time periods are consecutive, they can be entered as X along with the time-series data (Y) into a regression analysis. The linear model explored in this example is

$$Y_i = \beta_0 + \beta_1 X_{ti} + \epsilon_i$$

where

Y_i = data value for period i
X_{ti} = ith time period

Figure 15.5 shows the Excel regression output for this example. By using the coefficients of the X variable and intercept, the equation of the trend line can be determined to be

$$\hat{Y} = 37.4161 - .0614X_t$$

The slope indicates that for every unit increase in time period, X_t, a predicted decrease of .0614 occurs in the length of the average workweek in manufacturing. Because the workweek is measured in hours, the length of the average workweek decreases by an average of (.0614)(60 minutes) = 3.7 minutes each year in Canada in manufacturing. The Y intercept, 37.4161, indicates that in the year prior to the first period of these data the average workweek was 37.4161 hours.

The probability of the t ratio (.00000003) indicates that significant linear trend is present in the data. In addition, $R^2 = .611$ indicates considerable predictability in the model. Inserting the various period values (1, 2, 3, . . . , 35) into the preceding regression equation produces the predicted values of Y that are the trend. For example, for period 23 the predicted value is

$$\hat{Y} = 37.4161 - .0614(23) = 36.0 \text{ hours}$$

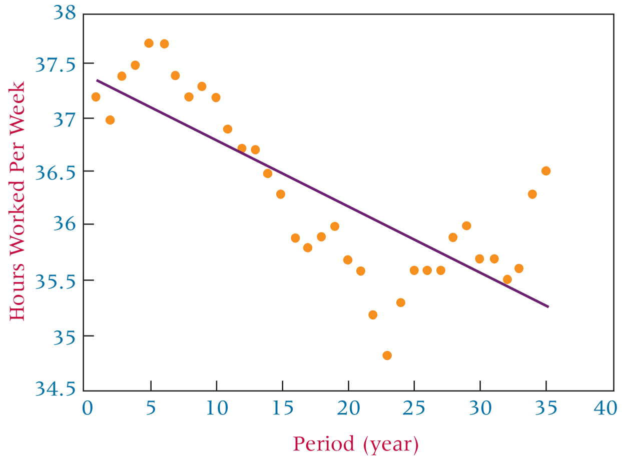
The model was developed with 35 periods (years). From this model, the average workweek in Canada in manufacturing for period 41 (the 41st year) can be forecast:

$$\hat{Y} = 37.4161 - .0614(41) = 34.9 \text{ hours}$$

Figure 15.6 presents an Excel scatter plot of the average workweek lengths over the 35 periods (years). In this Excel plot, the trend line has been fitted through the points. Observe the general downward trend of the data, but also note the somewhat cyclical nature of the points. Because of this pattern, a forecaster might want to determine whether a quadratic model is a better fit for trend.

FIGURE 15.6

Excel Graph of Canadian Manufacturing Data with Trend Line



Regression Trend Analysis Using Quadratic Models

In addition to linear regression, forecasters can explore using quadratic regression models to predict data by using the time-series periods. The quadratic regression model is

$$Y_i = \beta_0 + \beta_1 X_{ti} + \beta_2 X_{ti}^2 + \epsilon_i$$

where

Y_i = the time-series data value for period i
X_{ti} = the ith period
X_{ti}^2 = the square of the ith period

This model can be implemented in time-series trend analysis by using the time periods squared as an additional predictor. Thus, in the hours worked example, besides using $X_t = 1, 2, 3, 4, \ldots, 35$ as a predictor, we would also use $X_t^2 = 1, 4, 9, 16, \ldots, 1225$ as a predictor.

Table 15.6 provides the data needed to compute a quadratic regression trend model on the manufacturing workweek data. Note that the table includes the original data, the time periods, and the time periods squared.

TABLE 15.6

Data for Quadratic Fit of Manufacturing Workweek Example

Time Period	(Time Period)^2	Hours	Time Period	(Time Period)^2	Hours
1	1	37.2	19	361	36.0
2	4	37.0	20	400	35.7
3	9	37.4	21	441	35.6
4	16	37.5	22	484	35.2
5	25	37.7	23	529	34.8
6	36	37.7	24	576	35.3
7	49	37.4	25	625	35.6
8	64	37.2	26	676	35.6
9	81	37.3	27	729	35.6
10	100	37.2	28	784	35.9
11	121	36.9	29	841	36.0
12	144	36.7	30	900	35.7
13	169	36.7	31	961	35.7
14	196	36.5	32	1024	35.5
15	225	36.3	33	1089	35.6
16	256	35.9	34	1156	36.3
17	289	35.8	35	1225	36.5
18	324	35.9			

Source: Data prepared by the U.S. Bureau of Labor Statistics, Office of Productivity and Technology.

FIGURE 15.7

Excel Regression Output for Canadian Manufacturing Example with Quadratic Trend

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.873
R Square	0.761
Adjusted R Square	0.747
Standard Error	0.4049
Observations	35

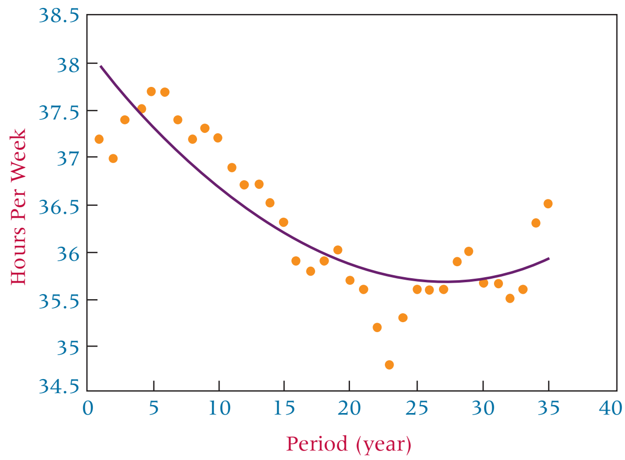
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	16.7483	8.3741	51.07	0.0000000001
Residual	32	5.2472	0.1640		
Total	34	21.9954			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	38.1644	0.2177	175.34	0.0000000
Time Period	–0.1827	0.0279	–6.55	0.0000002
(Time Period)Sq	0.0034	0.0008	4.49	0.0000876

FIGURE 15.8

Excel Graph of Canadian Manufacturing Data with a Second-Degree Polynomial Fit



The Excel computer output for this quadratic trend regression analysis is shown in Figure 15.7. We see that the quadratic regression model produces an R^2 of .761 with both X_t and X_t^2 in the model. The linear model produced an R^2 of .611 with X_t alone. The quadratic regression seems to add some predictability to the trend model. Figure 15.8 displays an Excel scatter plot of the week work data with a second-degree polynomial fit through the data.

DEMONSTRATION PROBLEM 15.4

Demonstration Problem

Following are data on the employed U.S. civilian labor force (100,000) for 1991 through 2007, obtained from the U.S. Bureau of Labor Statistics. Use regression analysis to fit a trend line through the data. Explore a quadratic regression trend also. Does either model do well? Compare the two models.

Year	Labor Force (100,000)
1991	117.72
1992	118.49
1993	120.26
1994	123.06
1995	124.90
1996	126.71
1997	129.56
1998	131.46
1999	133.49
2000	136.89
2001	136.93
2002	136.49
2003	137.74
2004	139.25
2005	141.73
2006	144.43
2007	146.05

Solution

Recode the time periods as 1 through 17 and let that be *X*. Run the regression analysis with the labor force members as *Y*, the dependent variable, and the time period as the independent variable. Now square all the X values, resulting in 1, 4, 9, . . . , 225, 256, 289 and let those formulate a second predictor (X^2). Run the regression analysis to predict the number in the labor force with both the time period variable (*X*) and the (time period)^2 variable. The Minitab output for each of these regression analyses follows.

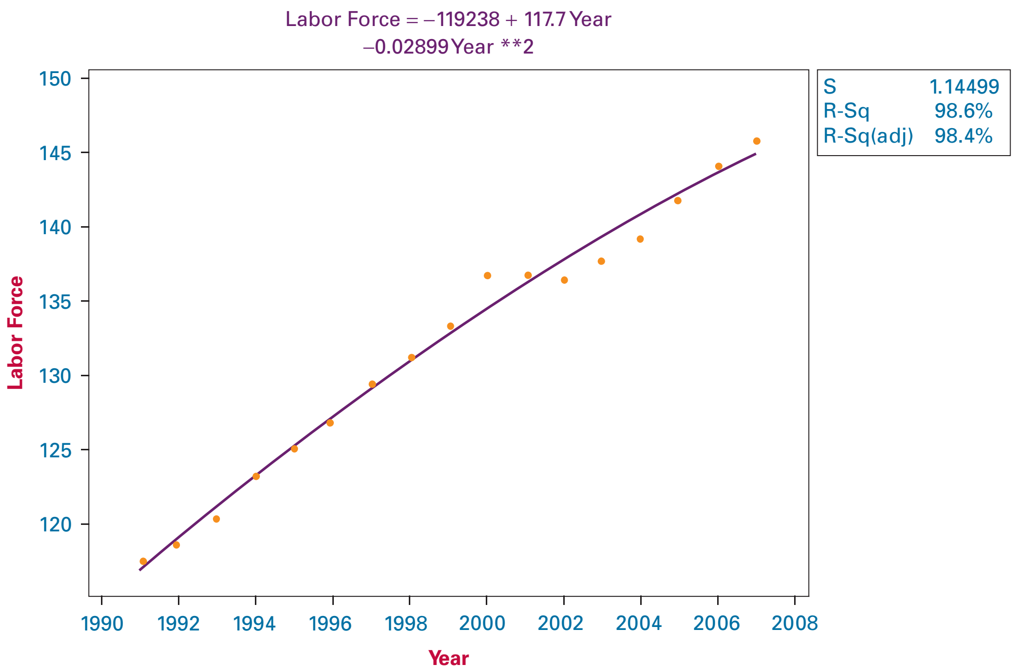
Regression Analysis: Labor Force Versus Year

```
The regression equation is
Labor Force =-3390 +1.76 Year

Predictor        Coef  SE Coef       T           P
Constant      -3390.0    127.4  -26.60       0.000
Year          1.76191  0.06375   27.64       0.000

S = 1.28762     R-Sq = 98.1%    R-Sq(adj) = 97.9%

Analysis of Variance

Source           DF       SS       MS        F        P
Regression        1   1266.6   1266.6   763.93    0.000
Residual Error   15     24.9      1.7
Total            16   1291.4
```

Regression Analysis: Labor Force Versus Year, Year Sq

```
The regression equation is
Labor Force =-119238 +118 Year - 0.0290 Year Sq

Predictor        Coef  SE Coef       T           P
Constant      -119238    51966   -2.29       0.038
Year           117.67    51.99    2.26       0.040
Year Sq      -0.02899  0.01300   -2.23       0.043

S = 1.14499     R-Sq = 98.6%    R-Sq(adj) = 98.4%

Analysis of Variance

Source           DF       SS       MS        F        P
Regression        2  1273.08   636.54   485.54    0.000
Residual Error   14    18.35     1.31
Total            16  1291.44
```

A comparison of the models shows that the linear model accounts for over 98% of the variability in the labor force figures, and the quadratic model only increases that predictability to 98.6%. Shown next are Minitab scatter plots of the data. First is the linear model, and then the quadratic model is presented.



Holt's Two-Parameter Exponential Smoothing Method

The exponential smoothing technique presented in Section 15.2 (single exponential smoothing) is appropriate to use in forecasting stationary time-series data but is ineffective in forecasting time-series data with a trend because the forecasts will lag behind the trend. However, another exponential smoothing technique, Holt's two-parameter exponential smoothing method, can be used for trend analysis. Holt's technique uses weights (β) to smooth the trend in a manner similar to the smoothing used in single exponential smoothing (α). Using these two weights and several equations, Holt's method is able to develop forecasts that include both a smoothing value and a trend value. A more detailed explanation of Holt's two-parameter exponential smoothing method, along with examples and practice problems, can be accessed at WileyPLUS and at the Wiley Web site for this text.

15.3 PROBLEMS

15.10 The "Economic Report to the President of the United States" included data on the amounts of manufacturers' new and unfilled orders in millions of dollars. Shown here are the figures for new orders over a 21-year period. Use a computer to develop a regression model to fit the trend effects for these data. Use a linear model and then try a quadratic model. How well does either model fit the data?

Year	Total Number of New Orders	Year	Total Number of New Orders
1	55,022	12	168,025
2	55,921	13	162,140
3	64,182	14	175,451
4	76,003	15	192,879
5	87,327	16	195,706
6	85,139	17	195,204
7	99,513	18	209,389
8	115,109	19	227,025
9	131,629	20	240,758
10	147,604	21	243,643
11	156,359		

15.11 The following data on the number of union members in the United States for the years 1984 through 2008 are provided by the U.S. Bureau of Labor Statistics. Using regression techniques discussed in this section, analyze the data for trend. Develop a scatter plot of the data and fit the trend line through the data. Discuss the strength of the model.

Year	Union Members (1000s)	Year	Union Members (1000s)
1984	17,340	1997	16,110
1985	16,996	1998	16,211
1986	16,975	1999	16,477
1987	16,913	2000	16,334
1988	17,002	2001	16,305
1989	16,960	2002	16,145
1990	16,740	2003	15,776
1991	16,568	2004	15,472
1992	16,390	2005	15,685
1993	16,598	2006	15,359
1994	16,748	2007	15,670
1995	16,360	2008	16,098
1996	16,269		

15.12 Shown below are dollar figures for commercial and industrial loans at all commercial banks in the United States as recorded for the month of April during a recent 9-year period and published by the Federal Reserve Bank of St. Louis. Plot the data, fit a trend line, and discuss the strength of the regresssion model. In addition, explore a quadratic trend and compare the results of the two models.

Year	Loans ($ billions)
1	741.0
2	807.4
3	871.3
4	951.6
5	1,033.6
6	1,089.8
7	1,002.6
8	940.8
9	888.5

15.4 SEASONAL EFFECTS

Earlier in the chapter, we discussed the notion that time-series data consist of four elements: trend, cyclical effects, seasonality, and irregularity. In this section, we examine techniques for identifying seasonal effects. **Seasonal effects** are *patterns of data behavior that occur in periods of time of less than one year.* How can we separate out the seasonal effects?

Decomposition

One of the main techniques for isolating the effects of seasonality is **decomposition**. The decomposition methodology presented here uses the multiplicative model as its basis. The multiplicative model is:

$$T \cdot C \cdot S \cdot I$$

where

T = trend
C = cyclicality
S = seasonality
I = irregularity

TABLE 15.7

Shipments of Household Appliances

Year	Quarter	Shipments
1	1	4009
	2	4321
	3	4224
	4	3944
2	1	4123
	2	4522
	3	4657
	4	4030
3	1	4493
	2	4806
	3	4551
	4	4485
4	1	4595
	2	4799
	3	4417
	4	4258
5	1	4245
	2	4900
	3	4585
	4	4533

To illustrate the decomposition process, we will use the 5-year quarterly time-series data on U.S. shipments of household appliances given in Table 15.7. Figure 15.9 provides a graph of these data.

According to the multiplicative time-series model, $T \cdot C \cdot S \cdot I$, the data can contain the elements of trend, cyclical effects, seasonal effects, and irregular fluctuations. The process of isolating the seasonal effects begins by determining $T \cdot C$ for each value and dividing the time-series data ($T \cdot C \cdot S \cdot I$) by $T \cdot C$. The result is

$$\frac{T \cdot C \cdot S \cdot I}{T \cdot C} = S \cdot I$$

The resulting expression contains seasonal effects along with irregular fluctuations. After reducing the time-series data to the effects of SI (seasonality and irregularity), a method for eliminating the irregular fluctuations can be applied, leaving only the seasonal effects.

Suppose we start with time-series data that cover several years and are measured in quarterly increments. If we average the data over four quarters, we will have "dampened" the seasonal effects of the data because the rise and fall of values during the quarterly periods will have been averaged out over the year.

We begin by computing a 4-quarter moving average for quarter 1 through quarter 4 of year 1, using the data from Table 15.7.

$$\text{4-quarter average} = \frac{4{,}009 + 4{,}321 + 4{,}224 + 3{,}944}{4} = 4{,}124.5$$

FIGURE 15.9

Minitab Time-Series Graph of Household Appliance Data

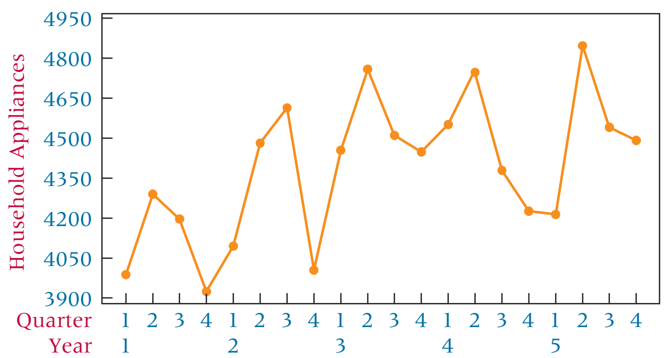


TABLE 15.8

Development of 4-Quarter Moving Averages for the Household Appliance Data

Quarter	Actual Values $(T \cdot C \cdot S \cdot I)$	4-Quarter Moving Total	4-Quarter 2-Year Moving Total	Centered Moving Average $(T \cdot C)$	Ratios of Actual Values to Moving Averages $(S \cdot I) \cdot (100)$
1 (year 1)	4,009				
2	4,321				
		16,498			
3	4,224		33,110	4,139	102.05
		16,612			
4	3,944		33,425	4,178	94.40
		16,813			
1 (year 2)	4,123		34,059	4,257	96.85
		17,246			
2	4,522		34,578	4,322	104.63
		17,332			
3	4,657		35,034	4,379	106.35
		17,702			
4	4,030		35,688	4,461	90.34
		17,986			
1 (year 3)	4,493		35,866	4,483	100.22
		17,880			
2	4,806		36,215	4,527	106.16
		18,335			
3	4,551		36,772	4,597	99.00
		18,437			
4	4,485		36,867	4,608	97.33
		18,430			
1 (year 4)	4,595		36,726	4,591	100.09
		18,296			
2	4,799		36,365	4,546	105.57
		18,069			
3	4,417		35,788	4,474	98.73
		17,719			
4	4,258		35,539	4,442	95.86
		17,820			
1 (year 5)	4,245		35,808	4,476	94.84
		17,988			
2	4,900		36,251	4,531	108.14
		18,263			
3	4,585				
4	4,533				

The 4-quarter moving average for quarter 1 through quarter 4 of year 1 is 4,124.5 ($ million) worth of shipments. Because the 4-quarter average is in the middle of the four quarters, it would be placed in the decomposition table between quarter 2 and quarter 3.

Quarter 1
Quarter 2
——— 4,124.5
Quarter 3
Quarter 4

To remove seasonal effects, we need to determine a value that is "centered" with each month. To find this value, instead of using a 4-quarter moving average, we use 4-quarter moving totals and then sum two consecutive moving totals. This 8-quarter total value is divided by 8 to produce a "centered" 4-quarter moving average that lines up across from a quarter. Using this method is analogous to computing two consecutive 4-quarter moving averages and averaging them, thus producing a value that falls on line with a quarter, in between the two averages. The results of using this procedure on the data from Table 15.7 are shown in Table 15.8 in column 5.

A 4-quarter moving total can be computed on these data starting with quarter 1 of year 1 through quarter 4 of year 1 as follows:

$$\text{First Moving Total} = 4{,}009 + 4{,}321 + 4{,}224 + 3{,}944 = 16{,}498$$

In Table 15.8, 16,498 is between quarter 2 and quarter 3 of year 1. The 4-month moving total for quarter 2 of year 1 through quarter 1 of year 2 is

$$\text{Second Moving Total} = 4{,}321 + 4{,}224 + 3{,}944 + 4{,}123 = 16{,}612$$

In Table 15.8, this value is between quarter 3 and quarter 4 of year 1. The 8-quarter (2-year) moving total is computed for quarter 3 of year 1 as

$$\text{8-Quarter Moving Total} = 16{,}498 + 16{,}612 = 33{,}110$$

TABLE 15.9

Seasonal Indexes for the Household Appliance Data

Quarter	Year 1	Year 2	Year 3	Year 4	Year 5
1	—	96.85	100.22	100.09	94.84
2	—	104.63	106.16	105.57	108.14
3	102.05	106.35	99.00	98.73	—
4	94.40	90.34	97.33	95.86	—

Notice that in Table 15.8 this value is centered with quarter 3 of year 1 because it is between the two adjacent 4-quarter moving totals. Dividing this total by 8 produces the 4-quarter moving average for quarter 3 of year 1 shown in column 5 of Table 15.8.

$$\frac{33{,}110}{8} = 4{,}139$$

Column 3 contains the uncentered 4-quarter moving totals, column 4 contains the 2-year centered moving totals, and column 5 contains the 4-quarter centered moving averages.

The 4-quarter centered moving averages shown in column 5 of Table 15.8 represent $T \cdot C$. Seasonal effects have been removed from the original data (actual values) by summing across the 4-quarter periods. Seasonal effects are removed when the data are summed across the time periods that include the seasonal periods and the irregular effects are smoothed, leaving only trend and cycle.

Column 2 of Table 15.8 contains the original data (actual values), which include all effects $(T \cdot C \cdot S \cdot I)$. Column 5 contains only the trend and cyclical effects, $T \cdot C$. If column 2 is divided by column 5, the result is $S \cdot I$, which is displayed in column 6 of Table 15.8.

The values in column 6, sometimes called ratios of actuals to moving average, have been multiplied by 100 to index the values. These values are thus seasonal indexes. An **index number** is *a ratio of a measure taken during one time frame to that same measure taken during another time frame, usually denoted as the time period*. Often the ratio is multipled by 100 and expressed as a percentage. Index numbers will be discussed more fully in section 15.6. Column 6 contains the effects of seasonality and irregular fluctuations. Now we must remove the irregular effects.

Table 15.9 contains the values from column 6 of Table 15.8 organized by quarter and year. Each quarter in these data has four seasonal indexes. Throwing out the high and low index for each quarter eliminates the extreme values. The remaining two indexes are averaged as follows for quarter 1.

Quarter 1: 96.85 100.22 100.09 94.84
Eliminate: 94.84 and 100.22
Average the Remaining Indexes:

$$\bar{X}_{Q1\,index} = \frac{96.85 + 100.09}{2} = 98.47$$

Table 15.10 gives the final seasonal indexes for all the quarters of these data.

TABLE 15.10

Final Seasonal Indexes for the Household Appliance Data

Quarter	Index
1	98.47
2	105.87
3	100.53
4	95.13

After the final adjusted seasonal indexes are determined, the original data can be **deseasonalized**. The deseasonalization of actual values is relatively common with data published by the government and other agencies. Data can be deseasonalized by dividing the actual values, which consist of $T \cdot C \cdot S \cdot I$, by the final adjusted seasonal effects.

$$\text{Deseasonalized Data} = \frac{T \cdot C \cdot S \cdot I}{S} = T \cdot C \cdot I$$

Because the seasonal effects are in terms of index numbers, the seasonal indexes must be divided by 100 before deseasonalization. Shown here are the computations for deseasonalizing the household appliance data from Table 15.7 for quarter 1 of year 1.

$$\text{Year 1 Quarter 1 Actual} = 4{,}009$$
$$\text{Year 1 Quarter 1 Seasonal Index} = 98.47$$
$$\text{Year 1 Quarter 1 Deseasonalized Value} = \frac{4{,}009}{.9847} = 4{,}071.3$$

Table 15.11 gives the deseasonalized data for this example for all years.

Figure 15.10 is a graph of the deseasonalized data.

TABLE 15.11
Deseasonalized Household Appliance Data

Year	Quarter	Shipments Actual Values $(T \cdot C \cdot S \cdot I)$	Seasonal Indexes S	Deseasonalized Data $T \cdot C \cdot I$
1	1	4,009	98.47	4,071
	2	4,321	105.87	4,081
	3	4,224	100.53	4,202
	4	3,944	95.13	4,146
2	1	4,123	98.47	4,187
	2	4,522	105.87	4,271
	3	4,657	100.53	4,632
	4	4,030	95.13	4,236
3	1	4,493	98.47	4,563
	2	4,806	105.87	4,540
	3	4,551	100.53	4,527
	4	4,485	95.13	4,715
4	1	4,595	98.47	4,666
	2	4,799	105.87	4,533
	3	4,417	100.53	4,394
	4	4,258	95.13	4,476
5	1	4,245	98.47	4,311
	2	4,900	105.87	4,628
	3	4,585	100.53	4,561
	4	4,533	95.13	4,765

Finding Seasonal Effects with the Computer

Through Minitab, decomposition can be performed on the computer with relative ease. The commands for this procedure are given at the end of the chapter in the Using the Computer section. Figure 15.11 displays Minitab output for seasonal decomposition of the household appliance example. Note that the seasonal indexes are virtually identical to those shown in Table 15.10 computed by hand.

Winters' Three-Parameter Exponential Smoothing Method

Holt's two-parameter exponential smoothing method can be extended to include seasonal analysis. This technique, referred to as Winters' method, not only smoothes observations

FIGURE 15.10
Graph of the Deseasonalized Household Appliance Data

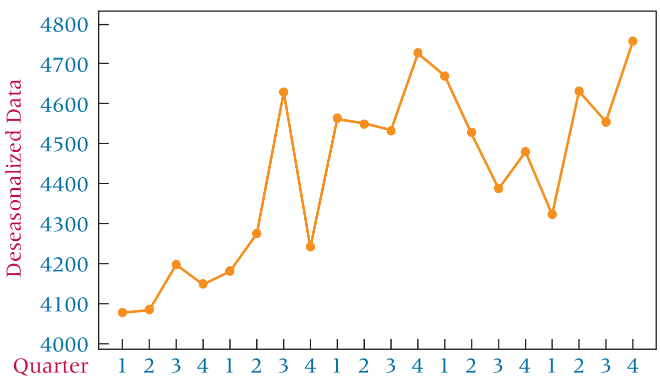


FIGURE 15.11

Minitab Output for Seasonal Decomposition of the Household Appliance Data

```
Time-Series Decomposition for Shipments

Multiplicative Model

Data       Shipments
Length     20
NMissing   0

Seasonal   Indices

  Period     Index
       1   0.98469
       2   1.05871
       3   1.00536
       4   0.95124
```

and trend but also smoothes the seasonal effects. In addition to the single exponential smoothing weight of α and the trend weight of β, Winters' method introduces γ, a weight for seasonality. Using these three weights and several equations, Winters' method is able to develop forecasts that include a smoothing value for observations, a trend value, and a seasonal value. A more detailed explanation of Winters' three-parameter exponential smoothing method along with examples and practice problems is presented in WileyPLUS and at the Wiley Web site for this text.

15.4 PROBLEMS

15.13 The U.S. Department of Agriculture publishes statistics on the production of various types of food commodities by month. Shown here are the production figures on broccoli for January of a recent year through December of the next year. Use these data to compute 12-month centered moving averages ($T \cdot C$). Using these computed values, determine the seasonal effects ($S \cdot I$).

Month	Broccoli (million pounds)	Month	Broccoli (million pounds)
January (1st year)	132.5	January (2nd year)	104.9
February	164.8	February	99.3
March	141.2	March	102.0
April	133.8	April	122.4
May	138.4	May	112.1
June	150.9	June	108.4
July	146.6	July	119.0
August	146.9	August	119.0
September	138.7	September	114.9
October	128.0	October	106.0
November	112.4	November	111.7
December	121.0	December	112.3

15.14 The U.S. Department of Commerce publishes census information on manufacturing. Included in these figures are monthly shipment data for the paperboard container and box industry shown on the next page for 6 years. The shipment figures are given in millions of dollars. Use the data to analyze the effects of seasonality, trend, and cycle. Develop the trend model with a linear model only.

Month	Shipments	Month	Shipments
January (year 1)	1,891	January (year 4)	2,336
February	1,986	February	2,474
March	1,987	March	2,546
April	1,987	April	2,566
May	2,000	May	2,473
June	2,082	June	2,572
July	1,878	July	2,336
August	2,074	August	2,518
September	2,086	September	2,454
October	2,045	October	2,559
November	1,945	November	2,384
December	1,861	December	2,305
Month	**Shipments**	**Month**	**Shipments**
January (year 2)	1,936	January (year 5)	2,389
February	2,104	February	2,463
March	2,126	March	2,522
April	2,131	April	2,417
May	2,163	May	2,468
June	2,346	June	2,492
July	2,109	July	2,304
August	2,211	August	2,511
September	2,268	September	2,494
October	2,285	October	2,530
November	2,107	November	2,381
December	2,077	December	2,211
Month	**Shipments**	**Month**	**Shipments**
January (year 3)	2,183	January (year 6)	2,377
February	2,230	February	2,381
March	2,222	March	2,268
April	2,319	April	2,407
May	2,369	May	2,367
June	2,529	June	2,446
July	2,267	July	2,341
August	2,457	August	2,491
September	2,524	September	2,452
October	2,502	October	2,561
November	2,314	November	2,377
December	2,277	December	2,277

15.5 AUTOCORRELATION AND AUTOREGRESSION

Data values gathered over time are often correlated with values from past time periods. This characteristic can cause problems in the use of regression in forecasting and at the same time can open some opportunities. One of the problems that can occur in regressing data over time is autocorrelation.

Autocorrelation

Autocorrelation, or **serial correlation**, occurs in data *when the error terms of a regression forecasting model are correlated.* The likelihood of this occurring with business data increases over time, particularly with economic variables. Autocorrelation can be a problem in using regression analysis as the forecasting method because one of the assumptions underlying regression analysis is that the error terms are independent or random (not correlated). In most business analysis situations, the correlation of error terms is likely to occur as positive autocorrelation (positive errors are associated with positive errors of comparable magnitude and negative errors are associated with negative errors of comparable magnitude).

When autocorrelation occurs in a regression analysis, several possible problems might arise. First, the estimates of the regression coefficients no longer have the minimum variance property and may be inefficient. Second, the variance of the error terms may be greatly underestimated by the mean square error value. Third, the true standard deviation of the estimated regression coefficient may be seriously underestimated. Fourth, the confidence intervals and tests using the t and F distributions are no longer strictly applicable.

First-order autocorrelation results from correlation between the error terms of adjacent time periods (as opposed to two or more previous periods). If first-order autocorrelation is present, the error for one time period, e_t, is a function of the error of the previous time period, e_{t-1}, as follows.

$$e_t = \rho e_{t-1} + \nu_t$$

The first-order autocorrelation coefficient, ρ, measures the correlation between the error terms. It is a value that lies between −1 and 0 and +1, as does the coefficient of correlation discussed in Chapter 12. ν_t is a normally distributed independent error term. If positive autocorrelation is present, the value of ρ is between 0 and +1. If the value of ρ is 0, $e_t = \nu_t$, which means there is no autocorrelation and e_t is just a random, independent error term.

One way to *test to determine whether autocorrelation is present in a time-series regression analysis* is by using the **Durbin-Watson test** for autocorrelation. Shown next is the formula for computing a Durbin-Watson test for autocorrelation.

DURBIN-WATSON TEST

$$D = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

where

n = the number of observations

Note from the formula that the Durbin-Watson test involves finding the difference between successive values of error $(e_t - e_{t-1})$. If errors are positively correlated, this difference will be smaller than with random or independent errors. Squaring this term eliminates the cancellation effects of positive and negative terms.

The null hypothesis for this test is that there is *no* autocorrelation. For a two-tailed test, the alternative hypothesis is that there *is* autocorrelation.

$$H_0\text{: } \rho = 0$$
$$H_a\text{: } \rho \neq 0$$

As mentioned before, most business forecasting autocorrelation is positive autocorrelation. In most cases, a one-tailed test is used.

$$H_0\text{: } \rho = 0$$
$$H_a\text{: } \rho > 0$$

In the Durbin-Watson test, D is the observed value of the Durbin-Watson statistic using the residuals from the regression analysis. A critical value for D can be obtained from the values of α, n, and k by using Table A.9 in the appendix, where α is the level of significance, n is the number of data items, and k is the number of predictors. Two Durbin-Watson tables are given in the appendix. One table contains values for $\alpha = .01$ and the other for $\alpha = .05$. The Durbin-Watson tables in Appendix A include values for d_U and d_L. These values range from 0 to 4. If the observed value of D is above d_U, we fail to reject the null hypothesis and there is no significant autocorrelation. If the observed value of D is below d_L, the null hypothesis is rejected and there is autocorrelation. Sometimes the observed statistic, D, is between the values of d_U and d_L. In this case, the Durbin-Watson test is inconclusive.

TABLE 15.12

U.S. Crude Oil Production and Natural Gas Withdrawals over a 25-Year Period

Year	Crude Oil Production (1000s)	Natural Gas Withdrawals from Natural Gas Wells (1000s)
1	8.597	17.573
2	8.572	17.337
3	8.649	15.809
4	8.688	14.153
5	8.879	15.513
6	8.971	14.535
7	8.680	14.154
8	8.349	14.807
9	8.140	15.467
10	7.613	15.709
11	7.355	16.054
12	7.417	16.018
13	7.171	16.165
14	6.847	16.691
15	6.662	17.351
16	6.560	17.282
17	6.465	17.737
18	6.452	17.844
19	6.252	17.729
20	5.881	17.590
21	5.822	17.726
22	5.801	18.129
23	5.746	17.795
24	5.681	17.819
25	5.430	17.739

As an example, consider Table 15.12, which contains crude oil production and natural gas withdrawal data for the United States over a 25-year period published by the Energy Information Administration in their Annual Energy Review. A regression line can be fit through these data to determine whether the amount of natural gas withdrawals can be predicted by the amount of crude oil production. The resulting errors of prediction can be tested by the Durbin-Watson statistic for the presence of significant positive autocorrelation by using $\alpha = .05$. The hypotheses are

$$H_0\colon \rho = 0$$
$$H_a\colon \rho > 0$$

The following regression equation was obtained by means of a Minitab computer analysis.

$$\text{Natural Gas Withdrawals} = 22.7372 - 0.8507\ \text{Crude Oil Production}$$

Using the values for crude oil production (X) from Table 15.12 and the regression equation shown here, predicted values of Y (natural gas withdrawals) can be computed. From the predicted values and the actual values, the errors of prediction for each time interval, e_t, can be calculated. Table 15.13 shows the values of $\hat{Y}$, e_t, e_t^2, $(e_t - e_{t-1})$, and $(e_t - e_{t-1})^2$ for this example. Note that the first predicted value of Y is

$$\hat{Y}_1 = 22.7372 - 0.8507(8.597) = 15.4237$$

The error for year 1 is

$$\text{Actual}_1 - \text{Predicted}_1 = 17.573 - 15.4237 = 2.1493$$

The value of $e_t - e_{t-1}$ for year 1 and year 2 is computed by subtracting the error for year 1 from the error of year 2.

$$e_{\text{year2}} - e_{\text{year1}} = 1.8920 - 2.1493 = -0.2573$$

TABLE 15.13

Predicted Values and Error Terms for the Crude Oil Production and Natural Gas Withdrawal Data

Year	$\hat{Y}$	e_t	e_t^2	$e_t - e_{t-1}$	$(e_t - e_{t-1})^2$
1	15.4237	2.1493	4.6195	—	—
2	15.4450	1.8920	3.5797	−0.2573	0.0662
3	15.3795	0.4295	0.1845	−1.4625	2.1389
4	15.3463	−1.1933	1.4240	−1.6228	2.6335
5	15.1838	0.3292	0.1084	1.5225	2.3180
6	15.1056	−0.5706	0.3256	−0.8998	0.8096
7	15.3531	−1.1991	1.4378	−0.6285	0.3950
8	15.6347	−0.8277	0.6851	0.3714	0.1379
9	15.8125	−0.3455	0.1194	0.4822	0.2325
10	16.2608	−0.5518	0.3045	−0.2063	0.0426
11	16.4803	−0.4263	0.1817	0.1255	0.0158
12	16.4276	−0.4096	0.1678	0.0167	0.0003
13	16.6368	−0.4718	0.2226	−0.0622	0.0039
14	16.9125	−0.2215	0.0491	0.2503	0.0627
15	17.0698	0.2812	0.0791	0.5027	0.2527
16	17.1566	0.1254	0.0157	−0.1558	0.0243
17	17.2374	0.4996	0.2496	0.3742	0.1400
18	17.2485	0.5955	0.3546	0.0959	0.0092
19	17.4186	0.3104	0.0963	−0.2851	0.0813
20	17.7342	−0.1442	0.0208	−0.4546	0.2067
21	17.7844	−0.0584	0.0034	0.0858	0.0074
22	17.8023	0.3267	0.1067	0.3851	0.1483
23	17.8491	−0.0541	0.0029	−0.3808	0.1450
24	17.9044	−0.0854	0.0073	−0.0313	0.0010
25	18.1179	−0.3789	0.1436	−.02935	0.0861
			$\Sigma e_t^2 = 14.4897$		$\Sigma(e_t - e_{t-1})^2 = 9.9589$

The Durbin-Watson statistic can now be computed:

$$D = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} = \frac{9.9589}{14.4897} = .6873$$

Because we used a simple linear regression, the value of k is 1. The sample size, n, is 25, and $\alpha = .05$. The critical values in Table A.9 are

$$d_U = 1.45 \text{ and } d_L = 1.29$$

Because the computed D statistic, .6873, is less than the value of $d_L = 1.29$, the null hypothesis is rejected. A positive autocorrelation is present in this example.

Ways to Overcome the Autocorrelation Problem

Several approaches to data analysis can be used when autocorrelation is present. One uses additional independent variables and another transforms the independent variable.

Addition of Independent Variables

Often the reason autocorrelation occurs in regression analyses is that one or more important predictor variables have been left out of the analysis. For example, suppose a researcher develops a regression forecasting model that attempts to predict sales of new homes by sales of used homes over some period of time. Such a model might contain significant autocorrelation. The exclusion of the variable "prime mortgage interest rate" might be a factor driving the autocorrelation between the other two variables. Adding this variable to the regression model might significantly reduce the autocorrelation.

Transforming Variables

When the inclusion of additional variables is not helpful in reducing autocorrelation to an acceptable level, transforming the data in the variables may help to solve the problem. One such method is the **first-differences approach**. With the first-differences approach, *each value of X is subtracted from each succeeding time period value of X*; these "differences" become the new and transformed X variable. The same process is used to transform the Y variable. The regression analysis is then computed on the transformed X and transformed Y variables to compute a new model that is hopefully free of significant autocorrelation effects.

Another way is to generate new variables by using the percentage changes from period to period and regressing these new variables. A third way is to use autoregression models.

Autoregression

A forecasting technique that takes advantage of the relationship of values (Y_t) to previous-period values ($Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots$) is called autoregression. **Autoregression** is *a multiple regression technique in which the independent variables are time-lagged versions of the dependent variable*, which means we try to predict a value of Y from values of Y from previous time periods. The independent variable can be lagged for one, two, three, or more time periods. An autoregressive model containing independent variables for three time periods looks like this:

$$\hat{Y} = b_0 + b_1 Y_{t-1} + b_2 Y_{t-2} + b_3 Y_{t-3}$$

TABLE 15.14
Time-Lagged Natural Gas Data

Year	Nat. Gas Withdrawal Y_t	One Period Lagged $Y_{t-1}(X_1)$	Two Period Lagged $Y_{t-2}\ (X_2)$
1	17.573	—	—
2	17.337	17.573	—
3	15.809	17.337	17.573
4	14.153	15.809	17.337
5	15.513	14.153	15.809
6	14.535	15.513	14.153
7	14.154	14.535	15.513
8	14.807	14.154	14.535
9	15.467	14.807	14.154
10	15.709	15.467	14.807
11	16.054	15.709	15.467
12	16.018	16.054	15.709
13	16.165	16.018	16.054
14	16.691	16.165	16.018
15	17.351	16.691	16.165
16	17.282	17.351	16.691
17	17.737	17.282	17.351
18	17.844	17.737	17.282
19	17.729	17.844	17.737
20	17.590	17.729	17.844
21	17.726	17.590	17.729
22	18.129	17.726	17.590
23	17.795	18.129	17.726
24	17.819	17.795	18.129
25	17.739	17.819	17.795

FIGURE 15.12

Excel Autoregression Results for Natural Gas Withdrawal Data

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.864
R Square	0.746
Adjusted R Square	0.721
Standard Error	0.693
Observations	23

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	28.3203	14.1602	29.44	0.0000011
Residual	20	9.6187	0.4809		
Total	22	37.9390			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2.4081	1.9608	1.23	0.233658
Lagged 1	0.9678	0.2221	4.36	0.000306
Lagged 2	–0.1128	0.2239	–0.50	0.620075

As an example, we shall attempt to predict the volume of natural gas withdrawal, displayed in Table 15.12, by using data lagged for both one and two time periods. The data used in this analysis are displayed in Table 15.14. Using Excel, a multiple regression model is developed to predict the values of Y_t by the values of Y_{t-1} and Y_{t-2}. The results appear in Figure 15.12. Note that the regression analysis does not use data from years 1 and 2 of Table 15.14 because there are no values for the two lagged variables for one or both of those years.

The autoregression model is

$$Y_t = 2.4081 + 0.9678Y_{t-1} - 0.1128Y_{t-2}$$

The relatively high value of R^2 (74.6%) and relatively small value of s_e (0.693) indicate that this regression model has fairly strong predictability. Interestingly, the one-period lagged variable is quite significant ($t = 4.36$ with a p-value of .000306), but the two-period lagged variable is not significant ($t = -0.50$ with a p-value of 0.62), indicating the presence of first-order autocorrelation.

Autoregression can be a useful tool in locating seasonal or cyclical effects in time series data. For example, if the data are given in monthly increments, autoregression using variables lagged by as much as 12 months can search for the predictability of previous monthly time periods. If data are given in quarterly time periods, autoregression of up to four periods removed can be a useful tool in locating the predictability of data from previous quarters. When the time periods are in years, lagging the data by yearly periods and using autoregression can help in locating cyclical predictability.

15.5 PROBLEMS

15.15 The U.S. Department of Labor publishes consumer price indexes (CPIs) on many commodities. Following are the percentage changes in the CPIs for food and for shelter for the years 1980 through 2008. Use these data to develop a linear regression model to forecast the percentage change in food CPIs by the percentage change in housing CPIs. Compute a Durbin-Watson statistic to determine whether significant autocorrelation is present in the model. Let $\alpha = .05$.

Year	Food	Housing	Year	Food	Housing
1980	8.5	15.7	1995	2.8	2.6
1981	7.8	11.5	1996	3.2	2.9
1982	4.1	7.2	1997	2.6	2.6
1983	2.3	2.7	1998	2.2	2.3
1984	3.7	4.1	1999	2.2	2.2
1985	2.3	4.0	2000	2.3	3.5
1986	3.3	3.0	2001	2.8	4.2
1987	4.0	3.0	2002	1.5	3.1
1988	4.1	3.8	2003	3.6	2.2
1989	5.7	3.8	2004	2.7	3.0
1990	5.8	4.5	2005	2.3	4.0
1991	3.6	4.0	2006	3.3	2.1
1992	1.4	2.9	2007	3.0	4.9
1993	2.1	2.7	2008	2.4	5.9
1994	2.3	2.5			

15.16 Use the data from Problem 15.15 to create a regression forecasting model using the first-differences data transformation. How do the results from this model differ from those obtained in Problem 15.15?

15.17 The Federal Deposit Insurance Corporation (FDIC) releases data on bank failures. Following are data on the number of U.S. bank failures in a given year and the total amount of bank deposits (in $ millions) involved in such failures for a given year. Use these data to develop a simple regression forecasting model that attempts to predict the failed bank assets involved in bank closings by the number of bank failures. Compute a Durbin-Watson statistic for this regression model and determine whether significant autocorrelation is present. Let $\alpha = .05$.

Year	Failures	Failed Bank Assets
1	11	8,189
2	7	104
3	34	1,862
4	45	4,137
5	79	36,394
6	118	3,034
7	144	7,609
8	201	7,538
9	221	56,620
10	206	28,507
11	159	10,739
12	108	43,552
13	100	16,915
14	42	2,588
15	11	825
16	6	753
17	5	186
18	1	27

15.18 Use the data in Problem 15.17 to compute a regression model after recoding the data by the first-differences approach. Compute a Durbin-Watson statistic to determine whether significant autocorrelation is present in this first-differences model. Compare this model with the model determined in Problem 15.17, and compare the significance of the Durbin-Watson statistics for the two problems. Let $\alpha = .05$.

15.19 *Current Construction Reports* from the U.S. Census Bureau contain data on new privately owned housing units. Data on new privately owned housing units (1000s) built in the West between 1980 and 2006 follow. Use these time-series data to develop an autoregression model with a one-period lag. Now try an autoregression model with a two-period lag. Discuss the results and compare the two models.

Year	Housing Starts (1000)	Year	Housing Starts (1000)
1980	333.0	1994	453.0
1981	270.4	1995	430.3
1982	281.1	1996	468.5
1983	443.0	1997	464.2
1984	432.3	1998	521.9
1985	428.9	1999	550.4
1986	443.2	2000	529.7
1987	413.1	2001	556.9
1988	391.6	2002	606.5
1989	361.5	2003	670.1
1990	318.1	2004	745.5
1991	308.4	2005	756.1
1992	382.2	2006	826.8
1993	419.5		

15.20 The U.S. Department of Agriculture publishes data on the production, utilization, and value of fruits in the United States. Shown here are the amounts of noncitrus fruit processed into juice (in kilotons) for a 25-year period. Use these data to develop an autoregression forecasting model with a two-period lag. Discuss the results of this analysis.

Year	Processed Juice	Year	Processed Juice
1	598	14	1135
2	768	15	1893
3	863	16	1372
4	818	17	1547
5	841	18	1450
6	1140	19	1557
7	1285	20	1742
8	1418	21	1736
9	1235	22	1886
10	1255	23	1857
11	1445	24	1582
12	1336	25	1675
13	1226		

15.6 INDEX NUMBERS

One particular type of descriptive measure that is useful in allowing comparisons of data over time is the index number. An index number is, in part, a ratio of a measure taken during one time frame to that same measure taken during another time frame, usually denoted as the base period. Often the ratio is multiplied by 100 and is expressed as a percentage. When expressed as a percentage, index numbers serve as an alternative to comparing raw numbers. Index number users become accustomed to interpreting measures for a given time period in light of a base period on a scale in which the base period has an index of 100(%). Index numbers are used to compare phenomena from one time period to another and are especially helpful in highlighting interperiod differences.

Index numbers are widely used around the world to relate information about stock markets, inflation, sales, exports and imports, agriculture, and many other things. Some examples of specific indexes are the employment cost index, price index for construction, index of manufacturing capacity, producer price index, consumer price index, Dow Jones industrial average, index of output, and Nikkei 225 average. This section, although recognizing the importance of stock indexes and others, will focus on price indexes.

The motivation for using an index number is to reduce data to an easier-to-use, more convenient form. As an example, examine the raw data on number of business bankruptcies in the United States from 1987 through 2008 shown in Table 15.15. An analyst can describe these data by observing that, in general, the number of business bankruptcies has been decreasing since 1987. How do the number of business bankruptcies in 1997 compare to 1987? How do the number of business bankruptcies in 2000 compare to 1990 or 1992? To answer these questions without index numbers, a business researcher would probably resort to subtracting the number of business bankruptcies for the years of interest and comparing the corresponding increases or decreases. This process can be tedious and frustrating for decision makers who must maximize their effort in minimal time. Using simple index numbers, the business researcher can transform these data into values that are more usable and make it easier to compare other years to one particular key year.

Simple Index Numbers

How are index numbers computed? The equation for computing a **simple index number** follows.

SIMPLE INDEX NUMBER

$$I_i = \frac{X_i}{X_0}(100)$$

where

X_0 = the quantity, price, or cost in the base year
X_i = the quantity, price, or cost in the year of interest
I_i = the index number for the year of interest

TABLE 15.15

Business Bankruptcies in the United States

Year	Business Bankruptcies
1987	81,463
1988	62,845
1989	62,449
1990	63,912
1991	70,605
1992	69,848
1993	62,399
1994	50,845
1995	50,516
1996	53,200
1997	53,819
1998	44,197
1999	37,639
2000	35,219
2001	39,719
2002	38,155
2003	35,037
2004	34,317
2005	39,201
2006	19,695
2007	28,322
2008	43,546

Suppose bankruptcy researchers examining the data from Table 15.15 decide to compute index numbers using 1987 as the base year. The index number for the year 2000 is

$$I_{2000} = \frac{X_{2000}}{X_{1987}}(100) = \frac{35{,}219}{81{,}463}(100) = 43.2$$

Table 15.16 displays all the index numbers for the data in Table 15.15, with 1987 as the base year, along with the raw data. A cursory glance at these index numbers reveals a decrease in the number of bankruptcies for most of the years since 1987 (because the index has been going down). In particular, the greatest drop in number seems to have occurred between 2005 and 2006—a drop of nearly 24 in the index. Because most people are easily able to understand the concept of 100%, it is likely that decision makers can make quick judgments on the number of business bankruptcies in the United states from one year relative to another by examining the index numbers over this period.

Unweighted Aggregate Price Index Numbers

The use of simple index numbers makes possible the conversion of prices, costs, quantities, and so on for different time periods into a number scale with the base year equaling 100%. One of the drawbacks of simple index numbers, however, is that each time period is represented by only one item or commodity. When multiple items are involved, multiple sets of index numbers are possible. Suppose a decision maker is interested in combining or pooling the prices of several items, creating a "market basket" in order to compare the prices for several years. Fortunately, a technique does exist for combining several items and determining index numbers for the total (aggregate). Because this technique is used mostly in determining price indexes, the focus in this section is on developing aggregate price indexes. The formula for constructing the **unweighted aggregate price index number** follows.

UNWEIGHTED AGGREGATE PRICE INDEX NUMBER

$$I_i = \frac{\Sigma P_i}{\Sigma P_0}(100)$$

where

P_i = the price of an item in the year of interest (i)
P_0 = the price of an item in the base year (0)
I_i = the index number for the year of interest (i)

TABLE 15.16

Index Numbers for Business Bankruptcies in the United States

Year	Business Bankruptcies	Index Number
1987	81,463	100.0
1988	62,845	77.1
1989	62,449	76.7
1990	63,912	78.5
1991	70,605	86.7
1992	69,848	85.7
1993	62,399	76.6
1994	50,845	62.4
1995	50,516	62.0
1996	53,200	65.3
1997	53,819	66.1
1998	44,197	54.3
1999	37,639	46.2
2000	35,219	43.2
2001	39,719	48.8
2002	38,155	46.8
2003	35,037	43.0
2004	34,317	42.1
2005	39,201	48.1
2006	19,695	24.2
2007	28,322	34.8
2008	43,546	53.5

Suppose a state's department of labor wants to compare the cost of family food buying over the years. Department officials decide that instead of using a single food item to do this comparison, they will use a food basket that consists of five items: eggs, milk, bananas, potatoes, and sugar. They gathered price information on these five items for the years 1995, 2000, and 2008. The items and the prices are listed in Table 15.17.

From the data in Table 15.17 and the formula, the unweighted aggregate price indexes for the years 1995, 2000, and 2008 can be computed by using 1995 as the base year. The first step is to add together, or aggregate, the prices for all the food basket items in a given year. These totals are shown in the last row of Table 15.17. The index numbers are constructed by using these totals (not individual item prices): $\Sigma P_{1995} = 2.91$, $\Sigma P_{2000} = 3.44$, and $\Sigma P_{2008} = 3.93$. From these figures, the unweighted aggregate price index for 2000 is computed as follows.

$$\text{For 2000: } I_{2000} = \frac{\Sigma P_{2000}}{\Sigma P_{1995}}(100) = \frac{3.44}{2.91}(100) = 118.2$$

Weighted Aggregate Price Index Numbers

A major drawback to unweighted aggregate price indexes is that they are *unweighted*—that is, equal weight is put on each item by assuming the market basket contains only one of each item. This assumption may or may not be true. For example, a household may consume 5 pounds of bananas per year but drink 50 gallons of milk. In addition, unweighted aggregate index numbers are dependent on the units selected for various items. For example, if milk is measured in quarts instead of gallons, the price of milk used in determining the index numbers is considerably lower. A class of index numbers that can be used to avoid these problems is weighted aggregate price index numbers.

Weighted aggregate price index numbers are *computed by multiplying quantity weights and item prices in determining the market basket worth for a given year.* Sometimes when price and quantity are multiplied to construct index numbers, the index numbers are referred to as *value indexes.* Thus, weighted aggregate price index numbers are also value indexes.

Including quantities eliminates the problems caused by how many of each item are consumed per time period and the units of items. If 50 gallons of milk but only 5 pounds of bananas are consumed, weighted aggregate price index numbers will reflect those weights. If the business researcher switches from gallons of milk to quarts, the prices will change downward but the quantity will increase fourfold (4 quarts in a gallon).

In general, weighted aggregate price indexes are constructed by multiplying the price of each item by its quantity and then summing these products for the market basket over a given time period (often a year). The ratio of this sum for one time period of interest

TABLE 15.17

Prices for a Basket of Food Items

Item	Year 1995	2000	2008
Eggs (dozen)	.78	.86	1.06
Milk (1/2 gallon)	1.14	1.39	1.59
Bananas (per lb.)	.36	.46	.49
Potatoes (per lb.)	.28	.31	.36
Sugar (per lb.)	.35	.42	.43
Total of Items	2.91	3.44	3.93

(year) to a base time period of interest (base year) is multiplied by 100. The following formula reflects a weighted aggregate price index computed by using quantity weights from each time period (year).

$$I_i = \frac{\Sigma P_i Q_i}{\Sigma P_0 Q_0}(100)$$

One of the problems with this formula is the implication that new and possibly different quantities apply for each time period. However, business researchers expend much time and money ascertaining the quantities used in a market basket. Redetermining quantity weights for each year is therefore often prohibitive for most organizations (even the government). Two particular types of weighted aggregate price indexes offer a solution to the problem of which quantity weights to use. The first and most widely used is the Laspeyres price index. The second and less widely used is the Paasche price index.

Laspeyres Price Index

The **Laspeyres price index** is *a weighted aggregate price index computed by using the quantities of the base period (year) for all other years.* The advantages of this technique are that the price indexes for all years can be compared, and new quantities do not have to be determined for each year. The formula for constructing the Laspeyres price index follows.

LASPEYRES PRICE INDEX

$$I_L = \frac{\Sigma P_i Q_0}{\Sigma P_0 Q_0}(100)$$

Notice that the formula requires the base period quantities (Q_0) in both the numerator and the denominator.

In Table 15.17, a food basket is presented in which aggregate price indexes are computed. This food basket consisted of eggs, milk, bananas, potatoes, and sugar. The prices of these items were combined (aggregated) for a given year and the price indexes were computed from these aggregate figures. The unweighted aggregate price indexes computed on these data gave all items equal importance. Suppose that the business researchers realize that applying equal weight to these five items is probably not a representative way to construct this food basket and consequently ascertain quantity weights on each food item for one year's consumption. Table 15.18 lists these five items, their prices, and their quantity usage weights for the base year (1995). From these data, the business researchers can compute Laspeyres price indexes.

The Laspeyres price index for 2008 with 1995 as the base year is:

$$\begin{aligned}
\Sigma P_i Q_0 &= \Sigma P_{2008} Q_{1995} \\
&= \Sigma[(1.06)(45) + (1.59)(60) + (.49)(12) + (.36)(55) + (.43)(36)] \\
&= 47.70 + 95.40 + 5.88 + 19.80 + 15.48 = 184.26 \\
\Sigma P_0 Q_0 &= \Sigma P_{1995} Q_{1995} \\
&= \Sigma[(.78)(45) + (1.14)(60) + (.36)(12) + (.28)(55) + (.35)(36)] \\
&= 35.10 + 68.40 + 4.32 + 15.40 + 12.60 = 135.82 \\
I_{2008} &= \frac{\Sigma P_{2008} Q_{1995}}{\Sigma P_{1995} Q_{1995}}(100) = \frac{184.26}{135.85}(100) = 135.7
\end{aligned}$$

TABLE 15.18
Food Basket Items with Quantity Weights

Item	Quantity	Price 1995	Price 2008
Eggs (dozen)	45	.78	1.06
Milk (1/2 gal.)	60	1.14	1.59
Bananas (per lb.)	12	.36	.49
Potatoes (per lb.)	55	.28	.36
Sugar (per lb.)	36	.35	.43

TABLE 15.19
Food Basket Items with Yearly Quantity Weights for 1995 and 2008

Item	P_{1995}	Q_{1995}	P_{2008}	Q_{2008}
Eggs (dozen)	.78	45	1.06	42
Milk (1/2 gal.)	1.14	60	1.59	57
Bananas (per lb.)	.36	12	.49	13
Potatoes (per lb.)	.28	55	.36	52
Sugar (per lb.)	.35	36	.43	36

Paasche Price Index

The **Paasche price index** is *a weighted aggregate price index computed by using the quantities for the year of interest in computations for a given year.* The advantage of this technique is that it incorporates current quantity figures in the calculations. One disadvantage is that ascertaining quantity figures for each time period is expensive. The formula for computing Paasche price indexes follows.

PAASCHE PRICE INDEX

$$I_P = \frac{\Sigma P_i Q_i}{\Sigma P_0 Q_i}(100)$$

Suppose the yearly quantities for the basket of food items listed in Table 15.18 are determined. The result is the quantities and prices shown in Table 15.19 for the years 1995 and 2008 that can be used to compute Paasche price index numbers.

The Paasche price index numbers can be determined for 2008 by using a base year of 1995 as follows.

For 2008:

$$\begin{aligned}\Sigma P_{2008}Q_{2008} &= \Sigma[(1.06)(42) + (1.59)(57) + (.49)(13) + (.36)(52) + (.43)(36)]\\ &= 44.52 + 90.63 + 6.37 + 18.72 + 15.48 = 175.72\\ \Sigma P_{1995}Q_{2008} &= [(.78)(42) + (1.14)(57) + (.36)(13) + (.28)(52) + (.35)(36)]\\ &= 32.76 + 64.98 + 4.68 + 14.56 + 12.60\\ &= 129.58\end{aligned}$$

$$I_{2008} = \frac{\Sigma P_{2008}Q_{2008}}{\Sigma P_{1995}Q_{2008}}(100) = \frac{175.72}{129.58}(100) = 135.6$$

DEMONSTRATION PROBLEM 15.5

The Arapaho Valley Pediatrics Clinic has been in business for 18 years. The office manager noticed that prices of clinic materials and office supplies fluctuate over time. To get a handle on the price trends for running the clinic, the office manager examined prices of six items the clinic uses as part of its operation. Shown here are the items, their prices, and the quantities for the years 2008 and 2009. Use these data to develop unweighted aggregate price indexes for 2009 with a base year of 2008. Compute the Laspeyres price index for the year 2009 using 2008 as the base year. Compute the Paasche index number for 2009 using 2008 as the base year.

	2008		2009	
Item	**Price**	**Quantity**	**Price**	**Quantity**
Syringes (dozen)	6.70	150	6.95	135
Cotton swabs (box)	1.35	60	1.45	65
Patient record forms (pad)	5.10	8	6.25	12
Children's Tylenol (bottle)	4.50	25	4.95	30
Computer paper (box)	11.95	6	13.20	8
Thermometers	7.90	4	9.00	2
Totals	37.50		41.80	

Solution

Unweighted Aggregate Index for 2009:

$$I_{2009} = \frac{\Sigma P_{2009}}{\Sigma P_{2008}}(100) = \frac{41.80}{37.50}(100) = 111.5$$

Laspeyres Index for 2009:

$$\begin{aligned}\Sigma P_{2009}Q_{2008} &= [(6.95)(150) + (1.45)(60) + (6.25)(8) + (4.95)(25) + (13.20)(6) \\ &\quad + (9.00)(4)] \\ &= 1{,}042.50 + 87.00 + 50.00 + 123.75 + 79.20 + 36.00 \\ &= 1{,}418.45\end{aligned}$$

$$\begin{aligned}\Sigma P_{2008}Q_{2008} &= [(6.70)(150) + (1.35)(60) + (5.10)(8) + (4.50)(25) + (11.95)(6) \\ &\quad + (7.90)(4)] \\ &= 1{,}005.00 + 81.00 + 40.80 + 112.50 + 71.70 + 31.60 \\ &= 1{,}342.6\end{aligned}$$

$$I_{2009} = \frac{\Sigma P_{2009}Q_{2008}}{\Sigma P_{2008}Q_{2008}}(100) = \frac{1{,}418.45}{1{,}342.6}(100) = 105.6$$

Passache Index for 2009:

$$\begin{aligned}\Sigma P_{2009}Q_{2009} &= [(6.95)(135) + (1.45)(65) + (6.25)(12) + (4.95)(30) + (13.20)(8) \\ &\quad + (9.00)(2)] \\ &= 938.25 + 94.25 + 75.00 + 148.50 + 105.60 + 18.00 \\ &= 1{,}379.60\end{aligned}$$

$$\begin{aligned}\Sigma P_{2008}Q_{2009} &= [(6.70)(135) + (1.35)(65) + (5.10)(12) + (4.50)(30) + (11.95)(8) \\ &\quad + (7.90)(2)] \\ &= 904.50 + 87.75 + 61.20 + 135.00 + 95.60 + 15.80 \\ &= 1{,}299.85\end{aligned}$$

$$I_{2009} = \frac{\Sigma P_{2009}Q_{2009}}{\Sigma P_{2008}Q_{2009}}(100) = \frac{1{,}379.60}{1{,}299.85}(100) = 106.1$$

15.6 PROBLEMS

15.21 Suppose the following data represent the price of 20 reams of office paper over a 50-year time frame. Find the simple index numbers for the data.

a. Let 1950 be the base year.

b. Let 1980 be the base year.

Year	Price	Year	Price
1950	$ 22.45	1980	$ 69.75
1955	31.40	1985	73.44
1960	32.33	1990	80.05
1965	36.50	1995	84.61
1970	44.90	2000	87.28
1975	61.24	2005	89.56

15.22 The U.S. Patent and Trademark Office reports fiscal year figures for patents issued in the United States. Following are the numbers of patents issued for the years 1980 through 2007. Using these data and a base year of 1990, determine the simple index numbers for each year.

Year	Number of Patents (1000s)	Year	Number of Patents (1000s)
1980	66.2	1994	113.6
1981	71.1	1995	113.8
1982	63.3	1996	121.7
1983	62.0	1997	124.1
1984	72.7	1998	163.1
1985	77.2	1999	169.1
1986	76.9	2000	176.0
1987	89.4	2001	184.0
1988	84.3	2002	184.4
1989	102.5	2003	187.0
1990	99.1	2004	181.3
1991	106.7	2005	157.7
1992	107.4	2006	196.4
1993	109.7	2007	182.9

15.23 Using the data that follow, compute the aggregate index numbers for the four types of meat. Let 1995 be the base year for this market basket of goods.

		Year	
Items	**1995**	**2002**	**2009**
Ground beef (per lb.)	1.53	1.40	2.17
Sausage (per lb.)	2.21	2.15	2.51
Bacon (per lb.)	1.92	2.68	2.60
Round steak (per lb.)	3.38	3.10	4.00

15.24 Suppose the following data are prices of market goods involved in household transportation for the years 2001 through 2009. Using 2003 as a base year, compute aggregate transportation price indexes for this data.

					Year				
Items	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Gasoline (per gal.)	1.10	1.16	1.23	1.23	1.08	1.56	1.85	2.59	2.89
Oil (per qt.)	1.58	1.61	1.78	1.77	1.61	1.71	1.90	2.05	2.08
Transmission fluid (per qt.)	1.80	1.82	1.98	1.96	1.94	1.90	1.92	1.94	1.96
Radiator coolant (per gal.)	7.95	7.96	8.24	8.21	8.19	8.05	8.12	8.10	8.24

15.25 Calculate Laspeyres price indexes for 2007–2009 from the following data. Use 2000 as the base year.

	Quantity	Price			
Item	**2000**	**2000**	**2007**	**2008**	**2009**
1	21	$0.50	$0.67	$0.68	$0.71
2	6	1.23	1.85	1.90	1.91
3	17	0.84	.75	.75	.80
4	43	0.15	.21	.25	.25

15.26 Calculate Paasche price indexes for 2008 and 2009 using the following data and 2000 as the base year.

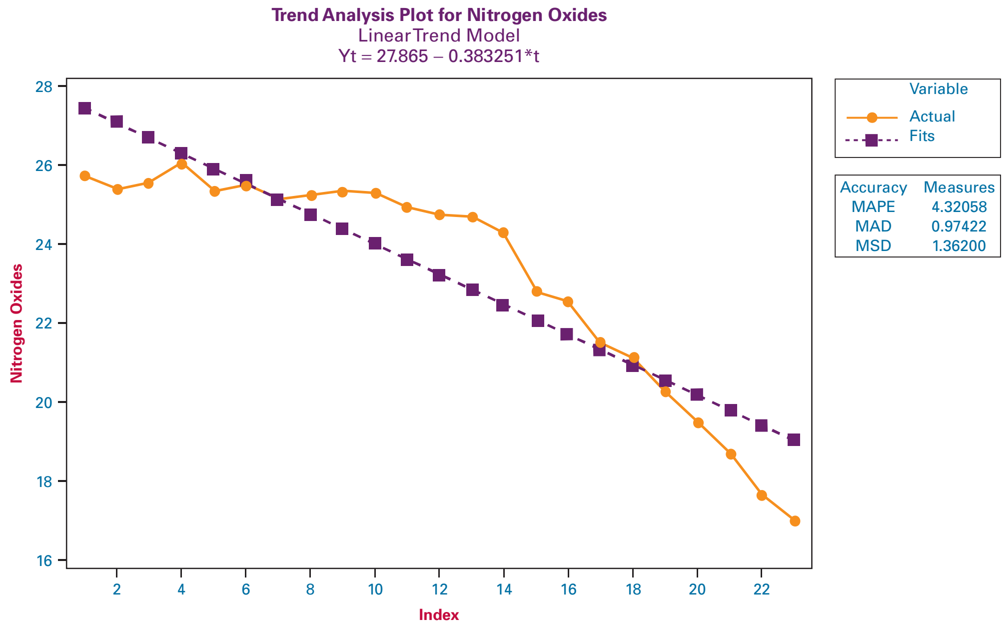
		2008		2009	
Item	**2000 Price**	**Price**	**Quantity**	**Price**	**Quantity**
1	$22.50	$27.80	13	$28.11	12
2	10.90	13.10	5	13.25	8
3	1.85	2.25	41	2.35	44

DECISION DILEMMA SOLVED

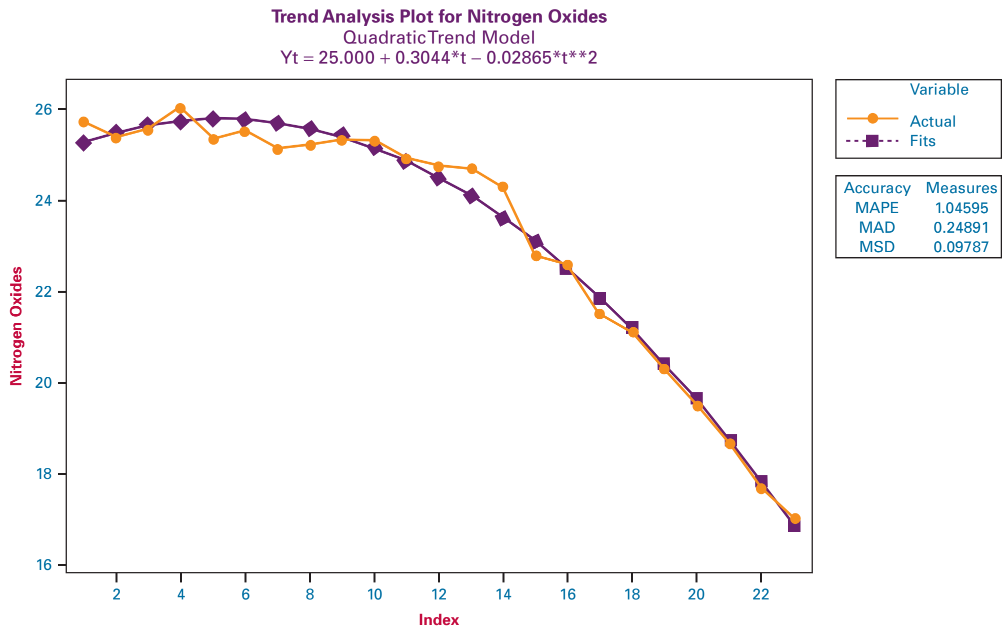
Forecasting Air Pollution

In searching for the most effective forecasting technique to use to forecast either the carbon monoxide emission or the nitrogen oxide, it is useful to determine whether a trend is evident in either set of time-series data. Minitab's trend analysis output is presented here for nitrogen oxides.

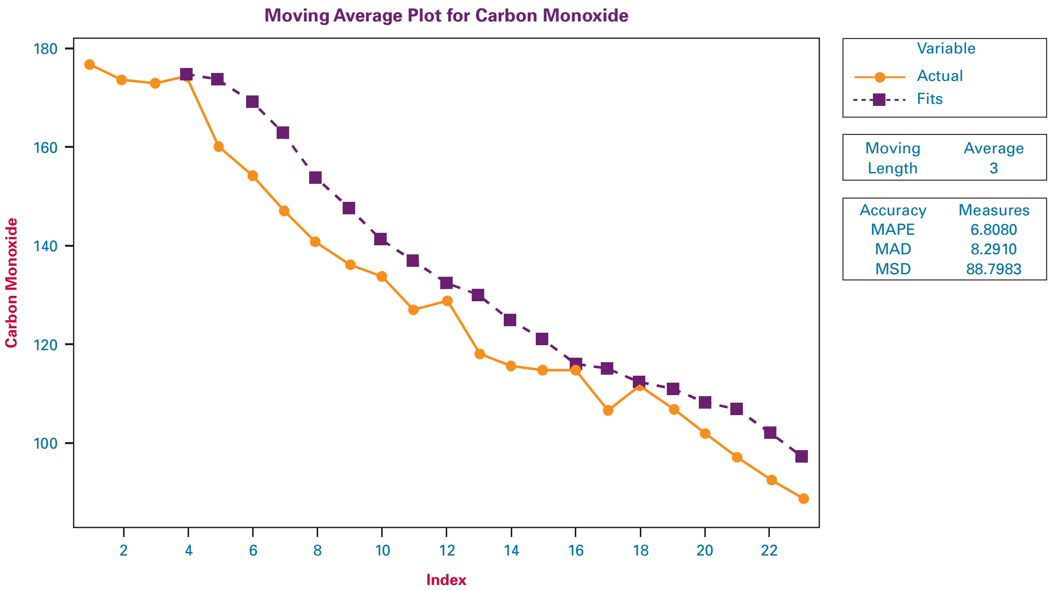
Decision Dilemma Solved



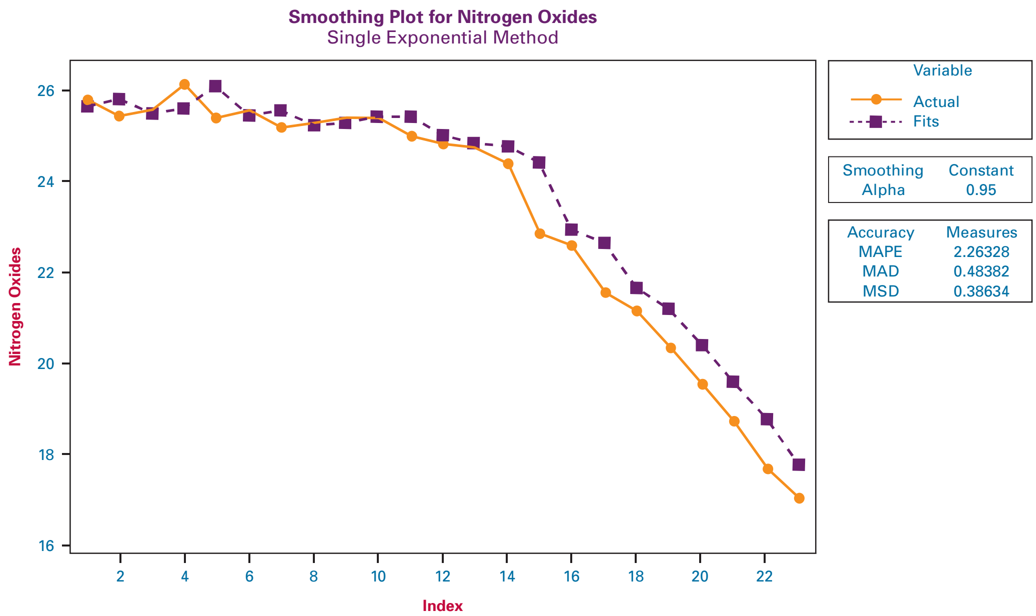
In observing the fit of this trend line and the time-series plot, it is evident that there appears to be more of a quadratic trend than a linear trend. Therefore, a Minitab-produced quadratic trend model was run and the results are presented below. Note that the error measures are all smaller for the quadratic model and that the curve fits the data much better than does the linear model.



Various smoothing techniques can be used to forecast time-series data. After exploring several moving average models to predict carbon monoxide emissions, it was determined that a 3-year moving average fits the data relatively well. The results of a Minitab moving average graphical analysis of carbon monoxide using a 3-year moving average is shown below. Note that the forecasts shadow the actual values quite well and actually intersect them in two locations.



The effectiveness of exponential smoothing as a forecasting tool for nitrogen oxide emissions was tested using Minitab for several values of α. Through this analysis, it was determined that the best forecasts were obtained for values of α near 1, indicating that the actual value for the previous time period was a much stronger contributor to the forecast than the previous time period's forecast. Shown below is a Minitab-produced graphical analysis of an exponential smoothing forecast of the nitrogen oxide data using an alpha of .95. You are encouraged to explore other methods for forecasting nitrogen oxide and carbon monoxide emissions.



ETHICAL CONSIDERATIONS

The true test of a forecast is the accuracy of the prediction. Until the actual value is obtained for a given time period, the accuracy of the forecast is unknown. Many forecasters make predictions in society, including card readers, religious leaders, and self-proclaimed prophets. The proof of the forecast is in the outcome. The same holds true in the business world. Forecasts are made about everything from market share to interest rates to number of international air travelers. Many businesses fail because of faulty forecasts.

Forecasting is perhaps as much an art as a science. To keep forecasting ethical, the consumer of the forecast should be given the caveats and limitations of the forecast. The forecaster should be honestly cautious in selling the predictions to a client. In addition, the forecaster should be constantly on the lookout for changes in the business setting being modeled and quickly translate and incorporate those changes into the forecasting model.

Unethical behavior can occur in forecasting when particular data are selected to develop a model that has been predetermined to produce certain results. As mentioned previously, statistics can be used to "prove" almost anything. The ethical forecaster lets the data drive the model and is constantly seeking honest input from new variables to revise the forecast. He or she strives to communicate the limitations of both the forecasts and the models to clients.

SUMMARY

Time-series data are data that have been gathered at regular intervals over a period of time. It is generally believed that time-series data are composed of four elements—trend, cyclical effects, seasonality, and irregularity. Trend is the long-term general direction of the time-series data. Cyclical effects are the business and economic cycles that occur over periods of more than 1 year. Seasonal effects are patterns or cycles of data behavior that occur over time periods of less than 1 year. Irregular fluctuations are unaccounted-for "blips" or variations that occur over short periods of time.

One way to establish the validity of a forecast is to examine the forecasting error. The error of a forecast is the difference between the actual value and the forecast value. Computing a value to measure forecasting error can be done in several different ways. This chapter presents mean absolute deviation and mean square error for this task.

Regression analysis with either linear or quadratic models can be used to explore trend. Regression trend analysis is a special case of regression analysis in which the dependent variable is the data to be forecast and the independent variable is the time periods numbered consecutively from 1 to k, where k is the number of time periods. For the quadratic model, a second independent variable is constructed by squaring the values in the first independent variable, and both independent variables are included in the analysis.

One group of time-series forecasting methods contains smoothing techniques. Among these techniques are naïve models, averaging techniques, and simple exponential smoothing. These techniques do much better if the time series data are stationary or show no significant trend or seasonal effects. Naïve forecasting models are models in which it is assumed that the more recent time periods of data represent the best predictions or forecasts for future outcomes.

Simple averages use the average value for some given length of previous time periods to forecast the value for the next period. Moving averages are time period averages that are revised for each time period by including the most recent value(s) in the computation of the average and deleting the value or values that are farthest away from the present time period. A special case of the moving average is the weighted moving average, in which different weights are placed on the values from different time periods.

Simple (single) exponential smoothing is a technique in which data from previous time periods are weighted exponentially to forecast the value for the present time period. The forecaster has the option of selecting how much to weight more recent values versus those of previous time periods.

Decomposition is a method for isolating the four possible effects in time-series data, trend, cyclical effects, seasonality, and irregular fluctuations.

Autocorrelation or serial correlation occurs when the error terms from forecasts are correlated over time. In regression analysis, this effect is particularly disturbing because one of the assumptions is that the error terms are independent. One way to test for autocorrelation is to use the Durbin-Watson test.

A number of methods attempt to overcome the effects of autocorrelation on the data. One way is to determine whether at least one independent variable is missing and, if so, include it or them in the model. Another way is to transform the variables. One transformation technique is the first-differences approach, in which each value of X is subtracted from the succeeding time period value of X and the differences are used as the values of the X variable. The same approach is used to transform the Y variable. The forecasting model is then developed from the transformed variables.

Autoregression is a forecasting technique in which time-series data are predicted by independent variables that are lagged versions of the original dependent variable data. A variable that is lagged one period is derived from values of the previous time period. Other variables can be lagged two or more periods.

Index numbers can be used to translate raw data into numbers that are more readily comparable. Simple index numbers are constructed by creating the ratio of the raw data value for a given time period to the raw data value for the base period and multiplying the ratio by 100. The index number for the base time period is designated to be 100.

Unweighted aggregate price index numbers are constructed by summing the prices of several items for a time period and comparing that sum to the sum of the prices of the same items during a base time period and multiplying the ratio by 100. Weighted aggregate price indexes are index numbers utilizing the prices of several items, and the items are weighted by their quantity usage.

The Laspeyres price index uses the quantity weights from the base year in all calculations. The Paasche price index uses the quantity weights for the current time period for both the current time period and the base time period in calculations.

KEY TERMS

Flash Cards

autocorrelation
autoregression
averaging models
cycles
cyclical effects
decomposition
deseasonalized data
Durbin-Watson test
error of an individual forecast
exponential smoothing
first-differences approach
forecasting
forecasting error
index number
irregular fluctuations
Laspeyres price index
mean absolute deviation (MAD)
mean square error (MSE)
moving average
naïve forecasting models
Paasche price index
seasonal effects
serial correlation
simple average
simple average model
simple index number
smoothing techniques
stationary
time-series data
trend
unweighted aggregate price index number
weighted aggregate price index numbers
weighted moving average

FORMULAS

Individual forecast error

$$e_t = X_t - F_t$$

Mean absolute deviation

$$\text{MAD} = \frac{\Sigma |e_i|}{\text{Number of Forecasts}}$$

Mean square error

$$\text{MSE} = \frac{\Sigma e_i^2}{\text{Number of Forecasts}}$$

Exponential smoothing

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

Durbin-Watson test

$$D = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

SUPPLEMENTARY PROBLEMS

CALCULATING THE STATISTICS

15.27 Following are the average yields of long-term new corporate bonds over a several-month period published by the Office of Market Finance of the U.S. Department of the Treasury.

Month	Yield	Month	Yield
1	10.08	7	9.37
2	10.05	8	8.55
3	9.24	9	8.36
4	9.23	10	8.59
5	9.69	11	7.99
6	9.55	12	8.12

(continued)

Month	Yield	Month	Yield
13	7.91	19	7.35
14	7.73	20	7.04
15	7.39	21	6.88
16	7.48	22	6.88
17	7.52	23	7.17
18	7.48	24	7.22

a. Explore trends in these data by using regression trend analysis. How strong are the models? Is the quadratic model significantly stronger than the linear trend model?

b. Use a 4-month moving average to forecast values for each of the ensuing months.
c. Use simple exponential smoothing to forecast values for each of the ensuing months. Let $\alpha = .3$ and then let $\alpha = .7$. Which weight produces better forecasts?
d. Compute MAD for the forecasts obtained in parts (b) and (c) and compare the results.
e. Determine seasonal effects using decomposition on these data. Let the seasonal effects have four periods. After determining the seasonal indexes, deseasonalize the data.

15.28 Compute index numbers for the following data using 1995 as the base year.

Year	Quantity	Year	Quantity
1995	2073	2003	2520
1996	2290	2004	2529
1997	2349	2005	2483
1998	2313	2006	2467
1999	2456	2007	2397
2000	2508	2008	2351
2001	2463	2009	2308
2002	2499		

15.29 Compute unweighted aggregate price index numbers for each of the given years using 2005 as the base year.

Item	2005	2006	2007	2008	2009
1	3.21	3.37	3.80	3.73	3.65
2	.51	.55	.68	.62	.59
3	.83	.90	.91	1.02	1.06
4	1.30	1.32	1.33	1.32	1.30
5	1.67	1.72	1.90	1.99	1.98
6	.62	.67	.70	.72	.71

15.30 Using the following data and 2006 as the base year, compute the Laspeyres price index for 2009 and the Paasche price index for 2008.

	2006		2007	
Item	Price	Quantity	Price	Quantity
1	$2.75	12	$2.98	9
2	0.85	47	0.89	52
3	1.33	20	1.32	28

	2008		2009	
Item	Price	Quantity	Price	Quantity
1	$3.10	9	$3.21	11
2	0.95	61	0.98	66
3	1.36	25	1.40	32

TESTING YOUR UNDERSTANDING

15.31 The following data contain the quantity (million pounds) of U.S. domestic fish caught annually over a 25-year period as published by the National Oceanic and Atmospheric Administration.
a. Use a 3-year moving average to forecast the quantity of fish for the years 1983 through 2004 for these data. Compute the error of each forecast and then determine the mean absolute deviation of error for the forecast.
b. Use exponential smoothing and $\alpha = .2$ to forecast the data from 1983 through 2004. Let the forecast for 1981 equal the actual value for 1980. Compute the error of each forecast and then determine the mean absolute deviation of error for the forecast.
c. Compare the results obtained in parts (a) and (b) using MAD. Which technique seems to perform better? Why?

Year	Quantity	Year	Quantity
1980	6,559	1993	10,209
1981	6,022	1994	10,500
1982	6,439	1995	9,913
1983	6,396	1996	9,644
1984	6,405	1997	9,952
1985	6,391	1998	9,333
1986	6,152	1999	9,409
1987	7,034	2000	9,143
1988	7,400	2001	9,512
1989	8,761	2002	9,430
1990	9,842	2003	9,513
1991	10,065	2004	10,085
1992	10,298		

15.32 The U.S. Department of Commerce publishes a series of census documents referred to as *Current Industrial Reports*. Included in these documents are the manufacturers' shipments, inventories, and orders over a 5-year period. Displayed here is a portion of these data representing the shipments of chemicals and allied products from January of year 1 through December of year 5. Use time-series decomposition methods to develop the seasonal indexes for these data.

Time Period	Chemicals and Allied Products ($ billion)	Time Period	Chemicals and Allied Products ($ billion)
January (year 1)	23.701	January (year 2)	23.347
February	24.189	February	24.122
March	24.200	March	25.282
April	24.971	April	25.426
May	24.560	May	25.185
June	24.992	June	26.486
July	22.566	July	24.088
August	24.037	August	24.672
September	25.047	September	26.072
October	24.115	October	24.328
November	23.034	November	23.826
December	22.590	December	24.373
January (year 3)	24.207	January (year 4)	25.316
February	25.772	February	26.435
March	27.591	March	29.346
April	26.958	April	28.983
May	25.920	May	28.424
June	28.460	June	30.149
July	24.821	July	26.746
August	25.560	August	28.966
September	27.218	September	30.783
October	25.650	October	28.594
November	25.589	November	28.762
December	25.370	December	29.018

Time Period	Chemicals and Allied Products ($ billion)
January (year 5)	28.931
February	30.456
March	32.372
April	30.905
May	30.743
June	32.794
July	29.342
August	30.765
September	31.637
October	30.206
November	30.842
December	31.090

15.33 Use the seasonal indexes computed to deseasonalize the data in Problem 15.32.

15.34 Determine the trend for the data in Problem 15.32 using the deseasonalized data from Problem 15.33. Explore both a linear and a quadratic model in an attempt to develop the better trend model.

15.35 Shown here are retail price figures and quantity estimates for five different food commodities over 3 years. Use these data and a base year of 2007 to compute unweighted aggregate price indexes for this market basket of food. Using a base year of 2007, calculate Laspeyres price indexes and Paasche price indexes for 2008 and 2009.

	2007		2008		2009	
Item	Price	Quantity	Price	Quantity	Price	Quantity
Margarine (lb.)	1.26	21	1.32	23	1.39	22
Shortening (lb.)	0.94	5	0.97	3	1.12	4
Milk (1/2 gal.)	1.43	70	1.56	68	1.62	65
Cola (2 liters)	1.05	12	1.02	13	1.25	11
Potato chips (12 oz.)	2.81	27	2.86	29	2.99	28

15.36 Given below are data on the number of business establishments (millions) and the self-employment rate (%) released by the Small Business Administration, Office of Advocacy, for a 21-year period of U.S. business activity. Develop a regression model to predict the self-employment rate by the number of business establishments. Use this model to predict the self-employment rate for a year in which there are 7.0 (million) business establishments. Discuss the strength of the regression model. Use these data and the regression model to compute a Durbin-Watson test to determine whether significant autocorrelation is present. Let alpha be .05.

Number of Establishments (millions)	Self-Employment Rate (%)
4.54317	8.1
4.58651	8.0
4.63396	8.1
5.30679	8.2
5.51772	8.2
5.70149	8.0
5.80697	7.9
5.93706	8.0
6.01637	8.2
6.10692	8.1
6.17556	8.0
6.20086	8.1
6.31930	7.8
6.40123	8.0
6.50907	8.1
6.61272	7.9
6.73848	7.8
6.89487	7.7
6.94182	7.5
7.00844	7.2
7.07005	6.9

15.37 Shown here are the consumer price indexes (CPIs) for housing for the years 1988 through 2005 from the Bureau of Labor Statistics Data Web site. Use the data to answer the following questions.

a. Compute the 4-year moving average to forecast the CPIs from 1992 through 2005.

b. Compute the 4-year weighted moving average to forecast the CPIs from 1992 through 2005. Weight the most recent year by 4, the next most recent year by 3, the next year by 2, and the last year of the four by 1.

c. Determine the errors for parts (a) and (b). Compute MSE for parts (a) and (b). Compare the MSE values and comment on the effectiveness of the moving average versus the weighted moving average for these data.

Year	Housing CPI	Year	Housing CPI
1988	118.5	1997	156.8
1989	123.0	1998	160.4
1990	128.5	1999	163.9
1991	133.6	2000	169.6
1992	137.5	2001	176.4
1993	141.2	2002	180.3
1994	144.8	2003	184.8
1995	148.5	2004	189.5
1996	152.8	2005	195.7

15.38 In the *Survey of Current Business,* the U.S. Department of Commerce publishes data on farm commodity prices. Given are the cotton prices from November of year 1 through February of year 4. The prices are indexes with a base of 100 from the period of 1910 through 1914. Use these data to develop autoregression models for a 1-month lag and a 4-month lag. Compare the results of these two models. Which model seems to yield better predictions? Why?

Time Period	Cotton Prices
November (year 1)	552
December	519
January (year 2)	505
February	512
March	541
April	549
May	552
June	526
July	531
August	545
September	549
October	570
November	576
December	568

(continued)

Time Period	Cotton Prices
January (year 3)	571
February	573
March	582
April	587
May	592
June	570
July	560
August	565
September	547
October	529
November	514
December	469
January (year 4)	436
February	419

15.39 The U.S. Department of Commerce publishes data on industrial machinery and equipment. Shown here are the shipments (in $ billions) of industrial machinery and equipment from the first quarter of year 1 through the fourth quarter of year 6. Use these data to determine the seasonal indexes for the data through time-series decomposition methods. Use the four-quarter centered moving average in the computations.

Time Period	Industrial Machinery and Equipment Shipments
1st quarter (year 1)	54.019
2nd quarter	56.495
3rd quarter	50.169
4th quarter	52.891
1st quarter (year 2)	51.915
2nd quarter	55.101
3rd quarter	53.419
4th quarter	57.236
1st quarter (year 3)	57.063
2nd quarter	62.488
3rd quarter	60.373
4th quarter	63.334
1st quarter (year 4)	62.723
2nd quarter	68.380
3rd quarter	63.256
4th quarter	66.446
1st quarter (year 5)	65.445
2nd quarter	68.011
3rd quarter	63.245
4th quarter	66.872
1st quarter (year 6)	59.714
2nd quarter	63.590
3rd quarter	58.088
4th quarter	61.443

15.40 Use the seasonal indexes computed to deseasonalize the data in Problem 15.39.

15.41 Use both a linear and quadratic model to explore trends in the deseasonalized data from Problem 15.40. Which model seems to produce a better fit of the data?

15.42 The Board of Governors of the Federal Reserve System publishes data on mortgage debt outstanding by type of property and holder. The following data give the amounts of residential nonfarm debt (in $ billions) held by savings institutions in the United States over a 10-year period. Use these data to develop an autoregression model with a one-period lag. Discuss the strength of the model.

Year	Debt
1	529
2	554
3	559
4	602
5	672
6	669
7	600
8	538
9	490
10	470

15.43 The data shown here, from the Investment Company Institute, show that the equity fund assets of mutual funds have been growing since 1981. At the same time, money market funds have been increasing since 1980. Use these data to develop a regression model to forecast the equity fund assets by money market funds. All figures are given in billion-dollar units. Conduct a Durbin-Watson test on the data and the regression model to determine whether significant autocorrelation is present. Let $\alpha = .01$.

Year	Equity Funds	Money Market Funds
1980	44.4	76.4
1981	41.2	186.2
1982	53.7	219.8
1983	77.0	179.4
1984	83.1	233.6
1985	116.9	243.8
1986	161.5	292.2
1987	180.7	316.1
1988	194.8	338.0
1989	249.0	428.1
1990	245.8	498.3
1991	411.6	542.4

(*continued*)

Year	Equity Funds	Money Market Funds
1992	522.8	546.2
1993	749.0	565.3
1994	866.4	611.0
1995	1,269.0	753.0
1996	1,750.9	901.8
1997	2,399.3	1,058.9
1998	2,978.2	1,351.7
1999	4,041.9	1,613.2
2000	3,962.3	1,845.3
2001	3,418.2	2,285.3
2002	2,662.5	2,272.0
2003	3,684.2	2,052.0
2004	4,384.1	1,913.2
2005	4,940.0	2,040.5

15.44 The purchasing-power value figures for the minimum wage in year 18 dollars for the years 1 through 18 are shown here. Use these data and exponential smoothing to develop forecasts for the years 2 through 18. Try $\alpha = .1, .5$, and .8, and compare the results using MAD. Discuss your findings. Select the value of alpha that worked best and use your exponential smoothing results to predict the figure for 19.

Year	Purchasing Power	Year	Purchasing Power
1	$6.04	10	$4.34
2	5.92	11	4.67
3	5.57	12	5.01
4	5.40	13	4.86
5	5.17	14	4.72
6	5.00	15	4.60
7	4.91	16	4.48
8	4.73	17	4.86
9	4.55	18	5.15

INTERPRETING THE OUTPUT

15.45 Shown on the following page is the Excel output for a regression analysis to predict the number of business bankruptcy filings over a 16-year period by the number of consumer bankruptcy filings. How strong is the model? Note the residuals. Compute a Durbin-Watson statistic from the data and discuss the presence of autocorrelation in this model.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.529
R Square	0.280
Adjusted R Square	0.228
Standard Error	8179.84
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	364069877.4	364069877.4	5.44	0.0351
Residual	14	936737379.6	66909812.8		
Total	15	1300807257			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	75532.43621	4980.08791	15.17	0.0000
Year	–0.01574	0.00675	–2.33	0.0351

RESIDUAL OUTPUT

Observation	*Predicted Bus. Bankruptcies*	*Residuals*
1	70638.58	–1338.6
2	71024.28	–8588.3
3	71054.61	–7050.6
4	70161.99	1115.0
5	68462.72	12772.3
6	67733.25	14712.8
7	66882.45	–3029.4
8	65834.05	–2599.1
9	64230.61	622.4
10	61801.70	9747.3
11	61354.16	9288.8
12	62738.76	–434.8
13	63249.36	–10875.4
14	61767.01	–9808.0
15	57826.69	–4277.7
16	54283.80	–256.8

ANALYZING THE DATABASES

see www.wiley.com/college/black and WileyPLUS

1. Use the Agricultural time-series database and the variable Green Beans to forecast the number of green beans for period 169 by using the following techniques.

a. Five-period moving average
b. Simple exponential smoothing with $\alpha = .6$
c. Time-series linear trend model
d. Decomposition

2. Use decomposition on Carrots in the Agricultural database to determine the seasonal indexes. These data actually represent 14 years of 12-month data. Do the seasonal indexes indicate the presence of some seasonal effects? Run an autoregression model to predict Carrots by a 1-month lag and another by a 12-month lag. Compare the two models. Because vegetables are somewhat seasonal, is the 12-month lag model significant?

3. Use the Energy database to forecast 2008 U.S. coal production by using simple exponential smoothing of previous U.S. coal production data. Let $\alpha = .2$ and $\alpha = .8$. Compare the forecast with the actual figure. Which of the two models produces the forecast with the least error?

4. Use the International Labor database to develop a regression model to predict the unemployment rate for Germany by the unemployment rate of Italy. Test for autocorrelation and discuss its presence or absence in this regression analysis.

CASE

DEBOURGH MANUFACTURING COMPANY

The DeBourgh Manufacturing Company was founded in 1909 as a metal-fabricating company in Minnesota by the four Berg brothers. In the 1980s, the company ran into hard times, as did the rest of the metal-fabricating industry. Among the problems that DeBourgh faced were declining sales, deteriorating labor relations, and increasing costs. Labor unions had resisted cost-cutting measures. Losses were piling up in the heavy job-shop fabrication division, which was the largest of the company's three divisions. A division that made pedestrian steel bridges closed in 1990. The remaining company division, producer of All-American lockers, had to move to a lower-cost environment.

In 1990, with the company's survival at stake, the firm made a risky decision and moved everything from its high-cost location in Minnesota to a lower-cost area in La Junta, Colorado. Eighty semitrailer trucks were used to move equipment and inventory 1000 miles at a cost of $1.2 million. The company was relocated to a building in La Junta that had stood vacant for 3 years. Only 10 of the Minnesota workers transferred with the company, which quickly hired and trained 80 more workers in La Junta. By moving to La Junta, the company was able to go nonunion.

DeBourgh also faced a financial crisis. A bank that had been loaning the company money for 35 years would no longer do so. In addition, a costly severance package was worked out with Minnesota workers to keep production going during the move. An internal stock-purchase "earnout" was arranged between company president Steven C. Berg and his three aunts, who were the other principal owners.

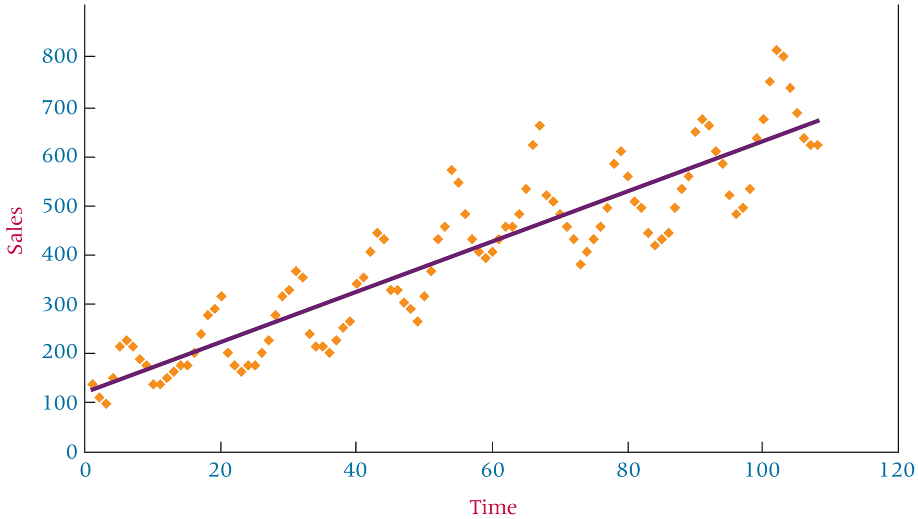
The roof of the building that was to be the new home of DeBourgh Manufacturing in La Junta was badly in need of repair. During the first few weeks of production, heavy rains fell on the area and production was all but halted. However, DeBourgh was able to overcome these obstacles. One year later, locker sales achieved record-high sales levels each month. The company is now more profitable than ever with sales topping $6 million. Much credit has been given to the positive spirit of teamwork fostered among its approximately 80 employees. Emphasis shifted to employee involvement in decision making, quality, teamwork, employee participation in compensation action, and shared profits. In addition, DeBourgh became a more socially responsible company by doing more for the town in which it is located and by using paints that are more environmentally friendly.

Discussion

1. After its move in 1990 to La Junta, Colorado, and its new initiatives, the DeBourgh Manufacturing Company began an upward climb of record sales. Suppose the figures shown here are the DeBourgh monthly sales figures from January 2001 through December 2009 (in $1,000s). Are any trends evident in the data? Does DeBourgh have a seasonal component to its sales? Shown after the sales figures is Minitab output from a decomposition analysis of the sales figures using 12-month seasonality. Next an Excel graph displays the data with a trend line. Examine the data, the output, and any additional analysis you feel is helpful, and write a short report on DeBourgh sales. Include a discussion of the general direction of sales and any seasonal tendencies that might be occurring.

Month	2001	2002	2003	2004	2005	2006	2007	2008	2009
January	139.7	165.1	177.8	228.6	266.7	431.8	381.0	431.8	495.3
February	114.3	177.8	203.2	254.0	317.5	457.2	406.4	444.5	533.4
March	101.6	177.8	228.6	266.7	368.3	457.2	431.8	495.3	635.0
April	152.4	203.2	279.4	342.9	431.8	482.6	457.2	533.4	673.1
May	215.9	241.3	317.5	355.6	457.2	533.4	495.3	558.8	749.3
June	228.6	279.4	330.2	406.4	571.5	622.3	584.2	647.7	812.8
July	215.9	292.1	368.3	444.5	546.1	660.4	609.6	673.1	800.1
August	190.5	317.5	355.6	431.8	482.6	520.7	558.8	660.4	736.6
September	177.8	203.2	241.3	330.2	431.8	508.0	508.0	609.6	685.8
October	139.7	177.8	215.9	330.2	406.4	482.6	495.3	584.2	635.0
November	139.7	165.1	215.9	304.8	393.7	457.2	444.5	520.7	622.3
December	152.4	177.8	203.2	292.1	406.4	431.8	419.1	482.6	622.3

```
Time-Series Decomposition
for Sales
Multiplicative Model
Data:      Sales
Length:    108
NMissing:  0
Fitted Trend Equation
Yt = 121.481 + 5.12862*t
Seasonal Indices
Period        Index
     1      0.79487
     2      0.85125
     3      0.92600
     4      1.02227
     5      1.11591
     6      1.24281
     7      1.31791
     8      1.16422
     9      0.99201
    10      0.91524
    11      0.85071
    12      0.80679
Accuracy Measures
MAPE:          8.04
 MAD:         29.51
 MSD:       1407.55
```



2. Suppose DeBourgh accountants computed a per-unit cost of lockers for each year since 1996, as reported here. Use techniques in this chapter to analyze the data. Forecast the per-unit labor costs through the year 2009. Use smoothing techniques, moving averages, trend analysis, and any others that seem appropriate. Calculate the error of the forecasts and determine which forecasting method seems to do the best job of minimizing error. Study the data and explain the behavior of the per-unit labor cost since 1996. Think about the company history and objectives since 1996.

Year	Per-Unit Labor Cost	Year	Per-Unit Labor Cost
1996	$80.15	2003	$59.84
1997	85.29	2004	57.29
1998	85.75	2005	58.74
1999	64.23	2006	55.01
2000	63.70	2007	56.20
2001	62.54	2008	55.93
2002	60.19	2009	55.60

Source: Adapted from "DeBourgh Manufacturing Company: A Move That Saved a Company," Real-World Lessons for America's Small Businesses: Insights from the Blue Chip Enterprise Initiative. Published by *Nation's Business* magazine on behalf of Connecticut Mutual Life Insurance Company and the U.S. Chamber of Commerce in association with the Blue Chip Enterprise Initiative, 1992. See also DeBourgh, available at http://www.debourgh.com, and the Web site containing Colorado Springs top business stories, available at http://www.csbj.com/1998/981113/top_stor.htm.

USING THE COMPUTER

EXCEL

- Excel has the capability of forecasting using several of the techniques presented in this chapter. Two of the forecasting techniques are accessed using the **Data Analysis** tool, and two other forecasting techniques are accessed using the **Insert Function**.
- To use the **Data Analysis** tool, begin by selecting the **Data** tab on the Excel worksheet. From the **Analysis** panel at the right top of the **Data** tab worksheet, click on **Data Analysis**. If your Excel worksheet does not show the **Data Analysis** option, then you can load it as an add-in following directions given in Chapter 2.
- To do exponential smoothing, select **Exponential Smoothing** from the **Data Analysis** pulldown menu. In the dialog box, input the location of the data to be smoothed in **Input Range**. Input the value of the dampening factor in **Damping factor**. Excel will default to .3. Input the location of the upper left cell of the output table in the **Output Range** space. The output consists of forecast values of the data. If you check **Standard Errors**, a second column of output will be given of standard errors.
- To compute moving averages, select **Moving Average** from the **Data Analysis** pulldown menu. In the dialog box, input the location of the data for which the moving averages are

to be computed in **Input Range**. Record how many values you want to include in computing the moving average in **Interval**. The default number is three values. Input the location of the upper left cell of the output table in **Output Range**. The output consists of the moving averages. If you check **Standard Errors**, a second column of output will be given of standard errors.

- To use the **Insert Function** (f_x) to compute forecasts and/or to fit a trend line, go to the **Formulas** tab on an Excel worksheet (top center tab). The **Insert Function** is on the far left of the menu bar. In the **Insert Function** dialog box at the top, there is a pulldown menu where it says **Or select a category**. From the pulldown menu associated with this command, select **Statistical**.
- To compute forecasts using linear regression, select **FORECAST** from the **Insert Function**'s **Statistical** menu. In the first line of the **FORECAST** dialog box, place the value of *x* for which you want a predicted value in **X**. An entry here is required. On the second line, place the location of the *y* values to be used in the development of the regression model in **Known_y's**. On the third line, place the location of the *x* values to be used in the development of the regression model in **Known_x's**. The output consists of the predicted value.
- To fit a trend line to data, select **TREND** from the **Insert Function**'s **Statistical** menu. On the first line of the **TREND** dialog box, place the location of the *y* values to be used in the development of the regression model in **Known_y's**. On the second line, place the location of the *x* values to be used in the development of the regression model in **Known_x's**. Note that the *x* values can consist of more than one column if you want to fit a polynomial curve. To accomplish this, place squared values of *x*, cubed values of *x*, and so on as desired in other columns, and include those columns in **Known_x**. On the third line, place the values for which you want to return corresponding *y* values in **New_x's**. In the fourth line, place **TRUE** in **Const** if you want to get a value for the constant as usual (default option). Place **FALSE** if you want to set b_0 to zero.

MINITAB

- There are several forecasting techniques available through Minitab. These techniques are accessed in the following way: select **Stat** from the menu bar, and from the ensuing pulldown menu, select **Time Series**. From this pulldown menu select one of several forecasting techniques as detailed below.
- To begin a **Time Series Plot**, select which of the four types of plots you want from **Simple**, **Multiple**, **With Groups**, **or Multiple with Groups**. Enter the column containing the values that you want to plot in **Series**. Other options include **Time/Scale**, where you can determine what time frame you want to use along the *x*-axis; **Labels**, where you input titles and data labels; **Data View**, where you can choose how you want the graph to appear with options of symbols, connect line, or project lines; **Multiple Graphs**; and **Data Options**.
- To begin a **Trend Analysis**, place the location of the time-series data in the **Variables** slot. Under **Model** Type, select the type of model you want to create from **Linear**, **Quadratic**, **Exponential growth**, or **S-Curve**. You can generate forecasts from your model by checking **Generate forecasts** and inserting how many forecasts you want and the starting point. Other options include **Time**, where you can determine what time frame you want to use along the *x*-axis; **Options**, where you input titles and data weights; **Storage**, where you can choose to store fits and/or residuals; **Graphs**, where you can choose from several graphical display options; and **Results**, which offers you three different ways to display the results.
- To begin a **Decomposition**, place the location of the time-series data in the **Variables** slot. Choose the **Model Type** by selecting from **Multiplicative**, **Additive**, **Trend plus seasonal**, or **Seasonal only**. You can generate forecasts from your model by checking **Generate forecasts** and inserting how many forecasts you want and the starting point. Other options include **Time**, where you can determine what time frame you want to use along the *x*-axis; **Options**, where you input the title and the seasonal location of the first observation; **Storage**, where you can choose to store trend line, detrended data, seasonals, seasonally adjusted data, fits, and residuals; **Graphs**, where you can choose from several graphical display options; and **Results**, which offers you three different ways to display the results.
- To begin a **Moving Average**, place the location of the time-series data in the **Variables** slot. Enter a positive integer to indicate desired length for the moving average in the **MA Length** slot. Check the **Center the moving averages** box if you want to place the moving average values at the period that is in the center of the range rather than at the end of the range. You can generate forecasts from your model by checking **Generate forecasts** and inserting how many forecasts you want and the starting point. Other options include **Time**, where you can determine what time frame you want to use along the *x*-axis; **Options**, where you input the title; **Storage**, where you can choose to store moving averages, fits, and residuals; **Graphs**, where you can choose from several graphical display options; and **Results**, which offers you three different ways to display the results.
- To begin **Single Exp Smoothing**, place the location of the time-series data in the **Variables** slot. Under **Weight to Use in Smoothing**, if you choose **Optimal ARIMA**, the forecasts will use the default weight, which Minitab computes by fitting an ARIMA (0, 1, 1) model to the data. With this option, Minitab calculates the initial smoothed value by

backcasting. If you choose **Use**, you can enter a specific weight that is between 0 and 2. You can generate forecasts from your model by checking **Generate forecasts** and inserting how many forecasts you want and the starting point. Other options include **Time**, where you can determine what time frame you want to use along the *x*-axis; **Options**, where you input the title; **Storage**, where you can choose to store smoothed data, fits, and residuals; **Graphs**, where you can choose from several graphical display options; and **Results**, which offers you three different ways to display the results.

- To begin **Differences**, enter the column containing the variable for which you want to compute differences in **Series.** Enter a storage column for the differences in the box beside **Store differences in**. In the box beside **Lag**, enter the value for the lag. The default lag value is 1.
- To begin **Lag**, enter the column containing the variable that you want to lag in **Series**. Enter the storage column for the lags in **Store lags in**. Enter the value for the lag in **Lag**. The default lag value is 1.
- To begin **Autocorrelation**, enter the column containing the time series in **Series.** If you want to use the default number of lags, choose **Default number of lags**. This number is $n/4$ for a series with less than or equal to 240 observations or $\sqrt{x} + 45$ for a series with more than 240 observations, where n is the number of observations in the series. By selecting **Number of lags**, you can enter the number of lags to use instead of the default. The maximum number of lags is $n - 1$. Check **Store ACF** to store the autocorrelation values in the next available column. Check **Store t Statistics** to store the t statistics. Check **Store Ljung-Box Q Statistics** to store the Ljung-Box Q statistics.