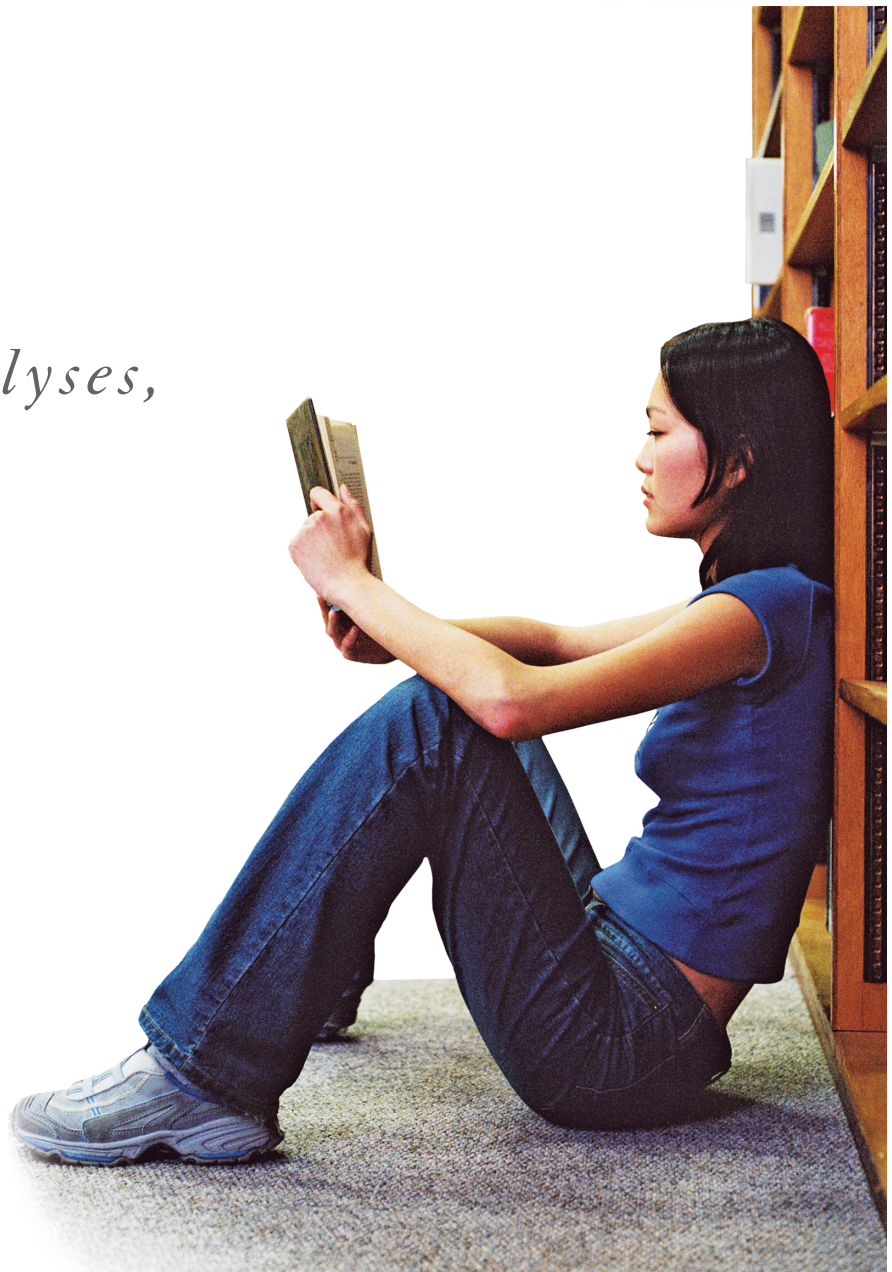


# CHAPTER 13

## Time Series

### *Descriptive Analyses, Models, and Forecasting*



#### CONTENTS

- 13.1** Descriptive Analysis: Index Numbers
- 13.2** Descriptive Analysis: Exponential Smoothing
- 13.3** Time Series Components
- 13.4** Forecasting: Exponential Smoothing
- 13.5** Forecasting Trends: The Holt-Winters Forecasting Model (Optional)
- 13.6** Measuring Forecast Accuracy: MAD and RMSE
- 13.7** Forecasting Trends: Simple Linear Regression
- 13.8** Seasonal Regression Models

- 13.9** Autocorrelation and the Durbin-Watson Test

#### STATISTICS IN ACTION

- Forecasting the Monthly Sales of a New Cold Medicine

#### USING TECHNOLOGY

- 13.1** Forecasting Using SPSS
- 13.2** Forecasting Using MINITAB
- 13.3** Forecasting Using Excel

#### WHERE WE'VE BEEN

- Discovered that variation in the output of processes is inevitable.
- Presented both managerial and statistical methods for continuously

improving processes and the quality of their output.

- Learned how to use control charts to monitor process variation and determine when action should be taken to improve a process.

#### WHERE WE'RE GOING

- Focus on methods for analyzing data generated by a process over time (i.e., *time series data*).
- Present descriptive methods for characterizing time series data.
- Present inferential methods for forecasting future values of time series data.

## Forecasting the Monthly Sales of a New Cold Medicine



*In the pharmaceutical industry, sales forecasting is critical to the success of the company. Accurate forecasts aid sales managers in improving decision making, the finance department in controlling and scheduling its operating costs and capital budget, the*

*human resources department in projecting staffing, and the purchasing department in controlling inventories and production schedules. Due to the critical life-and-death nature of the industry, pharmaceutical manufacturers rely on sophisticated analytical techniques to build the forecasts.*

Several years ago, a major pharmaceutical company based in New Jersey introduced a new cold medicine called Coldex. (For proprietary reasons, the actual name of the product is withheld.) Coldex is now sold regularly in drugstores and supermarkets across the United States. Prior to launching the product nationally, the company hired consultants from the Graduate School of Management at Rutgers University (The State University of New Jersey) to help the company build a monthly forecast model for Coldex. This Statistics in Action problem involves a portion of the analysis conducted by the consultants.

Consider the task of forecasting the sales of Coldex for the first 3 months of the third year of the product's existence. The company provided data on the monthly sales (in thousands of dollars) for the first 2 years of the product's life. The data, saved in the **COLDEX** file, are listed in Table SIA13.1. In the Statistics in Action Revisited sections listed below, we demonstrate several forecasting methods used by the consultants.



**COLDEX**

**TABLE SIA13.1 Coldex Monthly Sales Data**

Year	Month	Time	Sales
1	Jan	1	3394
	Feb	2	4010
	Mar	3	924
	Apr	4	205
	May	5	293
	Jun	6	1130
	Jul	7	1116
	Aug	8	4009
	Sep	9	5692
	Oct	10	3458
	Nov	11	2849
	Dec	12	3470
2	Jan	13	4568
	Feb	14	3710
	Mar	15	1675
	Apr	16	999
	May	17	986
	Jun	18	1786
	Jul	19	2253
	Aug	20	5237
	Sep	21	6679
	Oct	22	4116
	Nov	23	4109
	Dec	24	5124

*Source:* Personal communication from Carol Cowley, Carla Marchesini, and Ginny Wilson, Rutgers University, Graduate School of Management.

### Statistics in Action Revisited

- Forecasting Coldex Sales with Exponential Smoothing (p. 13-33)
- Forecasting Coldex Sales with Simple Linear Regression (p. 13-37)
- Forecasting Coldex Sales with a Seasonal Regression Model (p. 13-42)

**I**n the previous chapter, we were concerned with improving processes. In this chapter, our concern is not with the improvement of the internal workings of processes but with describing and predicting the output of processes. The process outputs on which we focus are the streams of data generated by processes over time. Recall from Chapters 2 and 12 that such data streams are called **time series** or **time series data**. For example, businesses generate time series data such as weekly sales, quarterly earnings, and yearly

profits that can be used to describe and evaluate the performance of the business. The U.S. economy can be thought of as a system that generates streams of data that include the gross domestic product, the Consumer Price Index, and the unemployment rate.

The methods of this chapter focus exclusively on the time series data generated by a process. Properly analyzed, these data reveal much about the past and future behavior of the process. Time series data, like other types of data we have discussed in previous chapters, are subjected to two kinds of analyses: **descriptive** and **inferential**. Descriptive analyses use graphical and numerical techniques to provide a clear understanding of any patterns that are present in the time series. After graphing the data, you will often want to use it to make inferences about the future values of the time series (i.e., you will want to **forecast** future values). For example, once you understand the past and present trends of the Dow Jones Industrial Average, you would probably want to forecast its future trend before making decisions about buying and selling stocks. Since significant amounts of money may be riding on the accuracy of your forecasts, you would be interested in measures of their reliability. Forecasts and their measures of reliability are examples of **inferential techniques** in time series analysis.

## 13.1 Descriptive Analysis: Index Numbers

The most common technique for characterizing a business or economic time series is to compute *index numbers*. Index numbers measure how a time series changes over time. Change is measured relative to a preselected time period, called the *base period*.

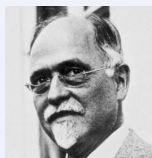
### Definition 13.1

An **index number** measures the change in a variable over time relative to the value of the variable during a specific **base period**.

Two types of indexes dominate business and economic applications: **price** and **quantity indexes**. Price indexes measure changes in the price of a commodity or group of commodities over time. The Consumer Price Index (CPI) is a price index because it measures price changes of a group of commodities that are intended to reflect typical purchases of American consumers. On the other hand, an index constructed to measure the change in the total number of automobiles produced annually by American manufacturers would be an example of a quantity index.

Methods of calculating index numbers range from very simple to extremely complex, depending on the numbers and types of commodities represented by the index. Several important types of index numbers are described in this section.

## Biography



**IRVING FISHER**  
(1867–1947)  
Index Numbers Expert

New York State native Irving Fisher, the son of a congregational minister, graduated from Yale University with a bachelor's degree in mathematics in 1888. Fisher continued at Yale with his graduate studies, earning the first PhD in economics ever awarded by the university in 1891. He had a long, distinguished career as a professor at Yale and became a very successful businessman. (Fisher made a fortune with his invention of a "visible card index" sys-

tem—known today as the Rolodex®.) Fisher is considered one of the most influential economists of the 19th and 20th centuries; he had an uncanny ability to explain and write clearly about the most technical economic theories. Fisher also had a reputation as a colorful eccentric. (To illustrate his price theory in his dissertation, Fisher constructed a remarkable machine equipped with pumps, wheels, levers, and pipes.) Fisher's best-known contribution to the field of statistics was as a pioneer in the construction and use of price indexes. A colleague at Yale, James Tobin, once called Fisher "the greatest expert of all time on index numbers."

## Simple Index Numbers

When an index number is based on the price or quantity of a single commodity, it is called a *simple index number*.

### Definition 13.2

A **simple index number** is based on the relative changes (over time) in the price or quantity of a single commodity.

For example, consider the price of silver (in dollars per fine ounce) between 1975 and 2005, shown in Table 13.1. To construct a simple index to describe the relative changes in silver prices, we must first choose a *base period*. The choice is important because the price for all other periods will be compared with the price during the base period. We select 1975 as the base period, a time just preceding the period of rapid economic inflation associated with dramatic oil price increases.

To calculate the simple index number for a particular year, we divide that year's price by the price during the base year and multiply the result by 100. Thus, for the 1980 silver price index number, we calculate

$$1980 \text{ index number} = \left( \frac{1980 \text{ silver price}}{1975 \text{ silver price}} \right) 100 = \left( \frac{20.64}{4.42} \right) 100 = 467.0$$

Similarly, the index number for 2005 is

$$2005 \text{ index number} = \left( \frac{2005 \text{ silver price}}{1975 \text{ silver price}} \right) 100 = \left( \frac{7.32}{4.42} \right) 100 = 165.6$$

The index number for the base period is always 100. In our example, we have

$$1975 \text{ index number} = \left( \frac{1975 \text{ silver price}}{1975 \text{ silver price}} \right) 100 = 100$$

Thus, the silver price has risen by 367% (the difference between the 1980 and 1975 index numbers) between 1975 and 1980, and by only 65.6% between 1975 and 2005. The simple index numbers for silver prices between 1975 and 2005 are given in Table 13.2 and are portrayed graphically in Figure 13.1. The steps for calculating simple index numbers are summarized in the next box.



### SILVER

**TABLE 13.1 Silver Prices, 1975–2005**

Year	Price (\$/oz.)	Year	Price (\$/oz.)	Year	Price (\$/oz.)
1975	4.42	1985	6.14	1995	5.20
1976	4.35	1986	5.47	1996	5.18
1977	4.62	1987	7.01	1997	4.89
1978	5.40	1988	6.53	1998	5.48
1979	11.09	1989	5.50	1999	5.22
1980	20.64	1990	4.82	2000	4.97
1981	10.52	1991	4.04	2001	4.37
1982	7.95	1992	3.94	2002	4.60
1983	11.44	1993	4.30	2003	4.80
1984	8.14	1994	5.30	2004	6.67
				2005	7.32

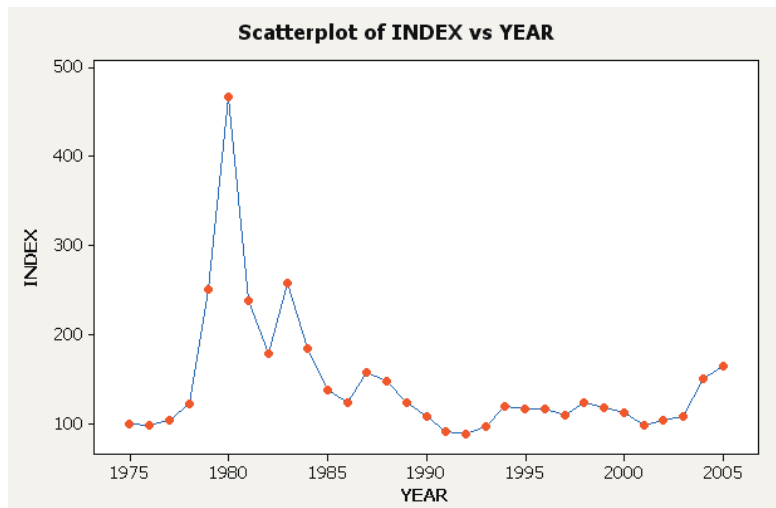
Source: Standard & Poor's, *Current Statistics*, Dec. 2005; [www.nma.org/enumerate/silver](http://www.nma.org/enumerate/silver).



## SILVER

**TABLE 13.2** Simple Index Numbers for Silver Prices (Base 1975)

Year	Index	Year	Index
1975	100.0	1991	91.4
1976	98.4	1992	89.1
1977	104.5	1993	97.3
1978	122.2	1994	119.9
1979	250.9	1995	117.6
1980	467.0	1996	117.2
1981	238.0	1997	110.6
1982	179.9	1998	124.0
1983	258.8	1999	118.1
1984	184.2	2000	112.4
1985	138.9	2001	98.9
1986	123.8	2002	104.1
1987	158.6	2003	108.6
1988	147.7	2004	150.9
1989	124.4	2005	165.6
1990	109.1		



**FIGURE 13.1**  
MINITAB graph of simple silver price index

### Now Work Exercise 13.6

#### Steps for Calculating a Simple Index Number

1. Obtain the prices or quantities for the commodity over the time period of interest.
2. Select a base period.
3. Calculate the index number for each period according to the formula

$$\text{Index number at time } t = \left( \frac{\text{Time series value at time } t}{\text{Time series value at base period}} \right) 100$$

Symbolically,

$$I_t = \left( \frac{Y_t}{Y_0} \right) 100$$

where  $I_t$  is the index number at time  $t$ ,  $Y_t$  is the time series value at time  $t$ , and  $Y_0$  is the time series value at the base period.

#### Composite Index Numbers

A **composite index number** represents combinations of the prices or quantities of several commodities. For example, suppose you want to construct an index for the total number of sales of the three major automobile manufacturers in the United States: General Motors, Ford, and Chrysler. The first step is to collect data on the sales of each manufacturer during the period in which you are interested, say 1980–2005. To summarize the information from all three time series in a single index, we add the sales of each manufacturer for each year—that is, we form a new time series consisting of the total number of automobiles sold by the three manufacturers. Then we construct a simple index for the *total* of the three series. The resulting index is called a *simple composite index*. We illustrate the construction of a simple composite index in Example 13.1.

**Definition 13.3**

A **simple composite index** is a simple index for a time series consisting of the total price or total quantity of two or more commodities.

### EXAMPLE 13.1

#### Constructing a Simple Composite Index for Hi-Tech Stocks

**Problem** One of the primary uses of index numbers is to characterize changes in stock prices over time. Stock market indexes have been constructed for many different types of companies and industries, and several composite indexes have been developed to characterize all stocks. These indexes are reported on a daily basis in the news media (e.g., Standard and Poor's 500 Stocks Index and Dow Jones 65 Stocks Index).

Consider the monthly closing prices (i.e., closing prices on the last day of each month) given in Table 13.3 for three high-technology company stocks listed on the New York Stock Exchange between 2004 and 2005. To see how this type of stock fared, construct a simple composite index using January 2004 as the base period. Graph the index, and comment on its implications.

**Solution** First, we calculate the total for the three stock prices each month. These totals are shown in the "TOTAL" column on the Excel workbook displayed in Figure 13.2. Then the simple composite index is calculated by dividing each monthly total by the January 2004 total. The index values are given in the last column of Figure 13.2, and a graph of the simple composite index is shown in Figure 13.3.

The plot of the simple composite index for these high-technology stocks shows a generally decreasing trend over the 2-year period. The composite price of these high-technology stocks dropped about 15% from January 2004 (Index = 100) to December 2005 (Index = 84.65).

**Look Back** The difference between two index numbers gives the percentage change in the value of the time series variable between the two time periods.

**HITECH****TABLE 13.3** Monthly Closing Prices of Three High-Technology Company Stocks

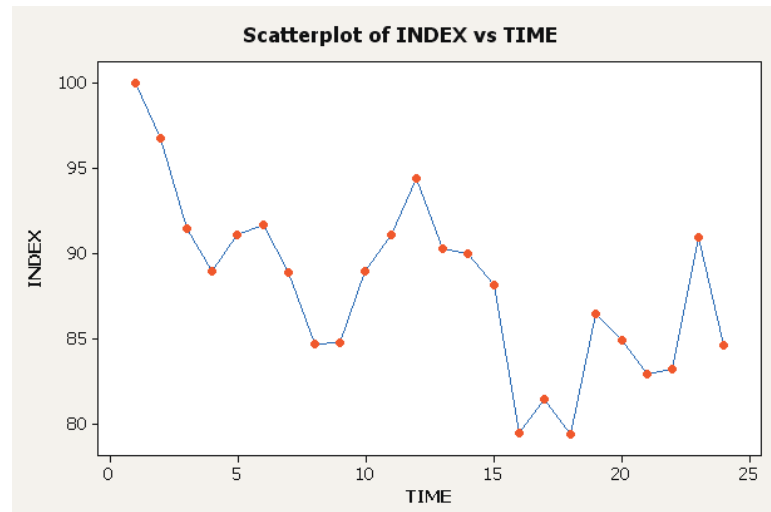
Year	Month	Time	IBM	Intel	Microsoft
2004	JAN	1	99.23	30.52	27.73
	FEB	2	96.50	29.20	26.61
	MAR	3	91.84	27.20	25.00
	APR	4	88.17	25.73	26.20
	MAY	5	88.59	28.55	26.30
	JUN	6	88.15	27.60	28.64
	JUL	7	87.07	24.38	28.57
	AUG	8	84.69	21.29	27.38
	SEP	9	85.74	20.06	27.73
	OCT	10	89.75	22.26	28.05
	NOV	11	94.24	22.38	26.81
	DEC	12	98.58	23.39	26.72
2005	JAN	13	93.42	22.45	26.28
	FEB	14	92.58	23.99	25.16
	MAR	15	91.38	23.23	24.17
	APR	16	76.38	23.52	25.30
	MAY	17	75.55	26.96	25.80
	JUN	18	74.20	26.02	24.84
	JUL	19	83.46	27.14	25.61
	AUG	20	80.62	25.72	27.38
	SEP	21	80.22	24.65	25.73
	OCT	22	81.88	23.50	25.70
	NOV	23	88.90	26.68	27.68
	DEC	24	82.20	24.96	26.15

Source: Standard & Poor's NYSE Daily Stock Price Record, 2004–2005.

	A	B	C	D	E	F	G	H
1	YEAR	MONTH	TIME	IBM	INTEL	MICROSOFT	TOTAL	INDEX
2	2004	JAN	1	99.23	30.52	27.73	157.48	100.00
3	2004	FEB	2	96.50	29.2	26.61	152.31	96.72
4	2004	MAR	3	91.84	27.2	25	144.04	91.47
5	2004	APR	4	88.17	25.73	26.2	140.10	88.96
6	2004	MAY	5	88.59	28.55	26.3	143.44	91.08
7	2004	JUN	6	88.15	27.6	28.64	144.39	91.69
8	2004	JUL	7	87.07	24.38	28.57	140.02	88.91
9	2004	AUG	8	84.69	21.29	27.38	133.36	84.68
10	2004	SEP	9	85.74	20.06	27.73	133.53	84.79
11	2004	OCT	10	89.75	22.26	28.05	140.06	88.94
12	2004	NOV	11	94.24	22.38	26.81	143.43	91.08
13	2004	DEC	12	98.58	23.39	26.72	148.69	94.42
14	2005	JAN	13	93.42	22.45	26.28	142.15	90.27
15	2005	FEB	14	92.58	23.99	25.16	141.73	90.00
16	2005	MAR	15	91.38	23.23	24.17	138.78	88.13
17	2005	APR	16	76.38	23.52	25.3	125.20	79.50
18	2005	MAY	17	75.55	26.96	25.8	128.31	81.48
19	2005	JUN	18	74.20	26.02	24.84	125.06	79.41
20	2005	JUL	19	83.46	27.14	25.81	136.21	86.49
21	2005	AUG	20	80.62	25.72	27.38	133.72	84.91
22	2005	SEP	21	80.22	24.65	25.73	130.60	82.93
23	2005	OCT	22	81.88	23.5	25.7	131.08	83.24
24	2005	NOV	23	88.90	26.68	27.68	143.26	90.97
25	2005	DEC	24	82.20	24.96	26.15	133.31	84.65

**FIGURE 13.2**

Excel workbook with simple composite index of stock prices

**FIGURE 13.3**

MINITAB graph of simple composite index for three stock prices

#### Now Work Exercise 13.10c, d

A simple composite price index has a major drawback: The quantity of the commodity that is purchased during each period is not taken into account. Only the price totals are used to calculate the index. We can remedy this situation by constructing a *weighted composite price index*.

#### Definition 13.4

A **weighted composite price index** weights the prices by quantities purchased prior to calculating totals for each time period. The weighted totals are then used to compute the index in the same way that the unweighted totals are used for simple composite indexes.

Since the quantities purchased change from time period to time period, the choice of which time period's quantities to use as the basis for the weighted composite index is an important one. A **Laspeyres index** uses the base period quantities as weights. The rationale is that the prices at each time period should be compared as if the same quantities were purchased each period as were purchased during the base period. This method measures price inflation (or deflation) by fixing the purchase quantities at their base period values. The method for calculating a Laspeyres index is given in the box.

### Steps for Calculating a Laspeyres Index

1. Collect price information for each of the  $k$  price series to be used in the composite index. Denote these series by  $P_{1t}, P_{2t}, \dots, P_{kt}$ .
2. Select a base period. Call this time period  $t_0$ .
3. Collect purchase quantity information for the base period. Denote the  $k$  quantities by  $Q_{1t_0}, Q_{2t_0}, \dots, Q_{kt_0}$ .
4. Calculate the weighted totals for each time period according to the formula

$$\sum_{i=1}^k Q_{it_0} P_{it}$$

5. Calculate the Laspeyres index,  $I_t$ , at time  $t$  by taking the ratio of the weighted total at time  $t$  to the base period weighted total and multiplying by 100—that is,

$$I_t = \frac{\sum_{i=1}^k Q_{it_0} P_{it}}{\sum_{i=1}^k Q_{it_0} P_{it_0}} \times 100$$

## EXAMPLE 13.2

### Constructing A Laspeyres Index

**Problem** The January 2004 and December 2005 prices for the three high-technology company stocks are given in Table 13.4. Suppose that, in January 2004, an investor purchased the quantities shown in the table. [Note: Only two prices are used to simplify the example. The same methods can be applied to calculate the index for other months.] Calculate the Laspeyres index for the investor's portfolio of high-technology stocks using January 2004 as the base period.

**TABLE 13.4** Prices of High-Technology Stocks with Quantities Purchased

	IBM	Intel	Microsoft
Shares purchased	500	100	1,000
January 2004 price	99.23	30.52	27.73
December 2005 price	82.20	24.96	26.15

**Solution** First, we calculate the weighted price totals for each time period, using the January 2004 quantities as weights. Thus,

$$\begin{aligned} \text{January 2004 weighted total} &= \sum_{i=1}^4 Q_{i, \text{Jan. 2004}} P_{i, \text{Jan. 2004}} \\ &= 500(99.23) + 100(30.52) + 1,000(27.73) \\ &= 80,397 \end{aligned}$$

$$\begin{aligned} \text{December 2005 weighted total} &= \sum_{i=1}^4 Q_{i, \text{Jan. 2004}} P_{i, \text{Dec. 2005}} \\ &= 500(82.20) + 100(24.96) + 1,000(26.15) \\ &= 69,746 \end{aligned}$$



Then the Laspeyres index is calculated by multiplying the ratio of each weighted total to the base period weighted total by 100. Thus,

$$I_{\text{Jan. 2004}} = \frac{\sum_{i=1}^4 Q_{i, \text{Jan. 2004}} P_{i, \text{Jan. 2004}}}{\sum_{i=1}^4 Q_{i, \text{Jan. 2004}} P_{i, \text{Jan. 2004}}} \times 100 = \frac{80,397}{80,397} \times 100 = 100$$

$$I_{\text{Dec. 2005}} = \frac{\sum_{i=1}^4 Q_{i, \text{Jan. 2004}} P_{i, \text{Dec. 2005}}}{\sum_{i=1}^4 Q_{i, \text{Jan. 2004}} P_{i, \text{Jan. 2005}}} \times 100 = \frac{69,746}{80,397} \times 100 = 86.75$$

**Look Back** The implication is that these stocks decreased in price by about  $(100 - 87\%) = 13\%$  from January 2004 to December 2005.

#### Now Work Exercise 13.14b

The Laspeyres index is appropriate when the base period quantities are reasonable weights to apply to all time periods. This is the case in applications such as that described in Example 13.2, where the base period quantities represent actual quantities of stock purchased and held for some period of time. Laspeyres indexes are also appropriate when the base period quantities remain reasonable approximations of purchase quantities in subsequent periods. However, it can be misleading when the relative purchase quantities change significantly from those in the base period.

Probably the best-known Laspeyres index is the all-items Consumer Price Index (CPI). This monthly composite index is made up of hundreds of item prices, and the U.S. Bureau of Labor Statistics (BLS) sampled over 30,000 families' purchases in 1982–1984 to determine the base period quantities. Thus, beginning in 1988, the all-items CPI published each month reflects quantities purchased in 1982–1984 by a sample of families across the United States. However, as prices increase for some commodities more quickly than for others, consumers tend to substitute less expensive commodities where possible. For example, as automobile and gasoline prices rapidly inflated in the mid-1970s, consumers began to purchase smaller cars. The net effect of using the base period quantities for the CPI is to overestimate the effect of inflation on consumers, because the quantities are fixed at levels that will actually change in response to price changes.

There are several solutions to the problem of purchase quantities that change relative to those of the base period. One is to change the base period regularly, so that the quantities are regularly updated. A second solution is to compute the index at each time period by using the purchase quantities of that period, rather than those of the base period. A **Paasche index** is calculated by using price totals weighted by the purchase quantities of the period the index value represents. The steps for calculating a Paasche index are given in the box.

#### Steps for Calculating a Paasche Index

1. Collect price information for each of the  $k$  price series to be used in the composite index. Denote these series by  $P_{1t}, P_{2t}, \dots, P_{kt}$ .
2. Select a base period. Call this time period  $t_0$ .
3. Collect purchase quantity information for every period. Denote the  $k$  quantities for period  $t$  by  $Q_{1t}, Q_{2t}, \dots, Q_{kt}$ .
4. Calculate the Paasche index for time  $t$  by multiplying the ratio of the weighted total at time  $t$  to the weighted total at time  $t_0$  (base period) by 100, where the weights used are the purchase quantities for time period  $t$ . Thus,

$$I_t = \frac{\sum_{i=1}^k Q_{it} P_{it}}{\sum_{i=1}^k Q_{it} P_{it_0}} \times 100$$

### EXAMPLE 13.3

#### Constructing a Paasche Index for Hi-Tech Stock Prices

**Problem** The January 2004 and December 2005 prices and volumes (actual quantities purchased) in millions of shares for the three high-technology company stocks are shown in Table 13.5. Calculate and interpret the Paasche index, using January 2004 as the base period.

**TABLE 13.5 Prices and Volumes of High-Technology Stocks**

	IBM		INTEL		MICROSOFT	
	Price	Volume	Price	Volume	Price	Volume
<b>January 2004</b>	99.23	6.1	30.52	59.5	27.73	63.6
<b>December 2005</b>	82.20	5.4	24.96	54.1	26.15	62.9

Source: Standard & Poor's. *NYSE Daily Stock Price Record*, Jan. 2004, 2005; <http://table.finance.yahoo.com>.

**Solution** The key to calculating a Paasche index is to remember that the weights (purchase quantities) change for each time period. Thus,

$$\begin{aligned}
 I_{\text{Jan. 2004}} &= \frac{\sum_{i=1}^4 Q_{i, \text{Jan. 2004}} P_{i, \text{Jan. 2004}}}{\sum_{i=1}^4 Q_{i, \text{Jan. 2004}} P_{i, \text{Jan. 2004}}} \times 100 = 100 \\
 I_{\text{Dec. 2005}} &= \frac{\sum_{i=1}^4 Q_{i, \text{Dec. 2005}} P_{i, \text{Dec. 2005}}}{\sum_{i=1}^4 Q_{i, \text{Dec. 2005}} P_{i, \text{Jan. 2005}}} \times 100 \\
 &= \frac{(5.4)(82.20) + (54.1)(24.96) + (62.9)(26.15)}{(5.4)(99.23) + (54.1)(30.52) + (62.9)(27.73)} \times 100 \\
 &= \frac{3,439.1}{3,931.2} \times 100 = 87.5
 \end{aligned}$$

The implication is that Dec. 2005 prices represent a  $(100 - 87.5)\% = 12.5\%$  decrease from Jan. 2004 prices, assuming the purchase quantities were at Dec. 2005 levels for *both* periods.

#### Now Work Exercise 13.14d

The Paasche index is most appropriate when you want to compare current prices to base period prices at *current* purchase levels. However, there are several major problems associated with the Paasche index. First, it requires that purchase quantities be known for every time period. This rules out a Paasche index for applications such as the CPI because the time and monetary resource expenditures required to collect quantity information are considerable. (Recall that more than 30,000 families were sampled to estimate purchase quantities in 1982–1984.) A second problem is that although each period is compared to the base period, it is difficult to compare the index at two other

periods because the quantities used are different for each period. Consequently, the change in the index is affected by changes in both prices *and* quantities. This fact makes it difficult to interpret the change in a Paasche index between periods when neither is the base period.

Although there are other types of indexes that use different weighting factors, the Laspeyres and Paasche indexes are the most popular composite indexes. Depending on the primary objective in constructing an index, one of them will probably be suitable for most purposes.

## Exercises 13.1–13.14

### Learning the Mechanics

**13.1** Explain in words how to construct a simple index.

**13.2** Explain in words how to calculate the following types of indexes:

- Simple composite index
- Weighted composite index
- Laspeyres index
- Paasche index

**13.3** Explain in words the difference between Laspeyres and Paasche indexes.

**13.4** The table below gives the prices for three products (A, B, and C) for the four quarters of last year.

Quarter	A	B	C
1	3.25	1.75	8.00
2	3.50	1.25	9.35
3	3.90	1.20	9.70
4	4.25	1.00	10.50

- Compute a simple index for the Quarter 4 price of product A, using Quarter 1 as the base period.
- Compute a simple index for the Quarter 2 price of product B, using Quarter 1 as the base period.
- Compute a simple composite index for the Quarter 4 price of all three products, using Quarter 1 as the base period.
- Compute a simple composite index for the Quarter 4 price of all three products, using Quarter 2 as the base period.

**13.5** Refer to Exercise 13.4. The next table gives the quantities purchased for three products (A, B, and C) for the four quarters of last year.

Quarter	A	B	C
1	100	20	50
2	200	25	35
3	250	50	25
4	300	100	20

- Compute a Laspeyres index for the Quarter 4 price of all three products, using Quarter 1 as the base period.
- Compute a Paasche index for the Quarter 4 price of all three products, using Quarter 2 as the base period.

### Applying the Concepts—Basic

**13.6 Annual median family income.** The next table lists the U.S. median annual family income every 5 years during the period 1975–2000. It also contains several values for each of two simple indexes for median family income.

- Calculate the missing values of each simple index.

- Interpret the index for 1990.

#### FAMINCOME

Year	Income (\$)	Base 1975 Index	Base 1980 Index
1975	13,719	–	65.26
1980	21,023	153.24	–
1985	27,735	202.16	–
1990	35,353	257.69	–
1995	40,611	–	–
2000	50,890	–	–

Source: U.S. Bureau of the Census. *Statistical Abstract of the United States*, 2003.

**13.7 Annual U.S. beer production.** The table below describes U.S. beer production (in millions of barrels) for the period 1980–2005.

- Use 1980 as the base period to compute the simple index for this time series. Interpret the value for 2005.
- Refer to part a. Is this an example of a quantity index or a price index?
- Recompute the simple index using 1991 as the base period. Plot the two indexes on the same graph. What pattern do you observe?

#### USBEEER

Year	Beer	Year	Beer	Year	Beer
1980	194	1989	199	1998	198
1981	194	1990	202	1999	198
1982	196	1991	204	2000	200
1983	196	1992	201	2001	199
1984	193	1993	202	2002	200
1985	194	1994	203	2003	195
1986	197	1995	200	2004	197
1987	195	1996	200	2005	195
1988	197	1997	199		

Source: 2005 *Brewer's Almanac*, U.S. Beer Institute.

**13.8 Quarterly single-family housing starts.** The quarterly numbers of single-family housing starts (in thousands of dwellings) in the United States from 2001 through 2005 are recorded in the table on p. 13-13.

- Using Quarter 1, 2001, as a base period, calculate the simple index for this quarterly time series.
- Interpret the simple index for Quarter 2, 2004.
- By what percentage did the number of housing starts increase between Quarter 1, 2001, and Quarter 4, 2005?
- By what percentage did the number of housing starts increase between Quarter 1, 2003, and Quarter 4, 2005?

### QTRHOUSE

Year	Quarter	Housing Starts
2001	1	274
	2	374
	3	341
2002	4	285
	1	293
	2	386
	3	361
2003	4	319
	1	304
	2	406
	3	412
2004	4	377
	1	345
	2	456
	3	440
2005	4	370
	1	369
	2	485
	3	471
	4	392

Source: U.S. Bureau of the Census, 2006.

**13.9 Price of natural gas.** The table below lists the price of natural gas (in dollars per 1,000 cubic feet) between 1980 and 2004.

### NATGAS

Year	Price	Year	Price	Year	Price
1980	3.68	1995	6.06	2001	9.63
1990	5.80	1996	6.34	2002	7.91
1991	5.82	1997	6.94	2003	9.52
1992	5.89	1998	6.82	2004	10.74
1993	6.16	1999	6.69		
1994	6.41	2000	7.71		

Source: U.S. Bureau of the Census, *Statistical Abstract of the United States*, 2006.

- Using 1980 as the base period, calculate and plot the simple index for the price of natural gas from 1990 through 2004.
- Use the simple index to interpret the trend in the price of natural gas.
- Is the index you constructed in part **a** a price or quantity index? Explain.

## Applying the Concepts—Intermediate

**13.10 Employment in farm and nonfarm categories.** Civilian employment is broadly classified by the federal government into two categories—agricultural and nonagricultural. Employment figures (in thousands of workers) for farm and nonfarm categories for selected years from 1980 to 2003 are given in the table below.

### CVEMPLOY

Year	Farm	Nonfarm	Year	Farm	Nonfarm
1980	3,364	95,938	1995	3,440	121,460
1985	3,179	103,971	2000	2,464	134,427
1990	3,223	115,570	2003	2,275	135,461

Source: U.S. Bureau of the Census, *Statistical Abstract of the United States*, 2005.

- Compute simple indexes for each of the two time series using 1980 as the base period.
  - Which segment has shown the greater percentage change in employment over the period shown?
- NW** **c.** Compute a simple composite index for total employment for the years 1980–2003. Use 1980 as a base period.
- NW** **d.** Refer to part **c**. Interpret the composite index value for 2003.

**13.11 GDP personal consumption expenditures.** The gross domestic product (GDP) is the total national output of goods and services valued at market prices. As such, the GDP is a commonly used barometer of the U.S. economy. One component of the GDP is personal consumption expenditures, which is itself the sum of expenditures for durable goods, nondurable goods, and services. The GDP for these components (in billions of dollars) is shown in the next table, in 5-year increments from 1960 to 2004.

- Using these three component values, construct a simple composite index for the personal consumption component of GDP. Use 1970 as the base year.
- Suppose we want to update the index by using 1980 as the base year. Update the index using only the index values you calculated in part **a**, without referring to the original data.
- Graph the personal consumption expenditure index for the years 1960–2004, first using 1970 as the base year and then using 1980 as the base year. What effect does changing the base year have on the graph of this index?

### GDP

Year	Durables	Nondurables	Services
1960	\$ 43.5	\$ 153.1	\$ 135.9
1965	63.5	191.9	189.2
1970	85.3	270.4	290.8
1975	134.3	416.0	474.5
1980	212.5	682.9	852.7
1985	352.9	919.4	1,395.1
1990	468.2	1,229.2	2,063.8
1995	589.7	1,497.3	2,882.0
2000	803.9	1,972.9	3,906.9
2004	993.9	2,377.0	4,859.0

Source: U.S. Bureau of the Census, *Statistical Abstract of the United States*, 2005; www.bea.gov.

**13.12 GDP personal consumption expenditures (cont'd).** Refer to Exercise 13.11. Suppose the output quantities in 1970, measured in billions of units purchased, are as follows:

*Durable goods:* 10.9

*Nondurable goods:* 14.02

*Services:* 42.6

- Use the outputs to calculate the Laspeyres index from 1960 to 2004 (same increments as in Exercise 13.11) with 1970 as the base period.
- Plot the simple composite index of Exercise 13.11 and Laspeyres index of part **a** on the same graph. Comment on the differences between the two indexes.

**13.13 Hourly earnings for nonsupervisory workers.** The next table presents the average hourly earnings and the average number of hours worked per week in 5-year increments from 1975 to 2000 for nonsupervisory workers in three different industries.

**NONSUPER (Data for Exercise 13.13)**

Year	MANUFACTURING		TRANSPORTATION AND PUBLIC UTILITIES		WHOLESALE TRADE	
	Hourly Earnings	Weekly Hours	Hourly Earnings	Weekly Hours	Hourly Earnings	Weekly Hours
1975	4.83	39.5	5.88	39.7	4.72	38.6
1980	7.27	39.7	8.87	39.6	6.95	38.4
1985	9.54	40.5	11.40	39.5	9.15	38.4
1990	10.83	40.8	12.97	38.9	10.79	38.1
1995	12.37	41.6	14.23	39.5	12.43	38.3
2000	14.38	41.7	16.22	39.6	15.20	38.4

Source: U.S. Bureau of the Census, *Statistical Abstract of the United States*, 2001.

- Compute a simple index for average hourly earnings for manufacturing workers over the period 1975–2000. Use 1975 as the base year. Do the same for transportation and public utilities workers.
  - Plot the two simple indexes on the same graph and interpret the results.
  - Compute simple composite indexes for hourly earnings and weekly hours for the 24-year period. Use 1975 as the base year.
  - Plot the two composite indexes, part **c**, on the same graph and interpret the results.
- 13.14 Production and price of lead, steel, and copper.** The level of price and production of metals in the United States is one measure of the strength of the industrial economy. The table below lists the 2004 prices (in dollars per ton) and production (in thousands of tons) for three metals important to U.S. industry.
- Compute simple composite price and quantity indexes for the 12-month period, using January as the base period.
  - NW** Compute the Laspeyres price index for the 12-month period, using January as the base period.
  - Plot the simple composite and Laspeyres indexes on the same graph. Comment on the differences.
  - Compute the Paasche price index for metals for the 12-month period, using January as the base period.
  - Plot the Laspeyres and Paasche indexes on the same graph. Comment on the differences.
  - Compare the Laspeyres and Paasche index values for September and December. Which index is more appropriate for describing the change in this 4-month period? Explain.

**METALS**

Month	COPPER		STEEL		LEAD	
	Price	Production	Price	Production	Price	Production
Jan	1133.0	104.0	187.75	8656	769.6	33.4
Feb	1380.0	98.9	219.92	8400	895.8	32.8
Mar	1380.0	105.0	250.85	9268	906.0	33.5
Apr	1386.4	111.0	224.55	8901	798.6	35.1
May	1520.0	109.0	181.90	9163	848.8	31.2
Jun	1520.0	113.0	180.00	9006	903.2	33.1
Jul	1532.8	104.0	222.50	9164	964.0	33.8
Aug	1800.0	107.0	249.32	9314	949.2	36.9
Sep	1800.0	112.0	217.38	9234	961.6	36.9
Oct	1800.0	111.0	237.62	9551	960.2	36.2
Nov	1800.0	104.0	248.00	8989	990.8	34.0
Dec	1800.0	114.0	223.57	8660	995.0	33.9

Source: *The CRB Commodity Yearbook*, 2005. New York: John Wiley & Sons, Inc.

## 13.2 Descriptive Analysis: Exponential Smoothing

As you have seen in the previous section, index numbers are useful for describing trends and changes in time series. However, time series often have such irregular fluctuations that trends are difficult to describe. Index numbers can be misleading in such cases because the series is changing so rapidly. Methods for removing the rapid fluctuations in a time series so the general trend can be seen are called **smoothing** techniques.

**Exponential smoothing** is one type of weighted average that assigns positive weights to past and current values of the time series. A single weight,  $w$ , called the **exponential smoothing constant**, is selected so that  $w$  is between 0 and 1. Then the exponentially smoothed series,  $E_t$ , is calculated as follows:

$$\begin{aligned} E_1 &= Y_1 \\ E_2 &= wY_2 + (1 - w)E_1 \\ E_3 &= wY_3 + (1 - w)E_2 \\ &\vdots \\ E_t &= wY_t + (1 - w)E_{t-1} \end{aligned}$$

Thus, the exponentially smoothed value at time  $t$  assigns the weight  $w$  to the current series value and the weight  $(1 - w)$  to the previous smoothed value.

For example, consider the silver price time series for 1975–2005 in Table 13.1 (p. 13-5). Suppose we want to calculate the exponentially smoothed series using a smoothing constant of  $w = .3$ . The calculations proceed as follows:

$$\begin{aligned} E_{1975} &= Y_{1975} = 4.42 \\ E_{1976} &= .3Y_{1976} + (1 - .3)E_{1975} = .3(4.35) + .7(4.42) = 4.40 \\ E_{1977} &= .3Y_{1977} + (1 - .3)E_{1976} = .3(4.62) + .7(4.40) = 4.47 \\ &\vdots \end{aligned}$$

All the exponentially smoothed values corresponding to  $w = .3$  are given in the MINITAB worksheet, Figure 13.4. (*Note:* MINITAB gives the value of  $E_t$  in row  $t + 1$ .)

The actual silver prices and exponentially smoothed prices are graphed in Figure 13.5. Like many averages, the exponentially smoothed series changes less rapidly than the time series itself. The choice of  $w$  affects the smoothness of  $E_t$ . The smaller (closer to 0) is the value of  $w$ , the smoother is  $E_t$ . Since small values of  $w$  give more weight to the past values of the time series, the smoothed series is not affected by rapid changes in the current values and, therefore, appears smoother than the original series. Conversely, choosing  $w$  near 1 yields an exponentially smoothed series that is much like the original series—that is, large values of  $w$  give more weight to the current value of the time series so the smoothed series looks like the original series. This concept is illustrated in Figure 13.6. The steps for calculating an exponentially smoothed series are given in the box.

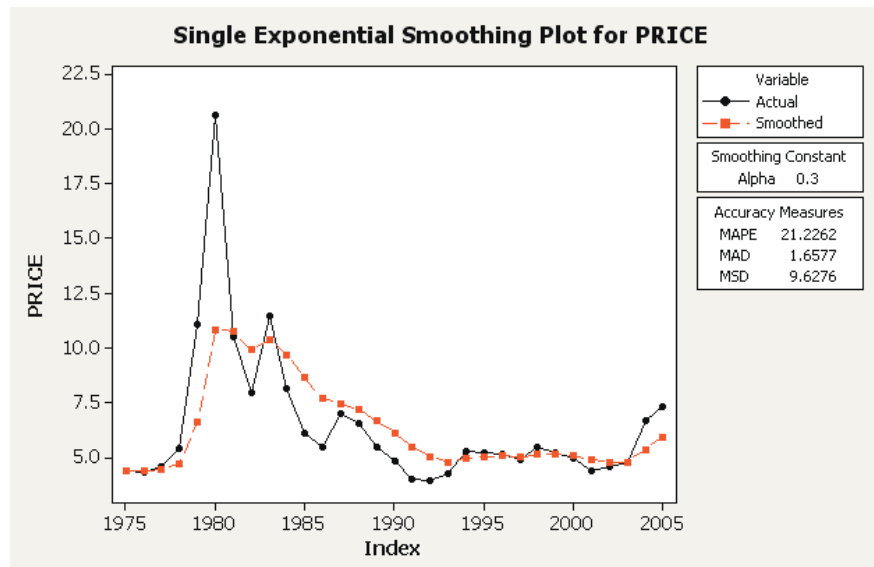
### Steps for Calculating an Exponentially Smoothed Series

1. Select an exponential smoothing constant,  $w$ , between 0 and 1. Remember that small values of  $w$  give less weight to the current value of the series and yield a smoother series. Larger choices of  $w$  assign more weight to the current value of the series and yield a more variable series.
2. Calculate the exponentially smoothed series  $E_t$  from the original time series  $Y_t$  as follows:

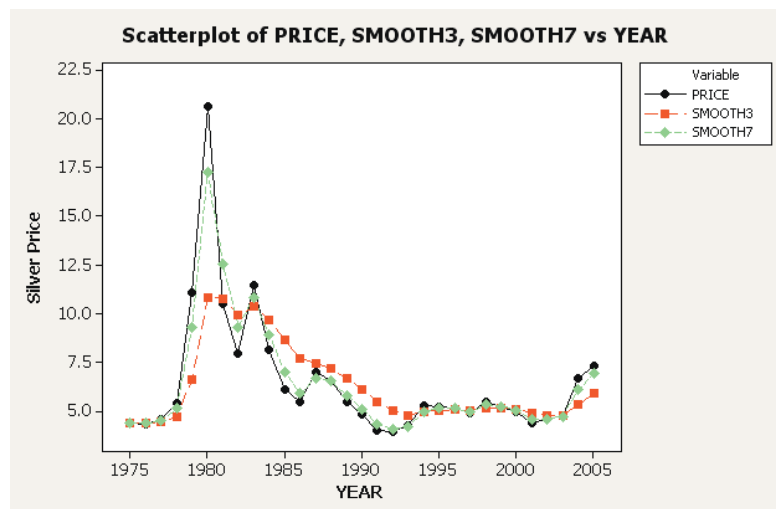
$$\begin{aligned} E_1 &= Y_1 \\ E_2 &= wY_2 + (1 - w)E_1 \\ E_3 &= wY_3 + (1 - w)E_2 \\ &\vdots \\ E_t &= wY_t + (1 - w)E_{t-1} \end{aligned}$$

SILVER.MTW ***			
↓	C1	C2	C3
	YEAR	PRICE	SMOOTH3
1	1975	4.42	4.4200
2	1976	4.35	4.3990
3	1977	4.62	4.4653
4	1978	5.40	4.7457
5	1979	11.09	6.6490
6	1980	20.64	10.8463
7	1981	10.52	10.7484
8	1982	7.95	9.9089
9	1983	11.44	10.3682
10	1984	8.14	9.6998
11	1985	6.14	8.6318
12	1986	5.47	7.8833
13	1987	7.01	7.4813
14	1988	6.53	7.1959
15	1989	5.50	6.8871
16	1990	4.82	6.1270
17	1991	4.04	5.5009
18	1992	3.94	5.0326
19	1993	4.30	4.8128
20	1994	5.30	4.9590
21	1995	5.20	5.0313
22	1996	5.18	5.0759
23	1997	4.89	5.0201
24	1998	5.48	5.1581
25	1999	5.22	5.1767
26	2000	4.97	5.1147
27	2001	4.37	4.8913
28	2002	4.60	4.8039
29	2003	4.80	4.8027
30	2004	6.67	5.3629
31	2005	7.32	5.9500

**FIGURE 13.4**  
MINITAB worksheet with exponentially smoothed ( $w = .3$ ) silver prices



**FIGURE 13.5**  
MINITAB graph of exponentially smoothed ( $w = .3$ ) silver prices



**FIGURE 13.6**  
MINITAB graph of exponentially smoothed ( $w = .3$  and  $w = .7$ ) silver prices

## EXAMPLE 13.4

### Constructing an Exponentially Smoothed Series for IBM Stock Prices

**Problem** Refer to Example 13.1 (p. 13-7). Consider the IBM common stock price from January 2004 to December 2005, shown in Table 13.3 (p. 13-7). Create the exponentially smoothed series using  $w = .5$ , and plot both series.

**Solution** To create the exponentially smoothed series with  $w = .5$ , we calculate

$$E_1 = Y_1 = 99.23$$

$$E_2 = wY_2 + (1 - w)E_1 = .5(96.50) + .5(99.23) = 97.87$$

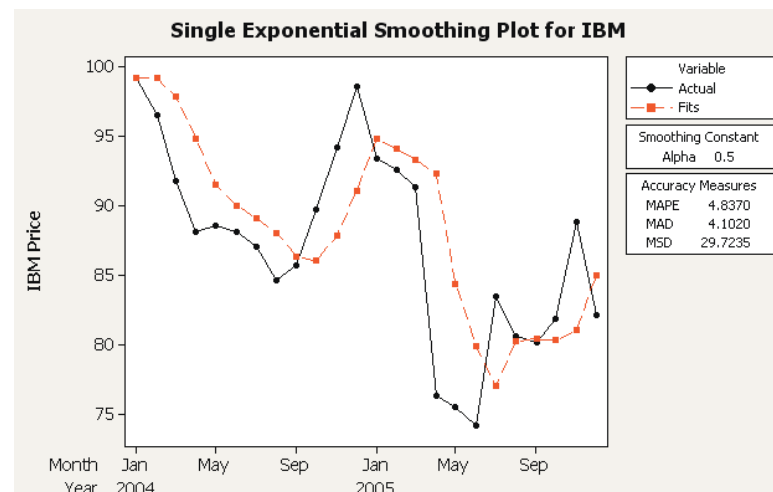
$$\vdots$$

$$E_{24} = wY_{24} + (1 - w)E_{23} = .5(82.20) + .5(85.0) = 83.60$$

These values, obtained using MINITAB, are shown on the MINITAB worksheet, Figure 13.7. The plot of the original and exponentially smoothed series is shown in Figure 13.8.

	YEAR	MONTH	TIME	IBM	SMOOTH5
1	2004	Jan	1	99.23	99.2300
2	2004	Feb	2	96.50	97.8650
3	2004	Mar	3	91.84	94.8525
4	2004	Apr	4	88.17	91.5113
5	2004	May	5	88.59	90.0506
6	2004	Jun	6	88.15	89.1003
7	2004	Jul	7	87.07	88.0852
8	2004	Aug	8	84.89	86.3876
9	2004	Sep	9	85.74	86.0638
10	2004	Oct	10	89.75	87.9069
11	2004	Nov	11	94.24	91.0734
12	2004	Dec	12	98.58	94.8267
13	2005	Jan	13	93.42	94.1234
14	2005	Feb	14	92.58	93.3517
15	2005	Mar	15	91.38	92.3658
16	2005	Apr	16	76.38	84.3729
17	2005	May	17	75.55	79.9615
18	2005	Jun	18	74.20	77.0807
19	2005	Jul	19	83.46	80.2704
20	2005	Aug	20	80.62	80.4452
21	2005	Sep	21	80.22	80.3326
22	2005	Oct	22	81.88	81.1063
23	2005	Nov	23	88.90	85.0031
24	2005	Dec	24	82.20	83.6016

**FIGURE 13.7**  
MINITAB worksheet with exponentially smoothed ( $w = .5$ ) IBM stock prices



**FIGURE 13.8**  
MINITAB graph of exponentially smoothed ( $w = .5$ ) IBM stock prices



**Look Back** The smoothed series provides a good picture of the general trend of the original series. Note, too, that the exponentially smoothed series will be less sensitive to any short-term deviations of the prices from the trend as occurred in 2004 and 2005.

**Now Work Exercise 13.16**

One of the primary uses of exponential smoothing is to forecast future values of a time series. Because only current and past values of the time series are used in exponential smoothing, it is easily adapted to forecasting. We demonstrate this application of exponentially smoothed series in Section 13.4.

## Exercises 13.15–13.22

### Learning the Mechanics

**13.15** Describe the effect of selecting an exponential constant of  $w = .2$ . Of  $w = .8$ . Which will produce a smoother trend?

**13.16** Consider the monthly time series shown in the table.

**NW**

**LM13\_16**

Month	$t$	$Y_t$	Exponentially Smoothed Series ( $w = .5$ )
Jan.	1	280	—
Feb.	2	281	—
Mar.	3	250	265.3
Apr.	4	246	255.6
May	5	239	—
Jun.	6	218	—
Jul.	7	218	—
Aug.	8	210	—
Sep.	9	205	—
Oct.	10	206	—
Nov.	11	200	—
Dec.	12	200	—

- Calculate the missing values in the exponentially smoothed series using  $w = .5$ .
- Graph the time series and the exponentially smoothed series on the same graph.

### Applying the Concepts—Basic

**USBEER**

**13.17 Annual U.S. beer production.** Refer to the annual U.S. beer production time series, Exercise 13.7. (p. 13-12).

- Calculate the exponentially smoothed series for U.S. beer production for the period 1980–2005 using  $w = .2$ .
- Calculate the exponentially smoothed series using  $w = .8$ .
- Plot the two exponentially smoothed series ( $w = .2$  and  $w = .8$ ) on the same graph. Which smoothed series best portrays the long-term trend?

**13.18 Foreign fish production.** Overfishing and pollution of U.S. coastal waters have resulted in an increased dependence by the United States on the fishing grounds of other countries. The next table describes the annual fish catch (in

thousands of metric tons) in all fishing areas of Brazil and Chile for selected years from 1990 to 2002.

**FISHTONS**

Year	Chile	Brazil
1990	5,195.4	802.9
1995	7,590.5	800.0
1998	3,265.3	706.8
1999	5,050.2	703.9
2000	4,300.0	766.8
2001	3,797.1	806.7
2002	4,271.5	822.1

Source: Statistical Division, Department of Economics and Social Information and Policy Analysis, *United Nations. Statistical Yearbook, 2005*.

- Compute an exponentially smoothed series for both Chile and Brazil, using a smoothing coefficient of  $w = .5$ .
- Plot both actual series and both smoothed series on the same graph. Describe the differences in variation of catches over time between the two countries. For example, do they move up and down together over time?

### Applying the Concepts—Intermediate

**13.19 Yearly price of gold.** The price of gold is used by some financial analysts as a barometer of investors' expectations of inflation, with the price of gold tending to increase as concerns about inflation increase. The table below shows the average annual price of gold (in dollars per ounce) from 1986 through 2005.

**GOLDYR**

Year	Price	Year	Price
1986	368	1996	389
1987	448	1997	333
1988	438	1998	294
1989	383	1999	278
1990	385	2000	279
1991	363	2001	273
1992	345	2002	310
1993	361	2003	363
1994	387	2004	409
1995	385	2005	444

Source: World Gold Council; www.gold.org.

- Compute an exponentially smoothed series for the gold price time series for the period from 1986 to 2005, using a smoothing coefficient of  $w = .8$ .
- Plot the original series and the exponentially smoothed series on the same graph. Comment on the trend observed.

**13.20 Personal consumption in transportation.** There has been phenomenal growth in the transportation sector of the economy since 1990. The personal consumption expenditure figures (in billions of dollars) are given in the table below.

- Compute exponentially smoothed values of this personal consumption time series, using the smoothing constants  $w = .2$  and  $w = .8$ .
- Plot the actual series and the two smoothed series on the same graph. Comment on the trend in personal consumption expenditure on transportation in the 2000s as compared to the 1990s.

**TRANSPRT**

Year	Expenditure on Transportation
1990	590.1
1991	548.9
1992	581.1
1993	607.6
1994	643.2
1995	654.6
1996	687.1
1997	727.4
1998	779.3
1999	831.6
2000	853.4
2001	872.0
2002	890.9
2003	912.3
2004	925.6
2005	931.5

Source: U.S. Bureau of the Census. *Statistical Abstract of the United States*, 2006.

**13.21 OPEC crude oil imports.** The data in the table below are the amounts of crude oil (millions of barrels) imported into the United States from the Organization of Petroleum Exporting Countries (OPEC) for the years 1974–2004.

**OPECOIL**

Year	$t$	Imports, $Y_t$	Year	$t$	Imports, $Y_t$	Year	$t$	Imports, $Y_t$
1974	1	926	1985	12	479	1996	23	1,258
1975	2	1,171	1986	13	771	1997	24	1,378
1976	3	1,663	1987	14	876	1998	25	1,522
1977	4	2,058	1988	15	987	1999	26	1,543
1978	5	1,892	1989	16	1,232	2000	27	1,664
1979	6	1,866	1990	17	1,282	2001	28	1,770
1980	7	1,414	1991	18	1,233	2002	29	1,490
1981	8	1,067	1992	19	1,247	2003	30	1,671
1982	9	633	1993	20	1,339	2004	31	1,833
1983	10	540	1994	21	1,307			
1984	11	553	1995	22	1,303			

Source: *Statistical Abstract of the United States*, U.S. Bureau of the Census, 2005.

- Construct two exponentially smoothed series for this time series, using  $w = .1$  and  $w = .9$ .
- Plot the original series and the two smoothed series on the same graph. Which smoothed series looks more like the original series? Why?

**13.22 S&P 500 Stock Index.** Standard & Poor's 500 Composite Stock Index (S&P 500) is a stock market index. Like the Dow Jones Industrial Average, it is an indicator of stock market activity. The next table contains end-of-quarter values of the S&P 500 for the years 1999–2006.

**SP500**

Year	Quarter	S&P 500	Year	Quarter	S&P 500
1999	1	1286.4	2003	1	848.2
	2	1372.7		2	974.5
	3	1282.7		3	996.0
	4	1469.2		4	1111.9
2000	1	1498.6	2004	1	1126.2
	2	1454.6		2	1140.8
	3	1436.5		3	1114.6
	4	1320.3		4	1211.9
2001	1	1160.3	2005	1	1180.6
	2	1224.4		2	1191.3
	3	1040.9		3	1228.8
	4	1148.1		4	1248.3
2002	1	1147.4	2006	1	1294.9
	2	989.8		2	1270.2
	3	815.3		3	1335.8
	4	879.8			

Source: Standard & Poor's Statistical Service: *Current Statistics*, 2006; [www.economagic.com](http://www.economagic.com).

- Calculate and plot the exponentially smoothed series for the quarterly S&P 500 using a smoothing constant of  $w = .3$ .
- Repeat part **a**, but use  $w = .7$ .
- Which exponentially smoothed series do you prefer for describing trends in the series? Explain.

## 13.3 Time Series Components

In the previous two sections, we showed how to use various *descriptive* techniques to obtain a picture of the behavior of a time series. Now we want to expand our coverage to include techniques that will let us make statistical inferences about the time series. These *inferential techniques* are generally focused on the problem of *forecasting* future values of the time series.

Before forecasts of future values of a time series can be made, some type of model that can be projected into the future must be used to describe the series. Time series models range in complexity from **descriptive models**, such as the exponential smoothing models discussed in the previous section, to **inferential models**, such as the combinations of regression and specialized time series models to be discussed later in this chapter. Whether the model is simple or complex, the objective is the same: to produce accurate forecasts of future values of the time series.

Many different algebraic representations of time series models have been proposed. One of the most widely used is an **additive model**\* of the form

$$Y_t = T_t + C_t + S_t + R_t$$

The **secular trend**,  $T_t$ , also known as the **long-term trend**, is a time series component that describes the long-term movements of  $Y_t$ . For example, if you want to characterize the secular trend of the production of automobiles since 1930, you would show  $T_t$  as an upward-moving time series over the period from 1930 to the present. This does not imply that the automobile production series has always moved upward from month to month and from year to year, but it does mean the long-term trend has been an increase over that period of time.

The **cyclical effect**,  $C_t$ , generally describes fluctuations of the time series about the secular trend that are attributable to business and economic conditions. For example, refer back to the monthly closing prices of three high-technology stocks for the years 2004–2005, Table 13.3. Recall that a plot of the simple composite index (Figure 13.3) showed a generally decreasing secular trend. However, during periods of recession, the index tends to lie below the secular trend, while in times of general economic expansion, it lies above the long-term trend line.

The **seasonal effect**,  $S_t$ , describes the fluctuations in the time series that recur during specific time periods. For example, quarterly power loads for a Florida utility company tend to be highest in the summer months (Quarter III), with another smaller peak in the winter months (Quarter I). The spring and fall (Quarters II and IV) seasonal effects are negative, meaning that the series tends to lie below the long-term trend line during those quarters.

The **residual effect**,  $R_t$ , is what remains of  $Y_t$  after the secular, cyclical, and seasonal components have been removed. Part of the residual effect may be attributable to unpredictable rare events (earthquake, presidential assassination, etc.) and part to the randomness of human actions. In any case, the presence of the residual component makes it impossible to forecast the future values of a time series without error. Thus, the presence of the residual effect emphasizes a point we first made in Chapter 10 in connection with regression models: No business phenomena should be described by deterministic models. All realistic business models, time series or otherwise, should include a residual component.

Each of the four components contributes to the determination of the value of  $Y_t$  at each time period. Although it will not always be possible to characterize each component separately, the component model provides a useful theoretical formulation that helps the time series analyst achieve a better understanding of the phenomena affecting the path followed by the time series.

\*Another useful form is the *multiplicative model*:  $Y_t = T_t C_t S_t R_t$ . This can be changed to an additive form by taking natural logarithms (i.e.,  $\ln Y_t = \ln T_t + \ln C_t + \ln S_t + \ln R_t$ ). See Section 13.8.

## 13.4 Forecasting: Exponential Smoothing

In Section 13.2, we discussed exponential smoothing as a method for describing a time series that involved removing the irregular fluctuations. In terms of the time series components discussed in the previous section, exponential smoothing tends to de-emphasize (or “smooth”) most of the residual effects. This, coupled with the fact that exponential smoothing uses only past and current values of the series, makes it a useful tool for forecasting time series.

Recall that the formula for exponential smoothing is

$$E_t = wY_t + (1 - w)E_{t-1}$$

where  $w$ , the *exponential smoothing constant*, is a number between 0 and 1. We learned that the selection of  $w$  controls the smoothness of  $E_t$ . A choice near 0 places more emphasis (weight) on *past* values of the time series and therefore yields a smoother series; a choice near 1 gives more weight to *current* values of the series.

Suppose the objective is to forecast the next value of the time series,  $Y_{t+1}$ . The **exponentially smoothed forecast** for  $Y_{t+1}$  is simply the smoothed value at time  $t$ :

$$F_{t+1} = E_t$$

where  $F_{t+1}$  is the **forecast** of  $Y_{t+1}$ . To help interpret this forecast formula, substitute the smoothing formula for  $E_t$ :

$$\begin{aligned} F_{t+1} = E_t &= wY_t + (1 - w)E_{t-1} \\ &= wY_t + (1 - w)F_t \\ &= F_t + w(Y_t - F_t) \end{aligned}$$

Note that we have substituted  $F_t$  for  $E_{t-1}$ , since the forecast for time  $t$  is the smoothed value for time  $(t - 1)$ . The final equation provides insight into the exponential smoothing forecast: The forecast for time  $(t + 1)$  is equal to the forecast for time  $t$ ,  $F_t$ , plus a correction for the error in the forecast for time  $t$ ,  $(Y_t - F_t)$ . This is why the exponentially smoothed forecast is called an **adaptive forecast**—the forecast for time  $(t + 1)$  is explicitly adapted for the error in the forecast for time  $t$ .

Because exponential smoothing consists of averaging past and present values, the smoothed values will tend to lag behind the series when a long-term trend exists. In addition, the averaging tends to smooth any seasonal component. Therefore, exponentially smoothed forecasts are appropriate only when the trend and seasonal components are relatively insignificant. Since the exponential smoothing model assumes that the time series has little or no trend or seasonal component, the forecast  $F_{t+1}$  is used to forecast not only  $Y_{t+1}$  but also all future values of  $Y_{t+1}$  that is, the forecast for two time periods ahead is

$$F_{t+2} = F_{t+1}$$

and for three time periods ahead is

$$F_{t+3} = F_{t+2} = F_{t+1}$$

The exponential smoothing forecasting technique is summarized in the box.

### Calculation of Exponentially Smoothed Forecasts

1. Given the observed time series  $Y_1, Y_2, \dots, Y_t$ , first calculate the exponentially smoothed values  $E_1, E_2, \dots, E_t$ , using

*Continued*

$$\begin{aligned}
 E_1 &= Y_1 \\
 E_2 &= wY_2 + (1 - w)E_1 \\
 &\vdots \\
 E_t &= wY_t + (1 - w)E_{t-1}
 \end{aligned}$$

2. Use the last smoothed value to forecast the next time series value:

$$F_{t+1} = E_t$$

3. Assuming that  $Y_t$  is relatively free of trend and seasonal components, use the same forecast for all future values of  $Y_t$ :

$$\begin{aligned}
 F_{t+2} &= F_{t+1} \\
 F_{t+3} &= F_{t+1} \\
 &\vdots
 \end{aligned}$$

Two important points must be made about exponentially smoothed forecasts:

1. The choice of  $w$  is crucial. If you decide that  $w$  will be small (near 0), you will obtain a smooth, slowly changing series of forecasts. On the other hand, the selection of a large value of  $w$  (near 1) will yield more rapidly changing forecasts that depend mostly on the current values of the series. In general, several values of  $w$  should be tried to determine how sensitive the forecast series is to the choice of  $w$ . Forecasting experience will provide the best basis for the choice of  $w$  for a particular application.
2. The farther into the future you forecast, the less certain you can be of accuracy. Since the exponentially smoothed forecast is constant for all future values, any changes in trend or seasonality are not taken into account. However, the uncertainty associated with future forecasts applies not only to exponentially smoothed forecasts, but also to all methods of forecasting. In general, time series forecasting should be confined to the short term.

### EXAMPLE 13.5

#### Forecasting Annual Silver Prices with Exponential Smoothing

**Problem** The annual silver prices from 1975 to 2005 are repeated in the MINITAB worksheet, Figure 13.9, along with the exponentially smoothed values using  $w = .3$  and  $w = .7$ . Apply the exponential smoothing technique to the data from 1975 to 2002 in order to forecast the silver prices from 2003 to 2005 using both  $w = .3$  and  $w = .7$ .

**Solution** First, we calculate the exponentially smoothed forecasts using  $w = .3$ . Following the steps outlined in the box, the forecast for 2003 is simply the smoothed price in 2002,

$$F_{2003} = E_{2002} = 4.80 \text{ (shaded on Figure 13.9)}$$

We use the same value as the forecast for 2004 and 2005:

$$\begin{aligned}
 F_{2004} &= E_{2002} = 4.80 \\
 F_{2005} &= E_{2002} = 4.80
 \end{aligned}$$

The same steps are repeated using  $w = .7$ . (The forecast for all 3 years is the smoothed value in 2002, 4.59.) Both sets of forecasts are shown in Table 13.6. Also shown are the actual silver prices from 2003 to 2005. The **forecast error**, defined as the actual value minus the forecast value, is given for each exponentially smoothed forecast.

**Look Back** Notice that the one-step-ahead forecasts for 2003 have smaller forecast errors than the two- and three-steps-ahead forecasts for 2004 and 2005. Neither the

↓	C1	C2	C3	C4
	YEAR	PRICE	SMOOTH3	SMOOTH7
3	1977	4.62	4.4653	4.5453
4	1978	5.40	4.7457	5.1436
5	1979	11.09	6.6490	9.3061
6	1980	20.64	10.8463	17.2398
7	1981	10.52	10.7484	12.5359
8	1982	7.95	9.9089	9.3258
9	1983	11.44	10.3682	10.8057
10	1984	8.14	9.6998	8.9397
11	1985	6.14	8.6318	6.9799
12	1986	5.47	7.8833	5.9230
13	1987	7.01	7.4813	6.6839
14	1988	6.53	7.1959	6.5762
15	1989	5.50	6.6871	5.8229
16	1990	4.82	6.1270	5.1209
17	1991	4.04	5.5009	4.3643
18	1992	3.94	5.0326	4.0673
19	1993	4.30	4.8128	4.2302
20	1994	5.30	4.9590	4.9791
21	1995	5.20	5.0313	5.1337
22	1996	5.18	5.0759	5.1661
23	1997	4.89	5.0201	4.9728
24	1998	5.48	5.1581	5.3279
25	1999	5.22	5.1767	5.2524
26	2000	4.97	5.1147	5.0547
27	2001	4.37	4.8913	4.5754
28	2002	4.80	4.8039	4.5926
29	2003	4.80	4.8027	4.7378
30	2004	6.67	5.3629	6.0903
31	2005	7.32	5.9500	6.9511

**FIGURE 13.9**  
MINITAB worksheet with exponentially smoothed ( $w = .5$  and  $w = .7$ ) silver prices

**TABLE 13.6 Silver Prices (2003–2005): Actual versus Forecast Values**

Year	Actual	Forecast ( $w = .3$ )	Forecast Error	Forecast ( $w = .7$ )	Forecast Error
2003	4.80	4.80	0	4.59	0.21
2004	6.67	4.80	1.87	4.59	2.08
2005	7.32	4.80	2.52	4.59	2.73

$w = .3$  nor the  $w = .7$  forecast projects the upturn in the silver prices in 2004 and 2005, because exponentially smoothed forecasts implicitly assume no trend exists in the time series. This example dramatically illustrates the risk associated with anything other than very short-term forecasting.

Many time series have long-term, or secular, trends. For such series, the exponentially smoothed forecast is inappropriate for all but the very short term. In the next section, we present an extension of the exponentially smoothed forecast—the *Holt-Winters forecast*—that allows for secular trend in the forecasts.

## 13.5 Forecasting Trends: The Holt-Winters Forecasting Model (Optional)

The exponentially smoothed forecasts for the silver prices in the previous section have large forecast errors, in part because they do not recognize the trend in the time series. In this section, we present an extension of the exponential smoothing method of forecasting that explicitly recognizes the trend in a time series. The **Holt-Winters forecasting model** consists of both an exponentially smoothed component ( $E_t$ ) and a trend component ( $T_t$ ). Consequently, the technique is sometimes called **double exponential smoothing**. The trend component is used in the calculation of the exponentially smoothed value. The following equations show that both  $E_t$  and  $T_t$  are weighted averages:

$$\begin{aligned} E_t &= wY_t + (1 - w)(E_{t-1} + T_{t-1}) \\ T_t &= v(E_t - E_{t-1}) + (1 - v)T_{t-1} \end{aligned}$$

Note that the equations require *two* smoothing constants,  $w$  and  $v$ , each of which is between 0 and 1. As before,  $w$  controls the smoothness of  $E_t$ ; a choice near 0 places more emphasis on past values of the time series, while a value of  $w$  near 1 gives more weight to current values of the series, and de-emphasizes the past.

The trend component of the series is estimated *adaptively*, using a weighted average of the most recent change in the level, represented by  $(E_t - E_{t-1})$ , and the trend estimate, represented by  $T_{t-1}$ , from the previous period. A choice of the weight  $v$  near 0 places more emphasis on the past estimates of trend, while a choice of  $v$  near 1 gives more weight to the current change in level.

The calculation of the Holt-Winters components, which proceeds much like the exponential smoothing calculations, is summarized in the box.

### Steps for Calculating Components of the Holt-Winters Model

1. Select an exponential smoothing constant  $w$  between 0 and 1. Small values of  $w$  give less weight to the current values of the time series and more weight to the past. Larger choices assign more weight to the current value of the series.
2. Select a trend smoothing constant  $v$  between 0 and 1. Small values of  $v$  give less weight to the current changes in the level of the series and more weight to the past trend. Larger values assign more weight to the most recent trend of the series and less to past trends.
3. Calculate the two components,  $E_t$  and  $T_t$ , from the time series  $Y_t$  beginning at time  $t = 2$  as follows:\*

$$\begin{aligned} E_2 &= Y_2 \\ T_2 &= Y_2 - Y_1 \\ E_3 &= wY_3 + (1 - w)(E_2 + T_2) \\ T_3 &= v(E_3 - E_2) + (1 - v)T_2 \\ &\vdots \end{aligned}$$

\*The calculation begins at time  $t = 2$  rather than at  $t = 1$  because the first two observations are needed to obtain the first estimate of trend,  $T_2$ . As an option, some statistical software packages use simple linear regression to estimate  $E_1$  and  $T_1$ ; for the model  $E(Y_t) = \beta_0 + \beta_1 t$ ,  $E_1 = \hat{\beta}_0$  and  $T_1 = \hat{\beta}_1$ .

$$E_t = wY_t + (1 - w)(E_{t-1} + T_{t-1})$$

$$T_t = v(E_t - E_{t-1}) + (1 - v)T_{t-1}$$

[Note:  $E_1$  and  $T_1$  are not defined.]

## EXAMPLE 13.6

### Applying the Holt-Winters Method to Annual Sales Data

**Problem** The yearly sales data for a firm's first 35 years of operation are given in Table 13.7. Calculate the Holt-Winters exponential smoothing and trend components for this time series using  $w = .7$  and  $v = .5$ . Show the data and the exponential smoothing component  $E_t$  on the same graph.



#### SALES35

**TABLE 13.7 A Firm's Yearly Sales Revenue (thousands of dollars)**

YEAR	SALES REVENUE	YEAR	SALES REVENUE	YEAR	SALES REVENUE
$t$	$Y_t$	$t$	$Y_t$	$t$	$Y_t$
1	4.8	13	48.4	25	100.3
2	4.0	14	61.6	26	111.7
3	5.5	15	65.6	27	108.2
4	15.6	16	71.4	28	115.5
5	23.1	17	83.4	29	119.2
6	23.3	18	93.6	30	125.2
7	31.4	19	94.2	31	136.3
8	46.0	20	85.4	32	146.8
9	46.1	21	86.2	33	146.1
10	41.9	22	89.9	34	151.4
11	45.5	23	89.2	35	150.9
12	53.5	24	99.1		

**Solution** Rather than perform the Holt-Winters calculations by hand, we used MINITAB to generate the values of  $E_t$  and  $T_t$  for the annual series. All the  $E_t$  and  $T_t$  values are given on the MINITAB printout, Figure 13.10.\* A graph of  $Y_t$  and  $E_t$  is shown in Figure 13.11. Note that the trend component  $T_t$  measures the general upward trend in  $Y_t$ .

**Look Back** The choice of  $v = .5$  gives equal weight to the most recent trend and to past trends in the sales of the firm. The result is that the exponential smoothing component  $E_t$  provides a smooth, upward-trending description of the firm's sales.

#### Now Work Exercise 13.24

Our objective is to use the Holt-Winters exponentially smoothed series to forecast the future values of the time series. For the one-step-ahead forecast, this is accomplished by adding the most recent exponentially smoothed component to the most recent trend component—that is, the forecast at time  $(t + 1)$ , given observed values up to time  $t$ , is

$$F_{t+1} = E_t + T_t$$

The idea is that we are constructing the forecast by combining the most recent smoothed estimate,  $E_t$ , with the estimate of the expected increase (or decrease) attributable to trend,  $T_t$ .

The forecast for two steps ahead is similar, except that we add estimated trend for two periods:

$$F_{t+2} = E_t + 2T_t$$

\*MINITAB uses simple linear regression to calculate the initial smoothed and trend values. See the footnote at the bottom of page 13-24.



↓	C1	C2	C3	C4
	T	SALES	SMOOTH	TREND
1	1	4.8	4.769	4.33163
2	2	4.0	5.530	2.54636
3	3	5.5	6.273	1.64455
4	4	15.6	13.295	4.33342
5	5	23.1	21.459	6.24838
6	6	23.3	24.622	4.70594
7	7	31.4	30.778	5.43113
8	8	46.0	43.063	8.85779
9	9	46.1	47.846	6.82056
10	10	41.9	45.730	2.35220
11	11	45.5	46.275	1.44842
12	12	53.5	51.767	3.47034
13	13	48.4	50.451	1.07730
14	14	61.6	58.579	4.60233
15	15	65.6	64.874	5.44902
16	16	71.4	71.077	5.82587
17	17	83.4	81.451	8.09987
18	18	93.6	92.385	9.51712
19	19	94.2	96.511	6.82130
20	20	85.4	90.780	0.54510
21	21	86.2	87.737	-1.24855
22	22	89.9	88.877	-0.05465
23	23	89.2	89.087	0.07765
24	24	99.1	96.119	3.55516
25	25	100.3	100.112	3.77411
26	26	111.7	109.356	6.50885
27	27	108.2	110.499	3.82618
28	28	115.5	115.148	4.23721
29	29	119.2	119.255	4.17250
30	30	125.2	124.668	4.79271
31	31	136.3	134.248	7.18632
32	32	146.8	145.190	9.06420
33	33	146.1	148.546	6.21009
34	34	151.4	152.407	5.03533
35	35	150.9	152.863	2.74553

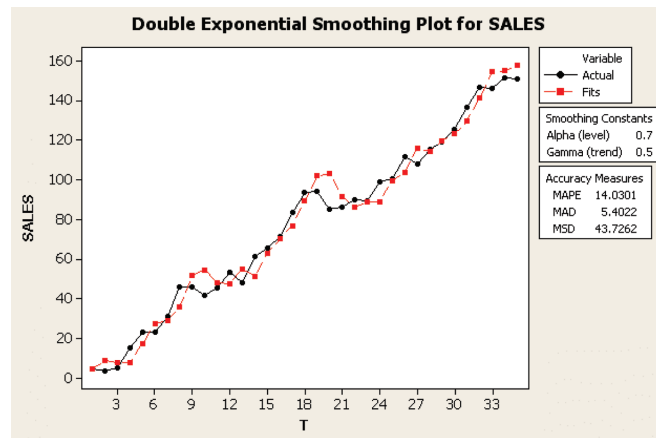
**FIGURE 13.10**  
MINITAB worksheet with  
Holt-Winters exponentially  
smoothed ( $w = .7$  and  $v = .5$ )  
sales data

Similarly, for the  $k$ -step-ahead forecast, we add the estimated increase (or decrease) in trend over  $k$  periods:

$$F_{t+k} = E_t + kT_t$$

The Holt-Winters forecasting methodology is summarized in the box.

**FIGURE 13.11**  
MINITAB graph of Holt-Winters exponentially smoothed ( $w = .7$  and  $v = 5$ ) sales data



### Holt-Winters Forecasting

1. Calculate the exponentially smoothed and trend components,  $E_t$  and  $T_t$ , for each observed value of  $Y_t (t \geq 2)$  using the formulas given in the previous box.
2. Calculate the one-step-ahead forecast using

$$F_{t+1} = E_t + T_t$$

3. Calculate the  $k$ -step-ahead forecast using

$$F_{t+k} = E_t + kT_t$$

### EXAMPLE 13.7

#### Obtaining Holt-Winters Forecasts for Annual Sales

**Problem** Refer to Example 13.6 and Figure 13.10, which lists the firm's 35 yearly sales figures, along with the Holt-Winters components using  $w = .7$  and  $v = .5$ . Use the Holt-Winters forecasting technique to forecast the firm's annual sales in years 36–40.

**Solution** From Figure 13.10, the smoothed and trend values for the last year are  $E_{35} = 152.86$  and  $T_{35} = 2.75$ . For year 36 we calculate

$$F_{36} = E_{35} + T_{35} = 152.86 + 2.75 = 155.61$$

The forecast 2 years ahead is

$$F_{37} = E_{35} + 2T_{35} = 152.86 + 2(2.75) = 158.36$$

For years 38–40 we find

$$F_{38} = 152.86 + 3(2.75) = 161.11$$

$$F_{39} = 152.86 + 4(2.75) = 163.86$$

$$F_{40} = 152.86 + 5(2.75) = 166.61$$

**Look Back** Note that the forecast values increase from year 36 to year 40. This upward trend in the forecast is a result of the Holt-Winters estimated trend component.

Now Work Exercise 13.25b

The selection of  $w = .7$  and  $v = .5$  as the smoothing and trend weights for the sales forecasts in Example 13.7 was based on the objectives of assigning more weight to recent series values in the exponentially smoothed component, and of assigning equal weights to the recent and past trend estimates. However, you may want to try several different combinations of weights when using the Holt-Winters forecasting model so that you can assess the sensitivity of the forecasts to the choice of weights. Experience with the particular time series and Holt-Winters forecasts will help in the selection of  $w$  and  $v$  in a practical application.

## Exercises 13.23–13.31

### Learning the Mechanics

**13.23** How does the choice of the smoothing constant  $w$  impact an exponentially smoothed forecast?

**13.24** Refer to Exercise 13.4 (p. 13-12). The table with the prices for product A for the four quarters of last year is reproduced below. The Holt-Winters smoothing method with  $w = .2$  and  $v = .6$  was applied to the data.

Quarter	A	Smoothed Value	Trend
1	3.25	—	—
2	3.50	3.50	0.25
3	3.90	3.78	?
4	4.25	?	0.29

- Find the missing trend value for Quarter 3.
- Find the missing smoothed value for Quarter 4.
- Give the Holt-Winters forecast for the price in Quarter 5.

- Calculate and plot the exponentially smoothed series for the CPI using a smoothing constant of  $w = .4$ . Use the exponentially smoothed values to forecast the CPI in 2006.
- Use the Holt-Winters forecasting model with trend to forecast the CPI in 2006. Use smoothing constants  $w = .4$  and  $v = .5$ .

#### CPI

Year	CPI	Year	CPI
1990	125.8	1998	163.0
1991	129.1	1999	166.6
1992	132.8	2000	171.5
1993	136.8	2001	177.1
1994	147.8	2002	179.9
1995	152.4	2003	184.0
1996	156.9	2004	188.9
1997	160.5	2005	195.3

Source: *Survey of Current Business*, U.S. Department of Commerce, Bureau of Economic Analysis.

### Applying the Concepts—Basic

#### USBEER

**13.25 Annual U.S. beer production.** Refer to Exercise 13.7 (p. 13-12) and the data on U.S. beer production (in millions of barrels) for the years 1980–2005.

- Use the 1980–2002 values to forecast the 2003–2005 production, using simple exponential smoothing with  $w = .3$ . With  $w = .7$ .
- Use the Holt-Winters model with  $w = .7$  and  $v = .3$  to forecast the 2003–2005 production. Repeat with  $w = .3$  and  $v = .7$ .

#### QTRHOUSE

**13.26 Quarterly single-family housing starts.** Refer to the quarterly housing start series, Exercise 13.8 (p. 13-13). Suppose you want to forecast the number of new housing starts in 2006 using data for 2004 and 2005.

- Calculate the exponentially smoothed values for 2004 and 2005 using  $w = .6$ .
- Plot the housing-starts series and the exponentially smoothed series on the same graph.
- Use the exponentially smoothed data from 2004–2005 to forecast the quarterly number of housing starts in 2006.

**13.27 Consumer Price Index.** The Consumer Price Index (CPI) measures the increase (or decrease) in the prices of goods and services relative to a base year. The CPI for the years 1990–2005 (using 1984 as a base period) is shown in the next table.

- Graph the time series. Do you detect a long-term trend?

#### OPECOIL

**13.28 OPEC crude oil imports.** Refer to the annual OPEC oil import data, Exercise 13.21. (p. 13-19).

- Use the exponentially smoothed ( $w = .9$ ) series you constructed in Exercise 13.21a to forecast OPEC oil imports in 2006.
- Forecast OPEC oil imports in 2006 using the Holt-Winters forecasting model with smoothing constants  $w = .3$  and  $v = .8$ .
- Annual OPEC crude oil imports in 2006 totaled 2,100 million barrels. Calculate the errors of the forecasts, parts **a** and **b**. Which method yields the smallest forecast error?

### Applying the Concepts—Intermediate

#### SP500

**13.29 S&P 500 Stock Index.** Refer to the quarterly Standard & Poor's 500 stock market index, Exercise 13.22 (p. 13-19).

- Use  $w = .7$  to smooth the series from 1999 through 2005. Then forecast the first three quarterly values in 2006 using *only* the information through the fourth quarter of 2005.
- Repeat part **a** using  $w = .3$ .

**13.30 S&P 500 Stock Index (cont'd).** Refer to Exercise 13.29. Suppose you want to use only the 2003–2005 S&P values to forecast the quarterly 2006 values. Calculate the forecasts using the Holt-Winters model with  $w = .3$  and  $v = .5$ . Repeat with  $w = .7$  and  $v = .5$ .

**13.31 Monthly gold prices.** The fluctuation of gold prices is a reflection of the strength or weakness of the U.S. dollar. The table below shows monthly gold prices from January 1998 to December 2005.

- a. Use exponential smoothing with  $w = .5$  to calculate monthly smoothed values from 1998 to 2004. Then forecast the monthly gold prices for 2005.

- b. Calculate 12 one-step-ahead forecasts for 2005 by updating the exponentially smoothed values with each month's actual value, and then forecasting the next month's value.
- c. Repeat parts a–b using the Holt-Winters technique with  $w = .5$  and  $v = .5$ .

 **GOLDMOM**

Month	1998	1999	2000	2001	2002	2003	2004	2005
Jan	304.9	285.4	284.3	265.5	281.7	356.9	414.0	424.2
Feb	297.4	287.1	299.9	261.9	295.5	359.0	405.3	423.4
Mar	301.0	279.5	286.4	263.0	294.0	340.6	406.7	434.2
Apr	310.7	286.6	279.9	260.5	302.7	328.2	403.0	428.9
May	293.6	269.0	275.3	272.4	314.5	355.7	383.4	421.9
Jun	296.3	262.6	285.7	270.2	321.2	356.5	392.0	430.7
Jul	288.9	255.6	281.6	267.5	313.3	351.0	398.1	424.5
Aug	273.4	254.8	274.5	272.4	310.3	359.8	400.5	437.9
Sep	293.9	307.5	273.7	283.4	319.2	378.9	405.3	456.0
Oct	294.0	299.1	270.0	283.1	316.6	378.9	420.5	469.9
Nov	294.7	291.4	266.0	276.2	319.2	389.9	439.4	476.7
Dec	287.8	290.3	271.5	275.9	333.4	407.6	441.7	509.8

Sources: Standard & Poor's Statistics, 2006; kitco.com Current Statistics, 2006.

## 13.6 Measuring Forecast Accuracy: MAD and RMSE

As demonstrated in Example 13.5, forecast error (i.e., the difference between the actual time series value and its forecast) can be used to evaluate the accuracy of the forecast. Knowledge of a forecast's accuracy aids in the selection of both the forecasting methodology to be utilized and the parameters of the forecast formula (e.g., the weights in the exponentially smoothed or Holt-Winters forecasts). Three popular measures of forecast accuracy, both based on forecast errors, are the *mean absolute deviation (MAD)*, the *mean absolute percentage error (MAPE)*, and the *root mean squared error (RMSE)* of the forecasts. Their formulas are given in the box.

### Measures of Forecast Accuracy for $m$ Forecasts

#### 1. Mean absolute deviation (MAD)

$$\text{MAD} = \frac{\sum_{t=1}^m |Y_t - F_t|}{m}$$

#### 2. Mean absolute percentage error (MAPE)

$$\text{MAPE} = \frac{\sum_{t=1}^m \left| \frac{Y_t - F_t}{Y_t} \right|}{m} \times 100$$

#### 3. Root mean squared error (RMSE)

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^m (Y_t - F_t)^2}{m}}$$

Note that all three measures require one or more actual values of the time series against which to compare the forecasts. Thus, we can either wait several time periods until the observed values are available, or we can hold out several of the values at the end of the time series, not using them to model the time series, but saving them for evaluating the forecasts obtained from the model.

**EXAMPLE 13.8**  
Comparing Measures of Forecast Accuracy

**Problem** Refer to the annual sales data of Examples 13.6 and 13.7. In Example 13.7 we used the Holt-Winters model with  $w = .7$  and  $v = .5$  to forecast annual sales for years 36–40. Consider two alternative forecasting models: exponential smoothing with  $w = .3$  and exponential smoothing with  $w = .7$ . MINITAB was used to obtain the forecasts for these alternative models. The MINITAB printouts shown in Figures 13.12 (a)–(c) give the forecasts (highlighted) for all three models. Suppose the actual sales values (in thousands of dollars) for years 36–40 are 150.2, 161.7, 159.3, 168.5, and 170.4, respectively. Find measures of forecast accuracy (MAD, MAPE, and RMSE) for each of the three forecasting models and use this information to evaluate the models.

**Solution** For ease of notation, we will number the forecasting models as follows.

*Model 1:* Exponential smoothing ( $w = .3$ )

*Model 2:* Exponential smoothing ( $w = .7$ )

*Model 3:* Holt-Winters ( $w = .7, v = .5$ )

The forecasts for the three models as well as forecast errors are listed in Table 13.8. These forecast errors are used to find the MAD, MAPE, and RMSE measures of forecast accuracy for each of the three models.

**Single Exponential Smoothing for SALES**

Data SALES  
Length 35

Smoothing Constant

Alpha 0.3

Accuracy Measures

MAPE 21.805  
MAD 13.179  
MSD 223.953

Forecasts

Period	Forecast	Lower	Upper
36	142.696	110.408	174.983
37	142.696	110.408	174.983
38	142.696	110.408	174.983
39	142.696	110.408	174.983
40	142.696	110.408	174.983

**FIGURE 13.12a**

MINITAB forecasts of annual sales—exponential smoothing model with  $w = .3$

**Single Exponential Smoothing for SALES**

Data SALES  
Length 35

Smoothing Constant

Alpha 0.7

Accuracy Measures

MAPE 12.8418  
MAD 6.7780  
MSD 68.9944

Forecasts

Period	Forecast	Lower	Upper
36	150.474	133.868	167.080
37	150.474	133.868	167.080
38	150.474	133.868	167.080
39	150.474	133.868	167.080
40	150.474	133.868	167.080

**FIGURE 13.12b**

MINITAB forecasts of annual sales—exponential smoothing model with  $w = .7$

## Double Exponential Smoothing for SALES

Data SALES  
Length 35

### Smoothing Constants

Alpha (level) 0.7  
Gamma (trend) 0.5

### Accuracy Measures

MAPE 14.0301  
MAD 5.4022  
MSD 43.7262

### Forecasts

Period	Forecast	Lower	Upper
36	155.608	142.373	168.843
37	158.354	141.644	175.063
38	161.099	140.614	181.585
39	163.845	139.421	188.269
40	166.590	138.133	195.048

**FIGURE 13.12c**

MINITAB forecasts of annual sales—Holt-Winters model  
with  $w = .7$  and  $v = .5$

**TABLE 13.8** Forecasts and Forecast Errors for Three Models of Annual Sales

Year	Actual Sales	MODEL 1		MODEL 2		MODEL 3	
		Forecast	Error	Forecast	Error	Forecast	Error
36	150.2	142.7	7.5	150.5	-0.3	155.6	-5.4
37	161.7	142.7	19.0	150.5	11.2	158.4	3.3
38	159.3	142.7	16.6	150.5	8.8	161.1	-1.8
39	168.5	142.7	25.8	150.5	18.0	163.8	4.7
40	170.4	142.7	27.7	150.5	19.9	166.6	3.8

Model 1:

$$\text{MAD}_1 = \frac{|7.5| + |19.0| + |16.6| + |25.8| + |27.7|}{5} = 19.32$$

$$\text{MAPE}_1 = \left\{ \frac{\left| \frac{7.5}{150.2} \right| + \left| \frac{19.0}{161.7} \right| + \left| \frac{16.6}{159.3} \right| + \left| \frac{25.8}{168.5} \right| + \left| \frac{27.7}{170.4} \right|}{5} \right\} 100 = 11.75$$

$$\text{RMSE}_1 = \sqrt{\frac{(7.5)^2 + (19.0)^2 + (16.6)^2 + (25.8)^2 + (27.7)^2}{5}} = 20.62$$

Model 2:

$$\text{MAD}_2 = \frac{|-.3| + |11.2| + |8.8| + |18.0| + |19.9|}{5} = 11.64$$

$$\text{MAPE}_2 = \left\{ \frac{\left| \frac{-.3}{150.2} \right| + \left| \frac{11.2}{161.7} \right| + \left| \frac{8.8}{159.3} \right| + \left| \frac{18.0}{168.5} \right| + \left| \frac{19.9}{170.4} \right|}{5} \right\} 100 = 7.00$$

$$\text{RMSE}_2 = \sqrt{\frac{(-.3)^2 + (11.2)^2 + (8.8)^2 + (18.0)^2 + (19.9)^2}{5}} = 13.59$$

Model 3:

$$\text{MAD}_3 = \frac{|-5.4| + |3.3| + |-1.8| + |4.7| + |3.8|}{5} = 3.80$$

$$\text{MAPE}_3 = \left\{ \frac{\left| \frac{-5.4}{150.2} \right| + \left| \frac{3.3}{161.7} \right| + \left| \frac{-1.8}{159.3} \right| + \left| \frac{4.7}{168.5} \right| + \left| \frac{3.8}{170.4} \right|}{5} \right\} 100 = 2.36$$

$$\text{RMSE}_3 = \sqrt{\frac{(-5.4)^2 + (3.3)^2 + (-1.8)^2 + (4.7)^2 + (3.8)^2}{5}} = 4.00$$

The model with the smallest mean absolute deviation (MAD) is Model 3. Similarly, Model 3 has the smallest mean absolute percentage error (MAPE) and the smallest root mean square error (RMSE). Of the three forecasting models, then, Model 3 (the Holt-Winters method with  $w = .7$  and  $v = .5$ ) yields the most accurate predictions of annual sales.

**Look Back** We expect the Holt-Winters method to yield more accurate forecasts for annual sales since it explicitly accounts for trend in the sales data. The exponential smoothing forecasts do not account for any increasing or decreasing trends in the data; hence, they are the same for all five forecasted years. The accuracy of all three forecasting methods, however, will decrease the farther we forecast into the future.

### Now Work Exercise 13.32

[Note: Most statistical software packages will automatically compute the values of MAPE, MAD, and RMSE (also called the *mean squared deviation, MSD*) for all  $n$  observations in the data set. For example, these statistics are shown in the middle of the MINITAB printouts, Figures 13.12(a)–(c).]

Criteria such as MAPE, MAD, and RMSE for assessing forecast accuracy require special care in interpretation. The number of time periods included in the evaluation is critical to the decision about which forecasting model is preferred. The choice depends on how many time periods ahead the analyst plans to forecast. With  $N$  time periods in your data, a good rule of thumb is to forecast ahead no more than  $N/2$  time periods. Remember, however, that long-term forecasts are generally less accurate than short-term forecasts.

We conclude this section with a comment. A major disadvantage of forecasting with smoothing techniques (exponential smoothing or the Holt-Winters model) is that no measure of the forecast error (or reliability) is known *prior* to observing the future value. Although forecast errors can be calculated *after* the future values of the time series have been observed (as in Example 13.8), we prefer to have some measure of the accuracy of the forecast *before* the actual values are observed. One option is to compute forecasts and forecast errors for all  $n$  observations in the data set and use these “past” forecast errors to estimate the standard deviation of all forecast errors (i.e., the *standard error of the forecast*). A rough estimate of this standard error is the value of RMSE, and an approximate 95% prediction interval for any future forecast is

$$F_t \pm 2(\text{RMSE})$$

[An interval like this is shown at the bottom of the MINITAB printouts, Figures 13.12(a)–(c).] However, since the theoretical distributional properties of the forecast errors with smoothing methods are unknown, many analysts regard smoothing methods as descriptive procedures rather than as inferential ones.

In the preceding chapters, we learned that predictions with inferential regression models are accompanied by well-known measures of reliability. The standard errors of the predicted values allow us to construct 95% prediction intervals. We discuss an inferential time series forecasting model in the next section.

## STATISTICS IN ACTION REVISITED

### Forecasting Coldex Sales with Exponential Smoothing



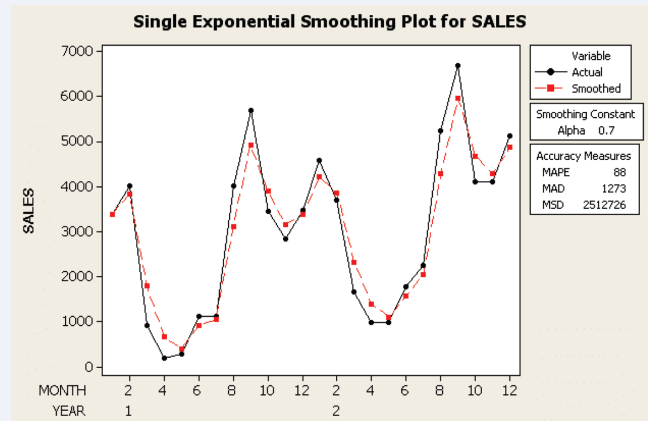
Recall that a pharmaceutical company hired consultants at Rutgers University to forecast monthly sales of a new brand of cold medicine called Coldex. The company provided monthly data on Coldex sales for the first 2 years of the product's life

and desires forecasts of sales for the first 3 months of the third year. (The data are saved in the **COLDEX** file.) One forecasting model considered by the consultants was an exponential smoothing model with a smoothing constant of  $w = .7$ . MINITAB was used to find the smoothed values of the monthly series. The MINITAB plot of both the actual monthly sales and smoothed sales values is shown in Figure SIA13.1, followed by the exponentially smoothed forecasts in Figure SIA13.2.

The exponentially smoothed sales forecast for each of the first 3 months of year 3 is the smoothed value for the last month of the series (month 24). This value, highlighted on Figure SIA13.2, is 4,870 thousand dollars. MINITAB also gives approximate 95% confidence bounds around the forecast. The interval (highlighted on the printout) is (1,750, 7,989). Thus, we are (approximately) 95% confident that the actual sales for the month will be between 1,750 and 7,989 thousand dollars. This wide interval was deemed unusable by the pharmaceutical company; consequently, the consultants searched for a better forecasting model. One of these models is presented in the next Statistics in Action Revisited (p. 13-37).

**Figure SIA13.2**

MINITAB forecasts of monthly Coldex sales using exponential smoothing ( $w = .7$ )



**Figure SIA13.1**

MINITAB plot of monthly Coldex sales with exponentially smoothed values ( $w = .7$ )

### Single Exponential Smoothing for SALES

Data SALES  
Length 24

Smoothing Constant

Alpha 0.7

Accuracy Measures

MAPE 88  
MAD 1273  
MSD 2512726

Forecasts

Period	Forecast	Lower	Upper
25	4869.91	1750.49	7989.34
26	4869.91	1750.49	7989.34
27	4869.91	1750.49	7989.34



## Exercises 13.32–13.37

### Applying the Concepts—Basic

#### USBEER

**13.32 Annual U.S. beer production.** Refer to the beer production forecasts, Exercise 13.25 (p. 13-28). In part **a** you obtained forecasts of 2003–2005 beer production using exponential smoothing with  $w = .3$  and  $w = .7$ .

- Calculate the forecast errors for the  $w = .3$  exponentially smoothed forecasts.
- Calculate the forecast errors for the  $w = .7$  exponentially smoothed forecasts.
- Calculate MAD, MAPE, and RMSE for the exponential smoothing forecasts using  $w = .3$ .
- Calculate MAD, MAPE, and RMSE for the exponential smoothing forecasts using  $w = .7$ .
- Refer to parts **c** and **d**. Which forecast method do you recommend?

**13.33 Annual U.S. beer production (cont'd).** Refer to the beer production forecasts, Exercise 13.25 (p. 13-28). In part **b** you obtained forecasts of 2003–2005 beer production using the Holt-Winters model with ( $w = .3, v = .7$ ) and ( $w = .7, v = .3$ ).

- Calculate the forecast errors for the  $w = .3, v = .7$  Holt-Winters forecasts.
- Calculate the forecast errors for the  $w = .7, v = .3$  Holt-Winters forecasts.
- Calculate MAD, MAPE, and RMSE for the  $w = .3, v = .7$  Holt-Winters forecasts.
- Calculate MAD, MAPE, and RMSE for the  $w = .7, v = .3$  Holt-Winters forecasts.
- Refer to parts **c** and **d**. Which forecast method do you recommend?

#### SP500

**13.34 S&P 500 Stock Index.** Refer to your exponential smoothing forecasts of the quarterly S&P 500 for 2006, Exercise 13.29 (p. 13-28).

- Calculate MAD, MAPE, and RMSE for the forecasts with  $w = .7$ .
- Calculate MAD, MAPE, and RMSE for the forecasts with  $w = .3$ .
- Compare MAD, MAPE, and RMSE for the two simple exponential smoothing forecast models. Which model leads to more accurate forecasts?

**13.35 S&P 500 Stock Index (cont'd).** Refer to your Holt-Winters forecasts of the quarterly S&P 500 for 2006, Exercise 13.30 (p. 13-28).

- Calculate MAD, MAPE, and RMSE for the forecasts with  $w = .3$  and  $v = .5$ .

- Calculate MAD, MAPE, and RMSE for the forecasts with  $w = .7$  and  $v = .5$ .
- Compare MAD, MAPE, and RMSE for the two Holt-Winters forecasts models. Which model leads to more accurate forecasts?

### Applying the Concepts—Intermediate

#### GOLDMON

**13.36 Monthly gold prices.** Refer to Exercise 13.31 (p. 13-29). Two models were used to forecast the monthly 2005 gold prices: an exponential smoothing model with  $w = .5$  and a Holt-Winters model with  $w = .5$  and  $v = .5$ .

- Use MAD, MAPE, and RMSE criteria to evaluate the two models' accuracy for forecasting the monthly 2005 values using the 1998–2004 data.
- Use the MAD, MAPE, and RMSE criteria to evaluate the two models' accuracy when making the 12 one-step-ahead forecasts, updating the models with each month's actual value before forecasting the next month's value.

**13.37 U.S. school enrollments.** The next table reports annual U.S. school enrollment (in thousands) for the period 1980–2005.

#### SCHOOLENROLL

Year	Enrollment	Year	Enrollment
1980	58,305	1993	63,241
1981	57,916	1994	63,986
1982	57,951	1995	64,764
1983	57,432	1996	65,743
1984	57,150	1997	66,470
1985	57,226	1998	66,983
1986	57,709	1999	67,667
1987	58,254	2000	68,146
1988	58,485	2001	69,936
1989	59,436	2002	71,215
1990	60,267	2003	71,442
1991	61,605	2004	71,688
1992	62,686	2005	72,075

Source: U.S. Census Bureau. *Statistical Abstract of the United States*, 2006.

- Use the 1980 to 2000 enrollments and simple exponential smoothing to forecast the 2001–2005 school enrollments. Use  $w = .8$ .
- Use the Holt-Winters model with  $w = .8$  and  $v = .7$  to forecast the 2001–2005 enrollments.
- Apply the MAD, MAPE, and RMSE criteria to evaluate the two forecasting models of parts **a** and **b**. Which model is better? Why?

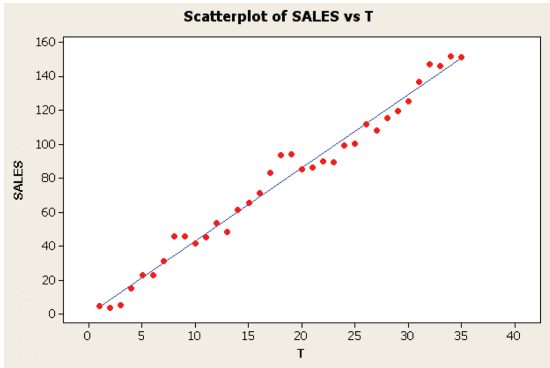
## 13.7 Forecasting Trends: Simple Linear Regression

Perhaps the simplest **inferential forecasting model** is one with which you are familiar: the simple linear regression model. A straight-line model is used to relate the time series,  $Y_t$ , to time,  $t$ , and the least squares line is used to forecast future values of  $Y_t$ .

Suppose a firm is interested in forecasting its sales revenues for each of the next 5 years. To make such forecasts and assess their reliability, a time series model must be

constructed. Refer again to the yearly sales data for a firm’s 35 years of operation given in Table 13.7 (p. 13-25). A MINITAB plot of the data (Figure 13.13) reveals a linearly increasing trend, so the model

$$E(Y_t) = \beta_0 + \beta_1 t$$



**FIGURE 13.13**  
MINITAB scatterplot of annual sales with least squares line

seems plausible for the secular trend. We fit the model to the data using SPSS; the resulting printout is shown in Figure 13.14. The least squares model (highlighted on the printout) is:

$$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 t = .402 + 4.296t \quad \text{with} \quad R^2 = .980$$

(This least squares line is shown on Figure 13.13.)

We can now forecast sales for years 36–40. The forecasts of sales and the corresponding 95% prediction intervals are shown (highlighted) on the SPSS printout Figure 13.15. For example, for  $t = 36$ , we have

$$\hat{Y}_{36} = 155.0$$

with the 95% prediction interval (141.3, 168.8). Similarly, we can obtain the forecasts and prediction intervals for years 37–40. Although it is not easily perceptible in the figure, the prediction intervals widen as we attempt to forecast farther into the future. This agrees with the intuitive notion that short-term forecasts should be more reliable than long-term forecasts.

**Model Summary<sup>b</sup>**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.990 <sup>a</sup>	.980	.979	6.3852

a. Predictors: (Constant), T  
b. Dependent Variable: SALES

**ANOVA<sup>b</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	65875.21	1	65875.208	1615.724	.000 <sup>a</sup>
	Residual	1345.454	33	40.771		
	Total	67220.66	34			

a. Predictors: (Constant), T  
b. Dependent Variable: SALES

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.402	2.206		.182	.857
	T	4.296	.107	.990	40.196	.000

a. Dependent Variable: SALES

**FIGURE 13.14**  
SPSS printout of least squares regression of annual sales

T	SALES	Unstandardized Predicted Value	95% L CI for SALES individual	95% U CI for SALES individual
1	4.8	4.69714	-8.98666	18.38094
2	4.0	8.99277	-4.63390	22.61945
3	5.5	13.28840	-.28440	26.86120
4	15.6	17.58403	4.06183	31.10624
5	23.1	21.87966	8.40474	35.35459
6	23.3	26.17529	12.74429	39.60630
7	31.4	30.47092	17.08046	43.86139
8	46.0	34.76655	21.41321	48.11990
9	46.1	39.06218	25.74251	52.38186
10	41.9	43.35782	30.06834	56.64729
11	45.5	47.65345	34.39068	60.91621
12	53.5	51.94908	38.70950	65.18865
13	48.4	56.24471	43.02478	69.46463
14	61.6	60.54034	47.33651	73.74416
15	65.6	64.83597	51.64468	78.02725
16	71.4	69.13160	55.94927	82.31392
17	83.4	73.42723	60.25028	86.60417
18	93.6	77.72286	64.54771	90.89801
19	94.2	82.01849	68.84154	95.19543
20	85.4	86.31412	73.13179	99.49644
21	86.2	90.60975	77.41846	103.80103
22	89.9	94.90538	81.70156	108.10920
23	89.2	99.20101	85.98108	112.42093
24	99.1	103.49664	90.25706	116.73622
25	100.3	107.79227	94.52950	121.05503
26	111.7	112.08790	98.79843	125.37737
27	108.2	116.38353	103.06386	129.70320
28	115.5	120.67916	107.32581	134.03250
29	119.2	124.97479	111.58432	138.36525
30	125.2	129.27042	115.83942	142.70142
31	136.3	133.56605	120.09112	147.04098
32	146.8	137.86168	124.33948	151.38388
33	146.1	142.15731	128.58451	155.73011
34	151.4	146.45294	132.82626	160.07962
35	150.9	150.74857	137.06477	164.43237
36	.	155.04420	141.30008	168.78832
37	.	159.33983	145.53223	173.14744
38	.	163.63546	149.76126	177.50967
39	.	167.93109	153.98721	181.87497
40	.	172.22672	158.21014	186.24330

**FIGURE 13.15**  
SPSS printout with 95%  
prediction intervals for annual  
sales

There are two problems associated with forecasting time series using a least squares model.

### Problem 1

We are using the least squares model to forecast values outside the region of observation of the independent variable,  $t$ —that is, we are forecasting for values of  $t$  between 36 and 40, but the observed sales are for  $t$  values between 1 and 35. As we noted in Chapters 10 and 11, it is extremely risky to use a least squares regression model for prediction outside the experimental region.

Problem 1 obviously cannot be avoided. Since forecasting always involves predictions about the future values of a time series, some or all of the independent variables will probably be outside the region of observation on which the model was developed. It is important that the forecaster recognize the dangers of this type of prediction. If underlying conditions change drastically after the model is estimated (e.g., if federal price

controls are imposed on the firm's products during the 36th year of operation), the forecasts and their confidence intervals are probably useless.

### Problem 2

Although the straight-line model may adequately describe the secular trend of the sales, we have not attempted to build any cyclical effects into the model. Thus, the effect of inflationary and recessionary periods will be to increase the error of the forecasts because the model does not anticipate such periods.

Fortunately, the forecaster often has some degree of control over problem 2, as we demonstrate in the remainder of the chapter.

In forming the prediction intervals for the forecasts, we made the standard regression assumptions (Chapters 10 and 11) about the random error component of the model. We assumed the errors have mean 0, constant variance, and normal probability distributions, and are *independent*. The latter assumption is dubious in time series models, especially in the presence of short-term trends. Often, if a year's value lies above the secular trend line, the next year's value has a tendency to be above the line also—that is, the errors tend to be correlated (see Figure 13.13).

We discuss how to deal with correlated errors in Section 13.9. For now, we can characterize the simple linear regression forecasting method as useful for discerning secular trends, but probably too simplistic for most time series. And, as with all forecasting methods, the simple linear regression forecasts should be applied only over the short term.

## STATISTICS IN ACTION REVISITED



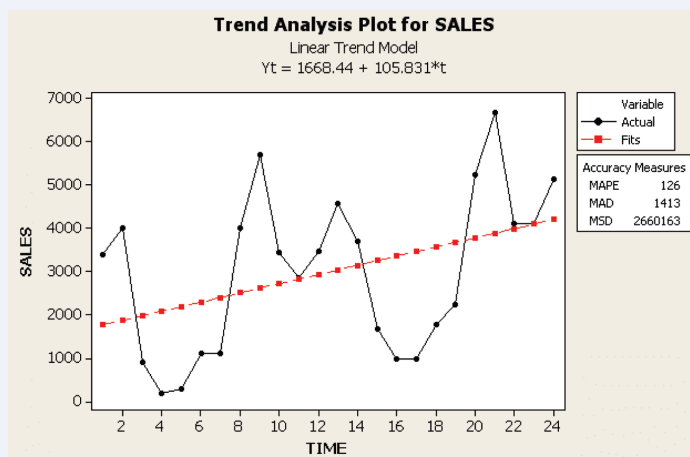
### Forecasting Coldex Sales with Simple Linear Regression

A second model considered by consultants to forecast monthly sales of a new cold medicine (Coldex) was a simple linear regression model with time ( $t$ ) as the independent variable, where  $t = 1, 2, 3, \dots, 24$ . The MINITAB graph of the least squares line is shown in Figure

SIA13.3, followed by the simple linear regression printout in Figure SIA13.4.

Note that the  $p$ -value for testing the slope coefficient is .047. Thus, at  $\alpha = .05$ , the model is statistically useful for predicting monthly sales. However, the coefficient of determination is low ( $R^2 = .168$ ); only about 17% of the sample variation in monthly sales can be explained by the linear time trend.

The simple linear forecasts for each of the first 3 months of year 3 are shown on the MINITAB worksheet, Figure SIA13.5. MINITAB also gives a 95% prediction interval for each forecast. The intervals for months 25, 26, and 27 are (480.5, 8147.9), (549.7, 8290.4), and (616.5, 8435.3), respectively. For the first month of year 3 (i.e., month 25), we are 95% confident



**Figure SIA13.3**  
MINITAB plot of least squares line for forecasting monthly Coldex sales

*Continued*

**Regression Analysis: SALES versus TIME**

The regression equation is  
 $SALES = 1668 + 106 TIME$

Predictor	Coef	SE Coef	T	P
Constant	1668.4	717.8	2.32	0.030
TIME	105.83	50.23	2.11	0.047

S = 1703.52    R-Sq = 16.8%    R-Sq(adj) = 13.0%

**Analysis of Variance**

Source	DF	SS	MS	F	P
Regression	1	12880305	12880305	4.44	0.047
Residual Error	22	63843919	2901996		
Total	23	76724223			

**Figure SIA13.4**  
 MINITAB simple linear regression printout for the linear trend forecasting model

↓	C1	C2	C3	C4	C5	C6
	YEAR	MONTH	TIME	FORECAST	PL95LOW	PL95UPP
1	3	1	25	4314.22	480.528	8147.92
2	3	2	26	4420.08	549.731	8290.38
3	3	3	27	4525.89	616.500	8435.27

**Figure SIA13.5**  
 MINITAB worksheet with simple linear regression forecasts of monthly sales

that the actual sales will fall between 480.5 and 8,147.9 thousand dollars. Similar interpretations are made for the other two forecasts.

As with the exponential smoothing model, these intervals were too wide to be of practical use by the pharmaceutical company. How can the forecasting model be improved? Examine Figure SIA13.3 and

note the cyclical trends in the monthly sales data. Neither the exponential smoothing model nor the linear regression model account for this cyclical variation. A forecasting model is needed that explicitly accounts for such trends. Such a model is presented in the next Statistics in Action Revisited (p. 13-42).

## 13.8 Seasonal Regression Models

Many time series have distinct seasonal patterns. Retail sales are usually highest around Christmas, spring, and fall, with lulls in the winter and summer periods. Energy usage is highest in summer and winter, and lowest in spring and fall. Teenage unemployment rises in summer months when schools are not in session and falls near Christmas when many businesses hire part-time help.

Multiple regression models can be used to forecast future values of a time series with strong seasonal components. To accomplish this, the mean value of the time series,  $E(Y_t)$ , is given a mathematical form that describes both the secular trend and seasonal components of the time series. Although the **seasonal model** can assume a wide variety of mathematical forms, the use of dummy variables to describe seasonal differences is common.



## QTRPOWER

**TABLE 13.9** Quarterly Power Loads (megawatts) for a Southern Utility Company, 1995–2006

Year	Quarter	Power Load	Year	Quarter	Power Load
1995	1	68.8	2001	1	130.6
	2	65.0		2	116.8
	3	88.4		3	144.2
	4	69.0		4	123.3
1996	1	83.6	2002	1	142.3
	2	69.7		2	124.0
	3	90.2		3	146.1
	4	72.5		4	135.5
1997	1	106.8	2003	1	147.1
	2	89.2		2	119.3
	3	110.7		3	138.2
	4	91.7		4	127.6
1998	1	108.6	2004	1	143.4
	2	98.9		2	134.0
	3	120.1		3	159.6
	4	102.1		4	135.1
1999	1	113.1	2005	1	149.5
	2	94.2		2	123.3
	3	120.5		3	154.4
	4	107.4		4	139.4
2000	1	116.2	2006	1	151.6
	2	104.4		2	133.7
	3	131.7		3	154.5
	4	117.9		4	135.1

For example, consider the power load data for a Southern utility company shown in Table 13.9. Data were obtained for each quarter from 1995 through 2006. A model that combines the expected growth in usage and the seasonal component is

$$E(Y_t) = \beta_0 + \beta_1 t + \beta_2 Q_1 + \beta_3 Q_2 + \beta_4 Q_3$$

where

$t$  = Time period, ranging from  $t = 1$  for quarter 1 of 1995 to  
 $t = 48$  for quarter 4 of 2006

$$Q_1 = \begin{cases} 1 & \text{if Quarter 1} \\ 0 & \text{if Quarter 2, 3, or 4} \end{cases}$$

$$Q_2 = \begin{cases} 1 & \text{if Quarter 2} \\ 0 & \text{if Quarter 1, 3, or 4} \end{cases}$$

$$Q_3 = \begin{cases} 1 & \text{if Quarter 3} \\ 0 & \text{if Quarter 1, 2, or 4} \end{cases}$$

The MINITAB printout in Figure 13.16 shows the least squares fit of this model to the data in Table 13.9.

Note that the model appears to fit well, with  $R^2 = .914$ , indicating that the model accounts for about 91% of the sample variability in power loads over the 12-year period. The global  $F = 114.88$  ( $p$ -value = .000) strongly supports the hypothesis that the model has predictive utility. The model standard deviation of 7.86 indicates that the model predictions will usually be accurate to within approximately  $\pm 2(7.86)$ , or about  $\pm 16$  megawatts. Furthermore,  $\hat{\beta}_1 = 1.64$  indicates an estimated average growth in load

The regression equation is

$$\text{LOAD} = 70.5 + 1.64 T + 13.7 Q_1 - 3.74 Q_2 + 18.5 Q_3$$

Predictor	Coef	SE Coef	T	P
Constant	70.509	3.116	22.63	0.000
T	1.63621	0.08214	19.92	0.000
Q1	13.659	3.217	4.25	0.000
Q2	-3.736	3.212	-1.16	0.251
Q3	18.470	3.209	5.76	0.000

S = 7.85795    R-Sq = 91.4%    R-Sq(adj) = 90.6%

#### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	28375.0	7093.7	114.88	0.000
Residual Error	43	2655.1	61.7		
Total	47	31030.1			

FIGURE 13.16

MINITAB printout of least squares fit to power load time series

of 1.64 megawatts per quarter. Finally, the seasonal dummy variables have the following interpretations (refer to Chapter 11):\*

$\hat{\beta}_2 = 13.66$     Quarter 1 loads average 13.66 megawatts more than Quarter 4 loads.

$\hat{\beta}_3 = -3.74$     Quarter 2 loads average 3.74 megawatts less than Quarter 4 loads.

$\hat{\beta}_4 = 18.47$     Quarter 3 loads average 18.47 megawatts more than Quarter 4 loads.

Thus, as expected, winter and summer loads exceed spring and fall loads, with the peak occurring during the summer months.

In order to forecast the 2007 power loads, we calculate the predicted value  $\hat{Y}_k$  for  $k = 49, 50, 51$ , and  $52$ , at the same time substituting the dummy variable appropriate for each quarter. Thus, for 2007,

$$\hat{Y}_{\text{Quarter 1}} = \hat{\beta}_0 + \hat{\beta}_1(49) + \hat{\beta}_2 = 70.51 + 1.636(49) + 13.66 = 164.3$$

$$\hat{Y}_{\text{Quarter 2}} = \hat{\beta}_0 + \hat{\beta}_1(50) + \hat{\beta}_3 = 148.6$$

$$\hat{Y}_{\text{Quarter 3}} = \hat{\beta}_0 + \hat{\beta}_1(51) + \hat{\beta}_4 = 172.4$$

$$\hat{Y}_{\text{Quarter 4}} = \hat{\beta}_0 + \hat{\beta}_1(52) = 155.6$$

The predicted values and 95% prediction intervals (highlighted) are given on the MINITAB worksheet, Figure 13.17. Also shown in Figure 13.17 are the actual 2007 quarterly power loads. Notice that all 2007 power loads fall inside the forecast intervals.

#### Now Work Exercise 13.39

The seasonal model used to forecast the power loads is an **additive model** because the secular trend component ( $\beta_1 t$ ) is added to the seasonal component ( $\beta_2 Q_1 + \beta_3 Q_2 + \beta_4 Q_3$ ) to form the model. A **multiplicative model** would have the same form, except that the dependent variable would be the natural logarithm of power load; that is,

\*These interpretations assume a fixed value of  $t$ . In practical terms this is unrealistic, since each quarter is associated with a different value of  $t$ . Nevertheless, the coefficients of the seasonal dummy variables provide insight into the seasonality of these time series data.

QTRPOWERforecasts.MTW ***										
↓	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
	T	YEAR	QTR	LOAD	Q1	Q2	Q3	PRED	LO95PLIM	HI95PLIM
1	49	2007	1	151.3	1	0	0	164.341	147.294	181.389
2	50	2007	2	132.9	0	1	0	148.583	131.536	165.630
3	51	2007	3	160.5	0	0	1	172.425	155.378	189.472
4	52	2007	4	161.0	0	0	0	155.591	138.544	172.639

**FIGURE 13.17**  
MINITAB worksheet with power load forecasts (megawatts)

$$\ln Y_t = \beta_0 + \beta_1 t + \beta_2 Q_1 + \beta_3 Q_2 + \beta_4 Q_3 + \epsilon$$

To see the multiplicative nature of this model, we take the antilogarithm of both sides of the equation to get

$$\begin{aligned}
 Y_t &= \exp\{\beta_0 + \beta_1 t + \beta_2 Q_1 + \beta_3 Q_2 + \beta_4 Q_3 + \epsilon\} \\
 &= \underbrace{\exp\{\beta_0\}}_{\text{Constant}} \underbrace{\exp\{\beta_1 t\}}_{\text{Secular trend}} \underbrace{\exp\{\beta_2 Q_1 + \beta_3 Q_2 + \beta_4 Q_3\}}_{\text{Seasonal component}} \underbrace{\exp\{\epsilon\}}_{\text{Residual component}}
 \end{aligned}$$

The multiplicative model often provides a better forecasting model when the time series is changing at an increasing rate over time.

When time series data are observed monthly, a regression forecasting model needs 11 dummy variables to describe monthly seasonality; three dummy variables can be used (as in the previous models) if the seasonal changes are hypothesized to occur quarterly. In general, this approach to seasonal modeling requires one dummy variable fewer than the number of seasonal changes expected to occur.

There are approaches besides the regression dummy variable method for forecasting seasonal time series. Trigonometric (sine and cosine) terms can be used in regression models to model periodicity. Other time series models (the Holt-Winters exponential smoothing model, for example) do not use the regression approach at all, and there are various methods for adding seasonal components to these models. We have chosen to discuss the regression approach because it makes use of the important modeling concepts covered in Chapter 11, and because the regression forecasts are accompanied by prediction intervals that provide some measure of the forecast reliability. While most other methods do not have explicit measures of reliability, many have proved their merit by providing good forecasts for particular applications. Consult the references at the end of the chapter for details of other seasonal models.

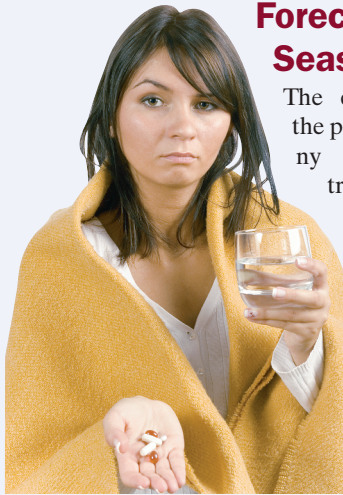
## ACTIVITY 13.1: Time Series

For this activity, select a recurring quantity from your own life for which you have monthly records for at least two years. This might be the cost of a utility bill, the number of cell phone minutes used, or even your income. If you do not have access to such records, use the Internet to find similar data, such as median, monthly home prices in your area for at least two years.

1. Which methods from this chapter might apply to your data? Does there appear to be a seasonal component affecting the data? If so, can you explain the seasonal effect in simple terms?
2. Use methods from this chapter to predict the value of your quantity for the next year. Be prepared to defend your choice of methods.



## STATISTICS IN ACTION REVISITED



### Forecasting Coldex Sales with a Seasonal Regression Model

The consultants hired by the pharmaceutical company detected a cyclical trend in the monthly sales data. They noted that sales of the cold medicine were higher during the winter and summer months as compared to the other months over the 2-year period. To account for this seasonal trend,

they created 11 dummy variables ( $x_1, x_2, \dots, x_{11}$ ) for the 12 months of the year. The seasonal forecasting model takes the form

$$E(Y_t) = \beta_0 + \beta_1 t + \beta_2 x_1 + \beta_3 x_2 + \beta_4 x_3 + \beta_5 x_4 + \beta_6 x_5 + \beta_7 x_6 + \beta_8 x_7 + \beta_9 x_8 + \beta_{10} x_9 + \beta_{11} x_{10} + \beta_{12} x_{11}$$

The MINITAB regression printout for the model is shown in Figure SIA13.6, followed by the model forecasts in Figure SIA13.7. The global  $F$ -test ( $p$ -value = .000) indicates that the model is statistically useful for predicting monthly sales, and the model coefficient of determination ( $R^2 = .983$ ) indicates that over 98% of the sample variation in monthly sales can be explained by the seasonal model. Statistically, this model is a tremendous improvement over the linear trend model.

The 95% prediction intervals for sales in months 25, 26, and 27 (highlighted on Figure SIA13.7) are (4285.8, 6349.2), (4164.8, 6228.2), and (1604.3, 3667.7), respectively. Thus, for the first month of year 3 (month 25), we are 95% confident that the actual sales will fall between 4,285.8 and 6,349.2 thousand dollars. (Similar interpretations are made for the other two forecasts.) These intervals are much narrower than those for the previous two forecasting models, and they also reflect the expected drop in sales in March (month 3) from the winter months. This seasonal model was used successfully by the pharmaceutical firm to forecast monthly sales.

The regression equation is

SALES = 2961 + 74.2 TIME + 501 JAN + 306 FEB - 2329 MAR - 3101 APR - 3138 MAY - 2393 JUN - 2241 JUL + 623 AUG + 2111 SEP - 361 OCT - 744 NOV

Predictor	Coef	SE Coef	T	P
Constant	2960.5	320.2	9.25	0.000
TIME	74.25	11.65	6.38	0.000
JAN	500.8	365.5	1.37	0.198
FEB	305.5	361.6	0.84	0.416
MAR	-2329.2	358.0	-6.51	0.000
APR	-3101.0	354.8	-8.74	0.000
MAY	-3137.7	351.9	-8.92	0.000
JUN	-2393.5	349.4	-6.85	0.000
JUL	-2241.2	347.2	-6.45	0.000
AUG	623.0	345.5	1.80	0.099
SEP	2111.3	344.1	6.14	0.000
OCT	-361.5	343.1	-1.05	0.315
NOV	-743.7	342.5	-2.17	0.053

S = 342.314    R-Sq = 98.3%    R-Sq(adj) = 96.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	12	75435259	6286272	53.65	0.000
Residual Error	11	1288964	117179		
Total	23	76724223			

Figure SIA13.6

MINITAB printout for seasonal regression model of monthly Coldex sales

↓	C1	C2	C3	C4	C5	C6
	YEAR	MONTH	TIME	FORECAST	PL95LOW	PL95UPP
1	3	1	25	5317.5	4285.83	6349.17
2	3	2	26	5196.5	4164.83	6228.17
3	3	3	27	2636.0	1604.33	3667.67

Figure SIA13.7

MINITAB worksheet with seasonal regression forecasts of monthly sales

## Exercises 13.38–13.44

### Learning the Mechanics

**13.38** The annual price of a finished product (in cents per pound) from 1991 to 2006 is given in the table below. The time variable  $t$  begins with  $t = 1$  in 1991 and is incremented by 1 for each additional year.

#### LM13\_38

Year	$t$	Price, $Y_t$	Year	$t$	Price, $Y_t$
1991	1	21.73	1999	9	24.42
1992	2	24.32	2000	10	25.49
1993	3	25.31	2001	11	26.19
1994	4	26.36	2002	12	27.31
1995	5	27.31	2003	13	24.40
1996	6	27.58	2004	14	24.24
1997	7	24.79	2005	15	25.87
1998	8	25.36	2006	16	26.86

- Fit the straight-line model,  $E(Y_t) = \beta_0 + \beta_1 t$ , to the data.
- Give the least squares estimates of the  $\beta$ 's.
- Use the least squares prediction equation to obtain the forecasts for 2003 and 2004.
- Find 95% forecast intervals for 2003 and 2004.

**13.39** Retail sales in Quarters 1–4 over a 10-year period for a department store are shown (in hundreds of thousands of dollars) in the table below.

NW

#### LM13\_39

Year	QUARTER			
	1	2	3	4
1	8.3	10.3	8.7	13.5
2	9.8	12.1	10.1	15.4
3	12.1	14.5	12.7	17.1
4	13.7	16.0	14.2	19.2
5	17.4	19.7	18.0	23.1
6	18.2	20.5	18.6	24.0
7	20.0	22.2	20.5	25.1
8	22.3	25.1	22.9	27.7
9	24.7	26.9	25.1	29.8
10	25.8	28.7	26.0	32.2

- Write a regression model that contains trend and seasonal components to describe the sales data.
- Use least squares regression to fit the model. Evaluate the fit of the model.

- Use the regression model to forecast the quarterly sales during year 11. Give 95% prediction intervals for the forecasts.

**13.40** What advantage do regression forecasts have over exponentially smoothed forecasts? Does this advantage ensure that regression forecasts will prove to be more accurate? Explain.

### Applying the Concepts—Basic

**13.41 Mortgage interest rates.** The level at which commercial lending institutions set mortgage interest rates has a significant effect on the volume of buying, selling, and construction of residential and commercial real estate. The data in the table are the annual average mortgage interest rates for conventional, fixed-rate, 30-year loans for the period 1980–2004.

#### INTRATE30

Year	Interest Rate (%)	Year	Interest Rate (%)
1980	14.30	1993	8.09
1981	16.54	1994	8.28
1982	16.83	1995	7.86
1983	13.92	1996	7.76
1984	13.71	1997	7.57
1985	12.91	1998	6.92
1986	11.33	1999	7.46
1987	10.46	2000	8.08
1988	10.86	2001	7.01
1989	12.07	2002	6.56
1990	11.78	2003	5.89
1991	11.14	2004	5.86
1992	9.29		

Source: Statistical Abstract of the United States, U.S. Bureau of the Census, 2006.

- Fit the simple regression model

$$E(Y_t) = \beta_0 + \beta_1 t$$

where  $t$  is the number of years since 1980 (i.e.,  $t = 0, 1, \dots, 24$ ).

- Forecast the average mortgage interest rate in 2007. Find a 95% prediction interval for this forecast.

#### NATGAS

**13.42 Price of natural gas.** Refer to Exercise 13.9 (p. 13-13) and the annual prices of natural gas from 1990 to 2004. A simple

linear regression model,  $E(Y_t) = \beta_0 + \beta_1 t$ , where  $t$  is the number of years since 1990, is proposed to forecast the annual price of natural gas.

- Give the least squares estimates of the  $\beta$ 's and interpret their values.
- Evaluate the model's fit.
- Find and interpret 95% prediction intervals for the years 2005 and 2006.
- Describe the problems associated with using a simple linear regression model to predict time series data.

### Applying the Concepts—Intermediate

**13.43 Life insurance policies in force.** The table below represents all life insurance policies (in millions) in force on the lives of U.S. residents for the years 1970 through 2003.

#### LIFEINS

Year	No. of Policies (in millions)	Year	No. of Policies (in millions)
1970	355	1987	395
1971	357	1988	391
1972	365	1989	394
1973	369	1990	389
1974	380	1991	375
1975	380	1992	366
1976	382	1993	363
1977	390	1994	371
1978	401	1995	392
1979	407	1996	355
1980	402	1997	351
1981	400	1998	358
1982	390	1999	367
1983	387	2000	369
1984	385	2001	377
1985	386	2002	375
1986	391	2003	370

Source: U.S. Bureau of the Census. *Statistical Abstract of the United States*, 2005.

- Use the method of least squares to fit a simple regression model to the data.
- Forecast the number of life insurance policies in force for 2004 and 2005.

- Construct 95% prediction intervals for the forecasts of part b.
- Check the accuracy of your forecasts by looking up the actual number of life insurance policies in force for 2004 and 2005 in the *Statistical Abstract of the United States*.

**13.44 Graphing calculator sales.** The next table presents the quarterly sales index for one brand of graphing calculator at a campus bookstore. The quarters are based on an academic year, so the first quarter represents fall; the second, winter; the third, spring; and the fourth, summer.

Define the time variable as  $t = 1$  for the first quarter of 2002,  $t = 2$  for the second quarter of 2002, etc. Consider the following seasonal dummy variables:

$$Q_1 = \begin{cases} 1 & \text{if Quarter 1} \\ 0 & \text{otherwise} \end{cases}$$

$$Q_2 = \begin{cases} 1 & \text{if Quarter 2} \\ 0 & \text{otherwise} \end{cases}$$

$$Q_3 = \begin{cases} 1 & \text{if Quarter 3} \\ 0 & \text{otherwise} \end{cases}$$

#### GRAPHICAL

Year	First Quarter	Second Quarter	Third Quarter	Fourth Quarter
2002	438	398	252	160
2003	464	429	376	216
2004	523	496	425	318
2005	593	576	456	398
2006	636	640	526	498

- Write a regression model for  $E(Y_t)$  as a function of  $t$ ,  $Q_1$ ,  $Q_2$ , and  $Q_3$ .
- Find and interpret the least squares estimates, and evaluate the usefulness of the model.
- Which of the assumptions about the random error component is in doubt when a regression model is fit to time series data?
- Find the forecasts and the 95% prediction intervals for the 2007 quarterly sales. Interpret the result.

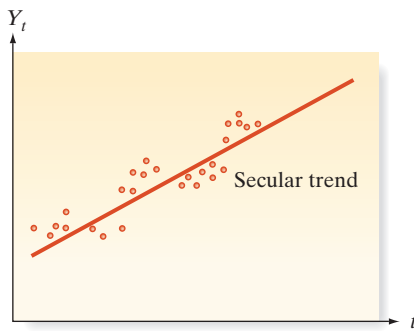
## 13.9 Autocorrelation and the Durbin-Watson Test

Recall that one of the assumptions we make when using a regression model for predictions is that the errors are independent. However, with time series data, this assumption is questionable. The cyclical component of a time series may result in deviations from the secular trend that tend to cluster alternately on the positive and negative sides of the trend, as shown in Figure 13.18.

The observed errors between the time series and the regression model for the secular trend (and seasonal component, if present) are called **time series residuals**. Thus, if the time series  $Y_t$  has an estimated trend of  $\hat{Y}_t$ , then the time series residual\* is

$$\hat{R}_t = Y_t - \hat{Y}_t$$

\*We use  $\hat{R}_t$  rather than  $\hat{\epsilon}$  to denote a time series residual because, as we shall see, time series residuals often do not satisfy the regression assumptions associated with the random component  $\epsilon$ .



**FIGURE 13.18**  
Illustration of cyclical errors

Note that time series residuals are defined just as the residuals for any regression model. However, we will usually plot time series residuals versus time to determine whether a cyclical component is apparent.

For example, consider the sales forecasting data in Table 13.7, to which we fit a simple straight-line regression model. The MINITAB plot of the data and model is repeated in Figure 13.19, and a plot of the time series residuals is shown in Figure 13.20.

Notice the tendency of the residuals to group alternately into positive and negative clusters—that is, if the residual for year  $t$  is positive, there is a tendency for the residual for year  $(t + 1)$  to be positive. These cycles are indicative of possible positive correlation between neighboring residuals. The correlation between time series residuals at different points in time is called *autocorrelation*, and the autocorrelation of neighboring residuals (time periods  $t$  and  $t + 1$ ) is called *first-order autocorrelation*.

### Definition 13.5

The correlation between time series residuals at different points in time is called **autocorrelation**. Correlation between neighboring residuals (at times  $t$  and  $t + 1$ ) is called **first-order autocorrelation**. In general, correlation between residuals at times  $t$  and  $t + d$  is called  $d$ th-order autocorrelation.

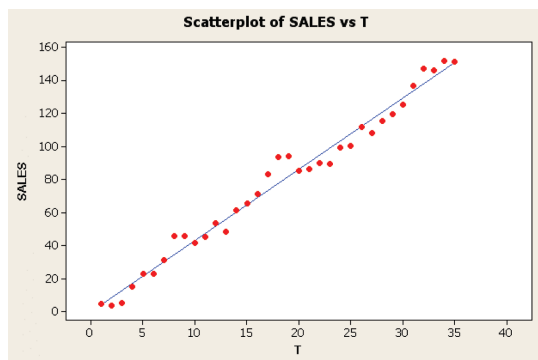
Rather than speculate about the presence of autocorrelation among time series residuals, we prefer to test for it. For most business and economic time series, the relevant test is for first-order autocorrelation. Other higher-order autocorrelations may indicate seasonality (e.g., fourth-order autocorrelation in a quarterly time series). However, when we use the term *autocorrelation* in this text, we are referring to first-order autocorrelation unless otherwise specified. So, we test

$H_0$ : No first-order autocorrelation of residuals

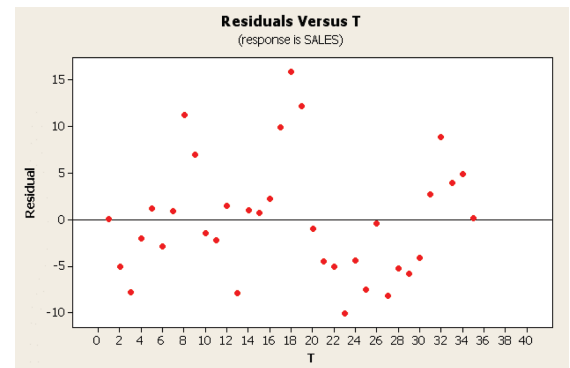
$H_a$ : Positive first-order autocorrelation of residuals

The **Durbin-Watson  $d$ -statistic** is used to test for the presence of first-order autocorrelation. The statistic is given by the formula

$$d = \frac{\sum_{t=2}^n (\hat{R}_t - \hat{R}_{t-1})^2}{\sum_{t=1}^n \hat{R}_t^2}$$



**FIGURE 13.19**  
MINITAB scatterplot of annual sales data



**FIGURE 13.20**  
MINITAB plot of residuals versus time for straight-line model of annual sales

where  $n$  is the number of observations (time periods) and  $(\hat{R}_t - \hat{R}_{t-1})$  represents the difference between a pair of successive time series residuals. The value of  $d$  always falls in the interval from 0 to 4. The interpretations of the values of  $d$  are given in the box. Most statistical software packages include a routine that calculates  $d$  for time series residuals.

### Interpretation of Durbin-Watson $d$ -Statistic

$$d = \frac{\sum_{t=2}^n (\hat{R}_t - \hat{R}_{t-1})^2}{\sum_{t=1}^n \hat{R}_t^2} \quad \text{Range of } d: 0 \leq d \leq 4$$

1. If the residuals are uncorrelated, then  $d \approx 2$ .
2. If the residuals are positively autocorrelated, then  $d < 2$ , and if the autocorrelation is very strong,  $d \approx 0$ .
3. If the residuals are negatively autocorrelated, then  $d > 2$ , and if the autocorrelation is very strong,  $d \approx 4$ .

## Biography

### GEOFFREY S. WATSON (1921–1998) The Durbin-Watson Test

Australian Geoff Watson was educated at the University of Melbourne, where he earned a mathematics degree in 1942. Following World War II, Watson moved to North Carolina State University to begin work on a graduate degree in statistics. He eventually earned his PhD in 1951. During his illustrious career as a statistics professor and researcher, Watson had appointments at Cambridge University, Australian National University, University of Toronto, Johns Hopkins University, and Princeton University (where he was chairman

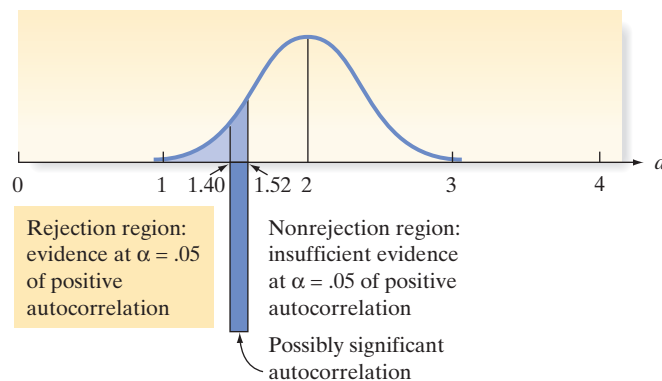
of the statistics department). While visiting Cambridge in the late 1940s, Watson collaborated with James Durbin of the London School of Economics to develop their well-known Durbin-Watson test for serial correlation. His research interests covered a wide spectrum of statistical applications all across the world, including estimating the size of the penguin population (Antarctica), paleontology problems (Sweden), probability in quantum mechanics (Rome), molecular biology (Italy), and ozone depletion (U.S. Energy Information Administration). Outside his professional life, Watson was a serious painter (landscapes and hills) and an accomplished tennis player (effective lob).

Durbin and Watson (1951) give tables for the lower-tail values of the  $d$ -statistic, which are shown in Tables XII ( $\alpha = .05$ ) and XIII ( $\alpha = .01$ ) of Appendix B. Part of Table XII is reproduced in Table 13.10. For the sales example, we have  $k = 1$  independent variable and  $n = 35$  observations. Using  $\alpha = .05$  for the one-tailed test for positive autocorrelation, we obtain the tabled values  $d_L = 1.40$  and  $d_U = 1.52$ . The meaning of these values is illustrated in Figure 13.21. Because of the complexity of the sampling distribution of  $d$ , it is not possible to specify a single point that acts as a boundary between the rejection and nonrejection regions, as we did for the  $z$ ,  $t$ ,  $F$ , and other test statistics. Instead, an upper ( $d_U$ ) and lower ( $d_L$ ) bound are specified. Thus a  $d$ -value less than  $d_L$  does provide strong evidence of positive autocorrelation at  $\alpha = .05$  (recall that small  $d$  values indicate positive autocorrelation); a  $d$  value greater than  $d_U$  does not provide evidence of positive autocorrelation at  $\alpha = .05$ ; and a value of  $d$  between  $d_L$  and  $d_U$  might or might not be significant at the  $\alpha = .05$  level. If  $d_L < d < d_U$ , more information is needed before we can reach any conclusion about the presence of autocorrelation.

Tests for negative autocorrelation and two-tailed tests can be conducted by making use of the symmetry of the sampling distribution of the  $d$ -statistic about its mean. The test procedure is summarized in the next box.

**TABLE 13.10** Reproduction of Part of Table XII of Appendix B: Critical Values for the Durbin-Watson  $d$ -Statistic,  $\alpha = .05$ 

$n$	$k = 1$		$k = 2$		$k = 3$		$k = 4$		$k = 5$	
	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79

**FIGURE 13.21**

Rejection region for the Durbin-Watson  $d$  test: Sales example

### Durbin-Watson $d$ -Test for Autocorrelation

#### One-Tailed Test

$H_0$ : No first-order autocorrelation  
 $H_a$ : Positive first-order autocorrelation  
 (or  $H_a$ : Negative first-order autocorrelation)

#### Two-Tailed Test

$H_0$ : No first-order autocorrelation  
 $H_a$ : Positive or negative first-order autocorrelation

#### Test statistic:

$$d = \frac{\sum_{t=2}^n (\hat{R}_t - \hat{R}_{t-1})^2}{\sum_{t=1}^n \hat{R}_t^2}$$

#### Rejection region:

$d < d_{L,\alpha}$   
 [or  $(4 - d) < d_{L,\alpha}$   
 if  $H_a$ : Negative first-order autocorrelation]

where  $d_{L,\alpha}$  is the lower tabled value corresponding to  $k$  independent variables and  $n$  observations. The corresponding upper value  $d_{U,\alpha}$  defines a “possibly significant” region between  $d_{L,\alpha}$  and  $d_{U,\alpha}$  (see Figure 13.21).

#### Rejection region:

$d < d_{L,\alpha/2}$  or  $(4 - d) < d_{L,\alpha/2}$

where  $d_{L,\alpha/2}$  is the lower tabled value corresponding to  $k$  independent variables and  $n$  observations. The corresponding upper value  $d_{U,\alpha/2}$  defines a “possibly significant” region between  $d_{L,\alpha/2}$  and  $d_{U,\alpha/2}$  (see Figure 13.21).

### Requirements for the Validity of the $d$ -Test

The residuals are normally distributed.

A portion of the Excel printout for the regression of annual sales is presented in Figure 13.22. It shows that the computed value of  $d$  is .821 (highlighted), which is less than the tabulated value of  $d_L = 1.40$  for  $\alpha = .05$ . Thus, we conclude that the residuals of the straight-line model for sales are positively autocorrelated.

	A	B
1	<b>Durbin-Watson Calculations</b>	
2		
3	Sum of Squared Difference of Residuals	1104.249775
4	Sum of Squared Residuals	1345.453546
5		
6	<b>Durbin-Watson Statistic</b>	<b>0.820726794</b>
7		
8		

**FIGURE 13.22**

Excel printout with Durbin-Watson statistics for annual sales model

### Now Work Exercise 13.47

Once strong evidence of first-order autocorrelation has been established, as in the case of the sales example, doubt is cast on the least squares results and any inferences drawn from them. Under these circumstances, a time series model that accounts for the autocorrelation of the random errors is needed. A useful model is the first-order autoregressive model. Consult the references at the end of the book.

## Exercises 13.45–13.54

### Learning the Mechanics

**13.45** Define autocorrelation. Explain why it is important in time series modeling and forecasting.

**13.46** What do the following Durbin-Watson statistics suggest about the autocorrelation of the time series residuals from which each was calculated?

- $d = 3.9$
- $d = .2$
- $d = 1.99$

**13.47** For each case, indicate the decision regarding the test of the null hypothesis of no first-order autocorrelation against the alternative hypothesis of positive first-order autocorrelation.

- $k = 2, n = 20, \alpha = .05, d = 1.1$
- $k = 2, n = 20, \alpha = .01, d = 1.1$
- $k = 5, n = 65, \alpha = .05, d = .95$
- $k = 1, n = 31, \alpha = .01, d = 1.35$

### Applying the Concepts—Basic

**13.48 Forecasting monthly car and truck sales.** Forecasts of automotive vehicle sales in the United States provide the basis for financial and strategic planning of large automotive corporations. The following forecasting model was developed for  $Y_t$ , total monthly passenger car and light truck sales (in thousands):

$$E(Y_t) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5$$

where  $x_1$  = average monthly retail price of regular gasoline,  $x_2$  = annual percentage change in GDP per quarter,  $x_3$  = monthly consumer confidence index,  $x_4$  = total number of vehicles scrapped (millions) per month, and  $x_5$  = vehicle seasonality. The model was fit to monthly data collected over a 12-year period (i.e.,  $n = 144$  months), with the following results:  $R^2 = .856$ , Durbin-Watson  $d = 1.01$ .

- Is there sufficient evidence to indicate that the overall model contributes information for the prediction of monthly passenger car and light truck sales? Test using  $\alpha = .05$ .
- Is there sufficient evidence to indicate that the regression errors are positively correlated? Test using  $\alpha = .05$ .
- Comment on the validity of the inference concerning model adequacy in light of the result of part **b**.

**13.49 Modeling the deposit share of a retail bank.** Exploratory research published in the *Journal of Professional Services Marketing* (Vol. 5, 1990) examined the relationship between deposit share of a retail bank and several marketing variables. Quarterly deposit share data were collected for 5 consecutive years for each of nine retail banking institutions. The model analyzed took the following form:

$$E(Y_t) = \beta_0 + \beta_1 P_{t-1} + \beta_2 S_{t-1} + \beta_3 D_{t-1}$$

where  $Y_t$  = deposit share of a bank in quarter  $t$  ( $t = 1, 2, \dots, 20$ ),  $P_{t-1}$  = expenditures on promotion-related activities in quarter  $t - 1$ ,  $S_{t-1}$  = expenditures on service-related activities in quarter  $t - 1$ , and  $D_{t-1}$  = expenditures on distribution-related activities in quarter  $t - 1$ . A separate model was fit for each bank with the results shown in the table.

Bank	$R^2$	p-Value for Global F-Test	Durbin-Watson $d$
1	.914	.000	1.3
2	.721	.004	3.4
3	.926	.000	2.7
4	.827	.000	1.9
5	.270	.155	.85
6	.616	.012	1.8
7	.962	.000	2.5
8	.495	.014	2.3
9	.500	.011	1.1

- Interpret the values of  $R^2$  for each bank.
- Test the overall adequacy of the model for each bank using  $\alpha = .01$ .
- Conduct the Durbin-Watson  $d$ -test for positive residual correlation for each bank at  $\alpha = .01$ . What conclusions do you draw about autocorrelation?

**13.50** The consumer purchasing value of the dollar,  $Y_t$ , from 1970 to 2004 is illustrated by the data in the next table. The buying power of the dollar (compared with 1982) is listed for each year. The first-order model

$$Y_t = \beta_0 + \beta_1 t + \epsilon$$

was fit to the data using the method of least squares. The MINITAB printout and a plot of the regression residuals are shown below.

- Examine the plot of the regression residuals against  $t$ . Is there a tendency for the residuals to have long positive and negative runs? To what do you attribute this phenomenon?
- Locate the Durbin-Watson  $d$ -statistic on the printout and test the null hypothesis that the time series residuals are uncorrelated. Use  $\alpha = .10$ .
- What assumption(s) must be satisfied in order for the test of part **b** to be valid?

Plot of regression residuals for Exercise 13.50



**BUYPOWER**

Year	$t$	Value, $Y_t$	Year	$t$	Value, $Y_t$
1970	1	2.545	1988	19	0.926
1971	2	2.469	1989	20	0.880
1972	3	2.392	1990	21	0.839
1973	4	2.193	1991	22	0.822
1974	5	1.901	1992	23	0.812
1975	6	1.718	1993	24	0.802
1976	7	1.645	1994	25	0.797
1977	8	1.546	1995	26	0.782
1978	9	1.433	1996	27	0.762
1979	10	1.289	1997	28	0.759
1980	11	1.136	1998	29	0.765
1981	12	1.041	1999	30	0.752
1982	13	1.000	2000	31	0.725
1983	14	0.984	2001	32	0.711
1984	15	0.964	2002	33	0.720
1985	16	0.955	2003	34	0.698
1986	17	0.969	2004	35	0.673
1987	18	0.949			

Source: Statistical Abstract of the United States, 2006.

**Applying the Concepts—Intermediate**

**INTRATE30**

**13.51 Mortgage interest rates.** Refer to the data on annual mortgage interest rate ( $Y_t$ ), Exercise 13.41 (p. 13-43). You fit the simple linear regression model,  $E(Y_t) = \beta_0 + \beta_1 t$ , to the data for the years 1980 to 2004 ( $t = 1, 2, \dots, 25$ ).

- Find and plot the regression residuals against  $t$ . Does the plot suggest the presence of autocorrelation? Explain.
- Conduct the Durbin-Watson test (at  $\alpha = .05$ ) to test formally for the presence of positively autocorrelated regression errors.
- Comment on the validity of the inference concerning model adequacy in light of the result of part **b**.

MINITAB Output for Exercise 13.50

**Regression Analysis: VALUE versus T**

The regression equation is  
 VALUE = 1.99 - 0.0466 T

Predictor	Coef	SE Coef	T	P
Constant	1.99235	0.09643	20.66	0.000
T	-0.046632	0.004672	-9.98	0.000

S = 0.279148 R-Sq = 75.1% R-Sq(adj) = 74.4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	7.7631	7.7631	99.62	0.000
Residual Error	33	2.5715	0.0779		
Total	34	10.3346			

Durbin-Watson statistic = 0.0627189



**NATGAS**

**13.52 Price of natural gas.** Refer to the annual data on natural gas price ( $Y_t$ ), Exercise 13.42 (p. 13-44). You fit the simple linear regression model,  $E(Y_t) = \beta_0 + \beta_1 t$ , to the data for the years 1990 to 2004 ( $t = 1, 2, \dots, 14$ ).

- a. Find and plot the regression residuals against  $t$ . Does the plot suggest the presence of autocorrelation? Explain.
- b. Conduct the Durbin-Watson test (at  $\alpha = .05$ ) to test formally for the presence of positively autocorrelated regression errors.
- c. Comment on the validity of the inference concerning model adequacy in light of the result of part b.

**13.53 Forecasting foreign exchange rates.** T. C. Chiang considered several time series forecasting models of future foreign exchange rates for U.S. currency (*The Journal of Financial Research*, Summer 1986). One popular theory among financial analysts is that the forward (90-day) exchange rate is a useful predictor of the future spot exchange rate. Using monthly data on exchange rates for the British pound for  $n = 81$  months, Chiang fit the model  $E(Y_t) = \beta_0 + \beta_1 x_{t-1}$ , where  $Y_t = \ln(\text{spot rate})$  in month  $t$ , and  $x_{t-1} = \ln(\text{forward rate})$  in month  $t - 1$ . The analysis yielded the following results:

$$t\text{-value} = 47.9, \quad s = .025, \quad R^2 = .957,$$

$$\text{Durbin-Watson } d = .962$$

- a. Is the model statistically useful for forecasting future spot exchange rates for the British pound? Test using  $\alpha = .05$ .
- b. Interpret the values of  $s$  and  $R^2$ .
- c. Is there evidence of positive autocorrelation among the residuals? Test using  $\alpha = .05$ .
- d. Based on the results of parts a–c, would you recommend using the model to forecast spot exchange rates?

**LIFEINS**

**13.54 Life insurance policies in force.** Refer to the annual data on number of life insurance policies in force ( $Y_t$ ), Exercise 13.43 (p. 13-44). You fit the simple linear regression model,  $E(Y_t) = \beta_0 + \beta_1 t$ , to the data for the years 1970 to 2003 ( $t = 1, 2, \dots, 34$ ).

- a. Find and plot the regression residuals against  $t$ . Does the plot suggest the presence of autocorrelation? Explain.
- b. Conduct the Durbin-Watson test (at  $\alpha = .05$ ) to test formally for the presence of positively autocorrelated regression errors.
- c. Comment on the validity of the inference concerning model adequacy in light of the result of part b.

**KEY TERMS**

Note: Starred (\*) items are from the optional section in this chapter.

Adaptive forecast 13-21

Additive model 13-20

Autocorrelation 13-45

Base period 13-4

Composite index number 13-6

Cyclical effect 13-20

Double exponential smoothing\* 13-24

Durbin-Watson  $d$ -statistic 13-45

Durbin-Watson test 13-47

Exponentially smoothed forecast 13-21

Exponential smoothing 13-15

Exponential smoothing constant 13-15

First-order autocorrelation 13-45

Forecast error 13-22

Holt-Winters forecasting model\* 13-24

Index number 13-4

Inferential forecasting model 13-34

Laspeyres index 13-9

Long-term trend 13-20

Mean absolute deviation 13-29

Mean absolute percentage error 13-29

Multiplicative model 13-40

Paasche index 13-10

Residual effect 13-20

Root mean squared error 13-29

Seasonal effect 13-20

Seasonal model 13-38

Secular trend 13-20

Simple composite index number 13-7

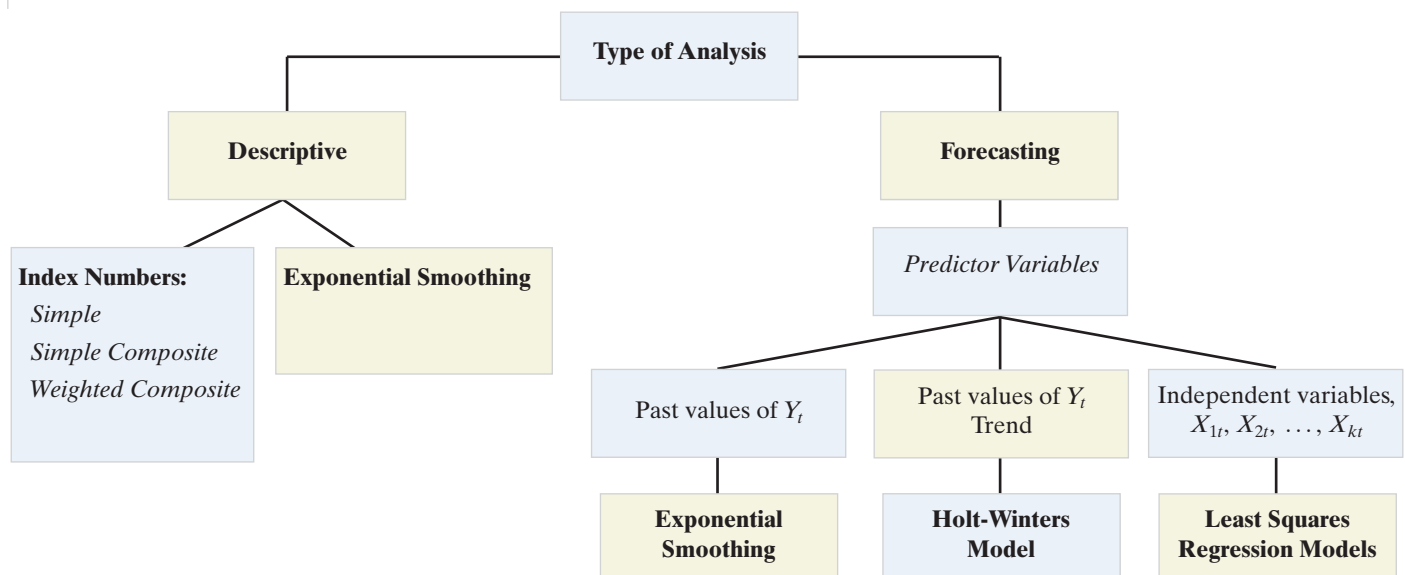
Simple index number 13-5

Time series 13-3

Time series residuals 13-44

Weighted composite price index 13-8

**GUIDE TO TIME SERIES ANALYSIS**



## CHAPTER NOTES

## Key Formulas

$$I_t = \left( \frac{Y_t}{Y_0} \right) 100$$

Simple index

$$I_t = \left( \frac{\text{Total of all } Y\text{-values at time } t}{\text{Total of all } Y\text{-values at time } t_0} \right) 100$$

Simple composite index

Weighted composite price indexes:

$$I_t = \left( \frac{\sum_{i=1}^k Q_{it_0} P_{it}}{\sum_{i=1}^k Q_{it_0} P_{i0}} \right) 100$$

Laspeyres

$$I_t = \left( \frac{\sum_{i=1}^k Q_{it} P_{it}}{\sum_{i=1}^k Q_{it} P_{i0}} \right) 100$$

Paasche

Exponential smoothing:

$$E_t = wY_t + (1 - w)E_{t-1}$$

$$\text{Note: } E_1 = Y_1$$

$$\text{Forecast: } F_{t+k} = E_t$$

Holt-Winters model\*:

$$E_t = wY_t + (1 - w)(E_{t-1} + T_{t-1})$$

$$\text{Note: } E_2 = Y_2, T_2 = Y_2 - Y_1$$

$$T_t = v(E_t - E_{t-1}) + (1 - v)T_{t-1}$$

$$\text{Forecast: } F_{t+k} = E_t + kT_t$$

$$\text{MAD} = \frac{\sum_{t=1}^m |Y_t - F_t|}{m}$$

Mean absolute deviation

$$\text{MAPE} = \frac{\sum_{t=1}^m \left| \frac{Y_t - F_t}{Y_t} \right|}{m} \times 100$$

Mean absolute percentage error

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^m (Y_t - F_t)^2}{m}}$$

Root mean square error

$$d = \frac{\sum_{t=2}^n (\hat{R}_t - \hat{R}_{t-1})^2}{\sum_{t=1}^n \hat{R}_t^2}$$

Durbin-Watson test statistic

**Time Series Data**

Data generated by processes over time.

**Index Number**

Measures the change in a variable over time relative to a base period.

Types of index numbers: (1) **simple index number**, (2) **simple composite index number**, and (3) **weighted composite number**.Types of weighted composite index numbers: (1) **Laspeyres index** and (2) **Paasche index**.

Time series components: (1) **secular (long-term) trend**, (2) **cyclical effect**, (3) **seasonal effect**, and (4) **residual effect**.

*Descriptive methods* of forecasting with smoothing: (1) **exponential smoothing** and (2) **Holt-Winters model**.

Measures of forecast accuracy: (1) **mean absolute deviation (MAD)**, (2) **mean absolute percentage error (MAPE)**, and (3) **root mean squared error (RMSE)**.

An *inferential* forecasting method: **least squares regression**.

Problems with least squares regression forecasting: (1) *prediction outside the experimental region* and (2) *regression errors are autocorrelated*.

### Autocorrelation

Correlation between time series residuals at different points in time.

A test for first-order autocorrelation: **Durbin-Watson test**.

## Key Symbols

$Y_t$	... Time series value at time $t$
$I_t$	... Index at time $t$
$P_t$	... Price series at time $t$
$Q_t$	... Quantity series at time $t$
$E_t$	... Exponentially smoothed value at time $t$
$T_t$	... Smoothed trend at time $t$
$F_{t+k}$	... $k$ -step-ahead forecast value
MAD	... Mean absolute deviation
MAPE	... Mean absolute percentage error
RMSE	... Root mean squared error
$\hat{R}_t$	... Residual at time $t$
$d$	... Value of Durbin-Watson test statistic
$d_L$	... Lower critical value of $d$
$d_U$	... Upper critical value of $d$

## SUPPLEMENTARY EXERCISES 13.55–13.68

### Applying the Concepts—Basic

**13.55 Average prices of steel.** The U.S. steel industry prices (in cents per pound) of three varieties of steel sheets are given in the table for selected years from 1995 to 2004.

#### STEEL3

Year	Cold Finished	Hot Rolled	Galvanized
1995	25.70	25.32	34.47
2000	23.08	15.67	21.38
2001	22.76	11.71	16.41
2002	23.26	16.46	22.00
2003	25.15	14.80	20.08
2004	38.67	30.84	36.69

Sources: Standard & Poor's Statistics: Metals, 2005, The CRB Community Yearbook, 2005.

- Calculate a simple composite index for the three steel price series using 1995 as the base period. Interpret the results.
- Is the index a price index or a quantity index?

- What information would you need in order to calculate a Laspeyres index with a base period of 1995? A Paasche index with a base period of 1995?

**13.56 Average prices of steel (cont'd).** Refer to Exercise 13.55.

- Compute the exponentially smoothed series corresponding to each of the price series using the smoothing constant  $w = .5$ .
- Plot the prices and their exponentially smoothed series on the same graph.
- Find the exponential smoothing forecasts of 2005 prices of the three varieties of steel. What are the drawbacks to these forecasts?

**13.57 Demand for emergency room services.** With the advent of managed care, U.S. hospitals have begun to operate like businesses. More than ever before, hospital administrators need to know and apply the theories and methods taught in business schools. Richmond Memorial Hospital in Richmond, Virginia, uses regression analysis to forecast the demand for emergency room services. Specifically, Richmond Memorial uses data on patient visits to the emergency room during each of the past 10 Augusts to

forecast next August’s demand. Data for the month of August in a recent 10-year period are shown below.

**ER**

Year $t$	Visits	Daily Average $Y_t$	Year $t$	Visits	Daily Average $Y_t$
1	1,367	44.09	6	3,019	97.38
2	1,642	52.96	7	2,794	90.12
3	1,780	57.41	8	2,846	91.80
4	2,060	66.45	9	3,001	96.80
5	2,257	72.80	10	3,548	114.45

Source: Adapted from Bolling, W. B. “Queuing Model of a Hospital Emergency Room.” *Industrial Engineering*, Sept. 1972.

- Use a straight-line regression model to construct a point forecast for emergency room demand for each of the next three Augusts.
- Provide 95% prediction intervals around the forecasts.
- Describe the potential dangers associated with using simple linear regression to forecast demand for emergency room services.
- Which other method described in this chapter would be appropriate for forecasting patient visits to the emergency room?

**13.58 Retail prices of food items.** In 1990, the average weekly food cost for a suburban family of four was estimated to be \$154.40. The table below presents the retail prices of four food items for selected years from 1990 to 2004. Assume a typical suburban family of four purchased the following quantities of food, on average, each week during 1990:

Spaghetti	Ground Beef	Eggs	Potatoes
2 lb.	5 lb.	1 doz.	10 lb.

**FOOD4**

Year	Spaghetti (\$/lb.)	Ground Beef (\$/lb.)	Eggs (\$/doz.)	Potatoes (\$/lb.)
1990	.85	1.63	1.00	.32
1995	.88	1.40	1.16	.38
2000	.88	1.63	.96	.35
2004	.95	2.14	.98	.51

Source: U.S. Bureau of the Census. *Statistical Abstract of the United States*, 2006.

- Calculate a Laspeyres price index for 1990 to 2004, using 1990 as the base year.
- According to your index, how much did the above “basket” of foods increase or decrease in price from 1990 to 2004?

**Applying the Concepts—Intermediate**

**13.59 Mortgage interest rates.** Refer to the annual interest rate time series, Exercise 13.41 (p. 13-43). Use  $w = .3$  and  $v = .7$  to compute the Holt-Winters forecasts for 2005–2007. Compare these to the linear regression forecasts obtained in Exercise 13.41 using MAD, MAPE, and RMSE. [Note: You will need to obtain the actual values of the time series for 2005–2007 to complete this exercise.]

**13.60 Price of Abbott Labs stock.** The stock of Abbott Laboratories has had the yearly closing prices shown in the next table.

**ABBLAB**

Year	Closing Price	Year	Closing Price
1980	56.50	1994	32.05
1981	27.00	1995	41.05
1982	38.75	1996	50.75
1983	45.25	1997	65.50
1984	41.75	1998	49.00
1985	68.37	1999	36.31
1986	45.62	2000	48.44
1987	48.02	2001	55.75
1988	48.01	2002	40.00
1989	64.03	2003	46.60
1990	45.00	2004	46.65
1991	68.07	2005	39.43
1992	30.03	2006	43.80
1993	29.05		

Sources: Standard & Poor’s. *NYSE Daily Stock Price Record*, 1980–2006.

- Use exponential smoothing with  $w = .8$  to forecast the 2007 and 2008 closing prices. If you buy at the end of 2006 and sell at the end of 2008, what is your expected gain (loss)?
- Repeat part a using the Holt-Winters model with  $w = .8$  and  $v = .5$ .
- In which forecast do you have more confidence? Explain.

**13.61 Price of Abbott Labs stock (cont’d).** Refer to Exercise 13.60.

- Fit a simple linear regression model to the stock price data.
- Plot the fitted regression line on a scattergram of the data.
- Forecast the 2007 and 2008 closing prices using the regression model.
- Construct 95% prediction intervals for the forecasts of part c. Interpret the intervals in the context of the problem.
- Obtain the time series residuals for the simple linear model, and use the Durbin-Watson  $d$ -statistic to test for the presence of autocorrelation.

**13.62 Assets of retirement mutual funds.** Annual mutual fund retirement assets (in billions of dollars) for two fund types are given in the table below.

- Compute simple indexes for each of the two time series using 1990 as the base period.
- Construct a time series plot that displays both indexes.
- Using the results of parts a and b, compare and contrast the two types of funds.

**FUND2**

Year	IRA	401(K)
1990	140	35
1994	350	184
1995	476	266
1996	598	346
1997	767	466
1998	960	616
1999	1,234	810
2000	1,232	815
2001	1,161	794
2002	1,034	706
2003	1,307	919
2004	1,490	1,086

Source: U.S. Census Bureau. *Statistical Abstract of the United States*, 2006.

**13.63 Quarterly GDP values.** The gross domestic product (GDP) is the total U.S. output of goods and services valued at market prices. The quarterly GDP values (in billions of dollars) for the period 2001–2005 are given in the accompanying table. Using weights  $w = .5$  and  $v = .5$ , calculate Holt-Winters forecasts for the four quarters of 2006.

🔗 **QTRGDP**

Year	Quarter	GDP	Year	Quarter	GDP
2001	I	9876	2004	I	10,612
	II	9906		II	10,704
	III	9871		III	10,809
	IV	9910		IV	10,897
2002	I	9977	2005	I	10,999
	II	10,032		II	11,089
	III	10,091		III	11,202
	IV	10,096		IV	11,248
2003	I	10,139			
	II	10,230			
	III	10,411			
	IV	10,503			

Source: Standard & Poor's Statistical Service: Current Statistics, 2006; www.bea.gov.

**13.64 Quarterly GDP values (cont'd).** Refer to Exercise 13.63.

- Use the simple linear regression model fit to the 2001–2005 data to forecast the 2006 quarterly GDP. Place 95% prediction limits on the forecasts.
- The GDP values given are *seasonally adjusted*, which means that an attempt to remove seasonality has been made prior to reporting the figures. Add quarterly dummy variables to the model. Use the partial  $F$ -test (discussed in Section 11.9) to determine whether the data indicate the significance of the seasonal component. Does the test support the assertion that the GDP figures are seasonally adjusted?
- Use the seasonal model to forecast the 2006 quarterly GDP values.
- Calculate the time series residuals for the seasonal model, and use the Durbin-Watson test to determine whether the residuals are autocorrelated. Use  $\alpha = .10$ .

**13.65 Quarterly GDP values (cont'd).** Refer to Exercises 13.63 and 13.64. For each of the forecasting models, apply the MAD, MAPE, and RMSE criteria to evaluate the forecasts for the four quarters of 2006. Which of the forecasting models performs best according to each criterion? (You will need to obtain the actual 2006 GDP values to complete this exercise.)

**13.66 Revolving credit loans.** A major portion of total consumer credit extended is in the category of revolving credit loans. Amounts outstanding (in billions of dollars) for the period 1980–2004 are given in the next table.

- Use a simple linear regression model to forecast the 2006 and 2007 values. Place 95% prediction limits on each forecast.
- Calculate the Holt-Winters forecasts for 2006 and 2007 using  $w = .7$  and  $v = .7$ . Compare the results with the simple linear regression forecasts of part **a**.

**13.67 Using the CPI to compute real income.** The number of dollars a person receives in a year is referred to as his or her *monetary* (or *money*) *income*. This figure can be adjusted to

🔗 **LOANS (for Exercise 13.66)**

Year	Revolving	Year	Revolving	Year	Revolving
1980	55	1991	245	2002	738
1981	61	1992	257	2003	759
1982	66	1993	288	2004	794
1983	79	1994	338		
1984	100	1995	443		
1985	122	1996	499		
1986	136	1997	530		
1987	153	1998	579		
1988	174	1999	608		
1989	198	2000	678		
1990	239	2001	722		

Source: U.S. Bureau of the Census. *Statistical Abstract of the United States*, 2006.

reflect the purchasing power of the dollars received relative to the purchasing power of dollars in some base period. The result is called a person's *real income*. The Consumer Price Index (CPI) can be used to adjust monetary income to obtain real income (in terms of 1984 dollars). To compute your real income for a specific year, simply divide your monetary income for that year by that year's CPI and multiply by 100. In Exercise 13.27 (p. 13-28), we listed the CPI for 1990 and 2005 as 125.8 and 195.3, respectively.

- Suppose your monetary income increased from \$50,000 in 1990 to \$95,000 in 2005. What were your real incomes in 1990 and 2005? Were you able to buy more goods and services in 1990 or 2005? Explain.
- What monetary income would have been required in 2005 to provide equivalent purchasing power to a 1990 monetary income of \$20,000?

🔗 **HITECH**

**13.68 IBM stock prices.** Refer to Example 13.4 (p. 13-16), and the monthly IBM stock prices from January 2004 to December 2005. The data are saved in the **HITECH** file.

- Use the exponentially smoothed series (with  $w = .5$ ) from January 2004 to September 2005 to forecast the monthly values of the IBM stock price from October to December 2005. Calculate the forecast errors.
- Use a simple linear regression model fit to the IBM stock prices from January 2004 to September 2005. Let time  $t$  range from 1 to 21, representing the 21 months in the sample. Interpret the least squares estimates.
- With what approximate precision do you expect to be able to predict the IBM stock price using the regression model?
- Give the simple linear regression forecasts and the 95% forecast intervals for the October–December 2005 prices. How does the precision of these forecasts agree with the approximation obtained in part **c**?
- Compare the exponential smoothing forecasts, part **a**, to the regression forecasts, part **d**, using MAD, MAPE, and RMSE.
- What assumptions does the random error component of the regression model have to satisfy in order to make the model inferences (such as the forecast intervals in part **c**) valid?
- Test to determine whether there is evidence of first-order positive autocorrelation in the random error component of the regression model. Use  $\alpha = .05$ . What can you infer about the validity of the model inferences?

## REFERENCES

- Abraham, B., and Ledholter, J. *Statistical Methods for Forecasting*. New York: Wiley, 1983.
- Anderson, T. W. *The Statistical Analysis of Time Series*. New York: Wiley, 1971.
- Box, G. E. P., and Jenkins, G. M. *Time Series Analysis: Forecasting and Control*, 2nd ed. San Francisco: Holden-Day, 1977.
- Chipman, J. S. “Efficiency of Least Squares Estimation of Linear Trend When Residuals Are Autocorrelated.” *Econometrica*, Vol. 47, 1979.
- Durbin, J., and Watson, G. S. “Testing for Serial Correlation in Least Squares Regression, I.” *Biometrika*, Vol. 37, 1950, pp. 409–428.
- Durbin, J., and Watson, G. S. “Testing for Serial Correlation in Least Squares Regression, II.” *Biometrika*, Vol. 38, 1951, pp. 159–178.
- Durbin, J., and Watson, G. S. “Testing for Serial Correlation in Least Squares Regression, III.” *Biometrika*, Vol. 58, 1971, pp. 1–19.
- Fuller, W. A. *Introduction to Statistical Time Series*. New York: Wiley, 1976.
- Granger, C. W. J., and Newbold, P. *Forecasting Economic Time Series*. New York: Academic Press, 1977.
- Greene, W. H. *Econometric Analysis*, 2nd ed. New York: Macmillan, 1993.
- Gross, C. W., and Patterson, R. J. *Business Forecasting*, 2nd ed. Boston: Houghton Mifflin, 1983.
- Hamilton, J. D. *Time Series Analysis*. Princeton: Princeton University Press, 1994.
- Harvey, A. *The Econometric Analysis of Time Series*, 2nd ed. Cambridge: MIT Press, 1990.
- Maddala, G. S. *Econometrics*. New York: McGraw-Hill, 1977.
- Makridakis, S., et al. *The Forecasting Accuracy of Major Time Series Methods*. New York: Wiley, 1984.
- Nelson, C. R. *Applied Time Series Analysis for Managerial Forecasting*. San Francisco: Holden-Day, 1983.
- Shively, T. S. “Fast Evaluation of the Distribution of the Durbin-Watson and Other Invariant Test Statistics in Time Series Regression.” *Journal of the American Statistical Association*, Vol. 85, 1990.
- Theil, H. *Principles of Econometrics*. New York: Wiley, 1971.
- White, K. J. “The Durbin-Watson Test for Autocorrelation in Nonlinear Models.” *Review of Economics and Statistics*, Vol. 74, 1992.
- Willis, R. E. *A Guide to Forecasting for Planners*. Englewood Cliffs, N.J.: Prentice Hall, 1987.

## Using Technology

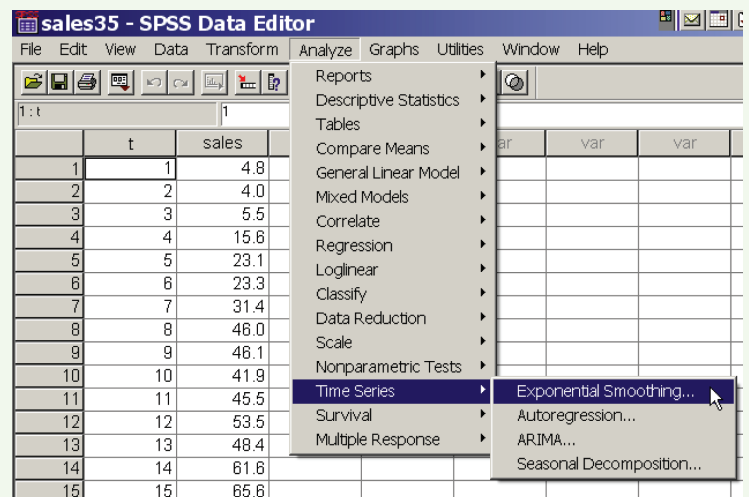
### 13.1 Forecasting Using SPSS



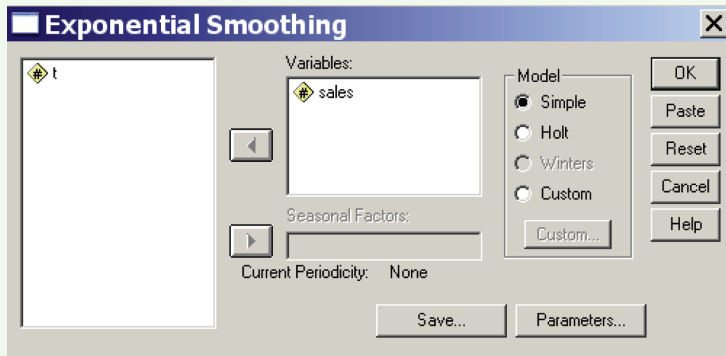
To produce forecasts using exponential smoothing or the Holt-Winters method, click the “Analyze” button on the SPSS main menu bar, then click on “Time Series”, and then click on

“Exponential Smoothing”, as shown in Figure 13.S.1. The resulting dialog box is shown in Figure 13.S.2.

Select the quantitative variable to be smoothed and place it in the “Variables” box. For exponential smoothing, select “Simple” in the “Model” box. For the Holt-Winters method, select “Holt” in the “Model” box. To set the value of the smoothing constants, click

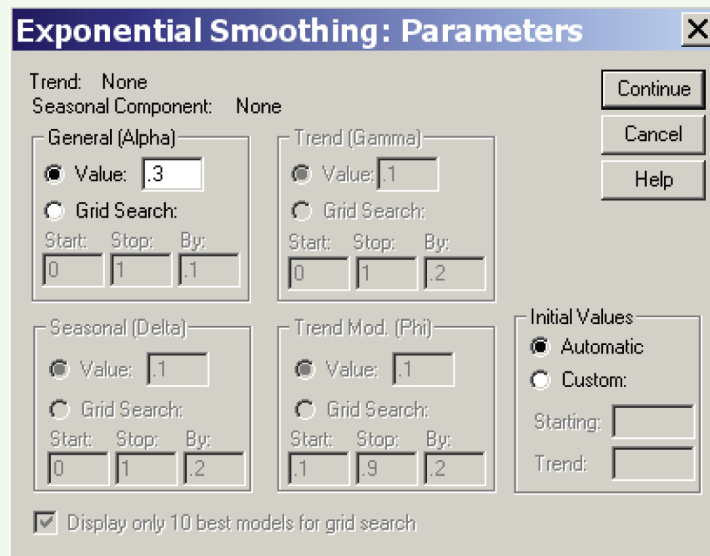


**Figure 13.S.1**  
SPSS options for exponential smoothing



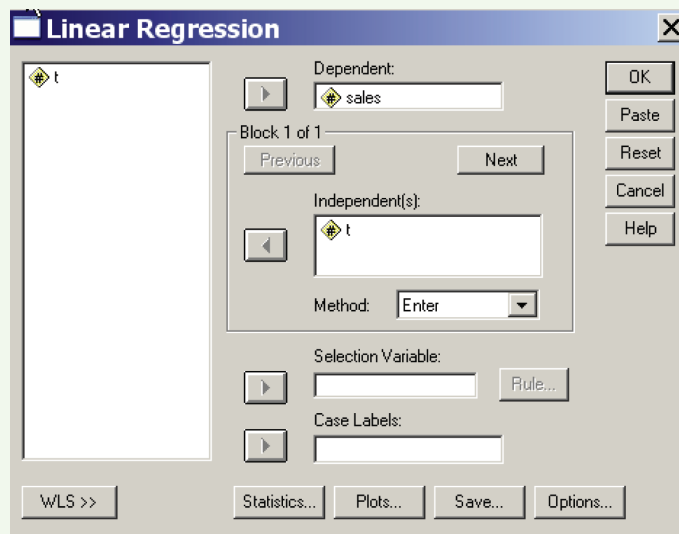
**Figure 13.S.2**  
SPSS exponential smoothing  
dialog box

the “Parameters” button and make your selections on the resulting menu screen, as shown in Figure 13.S.3. Click “Continue” to return to the “Exponential Smoothing” dialog box, then click “OK” to view the results. [Note: Forecasts for each time period in the data set will show up in a column on the SPSS spreadsheet screen.]



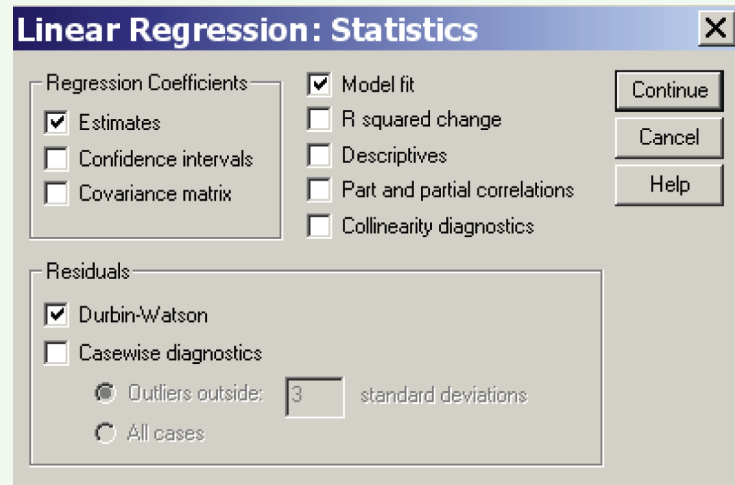
**Figure 13.S.3**  
Selecting exponential  
smoothing parameters

To produce forecasts using a regression model, click the “Analyze” button on the SPSS main menu bar, then click on “Regression” and “Linear”. The resulting dialog box is shown in Figure 13.S.4.



**Figure 13.S.4**  
SPSS linear regression dialog  
box

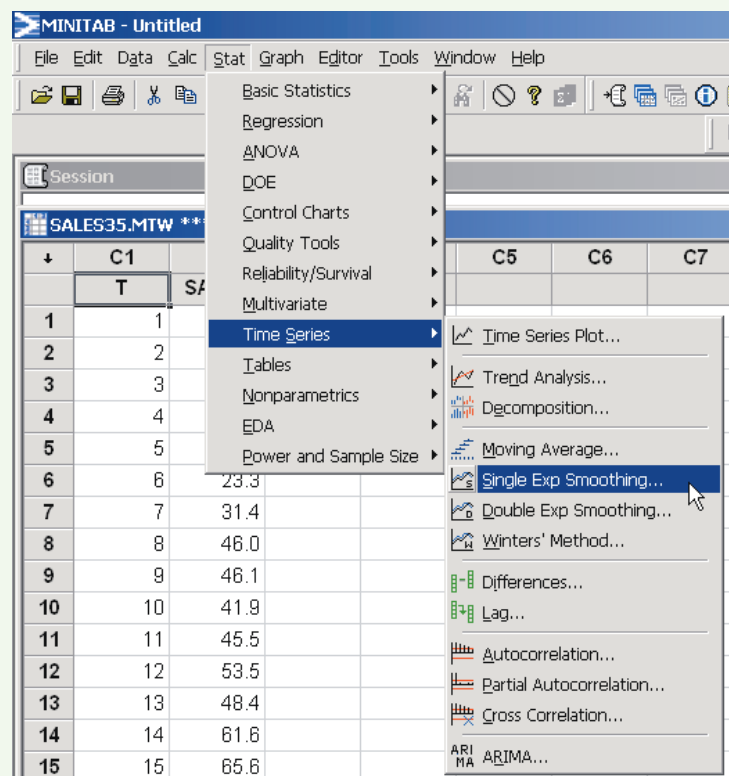
**Figure 13.S.5**  
SPSS linear regression  
statistics menu



Specify the dependent time series variable in the “Dependent” box and the independent variables in the model in the “Independent(s)” box. Click “Save” and make the appropriate menu selections to save the forecasted values as well as 95% prediction intervals. To conduct the Durbin-Watson test for autocorrelated errors, click on the “Statistics” button to obtain the menu shown in Figure 13.S.5. Check the “Durbin-Watson” box, then click “Continue” to return to the “Linear Regression” dialog box. Click “OK” to view the results.

### 13.2 Forecasting Using MINITAB

To produce forecasts using exponential smoothing or the Holt-Winters method, click the “Stat” button on the MINITAB main menu bar, then click on “Time Series”. This will produce the menu list shown in Figure 13.M.1.



**Figure 13.M.1**  
MINITAB options for time series analysis



Click on “Single Exp Smoothing” for the exponential smoothing method or “Double Exp Smoothing” for the Holt-Winter’s method with trend. For example, clicking “Single Exp Smoothing” will result in the dialog box shown in Figure 13.M.2.

Select the quantitative variable to be smoothed and place it in the “Variable” box and set the value of the smoothing constant in the “Weight to use in smoothing” box. Select the “Options” box and specify “1” where MINITAB asks for the number of observations to use for the initial smoothed value. As an option, you can store the forecast values by selecting “Storage” and making the appropriate selections. Click “OK” to view the results.

To produce forecasts using a regression model, click the “Stat” button on the MINITAB main menu bar, then click on “Regression” and “Regression” again. The resulting dialog box is shown in Figure 13.M.3.

Specify the dependent time series variable in the “Response” box and the independent variables in the model in the “Predictors” box. Click “Options” to display the Regression Options dialog box. As shown in Figure 13.M.4, you may select “Durbin-Watson statistic” to conduct a test for autocorrelated errors and/or make selections for producing a prediction interval for a future value of the time series variable. Click “Storage” and make the appropriate menu selections to save the forecasted values as well as 95% prediction intervals. (These values will appear on the MINITAB worksheet.) When all selections have been made, click “OK” on the Linear Regression dialog box to produce the forecasts.

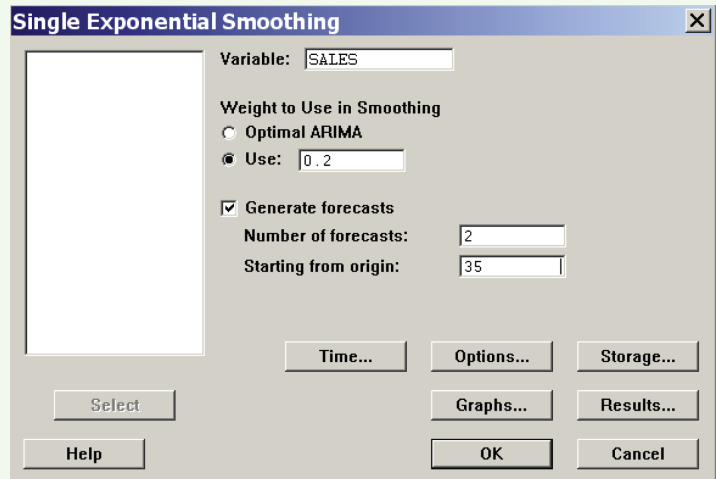


Figure 13.M.2  
MINITAB exponential smoothing dialog box

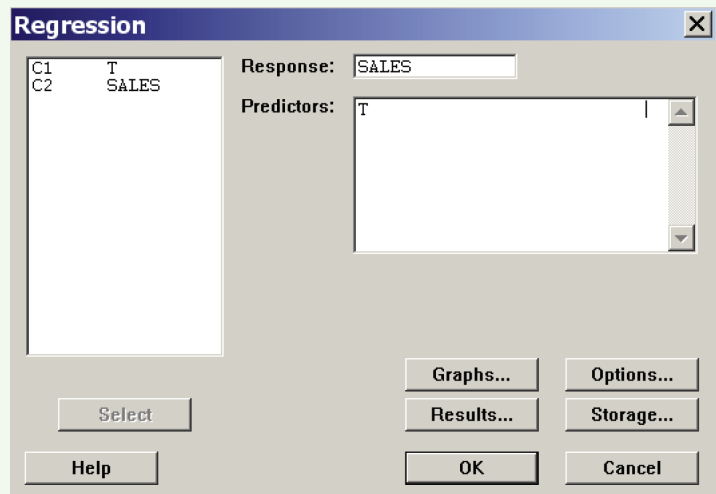


Figure 13.M.3  
MINITAB linear regression dialog box

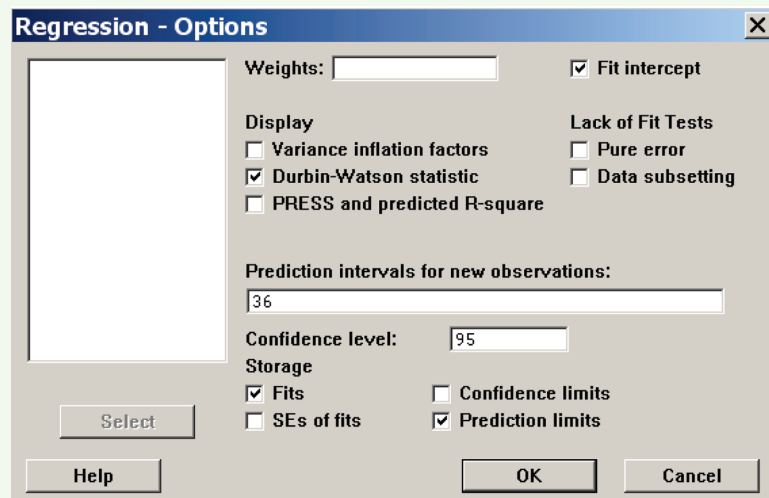


Figure 13.M.4  
MINITAB regression options

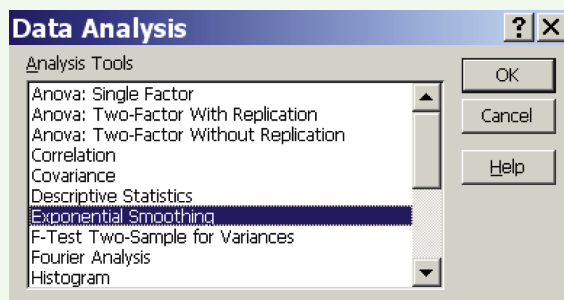
### 13.3 Forecasting Using Excel

To produce forecasts using exponential smoothing, click on the “Tools” button on the Excel main menu bar, then click on “Data Analysis”. Select “Exponential Smoothing” from the resulting menu list, as shown in Figure 13.E.1. Click “OK” to display the Excel Exponential Smoothing dialog box as shown in Figure 13.E.2.

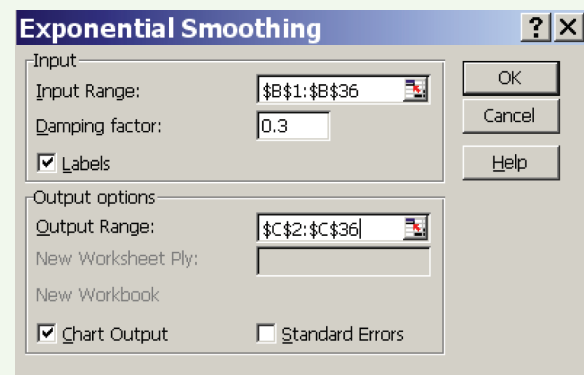
Specify the input range of the quantitative variable to be smoothed, the smoothing constant (“damping factor”), and the output range for where the smoothed values will appear on the Excel worksheet. As an option, you can check “Chart Output” to produce a plot of the smoothed series. Click “OK” to view the results. [Note: The Holt-Winters forecasting method is not currently available in Excel.]

To produce forecasts using a regression model, click the “PHStat” button on the Excel main menu bar, then click on “Regression” and either “Simple Linear Regression” or “Multiple Regression”. The resulting dialog box is similar to the one shown in Figure 13.E.3.

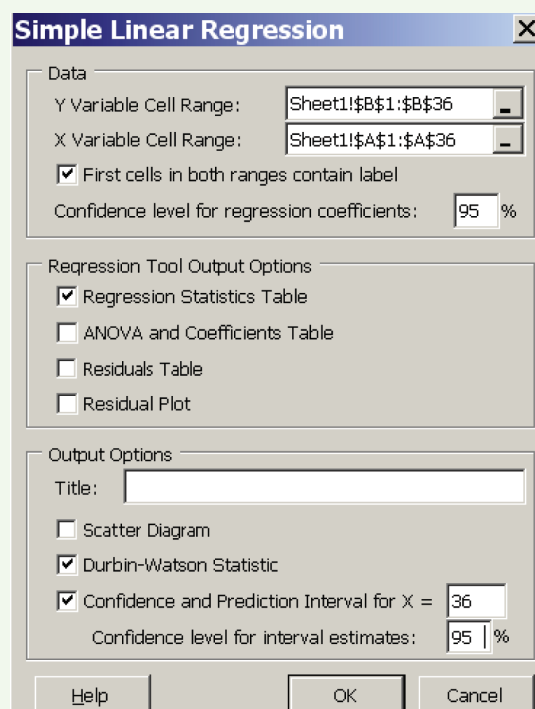
Specify the cell range for the dependent ( $y$ ) time series variable and the cell range for the independent ( $x$ ) variable(s). Check the “Durbin-Watson Statistic” option to conduct a test for autocorrelated errors and/or make selections for producing a prediction interval for a future value of the time series variable. Click “OK” to produce the regression results and forecasts.



**Figure 13.E.1**  
Excel data analysis options



**Figure 13.E.2**  
Excel exponential smoothing dialog box

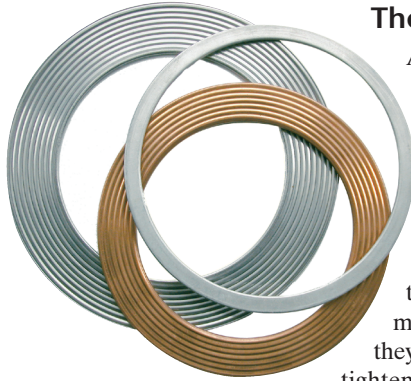


**Figure 13.E.3**  
Excel linear regression dialog box

# Real-World Case

## The Gasket Manufacturing Case

(A Case Covering Chapters 12 and 13)



### The Problem

A Midwestern manufacturer of gaskets for automotive and off-road vehicle applications was suddenly and unexpectedly notified by a major customer—a U.S. auto manufacturer—that they had significantly tightened the specification

limits on the overall thickness of a hard gasket used in their automotive engines. Although the current specification limits were by and large being met by the gasket manufacturer, their product did not come close to meeting the new specification.

The gasket manufacturer's first reaction was to negotiate with the customer to obtain a relaxation of the new specification. When these efforts failed, the customer-supplier relationship became somewhat strained. The gasket manufacturer's next thought was that if they waited long enough, the automotive company would eventually be forced to loosen the requirements and purchase the existing product. However, as time went on, it became clear that this was not going to happen and that some positive steps would have to be taken to improve the quality of their gaskets. But what should be done? And by whom?

### The Product

Figure C5.1 shows the product in question, a hard gasket. A hard gasket is comprised of two outer layers of soft gasket material and an inner layer consisting of a perforated piece of sheet metal. These three pieces are assembled, and some blanking and punching operations follow, after which metal rings are installed around the inside of the cylinder bore clearance holes and the entire outside periphery of the gasket. The quality characteristic of interest in this case is the assembly thickness.

### The Process

An initial study by the staff engineers revealed that the variation in the thickness of soft gasket material—the two outer layers of the hard gasket—was large and undoubted-

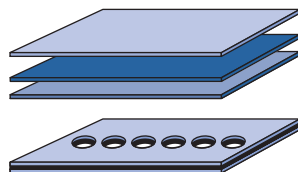


FIGURE C5.1

A hard gasket for automotive applications

ly responsible for much of the total variability in the final product. Figure C5.2 shows the roll mill process that fabricates the sheets of soft gasket material from which the two outer layers of the hard gasket are made. To manufacture a sheet of soft gasket material, an operator adds raw material, in a soft pelletlike form, to the gap—called the *knip*—between the two rolls. The larger roll rotates about its axis with no lateral movement; the smaller roll rotates and moves back and forth laterally to change the size of the knip. As the operator adds more and more material to the knip, the sheet is formed around the larger roll. When the smaller roll reaches a preset destination (i.e., final gap/sheet thickness), a bell rings and a red light goes on, telling the operator to stop adding raw material. The operator stops the rolls and cuts the sheet horizontally along the larger roll so that it may be pulled off the roll. The finished sheet, called a *pull*, is pulled onto a table where the operator checks its thickness with a micrometer. The operator can adjust the final gap if he or she believes that the sheets are coming out too thick or too thin relative to the prescribed nominal value (i.e., the target thickness).

### Process Operation

Investigation revealed that the operator runs the process in the following way. After each sheet is made, the operator measures the thickness with a micrometer. The thickness values for three consecutive sheets are averaged and the average is plotted on a piece of graph paper that, at the start of the shift, has only a solid horizontal line drawn on it to indicate the target thickness value for the particular soft gasket sheet the operator is making. Periodically, the operator reviews these evolving data and makes a decision as to whether or not the process mean—the sheet thickness—needs to be adjusted. This can be accomplished by stopping the machine, loosening some clamps on the small roll, and joggling the small roll laterally in or out by a few thousandths of an inch—whatever the operator feels is needed. The clamps are tightened, the gap is checked with a taper gauge, and if adjusted properly, the operator begins to make sheets again. Typically, this adjustment process takes

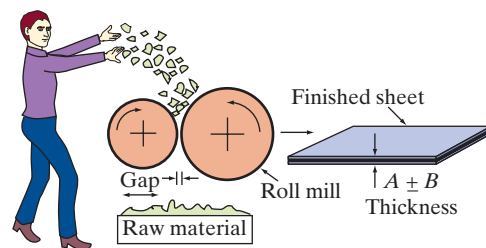


FIGURE C5.2

Roll mill for the manufacture of soft gasket material

10 to 15 minutes. The questions of when to make such adjustments and how much to change the roll gap for each adjustment are completely at the operator's discretion, based on the evolving plot of thickness averages.

Figure C5.3 shows a series of plots that detail the history of one particular work shift over which the operator made several process adjustments. (These data come from the same shift that the staff engineers used to collect data for a process capability study that is described later.) Figure C5.3(a) shows the process data after the first 12 sheets have been made—four averages of three successive sheet thicknesses. At this point the operator judged that the data were telling her that the process was running below the target, so she stopped the process and made an adjustment to slightly increase the final roll gap. She then proceeded to make more sheets. Figure C5.3(b) shows the state of the process somewhat later. Now it appeared to the operator that the sheets were coming out too thick, so she stopped and made another adjustment. As shown in Figure C5.3(c), the process seemed to run well for a while, but then an average somewhat below the target led the operator to believe that another adjustment was necessary. Figures C5.3(d) and C5.3(e) show points in time where other adjustments were made.

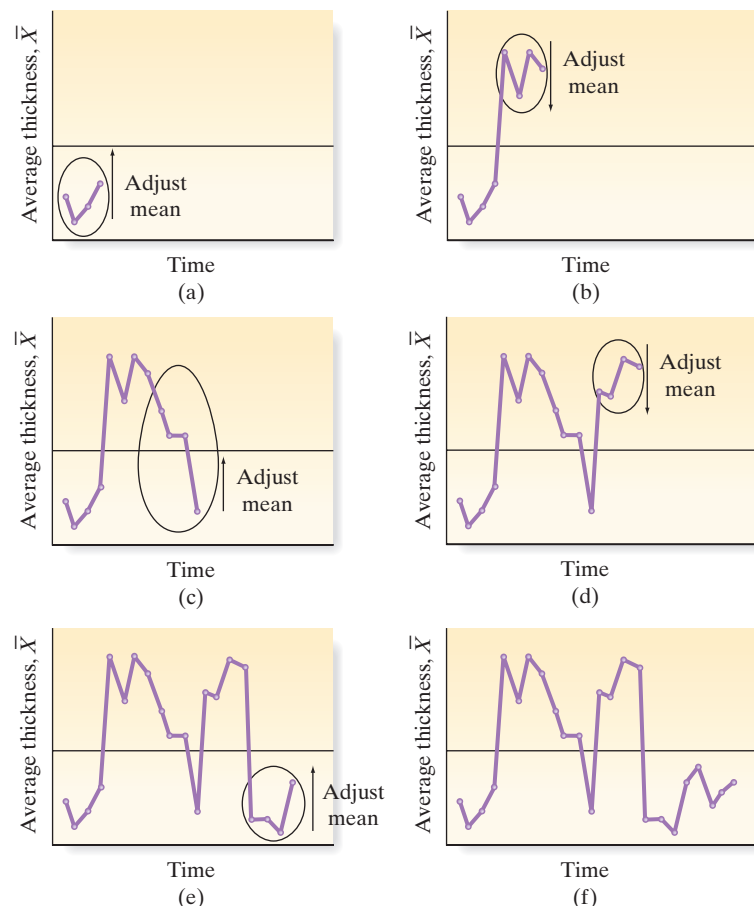
Figure C5.3(f) shows the complete history of the shift. A total of  $24 \times 3$ , or 72, sheets were made during this shift. When asked, the operator indicated that the history of this shift was quite typical of what happens on a day-to-day basis.

### The Company's Stop-Gap Solution

While the staff engineers were studying the problem to formulate an appropriate action plan, something had to be done to make it possible to deliver hard gaskets within the new specification limits. Management decided to increase product inspection and, in particular, to grade each piece of material according to thickness so that the wide variation in thickness could be balanced out at the assembly process. Extra inspectors were used to grade each piece of soft gasket material. Sheets of the same thickness were shipped in separate bundles on pallets to a sister plant for assembly. Thick and thin sheets were selected as needed to make a hard gasket that met the specification. The process worked pretty well, and there was some discussion about making it permanent. However, some felt it was too costly and did not get at the root cause of the problem.

### The Engineering Department's Analysis

Meanwhile, the staff engineers in the company were continuing to study the problem and came to the conclusion that the existing roll mill process equipment for making the soft gasket sheets simply was not capable of meeting the new specifications. This conclusion was reached as a result of the examination of production data and scrap logs over the past several months. They had researched some new equipment that had a track record for very good sheet-to-sheet consistency and had decided to write a proposal to replace the existing roll mills with this new equipment.



**FIGURE C5.3**  
Process adjustment history over one shift



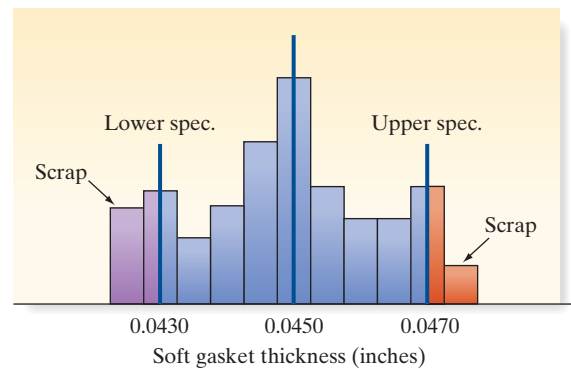
## GASKET

**TABLE C5.1** Measurements of Sheet Thickness

Sheet	Thickness (in.)	Sheet	Thickness (in.)	Sheet	Thickness (in.)
1	0.0440	25	0.0464	49	0.0427
2	0.0446	26	0.0457	50	0.0437
3	0.0437	27	0.0447	51	0.0445
4	0.0438	28	0.0451	52	0.0431
5	0.0425	29	0.0447	53	0.0448
6	0.0443	30	0.0457	54	0.0429
7	0.0453	31	0.0456	55	0.0425
8	0.0428	32	0.0455	56	0.0442
9	0.0433	33	0.0445	57	0.0432
10	0.0451	34	0.0448	58	0.0429
11	0.0441	35	0.0423	59	0.0447
12	0.0434	36	0.0442	60	0.0450
13	0.0459	37	0.0459	61	0.0443
14	0.0466	38	0.0468	62	0.0441
15	0.0476	39	0.0452	63	0.0450
16	0.0449	40	0.0456	64	0.0443
17	0.0471	41	0.0471	65	0.0423
18	0.0451	42	0.0450	66	0.0447
19	0.0472	43	0.0472	67	0.0429
20	0.0477	44	0.0465	68	0.0427
21	0.0452	45	0.0461	69	0.0464
22	0.0457	46	0.0462	70	0.0448
23	0.0459	47	0.0463	71	0.0451
24	0.0472	48	0.0471	72	0.0428

To strengthen the proposal, their boss asked them to include data that demonstrated the poor capability of the existing equipment. The engineers, confident that the equipment was not capable, selected what they thought was the best operator and the best roll mill (the plant has several roll mill lines) and took careful measurements of the thickness of each sheet made on an 8-hour shift. During that shift, a total of 72 sheets/pulls were made. This was considered quite acceptable since the work standard for this process is 70 sheets per shift. The measurements of the sheet thickness (in the order of manufacture) for the 72 sheets are given in Table C5.1. The engineers set out to use these data to conduct a process capability study.

Relying on a statistical methods course that one of the engineers had in college 10 years ago, the group decided to construct a frequency distribution from the data and use it to estimate the percentage of the measurements that fell within the specification limits. Their histogram is shown in Figure C5.4. Also shown in the figure are the upper and lower specification values. The dark shaded part of the histogram represents the amount of the product that lies outside of the specification limits. It is immediately apparent from the histogram that a large proportion of the output does not meet the customer's needs. Eight of the 72 sheets fall outside the specification limits. Therefore, in terms of percent conforming to specifications, the engineers estimated the process capability to be 88.8%. This was clearly unacceptable. This analysis confirmed the engineer's low opinion of the roll mill process equipment. They included it

**FIGURE C5.4**

Histogram of data from process capability study

in their proposal and sent their recommendation to replace the equipment to the president's office.

### Your Assignment

You have been hired as an external consultant by the company's president, Marilyn Carlson. She would like you to critique the engineers' analysis, conclusion, and recommendations.

Suspecting that the engineers' work may be flawed, President Carlson would also like you to conduct your own study and make your own recommendations concerning how to resolve the company's problem. She would like you



## GASKET

**TABLE C5.2** Measurements of Sheet Thickness for a Shift Run with No Operator Adjustment

Sheet	Thickness (in.)	Sheet	Thickness (in.)	Sheet	Thickness (in.)
1	0.0440	25	0.0464	49	0.0427
1	.0445	25	.0443	49	.0445
2	.0455	26	.0450	50	.0471
3	.0457	27	.0441	51	.0465
4	.0435	28	.0449	52	.0438
5	.0453	29	.0448	53	.0445
6	.0450	30	.0467	54	.0472
7	.0438	31	.0465	55	.0453
8	.0459	32	.0449	56	.0444
9	.0428	33	.0448	57	.0451
10	.0449	34	.0461	58	.0455
11	.0449	35	.0439	59	.0435
12	.0467	36	.0452	60	.0443
13	.0433	37	.0443	61	.0440
14	.0461	38	.0434	62	.0438
15	.0451	39	.0454	63	.0444
16	.0455	40	.0456	64	.0444
17	.0454	41	.0459	65	.0450
18	.0461	42	.0452	66	.0467
19	.0455	43	.0447	67	.0445
20	.0458	44	.0442	68	.0447
21	.0445	45	.0457	69	.0461
22	.0445	46	.0454	70	.0450
23	.0451	47	.0445	71	.0463
24	.0436	48	.0451	72	.0456

to use the data reported in Table C5.1 along with the data of Table C5.2, which she ordered be collected for you. These data were collected in the same manner as the data in Table C5.1. However, they were collected during a period of time when the roll mill operator was instructed *not* to adjust the sheet thickness. In your analysis, if you choose to construct control charts, use the same three-measurement subgrouping that the operators use.

Prepare an in-depth written report for the president that responds to her requests. It should begin with an executive summary and include whatever tables and figures are needed to support your analysis and recommendations. [The data file for this case is named **GASKET**. The file contains three variables: sheet number, thickness, and a code for operator adjustments (A) or no adjustments (N).]

This case is based on the experiences of an actual company whose identity is disguised for confidentiality reasons. The case was originally written by DeVor, Chang, and Sutherland (*Statistical Quality Design and Control* [New York: Macmillan Publishing Co., 1992] pp. 298–329) and has been adapted to focus on the material presented in Chapter 12.

## ANSWERS TO SELECTED EXERCISES

**13.5 a.** 127.63 **b.** 149.74 **13.7 a.** 100, 100, 101.03, 101.03, 99.48, 100, 101.55, 100.52, 101.55, 102.58, 104.12, 105.15, 103.61, 104.12, 104.64, 103.09, 103.09, 102.58, 102.06, 102.06, 103.09, 102.58, 103.09, 100.52, 101.55, 100.52 **b.** quantity **c.** 95.10, 95.10, 96.08, 96.08, 94.61, 95.10, 96.57, 95.59, 96.57, 97.55, 99.02, 100, 98.53, 99.02, 99.51, 98.04, 98.04, 97.55, 97.06, 97.06, 98.04, 97.55, 98.04, 95.59, 96.57, 95.59

**13.9 a.** 100.00, 157.61, 158.15, 160.05, 167.39, 174.18, 164.67, 172.28, 188.59, 185.33, 181.79, 209.51, 261.68, 214.95, 258.70, 291.85 **c.** price

**13.11 a.** 51.43, 68.77, 100.00, 158.52, 270.39, 412.59, 581.78, 768.60, 1033.83, 1272.99 **b.** 19.02, 25.43, 36.98, 58.62, 100.00, 152.59, 215.16, 284.26, 382.35, 470.80 **c.** flattens the graph **13.13 a.** Manufacturing: 100.00, 150.52, 197.52, 224.22, 256.11, 297.72; transportation: 100.00, 150.85, 193.88, 220.58, 242.01, 275.85 **c.** Earnings: 100.00, 149.64, 195.01, 224.17, 252.95, 296.82; hours: 100.00, 99.92, 100.51, 100.00, 101.36, 101.61 **13.15**  $\omega = .2$  **13.17 a.** 194.0, 194.0, 194.4, 194.7, 194.4, 194.3, 194.8, 194.9, 195.3, 196.0, 197.2, 198.6, 199.1, 199.7, 200.3, 200.3, 200.2, 200.0, 199.6, 199.3, 199.4, 199.3, 199.5, 198.6, 198.3, 197.6 **b.** 194.0, 194.0, 195.6, 195.9, 193.6, 193.9, 196.4, 195.3, 196.7, 198.5, 201.3, 203.5, 201.5, 201.9, 202.8, 200.6, 200.1, 199.2, 198.2, 198.0, 199.6, 199.1, 199.8, 196.0, 196.8, 195.4 **c.**  $\omega = .2$  series **13.19 a.** 368.00, 432.00, 436.80, 393.76, 386.75, 367.75, 349.55, 358.71, 381.34, 384.27, 388.05, 344.01, 304.00, 283.20, 279.84, 274.37, 302.87, 350.97, 397.39, 434.68 **13.21 a.**  $\omega = .1$ : 926.0, 950.5, 1021.8, 1125.4, 1202.0, 1268.4, 1283.0, 1261.4, 1198.6, 1132.7, 1074.7, 1015.2, 990.7, 979.3, 980.0, 1005.2, 1032.9, 1052.9, 1072.3, 1099.0, 1119.8, 1138.1, 1150.1, 1172.9, 1207.8, 1241.3, 1283.6, 1425.6;  $\omega = .9$ : 926.0, 1146.5, 1611.4, 2013.3, 1904.1, 1869.8, 1459.6, 1106.3, 680.3, 554.0, 553.1, 486.4, 742.5, 862.7, 974.6, 1206.3, 1274.4, 1237.1, 1246.0, 1329.7, 1309.3, 1303.6, 1262.6, 1366.5, 1506.4, 1539.3, 1651.5, 1815.3 **b.**  $\omega = .9$  series **13.25 a.**  $\omega = .3$ : 199.48;  $\omega = .7$ : 199.74 **b.**  $\omega = .7$  and  $v = .3$ :  $F_{2003} = 199.86$ ,  $F_{2004} = 200.02$ ,  $F_{2005} = 200.18$ ;  $\omega = .3$  and  $v = .7$ :  $F_{2003} = 198.58$ ,  $F_{2004} = 199.14$ ,  $F_{2005} = 199.70$  **13.27 a.** yes **b.**  $F_{2006} = 187.53$  **c.**  $F_{2006} = 198.11$  **13.29 a.** forecast for all 3 quarters = 1238.8 **b.** forecast for all 3 quarters = 1198.9 **13.31 a.** forecast for all 12 quarters = 433.47 **b.** forecasts: 433.47, 428.83, 426.12, 430.16, 429.53, 425.71, 428.21, 426.35, 432.13, 444.06, 456.98, 466.84 **c.** forecasts: 454.09, 466.55, 479.01, 491.47, 503.93, 516.39, 528.85, 541.31, 553.77, 566.23, 578.69, 591.15; one-step-ahead forecasts: 454.09, 444.12, 433.57, 433.84, 430.10, 422.67, 425.37, 423.40, 432.74, 452.27, 473.40, 488.19 **13.33 a.** -3.58, -2.14, -4.70 **b.** -4.86, -3.02, -5.18 **c.** MAD = 3.47, MAPE = 1.78, RMSE = 3.63 **d.** MAD = 4.35, MAPE = 2.23, RMSE = 4.46 **13.35 a.** MAD = 73.18, MAPE = 5.59, RMSE = 78.28 **b.** MAD = 24.62, MAPE = 1.89, RMSE = 25.29 **c.** Holt-Winters series with  $w = .7$  and  $v = .5$

**13.37 a.** forecasts for all 5 years = 68,017.2 **b.**  $F_{2001} = 68,718.4$ ,  $F_{2002} = 69,260.5$ ,  $F_{2003} = 69,802.6$ ,  $F_{2004} = 70,344.7$ ,  $F_{2005} = 70,886.8$  **c.** Holt-Winters **13.39 a.**  $E(Y_t) = \beta_0 + \beta_1 t + \beta_2 x_1 + \beta_3 x_2 + \beta_4 x_3$ , where  $x_1 = \{1 \text{ if Qtr. 1, 0 otherwise}\}$ ,  $x_2 = \{1 \text{ if Qtr. 2, 0 otherwise}\}$ ,  $x_3 = \{1 \text{ if Qtr. 3, 0 otherwise}\}$  **b.**  $\hat{Y}_t = 11.49 + .51t - 3.95x_1 - 2.09x_2 - 4.52x_3$ ;  $F = 1275.44$ , reject  $H_0$  **c.** Qtr. 1: (27.22, 29.67); Qtr. 2: (29.59, 32.04); Qtr. 3: (27.67, 30.12); Qtr. 4: (32.70, 35.15) **13.41 a.**  $\hat{Y}_t = 15.52 - .417t$  **b.** 3.84; (1.35, 6.33) **13.43 a.**  $\hat{Y}_t = 385.33 - .363t$  **b.** 372.61, 372.25 **c.** 2004: (340.12, 405.11); 2005: (339.59, 404.91) **13.47 a.** inconclusive **b.** inconclusive **c.** reject  $H_0$  **d.** fail to reject  $H_0$  **13.49 b.** Models statistically useful for Banks 1, 2, 3, 4, and 7 **c.** No evidence of positive autocorrelation for all 9 banks **13.51 a.** yes **b.**  $d = .905$ , reject  $H_0$  **c.** validity in question **13.53 a.** yes **c.** yes **d.** no **13.55 a.** 100.0, 70.3, 59.5, 72.2, 70.2, 124.2 **b.** price **c.** quantities of each product in 1995; quantities of each product in each year **13.57 a.**  $\hat{Y}_t = 38.17 + 7.32t$ ; forecasts: 118.68, 126.00, 133.32 **b.** Year 11: (100.61, 136.75); Year 12: (107.06, 144.94); Year 13: (113.40, 153.24)

**13.59**  $F_{2005} = 5.86$ ,  $F_{2006} = 5.50$ ,  $F_{2007} = 5.14$  **13.61 a.**  $\hat{Y}_t = 47.24 - .065t$  **c.**  $F_{2007} = 45.41$ ,  $F_{2008} = 45.35$  **d.** 2007: (19.94, 70.89); 2008: (19.68, 71.02) **13.63**  $F_{2006,1} = 11,356.88$ ;  $F_{2006,2} = 11,442.24$ ;  $F_{2006,3} = 11,527.60$ ;  $F_{2006,4} = 11,612.96$  **13.67 a.** \$39,745.63; \$48,643.11; 2005 **b.** \$31,049.28