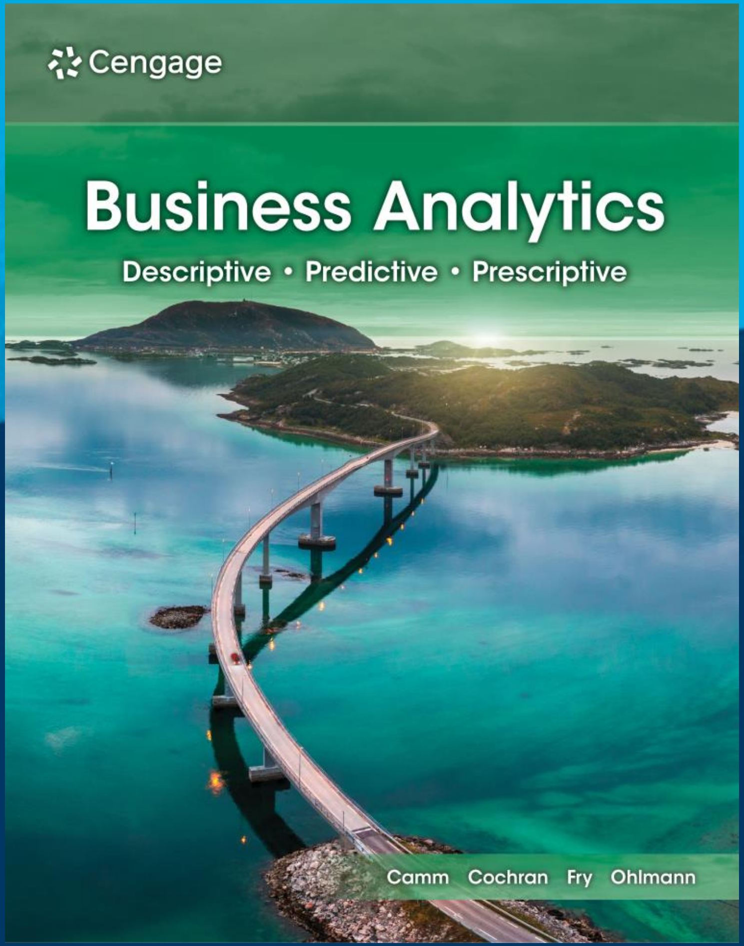


Business Analytics, 5e

Chapter 9 – Time Series Analysis and Forecasting

Camm, Cochran, Fry, Ohlmann, Business Analytics, 5th Edition. © 2024 Cengage Group. All Rights Reserved.
May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chapter Contents

CENGAGE
© 2024 Cengage Group. All Rights Reserved.

Learning Objectives (1 of 2)

After completing this chapter, you will be able to:

LO 9-1	Calculate measures of forecasting accuracy including mean absolute error, mean squared error, and mean absolute percentage error.
LO 9-2	Use measures of forecasting accuracy to choose an appropriate forecasting model.
LO 9-3	Identify an underlying pattern from a time-series plot of data.
LO 9-4	Use simple techniques such as the naïve method and the average of all historical data to forecast a time series that exhibits a horizontal pattern.

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

Learning Objectives (2 of 2)

LO 9-5	Use smoothing techniques such as moving average and exponential smoothing to forecast a time series that exhibits a horizontal pattern.
LO 9-6	Use simple linear regression to forecast a time series that exhibits a linear trend.
LO 9-7	Use multiple linear regression analysis to develop a forecast for a time series.

© 2024 Cengage Group. All Rights Reserved.

Introduction

Forecasting methods can be classified as qualitative or quantitative.

Qualitative methods involve the use of expert judgment to develop **forecasts** when historical data on the variable being forecast is either unavailable or not applicable.

Quantitative methods are used when past information about the forecast variable is available, the information can be quantified, and the past may be assumed as a prologue.

In this chapter, we will focus exclusively on quantitative forecasting methods.

When using historical data of the variable to be forecast, the forecasting procedure is called a time series method and the historical data are referred to as a time series.

CENGAGE
© 2024 Cengage Group. All Rights Reserved.

9.1 Time Series Patterns

A **time series** is a sequence of observations on a variable measured at successive points in time or over successive periods of time,

- such as every hour, day, week, month, or year, etc.

If a past data pattern can be expected to continue in the future, we can use it to guide us in selecting an appropriate forecasting method.

The underlying data pattern is visualized using a **time series plot**,

- a graphical presentation of the relationship with time graphed on the horizontal axis and the time series variable on the vertical axis.

The objective of time series analysis is to uncover a pattern in the historical data or time series.

CENGAGE
© 2024 Cengage Group. All Rights Reserved.

9.1 Horizontal Pattern

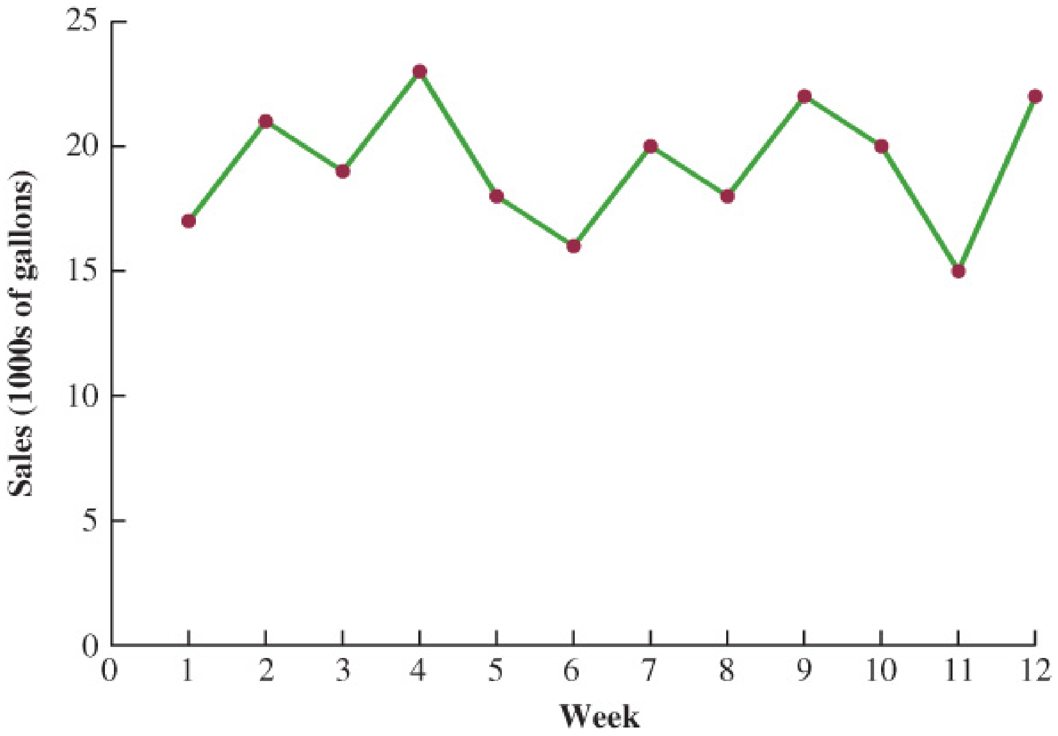
A horizontal pattern exists when the data fluctuate randomly around a constant mean over time.

- Consider the 12 weeks of data in the DATAfile: *gasoline.*

The data show the number of gallons of gasoline sold by a gasoline distributor in Bennington, VT, over the past 12 weeks.

The time series plot for these data shows how the data fluctuate around the sample mean of 19,250 gallons.

Although random variability is present, we would say that these data follow a horizontal pattern.



© 2024 Cengage Group. All Rights Reserved.

9.1 Stationary Time Series

The term **stationary time series** is used to denote a time series in which

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

A time series plot for a stationary time series will always exhibit a horizontal pattern with random fluctuations.

- However, not all horizontal patterns imply a stationary time series.
- Changes in business conditions often result in a time series with a horizontal pattern that shifts to a new level at some point in time.

More advanced texts discuss procedures for determining if a time series is stationary and provide methods for transforming a time series that is nonstationary into a stationary series.

© 2024 Cengage Group. All Rights Reserved.

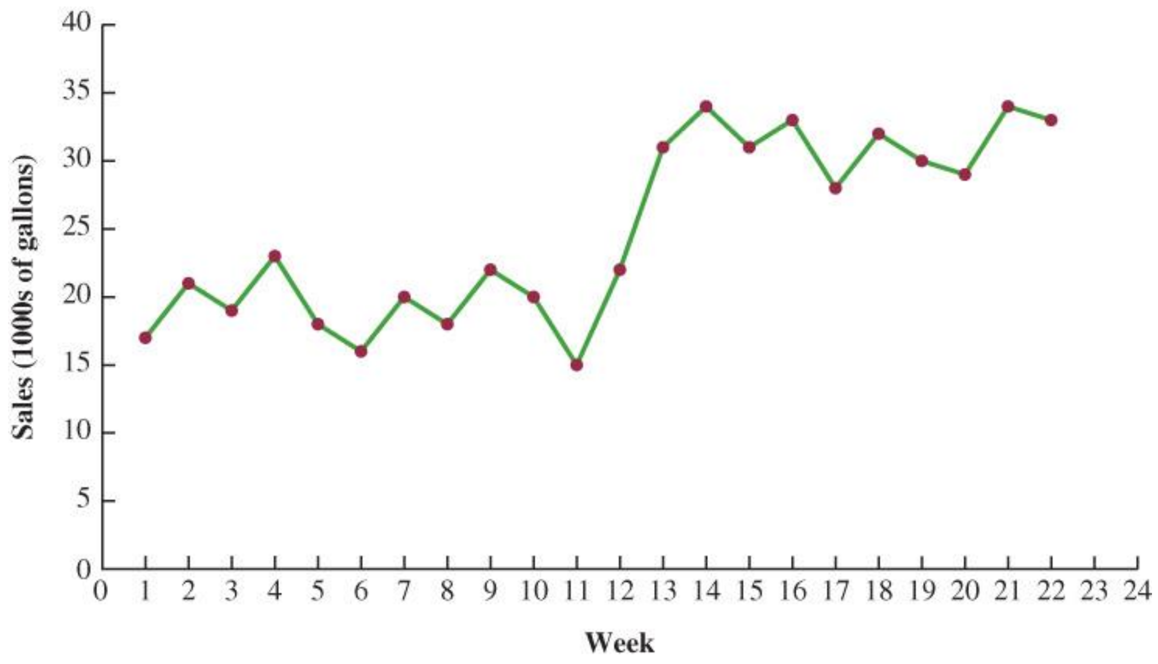
9.1 Shift in Horizontal Pattern

To illustrate a shift in horizontal pattern, consider the 22 weeks included in the DATAfile: *gasolinerevised*.

The gasoline distributor signed a contract with the Vermont State Police beginning on week 13, which is reflected in the time series plot.

This change in the level of the time series makes it more difficult to choose an appropriate forecasting method.

Selecting a forecasting method that adapts well to changes in the level of a time series is always an important consideration.



CENGAGE
© 2024 Cengage Group. All Rights Reserved.

9.1 Linear Trend Pattern

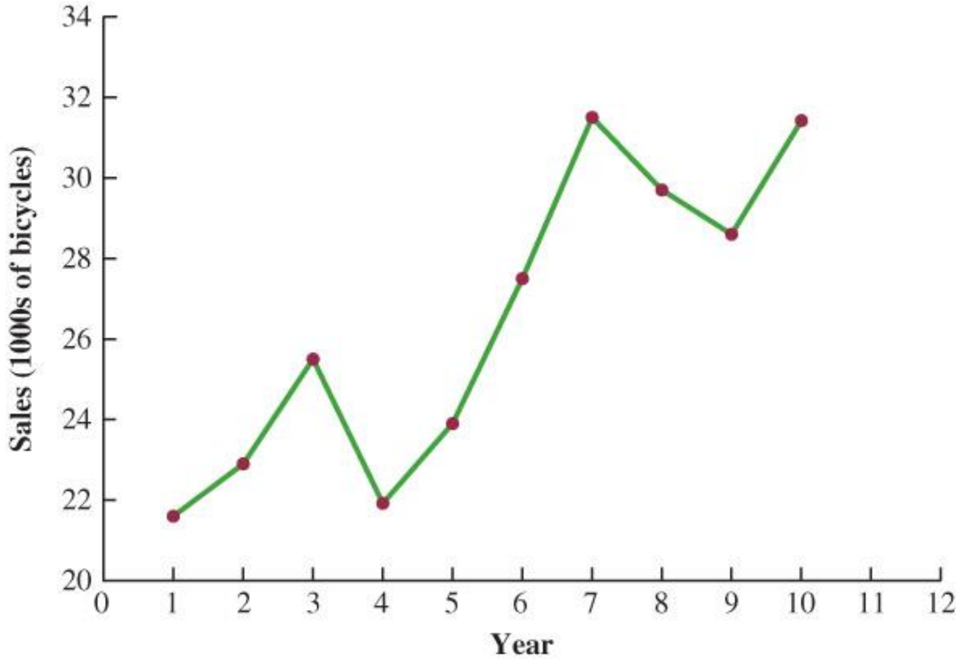
A time series that exhibit show gradual shifts or movements to relatively higher or lower values over a longer period of time it is said to have a **trend pattern**.

- e.g., the result of long-term factors such as population increases or changes in consumer preferences.

To illustrate a time series with a *linear trend pattern*, consider the 10 years of sales for a manufacturer over the past 10 years.

DATAfile: *bicycle*

Up and down movements are visible, but the time series has a systematically increasing upward trend.



CENGAGE

© 2024 Cengage Group. All Rights Reserved.

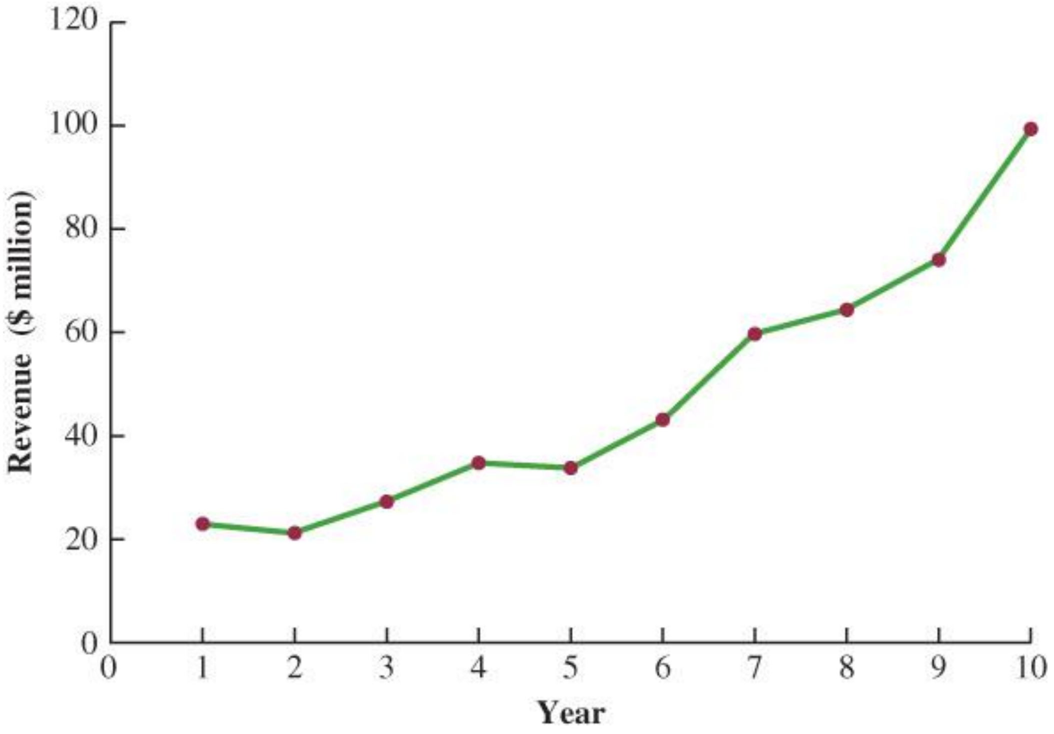
9.1 Non-Linear Trend Pattern

Sometimes, a trend can be described better by a *non-linear trend pattern*.

Consider the DATAfile: *cholesterol* containing the revenue for a cholesterol drug since the company won FDA approval for the drug 10 years ago.

The time series plot shows a nonlinear pattern in which the rate of change of the revenue increases each year.

Non-linear trend patterns in which the time series variable increases exponentially are appropriate when the percentage change from one period to the next is relatively constant.



CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.1 Seasonal Pattern

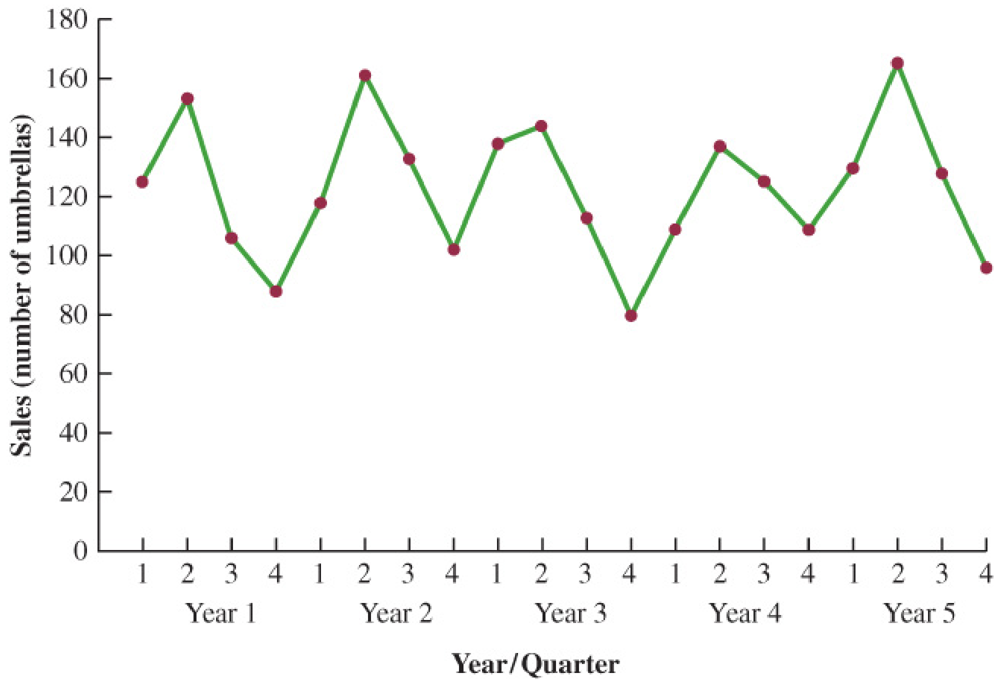
A **seasonal pattern** is a recurring pattern over successive periods of time.

Time series data may exhibit seasonal patterns of less than one year in duration, such as daily traffic flow or weekly restaurant sales.

DATAfile: *umbrella*

The quarterly sales of umbrellas at a clothing store over the past five years display a horizontal pattern with yearly seasonal fluctuations.

- Note how the highest sales occur on the second quarter and bottoming on the fourth quarter of each year.



CENGAGE

© 2024 Cengage Group. All Rights Reserved.

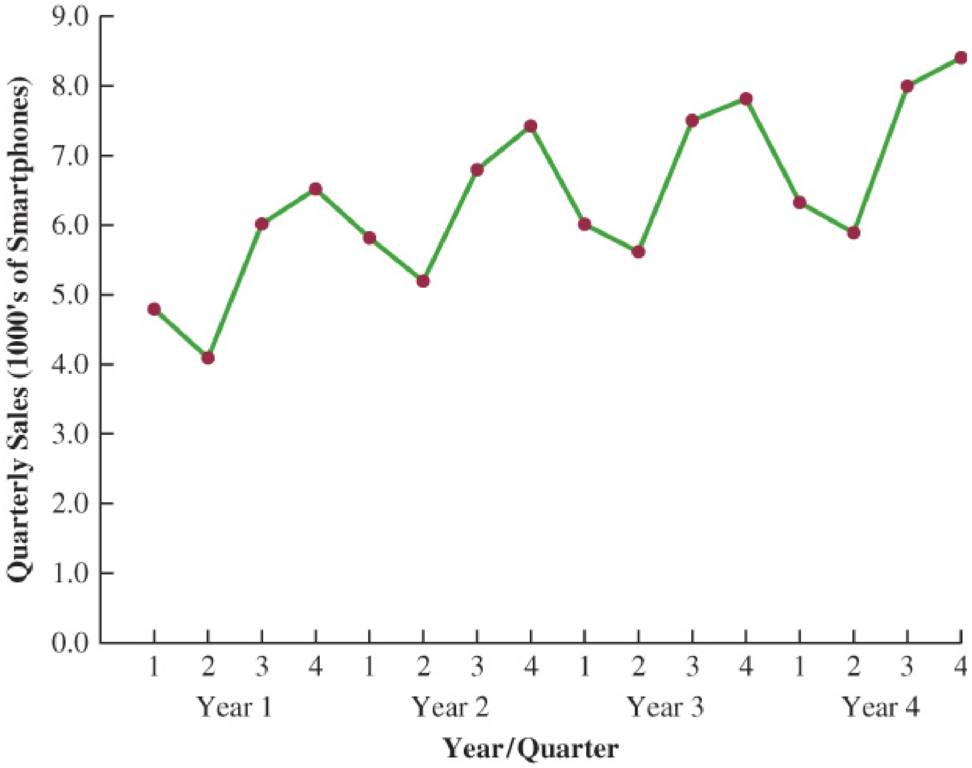
9.1 Trend and Seasonal Pattern

The smartphone sales data for a particular manufacturer over the past four years included in the DATAfile *smartphonesales*.

The time series plot exhibits quarterly sales with

- an increasing linear trend, and
- a seasonal pattern, lowest in quarters 1 and 2 and highest in quarters 3 and 4.

In such cases, we must use a forecasting method able to deal with both trends and seasonality.



CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.1 Cyclical Pattern

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

- Economic time series commonly exhibit cyclical behavior with regular runs of observations below and above the trend line.

The cyclical component may be due to multi-year business cycles.

- Periods of moderate inflation followed by periods of rapid inflation can lead to time series that alternate below and above a generally increasing trend line (e.g., a time series for housing costs.)

Because business cycles are extremely difficult, if not impossible, to forecast, cyclical effects are often combined with long-term trend effects and referred to as trend-cycle effects.

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.2 Naïve Forecasting Method

We begin developing forecasts for the gasoline time series included in the DATAfile: *gasoline* using the simplest of all the forecasting methods:

> We use the most recent week's sales volume as the forecast for the next week.

Because of its simplicity, this method is often referred to as a **naïve forecasting method** (column labeled *Forecast* in the table.)

To measure the accuracy of a forecasting method, in the next slides, we introduce several measures of forecasting accuracy.

Week	Time Series Value	Forecast
1	17	
2	21	17
3	19	21
4	23	19
5	18	23
6	16	18
7	20	16
8	18	20
9	22	18
10	20	22
11	15	20
12	22	15

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.2 Forecast Error

The **forecast error** for time period *t* is the difference between the actual value y_t and the forecasted value $\hat{y}_t$ for period *t*.

$$\boldsymbol{e_t = y_t - \hat{y}_t}$$

A simple measure of forecast accuracy is the **mean forecast error (MFE)**.

$$\boldsymbol{MFE} = \frac{\sum_{t=k+1}^{n} \boldsymbol{e_t}}{\boldsymbol{n-k}} = \frac{5}{12-1} = 0.45$$

Because $MFE > 0$, this forecast method is under-forecasting. See notes for details.

Week	Time Series Value	Forecast	Forecast Error
1	17		
2	21	17	4
3	19	21	-2
4	23	19	4
5	18	23	-5
6	16	18	-2
7	20	16	4
8	18	20	-2
9	22	18	4
10	20	22	-2
11	15	20	-5
12	22	15	7
		Total	5

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.2 Mean Absolute Error

The **mean absolute error (MAE)**, also referred to as MAD*, is a measure of forecast accuracy that averages the absolute values of the forecast errors.

$$MAE = \frac{\sum_{t=k+1}^{n} |e_t|}{n - k} = \frac{41}{12 - 1} = 3.73$$

We adjust the denominator from n to $n-k$, where k is the number of observations without a forecast.

For the naïve forecast of the gasoline data, we have $k = 1$ and $n-k = 11$.

Week	Time Series Value	Forecast	Forecast Error	Abs. Value Forecast Error
1	17			
2	21	17	4	4
3	19	21	-2	2
4	23	19	4	4
5	18	23	-5	5
6	16	18	-2	2
7	20	16	4	4
8	18	20	-2	2
9	22	18	4	4
10	20	22	-2	2
11	15	20	-5	5
12	22	15	7	7
		Totals	5	41

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.2 Mean Square Error

The **mean squared error (MSE)** also avoids the problem of offsetting positive and negative forecast errors by squaring them.

$$\boldsymbol{MSE} = \frac{\sum_{t=k+1}^{n} e_t^2}{n-k} = \frac{179}{12-1} = 16.27$$

Because MAE and MSE depend upon the scale of the data, they make it difficult to compare different time intervals or across different time series.

The next measure helps with that.

Week	Time Series Value	Forecast	Forecast Error	Squared Forecast Error
1	17			
2	21	17	4	16
3	19	21	-2	4
4	23	19	4	16
5	18	23	-5	25
6	16	18	-2	4
7	20	16	4	16
8	18	20	-2	4
9	22	18	4	16
10	20	22	-2	4
11	15	20	-5	25
12	22	15	7	49
		Totals	5	179

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.2 Mean Absolute Percentage Error

The **mean absolute percentage error (MAPE)** allows for comparisons that do not depend upon the scale of the data by computing the absolute error as a percentage of the actual value y_t.

$$\boldsymbol{MAPE} = \frac{\sum_{t=k+1}^{n} \left| \left(\frac{e_t}{y_t} \right) 100 \right|}{n-k} = \frac{211.69}{11}$$

$$= 19.24\%$$

Where $(e_t/y_t)100$ is the percentage error for the forecast at time *t*.

Week	Time Series Value	Forecast	Forecast Error	Abs. Value Forecast % Error
1	17			
2	21	17	4	19.05
3	19	21	-2	10.53
4	23	19	4	17.39
5	18	23	-5	27.78
6	16	18	-2	12.50
7	20	16	4	20.00
8	18	20	-2	11.11
9	22	18	4	18.18
10	20	22	-2	10.00
11	15	20	-5	33.33
12	22	15	7	31.82
		Totals	5	211.69

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.2 Average of All Past Values Forecast Error

When comparing different forecasts, we should select the forecasting method that fits best the historical time series data if

- we think that the historical pattern will continue into the future.

Suppose we forecast the gasoline time series data using the average of past values.

Thus, if we use the notation $\hat{y}_t$ for the forecast at time period *t*, we have

$\hat{y}_2 = y_1 = 17$

$\hat{y}_3 = (y_1 + y_2)/2 = 19$

$\hat{y}_4 = (y_1 + y_2 + y_3)/3 = 19$, and so on

Week	Time Series Value	Forecast	Forecast Error
1	17		
2	21	17	4
3	19	19	0
4	23	19	4
5	18	20	−2
6	16	19.6	−3.6
7	20	19	1
8	18	19.14	−1.14
9	22	19	3
10	20	19.33	0.67
11	15	19.4	−4.4
12	22	19	3.00
		Total	4.52

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.2 Compare Forecasting Models

We can now proceed with the calculations of MAE, MSE, and MAPE as we did for the naïve method. The results are summarized in the table below.

When comparing the accuracy of the two forecasting methods, we see that the average of past values method is more accurate for each measure of forecasting accuracy.

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

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.3 Smoothing Methods

In this section, we discuss two forecasting methods appropriate for a time series with a horizontal pattern.

- Moving averages
- Exponential smoothing

Because the objective of each of these methods is to “smooth out” random fluctuations in the time series, they are referred to as *smoothing methods*.

Smoothing methods are

- not appropriate when trend, cyclical, or seasonal effects are present,
- capable of adapting well to changes in the level of a horizontal pattern,
- easy to use, and generally provide a high level of accuracy for short-range forecasts, such as a forecast for the next time period.

© 2024 Cengage Group. All Rights Reserved.

9.3 Moving Averages

The moving averages method uses the average of the most recent *k* data values in the time series as the forecast for the next period.

A **moving average** forecast of order *k* is:

$$\boldsymbol{\hat{y}_{t+1}} = \frac{\sum(\textbf{most recent k data values})}{\boldsymbol{k}} = \frac{\sum_{i=t-k+1}^{t} y_i}{k} = \frac{y_{t-k+1} + \cdots + y_{t-1} + y_t}{k}$$

Where:

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

y_t is actual value of the time series in period *t*

k is the number of periods used to generate the forecast

© 2024 Cengage Group. All Rights Reserved.

9.3 Moving Average Considerations

The term ‘*moving*’ is used because every time a new observation becomes available for the time series, it replaces the oldest observation in the equation and a new average is computed.

- Thus, the periods over which the average is calculated, *move* with each ensuing period.

To use moving averages to forecast a time series, we must first select the order k, or the number of time series values to be included in the moving average.

The value of k is selected based on the number of past values that are considered relevant.

- The greater the number of relevant past values, the larger the value selected for k.

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.3 Moving Average of the Gasoline Time Series

A time series with a horizontal pattern can shift to a new level over time.

A moving average will adapt to the new level of the series and resume providing good forecasts in *k* periods.

- A smaller value of k will track shifts in a time series more quickly.
  - The naïve approach is actually a moving average with $k = 1$.
- Larger values of k are more effective to smooth out random fluctuations.

Thus, managerial judgment based on an understanding of the behavior of a time series is helpful in choosing an appropriate value of k.

A summary of a three-week ($k = 3$) moving average calculations for the gasoline time series data follows. See notes for Excel instructions.

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.3 Three-Week Moving Average Calculations

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

© 2024 Cengage Group. All Rights Reserved.

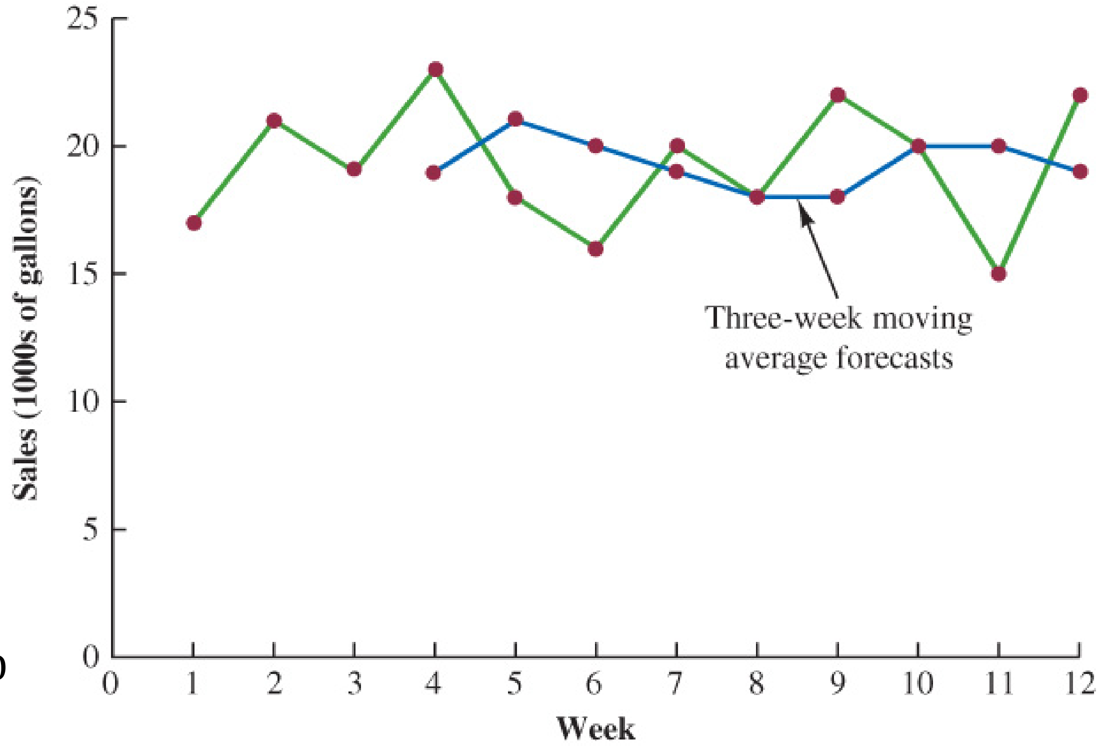
9.3 Moving Average Forecast Accuracy

The values for the three measures of forecast accuracy are:

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

$$MSE = \frac{\sum_{t=4}^{12} e_t^2}{12-3} = \frac{92}{9} = 10.22 \quad (\text{*see notes})$$

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



In each case, the three-week moving average approach provides a more accurate forecast than simply using the most recent observation as the forecast.

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.3 Exponential Smoothing

The **exponential smoothing forecast** uses a weighted average of past time series values as a forecast.

- The weights are computed automatically and become smaller as the observations move farther into the past.

$$\boldsymbol{\hat{y}_{t+1} = \alpha y_t + (1 - \alpha)\hat{y}_t}$$

Where:

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

y_t is the actual value of the time series in period t

$\hat{y}_t$ is the forecast of the time series for period t

α is the smoothing constant ($0 \leq \alpha \leq 1$).

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.3 Exponential Smoothing as Weighted Average

The exponential smoothing forecast for any period can be viewed as the weighted average of *all the previous actual values* of the time series.

Consider a time series involving three periods of data: y_1, y_2, and y_3. By setting $\hat{y}_1 = y_1$, the forecast for period 2 can be written as

$$\hat{y}_2 = \alpha y_1 + (1-\alpha)\hat{y}_1 = \alpha y_1 + (1-\alpha)y_1 = y_1$$

Using the forecast for period 2, the forecast for period 3 can be written as

$$\hat{y}_3 = \alpha y_2 + (1-\alpha)\hat{y}_2 = \alpha y_2 + (1-\alpha)y_1$$

Thus, the forecast for period 4 is

$$\hat{y}_4 = \alpha y_3 + (1-\alpha)\hat{y}_3 = \alpha y_3 + (1-\alpha)(\alpha y_2 + (1-\alpha)y_1)$$
$$= \alpha y_3 + \alpha(1-\alpha)y_2 + (1-\alpha)^2 y_1 \quad \text{where: } \alpha + \alpha(1-\alpha) + (1-\alpha)^2 = 1$$

$\hat{y}_4$ is a weighted average of y_1, y_2, and y_3, and the sum of the weighs equals 1.

CENGAGE
© 2024 Cengage Group. All Rights Reserved.

9.3 Exponential Smoothing of the Gasoline Time Series

Let us apply the exponential smoothing method to forecasting the gasoline sales time series using a smoothing constant $\alpha = 0.2$.

$$\hat{y}_2 = y_1 = 17.00$$

$$\hat{y}_3 = 0.2(21) + (1 - 0.2)(17.00) = 17.80$$

$$\hat{y}_4 = 0.2(19) + (1 - 0.2)(17.80) = 18.04$$

$$\vdots$$

$$\hat{y}_{12} = 0.2(15) + (1 - 0.2)(19.35) = 18.48$$

A summary of the exponential smoothing calculations for the gasoline time series data with smoothing constant $\alpha = 0.2$ is shown to the right. See notes for Excel.

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

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.3 Exponential Smoothing Forecast Accuracy

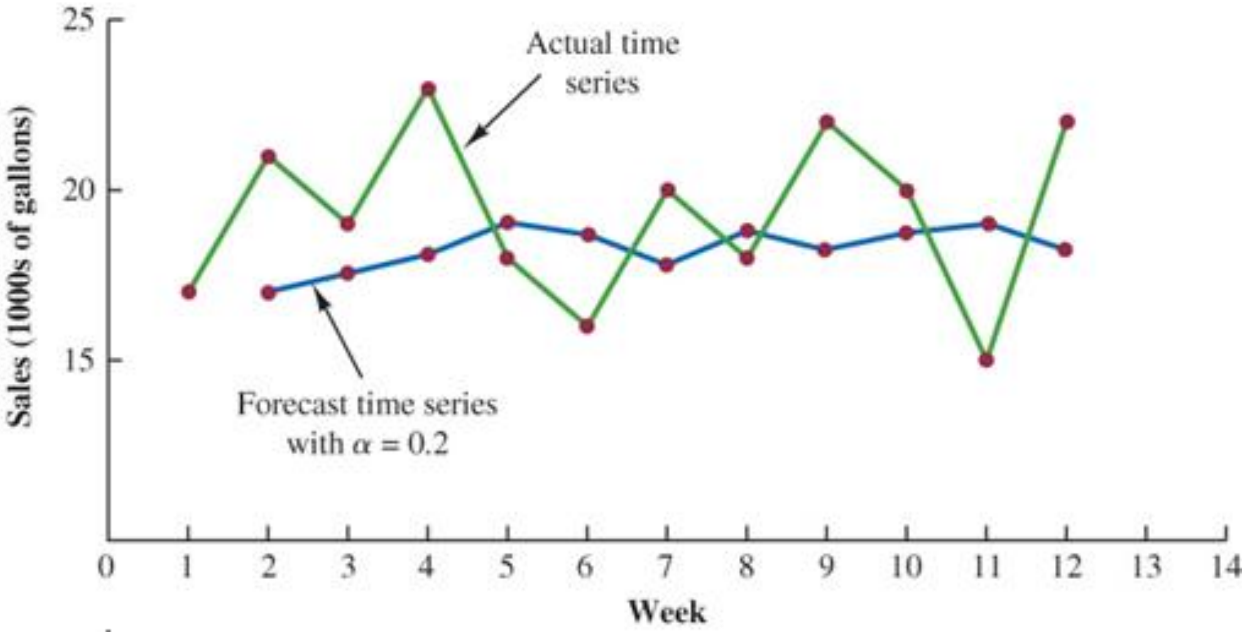
A more suitable value for α can be obtained by rewriting the basic exponential smoothing model as follows:

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t = \alpha y_t + \hat{y}_t - \alpha\hat{y}_t = \hat{y}_t + \alpha(Y_t - \hat{y}_t) = \hat{y}_t + \alpha e_t$$

Thus, $\hat{y}_{t+1}$ is equal to $\hat{y}_t$ plus the smoothing constant times the most recent forecast error.

See notes for details.

$MSE = 98.80/11 = 8.98$, which is lower than the $MSE = 10.22$ that we observed earlier for the three-week moving average forecast.



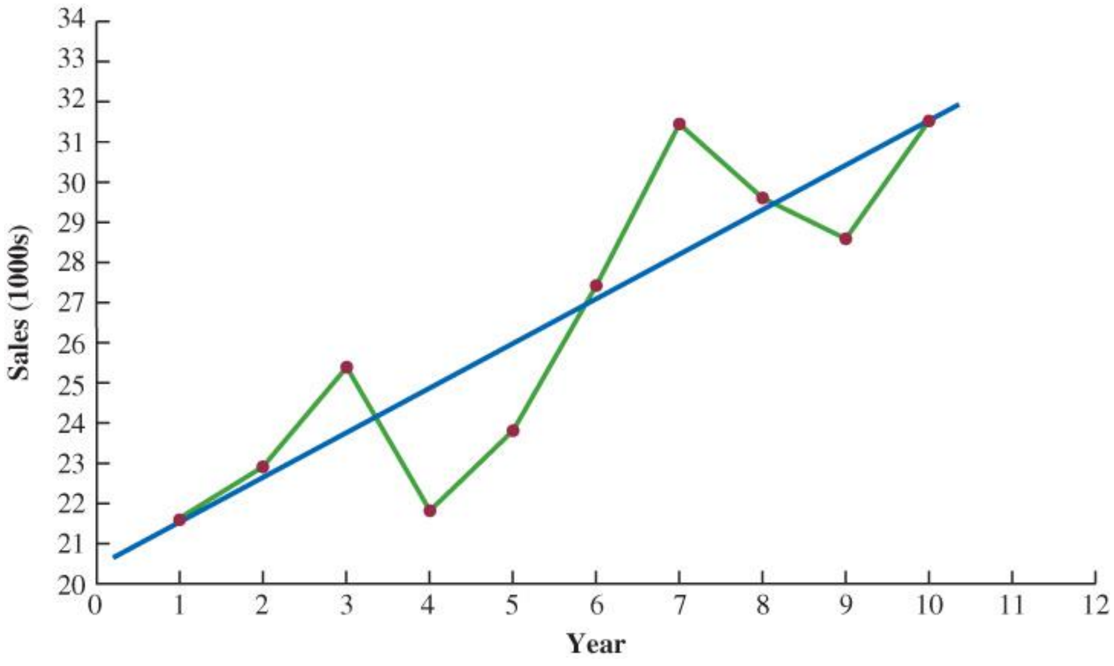
CENGAGE
© 2024 Cengage Group. All Rights Reserved.

9.4 Linear Trend Projection

We can use **linear regression analysis** to forecast a time series with a linear trend as is the case for the previously introduced bicycle sales time series (DATAfile: *bicycle.*)

Although, the time series plot shows up and down movement, a linear trend provides a reasonable description of the long-run movement in the bicycle sales time series data.

Estimating the linear relationship requires finding the values of b_0 and b_1 for the equation of a line so that the MSE is minimized.



CENGAGE
© 2024 Cengage Group. All Rights Reserved.

9.4 Linear Trend Equation: Line of Best Fit

The notation for the linear trend equation is:

$$\hat{y}_t = b_0 + b_1 t$$

Where:

$\hat{y}_t$ is the linear trend forecast in period t

t is the time period

b_0 is the intercept of the linear trend line

b_1 is the slope of the linear trend line.

Excel can be used to compute the estimated y-intercept b_0 and slope b_1.

- The Excel output for a regression analysis of the bicycle sales data is provided in the next slide.

© 2024 Cengage Group. All Rights Reserved.

9.4 Regression Output for the Bicycle Sales Data

	A	B	C	D	E	F	G	H	I
1	SUMMARY OUTPUT								
2									
3	*Regression Statistics*								
4	Multiple R	0.874526167							
5	R Square	0.764796016							
6	Adjusted R Square	0.735395518							
7	Standard Error	1.958953802							
8	Observations	10							
9									
10	ANOVA								
11		*df*	*SS*	*MS*	*F*	*Significance F*			
12	Regression	1	99.825	99.825	26.01302932	0.000929509			
13	Residual	8	30.7	3.8375					
14	Total	9	130.525						
15									
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 99.0%*	*Upper 99.0%*
17	Intercept	20.4	1.338220211	15.24412786	3.39989E-07	17.31405866	23.48594134	15.90975286	24.89024714
18	Year	1.1	0.215673715	5.100296983	0.000929509	0.60265552	1.59734448	0.376331148	1.823668852

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.4 Prediction with the Linear Trendline

The linear trend regression equation for the bicycle sales time series data is

$$\hat{y}_t = 20.4 + 1.10t$$

The slope, $b_1 = 1.10$ can be interpreted as an average growth of 1,100 units per year over the past 10 years. See notes for forecast accuracy.

We can then use the trend equation to project the trend component of the time series for year 11 as $\hat{y}_{11} = 20.4 + 1.10(11) = 32.5$

- The linear model yields a sales forecast of 32,500 bicycles for next year.

Using the linear trend equation, we can also develop annual forecasts for two and three years into the future as

$$\hat{y}_{12} = 20.4 + 1.10(12) = 33.6 \qquad \hat{y}_{13} = 20.4 + 1.10(13) = 34.7$$

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.4 Other Regression-Based Forecasting Models

We can use more complex regression models to fit nonlinear trends.

For example, to generate a forecast of a time series with a curvilinear trend, we could include t^2 and t^3 as independent variables in our model.

$$\hat{y}_t = b_0 + b_1 t + b_2 t^2 + b_3 t^3$$

We can also build linear regression-based forecasting models called **autoregressive models**, in which the independent variables are previous values of the same time series.

$$\hat{y}_t = b_0 + b_1 y_{t-1} + b_2 y_{t-2} + b_3 y_{t-3}$$

See notes for further information on autoregressive models.

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.4 Seasonality Without Trend

The quarterly sales of umbrellas at a store over the past five years are an example of a seasonal pattern with no long-term trend (DATAfile: *umbrella.*)

We can model a time series with a seasonal pattern by treating the season as a categorical variable with four levels, one for each quarter.

When a categorical variable has 4 levels, we need $4 - 1 = 3$ dummy variables.

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

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

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

© 2024 Cengage Group. All Rights Reserved.

9.4 Umbrella Sales Time Series with Dummy Variables

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

© 2024 Cengage Group. All Rights Reserved.

9.4 Seasonality Without Trend: Regression Equation

The general form of the equation relating the quarterly number of umbrellas sold is:

$$\hat{y}_t = b_0 + b_1 Qtr1 + b_2 Qtr2 + b_3\ Qtr3$$

We can use the coded values of the dummy variables and the computer output for multiple regression (see previous slide) to find the values of b_0, b_1, b_2, and b_3.

$$\hat{y}_t = 95.0 + 29.0\ Qtr1 + 57.0\ Qtr2 + 26.0\ Qtr3$$

And use it to forecast quarterly sales for next year:

$$\text{Qtr1 sales} = 95 + 29(1) + 57(0) + 26(0) = 124$$

$$\text{Qtr2 sales} = 95 + 29(0) + 57(1) + 26(0) = 152$$

$$\text{Qtr3 sales} = 95 + 29(0) + 57(0) + 26(1) = 121$$

$$\text{Qtr4 sales} = 95 + 29(0) + 57(0) + 26(0) = 95$$

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.4 Seasonal Models Based on Monthly Data

We showed how dummy variables can be used to account for the quarterly seasonal effects in the time series using three dummy variables.

For monthly data, *season* is a categorical variable with 12 levels. Thus, we need $12 - 1 = 11$ dummy variables coded as follows:

$$\text{Month1} = \begin{cases} 1 \text{ if Jan} \\ 0 \text{ if not Jan} \end{cases} \quad \text{Month2} = \begin{cases} 1 \text{ if Feb} \\ 0 \text{ if not Feb} \end{cases} \quad \ldots \quad \text{Month11} = \begin{cases} 1 \text{ if Nov} \\ 0 \text{ if not Nov} \end{cases}$$

December is obtained by setting all 11 dummy variables equal to 0.

Other than this change, the approach for handling seasonality remains the same. Time series data collected at other intervals can be handled in a similar manner.

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.4 Seasonality With Trend

The smartphone sales data over the past four years are an example of a seasonal pattern with a long-term trend (DATA file: *smartphonesales.*)

We can model the smart phone sales using three quarterly dummy variables with the addition of a time period *t* to account for the linear trend.

$$\hat{y}_t = b_0 + b_1 Qtr1 + b_2 Qtr2 + b_3\, Qtr3 + b_4\, t$$

Where:

t is the time period

$\hat{y}_t$ the forecast of sales in period t

$Qtr1$, $Qtr2$, $Qtr3$, are the seasonal dummy variables, defined as in the seasonality without trend example.

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.4 Smart Phone Sales Time Series with Dummy Variables

Period	Year	Quarter	Qtr1	Qtr2	Qtr3	Sales (1,000s)
1	1	1	1	0	0	4.8
2	1	2	0	1	0	4.1
3	1	3	0	0	1	6.0
4	1	4	0	0	0	6.5
5	2	1	1	0	0	5.8
6	2	2	0	1	0	5.2
7	2	3	0	0	1	6.8
8	2	4	0	0	0	7.4

Period	Year	Quarter	Qtr1	Qtr2	Qtr3	Sales (1,000s)
9	3	3	0	0	1	6.0
10	3	4	0	0	0	5.6
11	4	1	1	0	0	7.5
12	4	2	0	1	0	7.8
13	4	3	0	0	1	6.3
14	4	4	0	0	0	5.9
15	5	1	1	0	0	8.0
16	5	2	0	1	0	8.4

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.4 Seasonality With Trend: Regression Equation

The general form of the equation relating the quarterly sales of smart phones is:

$$\hat{y}_t = 6.069 - 1.363\, Qtr1 - 2.034\, Qtr1 - 0.304\, Qtr1 + 0.1456t$$

We use the regression equation to forecast quarterly smart phone sales for year 5.

$$\text{Qtr1 sales} = 6.069 - 1.363\,(1) - 2.034\,(0) - 0.304\,(0) + 0.1456(17) = 7.18$$

$$\text{Qtr2 sales} = 6.069 - 1.363\,(0) - 2.034\,(1) - 0.304\,(0) + 0.1456(18) = 6.66$$

$$\text{Qtr3 sales} = 6.069 - 1.363\,(0) - 2.034\,(0) - 0.304\,(1) + 0.1456(19) = 8.53$$

$$\text{Qtr4 sales} = 6.069 - 1.363\,(0) - 2.034\,(0) - 0.304\,(0) + 0.1456(20) = 8.98$$

Thus, accounting for the seasonal effects and the linear trend in smartphone sales, the estimates of quarterly sales in Year 5 are 7,180, 6,660, 8,530, and 8,980 units.

The slope of the trendline $b_4 = 0.1456$ indicates a consistent growth in sales of about 146 units per quarter.

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.4 Regression Analysis as a Causal Forecasting Method

The relationship of the variable to be forecast with other variables may also be used to develop a forecasting model.

- Advertising expenditures to forecast sales.
- The mortgage rate to forecast new housing construction.
- Grade point average to forecast starting salaries for recent graduates.
- The price of a product to forecast its demand.
- The value of the Dow Jones Industrial Average to forecast the value of an individual stock.
- Daily high temperature to forecast electricity usage.

Causal models are forecasting models that include only variables that are believed to cause changes in the variable to be forecast.

© 2024 Cengage Group. All Rights Reserved.

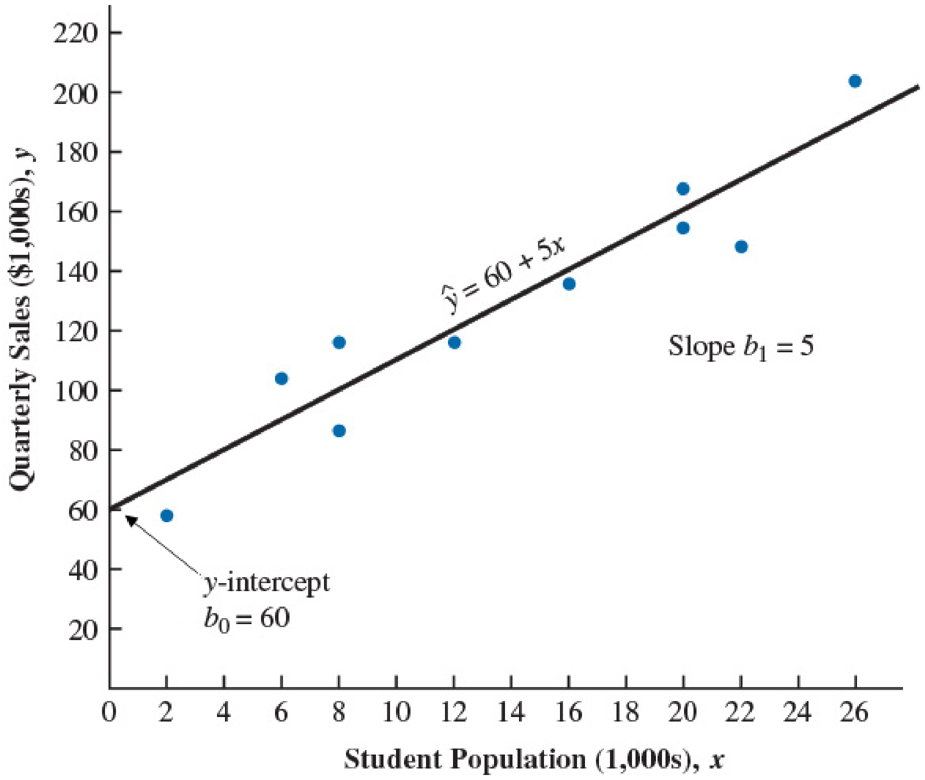
9.4 The Armand's Pizza Parlors Regression Equation

Management at Armand's Pizza Parlors wants to forecast sales for a new restaurant to be opened near a college campus using the student population at 10 colleges located near similar restaurants (DATAFile: *armands*.)

Because the relationship between student population x_i and quarterly sales y_i can be approximated by a straight line, simple linear regression is adopted.

$$\hat{y}_i = 60 + 5x_i$$

Given the slope b_1 , an additional 1,000 students are associated with an increase of $5,000 in expected quarterly sales.



© 2024 Cengage Group. All Rights Reserved.

9.4 Armand's Pizza Parlor Sales Forecast

The estimated y-intercept b_0 tells us that the location of a new restaurant with a local population of 0 students would expect quarterly sales of $60,000.

We can forecast quarterly sales for a new restaurant to be opened near a 16,000 students campus as follows.

$$\hat{y} = 60 + 5(16) = 140$$

Thus, quarterly sales are expected to be $140,000.

	A	B	C	D	E
1	SUMMARY OUTPUT				
2					
3	*Regression Statistics*				
4	Multiple R	0.950122955			
5	R Square	0.90273363			
6	Adjusted R Square	0.890575334			
7	Standard Error	13.82931669			
8	Observations	10			
9					
10	ANOVA				
11		*df*	*SS*	*MS*	*F*
12	Regression	1	14200	14200	74.24836601
13	Residual	8	1530	191.25	
14	Total	9	15730		
15					
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*
17	Intercept	60	9.22603481	6.503335532	0.000187444
18	Student Population (1000s)	5	0.580265238	8.616749156	2.54887E-05

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.4 Combining Causal Variables with Trend and Seasonality Effects

Regression models are very flexible and can incorporate both causal variables and time series effects.

Suppose we had a time series of several years of quarterly sales data and advertising expenditures for a single Armand's restaurant.

If we suspected that sales were related to the causal variable advertising expenditures and that sales showed trend and seasonal effects, we could incorporate each into a single model by combining the approaches we have outlined as follows.

$$\hat{y}_t = b_0 + b_1 Qtr1 + b_2 Qtr2 + b_3\, Qtr3 + b_4\, t + b_5\, x_t$$

Where x_t are quarterly advertising expenditures at the Armand's restaurant.

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.4 Considerations in Using Regression in Forecasting

Although linear regression analysis allows for the estimation of complex forecasting models, we must be cautious about using such models and guard against the potential for overfitting our model to the sample data.

Many research studies have also shown that quantitative forecasting models such as those presented in this chapter commonly outperform qualitative forecasts made by "experts."

Whether a regression approach provides a good forecast depends largely on how well we are able to identify and obtain data for independent variables that are closely related to the time series.

Part of the regression analysis procedure should focus on the selection of the set of independent variables that provides the best forecasting model.

CENGAGE

© 2024 Cengage Group. All Rights Reserved.

9.5 Determining the Best Forecasting Model to Use

A visual inspection can indicate whether seasonality appears to be a factor and whether a linear or nonlinear trend seems to exist.

For causal modeling, scatter charts can indicate whether strong relationships exist between each independent variable and the dependent variable.

- If certain relationships appear totally random, this may lead to exclude the corresponding independent variables from the model.

While working with large data sets, it is recommended to divide the data into training and validation sets, with training performed on the older data.

Based on the errors produced by the different models for the validation set, the model that minimizes MAE, MSE or MAPE is selected.

- There are software packages that will automatically select the best model.

CENGAGE
© 2024 Cengage Group. All Rights Reserved.

Summary

- We introduced basic methods for time series analysis and forecasting.
- First, we showed how to explain the behavior of a time series in terms of trend, seasonal, and/or cyclical components.
- Then, we discussed how smoothing methods can be used to forecast a time series that exhibits no significant trend, seasonal, or cyclical effect.
  - For time series that have only a long-term trend, we showed how linear regression analysis could be used to make trend projections.
  - For time series with seasonal influences, we showed how to incorporate the seasonality for more accurate forecasts.
- Finally, we showed how linear regression is used to develop causal forecast-ing models and provided guidance on how to select an appropriate model.

CENGAGE

© 2024 Cengage Group. All Rights Reserved.