Business Statistics: A Decision-Making Approach

6th Edition

Chapter 16

Analyzing and Forecasting Time-Series Data

Chapter Goals

After completing this chapter, you should be able to:

- Develop and implement basic forecasting models
- Identify the components present in a time series
- Compute and interpret basic index numbers
- Use smoothing-based forecasting models, including single and double exponential smoothing
- Apply trend-based forecasting models, including linear trend, nonlinear trend, and seasonally adjusted trend

The Importance of Forecasting

- Governments forecast unemployment, interest rates, and expected revenues from income taxes for policy purposes
- Marketing executives forecast demand, sales, and consumer preferences for strategic planning
- College administrators forecast enrollments to plan for facilities and for faculty recruitment
- Retail stores forecast demand to control inventory levels, hire employees and provide training

What is the difference between **planning** and **forecasting?**

- **Planning** is the process of determining how to deal with the future. **Forecasting** is the process of predicting the timing and magnitude of future events, predicting what the future will be like.
- Manufacturing firms must plan their production: an exercise known as aggregate production planning in which the firm states how many units to produce on a period by period basis and what level of employment to have over that time period. A forecast of the firm's demand is a necessary input to the production planning process.
- Likewise, service organizations will use a forecast in their budgeting and planning activities. Short term demand forecasts may be used as one input to determine the number of workers to schedule for a particular shift.
- Electric utility companies will use a demand forecast to plan their long-term capacity requirements, as well as plan and prepare for short-term needs.

Time-Series Data

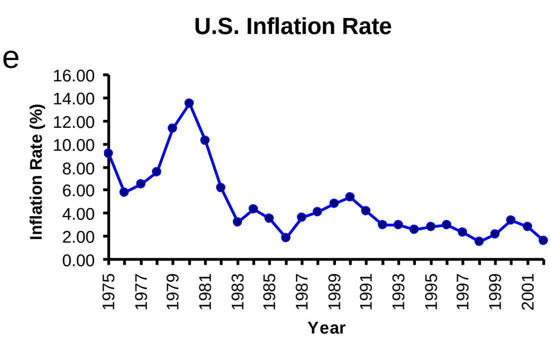
- Numerical data obtained at regular time intervals
- The time intervals can be annually, quarterly, daily, hourly, etc.
- Example:

Year:	1999	2000	2001	2002	2003
Sales:	75.3	74.2	78.5	79.7	80.2

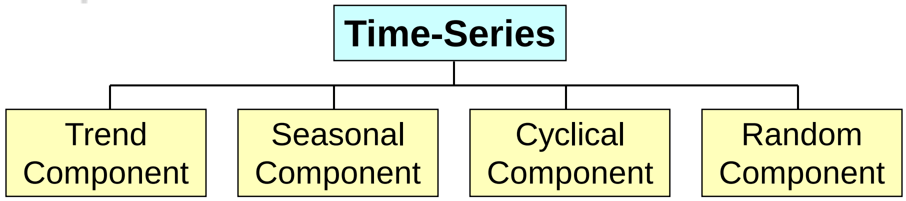
Time Series Plot

A time-series plot is a two-dimensional plot of time series data

- the vertical axis measures the variable of interest
- the horizontal axis corresponds to the time periods

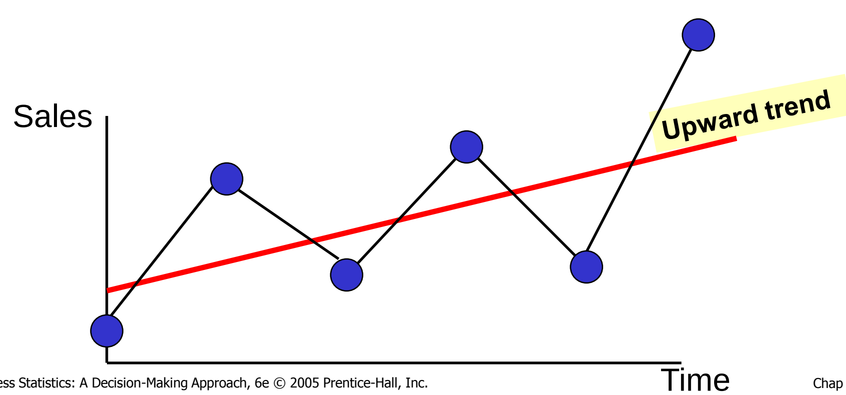


Time-Series Components



Trend Component

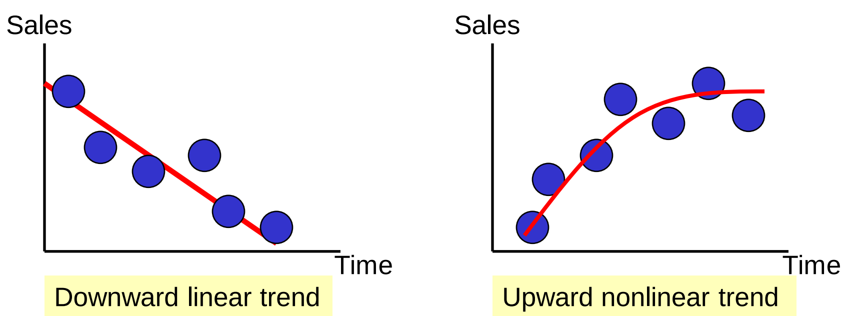
- Long-run increase or decrease over time (overall upward or downward movement)
- Data taken over a long period of time



Trend Component

(continued)

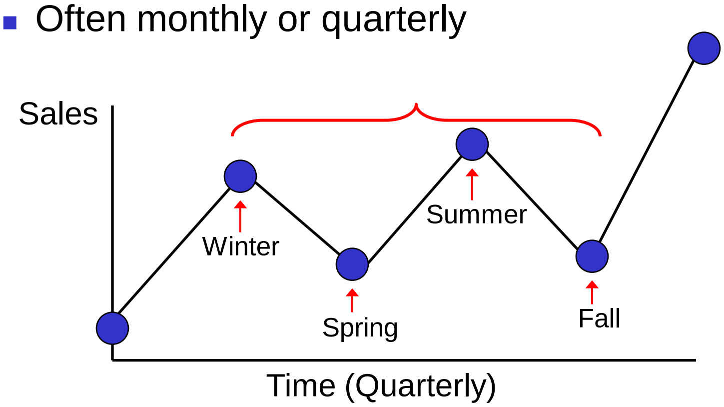
- Trend can be upward or downward
- Trend can be linear or non-linear



 Chap 15-9

Seasonal Component

- Short-term regular wave-like patterns
- Observed within 1 year
- Often monthly or quarterly

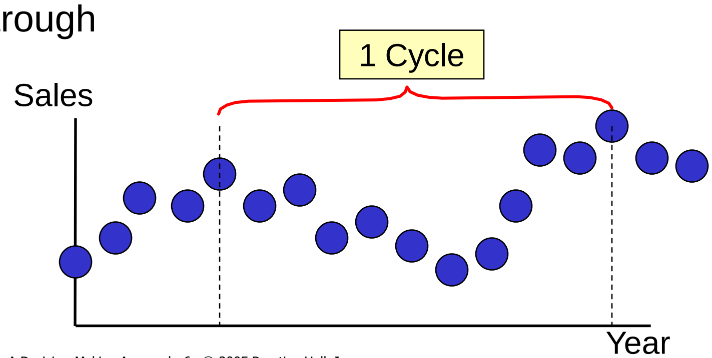


 Chap 15-10

- A seasonal component is a pattern in the time series that repeats itself with the same period of recurrence. While we often think of seasonal effects as being associated with the seasons (spring, summer, fall, winter) of the year, the seasonal pattern may be hourly, daily, weekly, or monthly. In fact a seasonal pattern can be any repeating pattern where the period of recurrence is at most one year. An example of a seasonal component that is not associated with **the seasons is the sales of tickets to a movie theater**. Ticket \sales may well be higher on Friday and Saturday evenings, than they are on Tuesday and Wednesday afternoons. If this pattern repeats itself over time, **the series is said to exhibit a daily seasonal effect**. Likewise, phone calls coming to a switchboard may be higher at certain hours of the day (between 9:00 a.m. and 10:00 a.m.) than at other times (between 3:00 p.m. and 4:00 p.m.). If this pattern repeats itself in a predictable way then we have an **hourly seasonal component**.

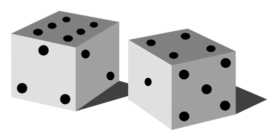
Cyclical Component

- Long-term wave-like patterns
- Regularly occur but may vary in length
- Often measured peak to peak or trough to trough



Random Component

- Unpredictable, random, “residual” fluctuations
- Due to random variations of
  - Nature
  - Accidents or unusual events
- “Noise” in the time series



Index Numbers

- Index numbers allow relative comparisons over time
- Index numbers are reported relative to a Base Period Index
- Base period index = 100 by definition
- Used for an individual item or measurement

Index Numbers

(continued)

- Simple Index number formula:

$$I_t = \frac{y_t}{y_0} 100$$

where

I_t = index number at time period t

y_t = value of the time series at time t

y_0 = value of the time series in the base period

Index Numbers: Example

- Company orders from 1995 to 2003:

Year	Number of Orders	Index (base year = 2000)
1995	272	85.0
1996	288	90.0
1997	295	92.2
1998	311	97.2
1999	322	100.6
2000	320	100.0
2001	348	108.8
2002	366	114.4
2003	384	120.0

$$I_{1996} = \frac{y_{1996}}{y_{2000}} 100 = \frac{288}{320}(100) = 90$$

Base Year:

$$I_{2000} = \frac{y_{2000}}{y_{2000}} 100 = \frac{320}{320}(100) = 100$$

$$I_{2003} = \frac{y_{2003}}{y_{2000}} 100 = \frac{384}{320}(100) = 120$$

Index Numbers: Interpretation

$$I_{1996} = \frac{y_{1996}}{y_{2000}} 100 = \frac{288}{320}(100) = 90$$

- Orders in 1996 were 90% of base year orders

$$I_{2000} = \frac{y_{2000}}{y_{2000}} 100 = \frac{320}{320}(100) = 100$$

- Orders in 2000 were 100% of base year orders (by definition, since 2000 is the base year)

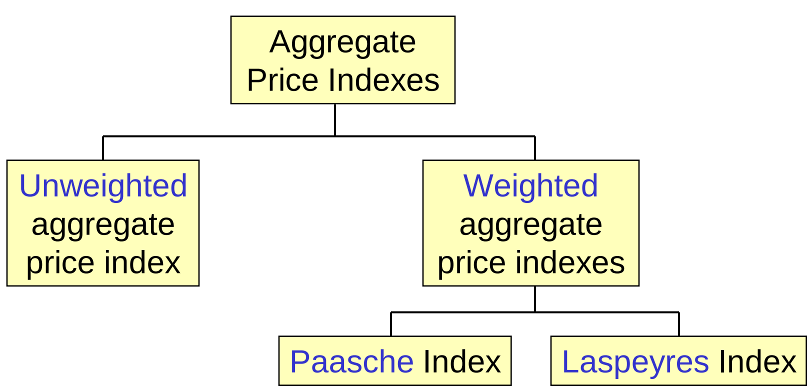
$$I_{2003} = \frac{y_{2003}}{y_{2000}} 100 = \frac{384}{320}(100) = 120$$

- Orders in 2003 were 120% of base year orders

 Chap 15-17

Aggregate Price Indexes

- An aggregate index is used to measure the rate of change from a base period for a group of items



 Chap 15-18

Unweighted Aggregate Price Index

- Unweighted aggregate price index formula:

$$I_t = \frac{\sum p_t}{\sum p_0}(100)$$

where

I_t = unweighted aggregate price index at time t

Σp_t = sum of the prices for the group of items at time t

Σp_0 = sum of the prices for the group of items in the base period

Unweighted Aggregate Price Index Example

	Automobile Expenses: Monthly Amounts ($):					
Year	**Lease payment**	**Fuel**	**Repair**	**Total**	**Index (2001=100)**	**Change**
2001	260	45	40	345	100.0	0%
2002	280	60	40	380	110.1	10.1%
2003	305	55	45	405	117.4	17.4%
2004	310	50	50	410	118.8	18.8%

$$I_{2004} = \frac{\sum p_{2004}}{\sum p_{2001}}(100) = \frac{410}{345}(100) = 118.8$$

- Combined expenses in 2004 were 18.8% higher in 2004 than in 2001

Weighted Aggregate Price Indexes

- Paasche index

$$I_t = \frac{\sum q_t p_t}{\sum q_t p_0}(100)$$

- Laspeyres index

$$I_t = \frac{\sum q_0 p_t}{\sum q_0 p_0}(100)$$

q_t = weighting percentage at time t

q_0 = weighting percentage at base period

p_t = price in time period t

p_0 = price in the base period

Commonly Used Index Numbers

- Consumer Price Index (CPI): *weighted aggregate index similar to Laspeyres Index, is based on items grouped into categories (such as food, housing, clothing, transportation, medical care, entertainment, and miscellaneous item)*
- Producer Price Index (PPI): *like the CPI, the PPI is a Laspeyres weighted aggregate Index.*
- Stock Market Indexes
  - Dow Jones Industrial Average
  - S&P 500 Index
  - NASDAQ Index

Deflating a Time Series

- Observed values can be adjusted to base year equivalent
- Allows uniform comparison over time
- Deflation formula:

$$y_{adj_t} = \frac{y_t}{I_t}(100)$$

where

y_{adj_t} = adjusted time series value at time t

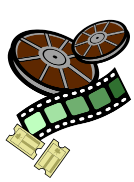
y_t = value of the time series at time t

I_t = index (such as CPI) at time t

Deflating a Time Series: Example

- Which movie made more money (in real terms)?

Year	Movie Title	Total Gross $
1939	Gone With the Wind	199
1977	Star Wars	461
1997	Titanic	601

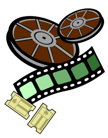


(Total Gross $ = Total domestic gross ticket receipts in $millions)

Deflating a Time Series: Example

(continued)

Year	Movie Title	Total Gross	CPI (base year = 1984)	Gross adjusted to 1984 dollars
1939	Gone With the Wind	199	13.9	1431.7
1977	Star Wars	461	60.6	760.7
1997	Titanic	601	160.5	374.5



$$GWTW_{adj-1984} = \frac{199}{13.9}(100) = 1431.7$$

- GWTW made about twice as much as Star Wars, and about 4 times as much as Titanic when measured in equivalent dollars

Trend-Based Forecasting

- Estimate a trend line using regression analysis

Year	Time Period (t)	Sales (y)
1999	1	20
2000	2	40
2001	3	30
2002	4	50
2003	5	70
2004	6	65

- Use time (t) as the independent variable:

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

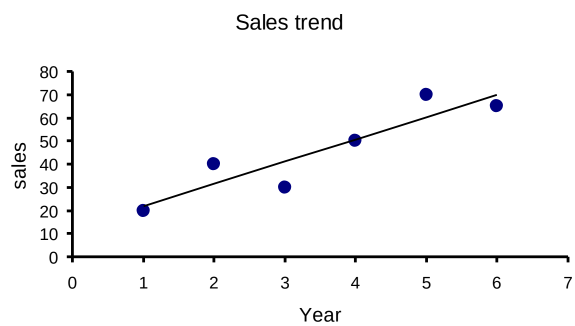
Trend-Based Forecasting

(continued)

Year	Time Period (t)	Sales (y)
1999	1	20
2000	2	40
2001	3	30
2002	4	50
2003	5	70
2004	6	65

- The linear trend model is:

$$\hat{y} = 12.333 + 9.5714\,t$$



Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 15-27

Trend-Based Forecasting

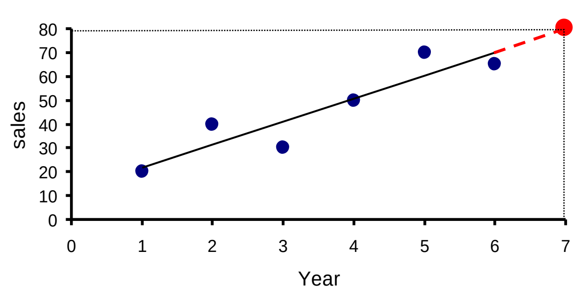
(continued)

Year	Time Period (t)	Sales (y)
1999	1	20
2000	2	40
2001	3	30
2002	4	50
2003	5	70
2004	6	65
2005	7	??

- Forecast for time period 7:

$$\hat{y} = 12.333 + 9.5714\,(7)$$

$$= \boxed{79.33}$$



Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 15-28

Comparing Forecast Values to Actual Data

- The forecast error or residual is the difference between the actual value in time t and the forecast value in time t:
- Error in time t:

$$e_t = y_t - F_t$$

Two common Measures of Fit

- Measures of fit are used to gauge how well the forecasts match the actual values

MSE (mean squared error)

- Average squared difference between y_t and F_t

MAD (mean absolute deviation)

- Average absolute value of difference between y_t and F_t
- Less sensitive to extreme values

MSE vs. MAD

Mean Square Error

$$MSE = \frac{\sum (y_t - F_t)^2}{n}$$

Mean Absolute Deviation

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

where:

y_t = Actual value at time t

F_t = Predicted value at time t

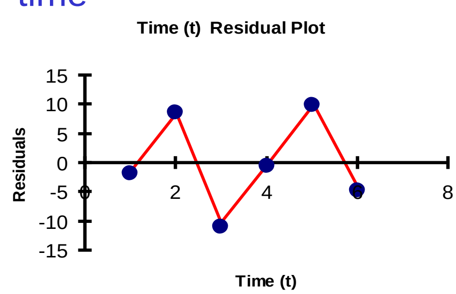
n = Number of time periods

Autocorrelation

(continued)

- Autocorrelation is correlation of the error terms (residuals) over time

- Here, residuals show a cyclic pattern, not random



- Violates the regression assumption that residuals are random and independent

Testing for Autocorrelation

- The Durbin-Watson Statistic is used to test for autocorrelation

H_0: ρ = 0 (residuals are not correlated)
H_A: ρ ≠ 0 (autocorrelation is present)

Durbin-Watson test statistic:

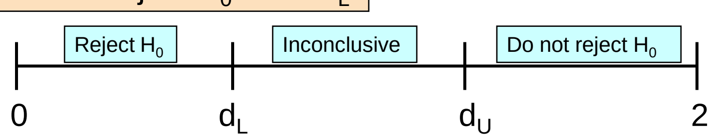
$$d = \frac{\sum_{t=1}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^{\,2}}$$

Testing for Positive Autocorrelation

H_0: ρ = 0 (positive autocorrelation does not exist)
H_A: ρ > 0 (positive autocorrelation is present)

- Calculate the Durbin-Watson test statistic = d
  (The Durbin-Watson Statistic can be found using PHStat or Minitab)
- Find the values d_L and d_U from the Durbin-Watson table
  (for sample size n and number of independent variables p)

Decision rule: reject H_0 if $d < d_L$



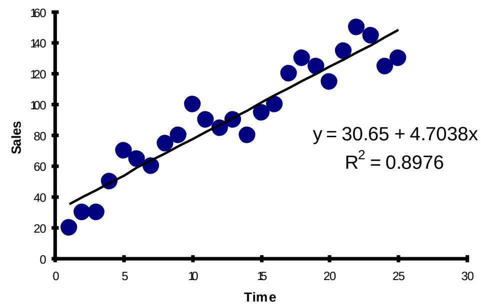
Testing for Positive Autocorrelation

(continued)

- Example with n = 25:

Excel/PHStat output:

Durbin-Watson Calculations	
Sum of Squared Difference of Residuals	3296.18
Sum of Squared Residuals	3279.98
Durbin-Watson Statistic	**1.00494**



$$d = \frac{\sum_{t=1}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^{\,2}} = \frac{3296.18}{3279.98} = 1.00494$$

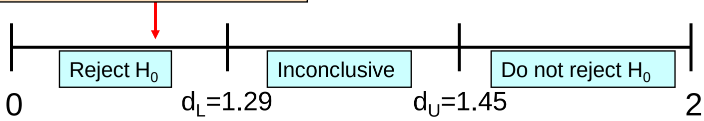
Testing for Positive Autocorrelation

(continued)

- Here, n = 25 and there is one independent variable
- Using the Durbin-Watson table, $d_L = 1.29$ and $d_U = 1.45$
- $d = 1.00494 < d_L = 1.29$, so reject H_0 and conclude that significant positive autocorrelation exists
- Therefore the linear model is not the appropriate model to forecast sales

Decision: **reject H_0** since

$d = 1.00494 < d_L$



Nonlinear Trend Forecasting

- A nonlinear regression model can be used when the time series exhibits a nonlinear trend
- One form of a nonlinear model:

$$y_t = \beta_0 + \beta_1 t^2 + \varepsilon_t$$

- Compare R^2 and s_ε to that of linear model to see if this is an improvement
- Can try other functional forms to get best fit

Multiplicative Time-Series Model

- Used primarily for forecasting
- Allows consideration of seasonal variation
- Observed value in time series is the product of components

$$y_t = T_t \times S_t \times C_t \times I_t$$

where T_t = Trend value at time t

S_t = Seasonal value at time t

C_t = Cyclical value at time t

I_t = Irregular (random) value at time t

Finding Seasonal Indexes

Ratio-to-moving average method:

- Begin by removing the seasonal and irregular components (S_t and I_t), leaving the trend and cyclical components (T_t and C_t)

- To do this, we need moving averages

Moving Average: averages of consecutive time series values

Moving Averages

- Used for smoothing
- Series of arithmetic means over time
- Result dependent upon choice of L (length of period for computing means)
- To smooth out seasonal variation, L should be equal to the number of seasons
  - For quarterly data, L = 4
  - For monthly data, L = 12

Moving Averages

(continued)

- Example: Four-quarter moving average
  - First average:

$$\text{Moving average}_1 = \frac{Q1 + Q2 + Q3 + Q4}{4}$$

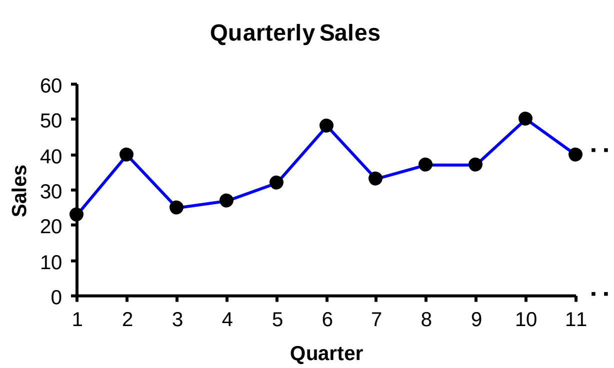
  - Second average:

$$\text{Moving average}_2 = \frac{Q2 + Q3 + Q4 + Q5}{4}$$

  - etc…

Seasonal Data

Quarter	Sales
1	23
2	40
3	25
4	27
5	32
6	48
7	33
8	37
9	37
10	50
11	40
etc…	etc…



Calculating Moving Averages

Quarter	Sales
1	23
2	40
3	25
4	27
5	32
6	48
7	33
8	37
9	37
10	50
11	40

etc…

Average Period	4-Quarter Moving Average
2.5	28.75
3.5	31.00
4.5	33.00
5.5	35.00
6.5	37.50
7.5	38.75
8.5	39.25
9.5	41.00

$$2.5 = \frac{1+2+3+4}{4}$$

$$28.75 = \frac{23+40+25+27}{4}$$

- Each moving average is for a consecutive block of 4 quarters

Centered Moving Averages

- Average periods of 2.5 or 3.5 don't match the original quarters, so we average two consecutive moving averages to get centered moving averages

Average Period	4-Quarter Moving Average
2.5	28.75
3.5	31.00
4.5	33.00
5.5	35.00
6.5	37.50
7.5	38.75
8.5	39.25
9.5	41.00

etc…

Centered Period	Centered Moving Average
3	29.88
4	32.00
5	34.00
6	36.25
7	38.13
8	39.00
9	40.13

Calculating the Ratio-to-Moving Average

- Now estimate the $S_t \times I_t$ value
- Divide the actual sales value by the centered moving average for that quarter
- Ratio-to-Moving Average formula:

$$S_t \times I_t = \frac{y_t}{T_t \times C_t}$$

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 15-45

Calculating Seasonal Indexes

Quarter	Sales	Centered Moving Average	Ratio-to-Moving Average
1	23		
2	40		
3	25	29.88	0.837
4	27	32.00	0.844
5	32	34.00	0.941
6	48	36.25	1.324
7	33	38.13	0.865
8	37	39.00	0.949
9	37	40.13	0.922
10	50	etc…	etc…
11	40	…	…
…	…	…	…

$$0.837 = \frac{25}{29.88}$$

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 15-46

Calculating Seasonal Indexes

(continued)

Quarter	Sales	Centered Moving Average	Ratio-to-Moving Average
1	23		
2	40		
Fall → 3	25	29.88	0.837
4	27	32.00	0.844
5	32	34.00	0.941
6	48	36.25	1.324
Fall → 7	33	38.13	0.865
8	37	39.00	0.949
9	37	40.13	0.922
10	50	etc…	etc…
Fall → 11	40	…	…
…	…	…	…

Average all of the Fall values to get Fall's seasonal index

Do the same for the other three seasons to get the other seasonal indexes

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 15-47

Interpreting Seasonal Indexes

- Suppose we get these seasonal indexes:

Season	Seasonal Index
Spring	0.825
Summer	1.310
Fall	0.920
Winter	0.945

$\Sigma = 4.000$ -- four seasons, so must sum to 4

- Interpretation:

Spring sales average 82.5% of the annual average sales

Summer sales are 31.0% higher than the annual average sales

etc…

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 15-48

Deseasonalizing

- The data is deseasonalized by dividing the observed value by its seasonal index

$$T_t \times C_t \times I_t = \frac{y_t}{S_t}$$

- This smooths the data by removing seasonal variation

Deseasonalizing

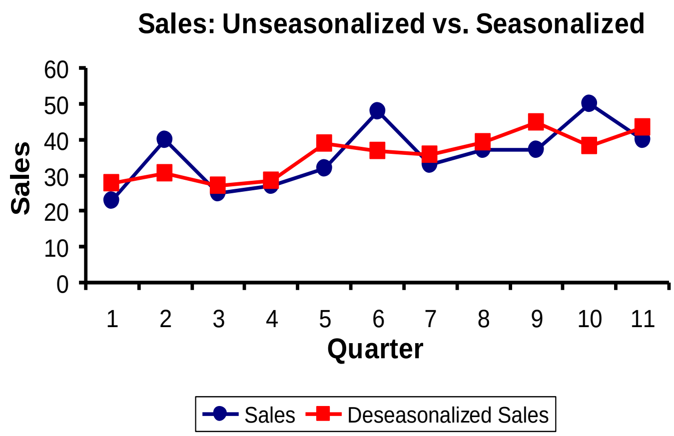
(continued)

Quarter	Sales	Seasonal Index	Deseasonalized Sales
1	23	0.825	27.88
2	40	1.310	30.53
3	25	0.920	27.17
4	27	0.945	28.57
5	32	0.825	38.79
6	48	1.310	36.64
7	33	0.920	35.87
8	37	0.945	39.15
9	37	0.825	44.85
10	50	1.310	38.17
11	40	0.920	43.48
…		…	…

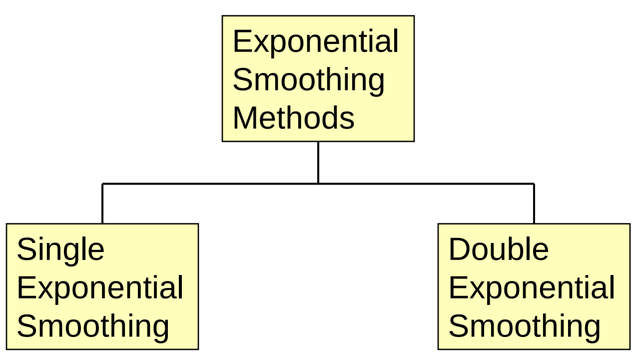
$$27.88 = \frac{23}{0.825}$$

etc…

Unseasonalized vs. Seasonalized



Forecasting Using Smoothing Methods



Single Exponential Smoothing

- A weighted moving average
  - Weights decline exponentially
  - Most recent observation weighted most
- Used for smoothing and short term forecasting

Single Exponential Smoothing

(continued)

- The weighting factor is α
  - Subjectively chosen
  - Range from 0 to 1
  - Smaller α gives more smoothing, larger α gives less smoothing
- The weight is:
  - Close to 0 for smoothing out unwanted cyclical and irregular components
  - Close to 1 for forecasting

Exponential Smoothing Model

- Single exponential smoothing model

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

or:

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

where:

F_{t+1} = forecast value for period t + 1
y_t = actual value for period t
F_t = forecast value for period t
α = alpha (smoothing constant)

Exponential Smoothing Example

- Suppose we use weight α = .2

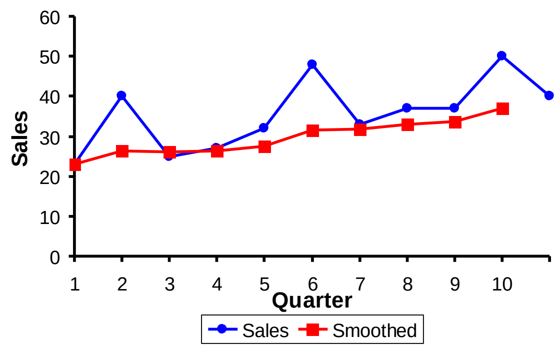
Quarter (t)	Sales (y_t)	Forecast from prior period	Forecast for next period (F_{t+1})
1	23	NA	23
2	40	23	(.2)(40)+(.8)(23)=26.4
3	25	26.4	(.2)(25)+(.8)(26.4)=26.12
4	27	26.12	(.2)(27)+(.8)(26.12)=26.296
5	32	26.296	(.2)(32)+(.8)(26.296)=27.437
6	48	27.437	(.2)(48)+(.8)(27.437)=31.549
7	33	31.549	(.2)(48)+(.8)(31.549)=31.840
8	37	31.840	(.2)(33)+(.8)(31.840)=32.872
9	37	32.872	(.2)(37)+(.8)(32.872)=33.697
10	50	33.697	(.2)(50)+(.8)(33.697)=36.958
etc…	etc…	etc…	etc…

$F_1 = y_1$ since no prior information exists

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

Sales vs. Smoothed Sales

- Seasonal fluctuations have been smoothed
- NOTE: the smoothed value in this case is generally a little low, since the trend is upward sloping and the weighting factor is only .2



Double Exponential Smoothing

- Double exponential smoothing is sometimes called exponential smoothing with trend
- If trend exists, single exponential smoothing may need adjustment
- Add a second smoothing constant to account for trend

Double Exponential Smoothing Model

$$C_t = \alpha y_t + (1-\alpha)(C_{t-1} + T_{t-1})$$

$$T_t = \beta(C_t - C_{t-1}) + (1-\beta)T_{t-1}$$

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

where:

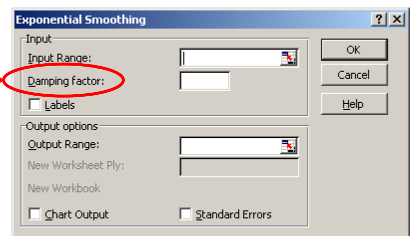
y_t = actual value in time t
α = constant-process smoothing constant
β = trend-smoothing constant
C_t = smoothed constant-process value for period t
T_t = smoothed trend value for period t
F_{t+1}= forecast value for period t + 1
t = current time period

Double Exponential Smoothing

- Double exponential smoothing is generally done by computer
- Use larger smoothing constants α and β when less smoothing is desired
- Use smaller smoothing constants α and β when more smoothing is desired

Exponential Smoothing in Excel

- Use tools / data analysis / exponential smoothing
  - The "damping factor" is $(1 - \alpha)$



Chapter Summary

- Discussed the importance of forecasting
- Addressed component factors present in the time-series model
- Computed and interpreted index numbers
- Described least square trend fitting and forecasting
  - linear and nonlinear models
- Performed smoothing of data series
  - moving averages
  - single and double exponential smoothing