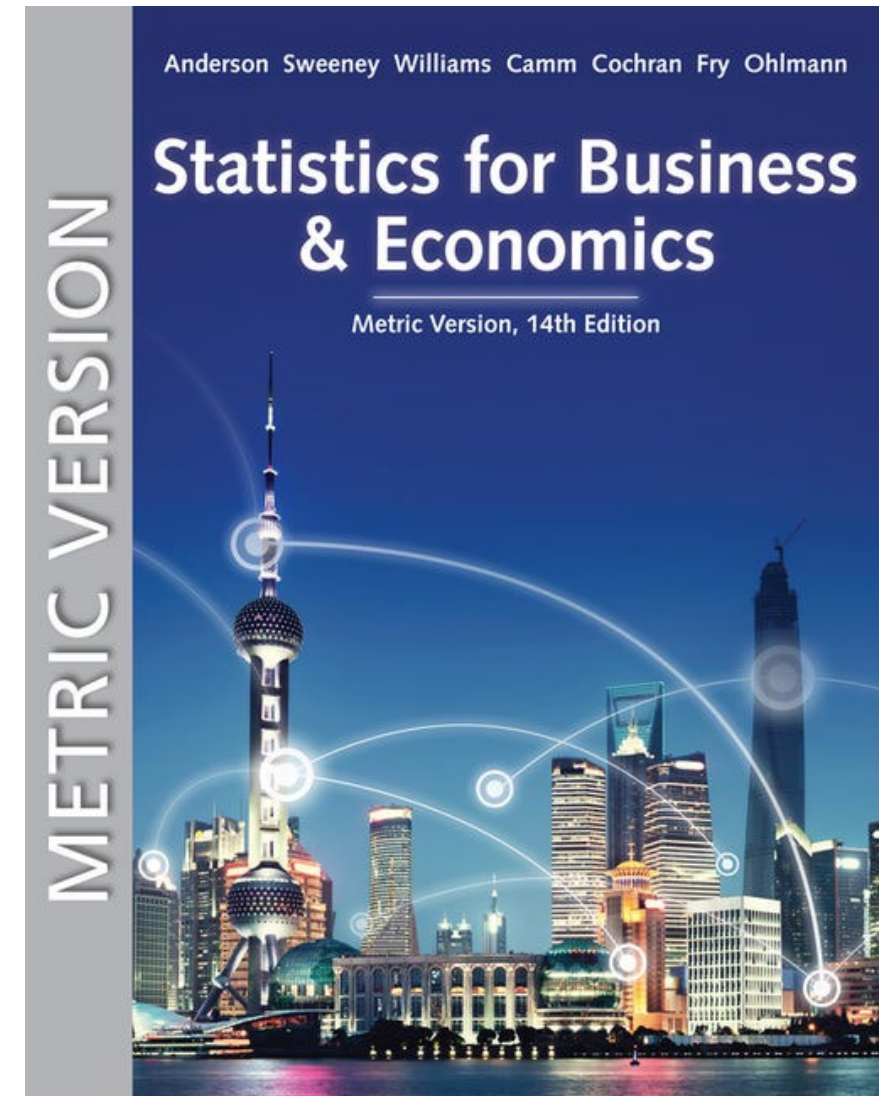


Statistics for Business and Economics (14e) Metric Version

Chapter 17 (預測)



“It’s tough to make predictions, especially about the future.”

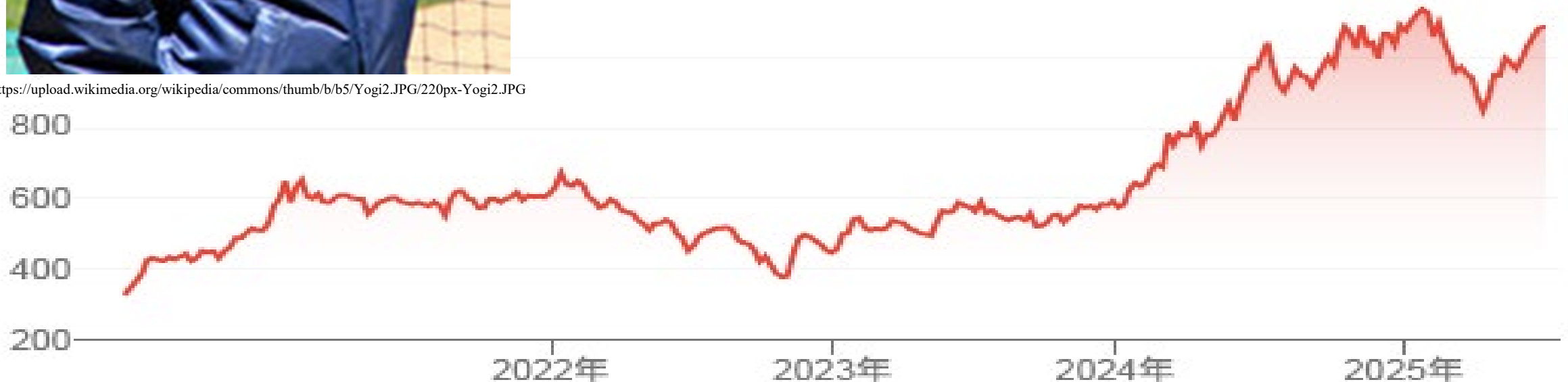
— Yogi Berra



<https://upload.wikimedia.org/wikipedia/commons/thumb/b/b5/Yogi2.JPG/220px-Yogi2.JPG>



<https://www.relexsolutions.com/resources/measuring-forecast-accuracy>



Chapter 17 - Time Series Analysis and Forecasting

17.1 – Time Series Patterns

17.2 – Forecast Accuracy

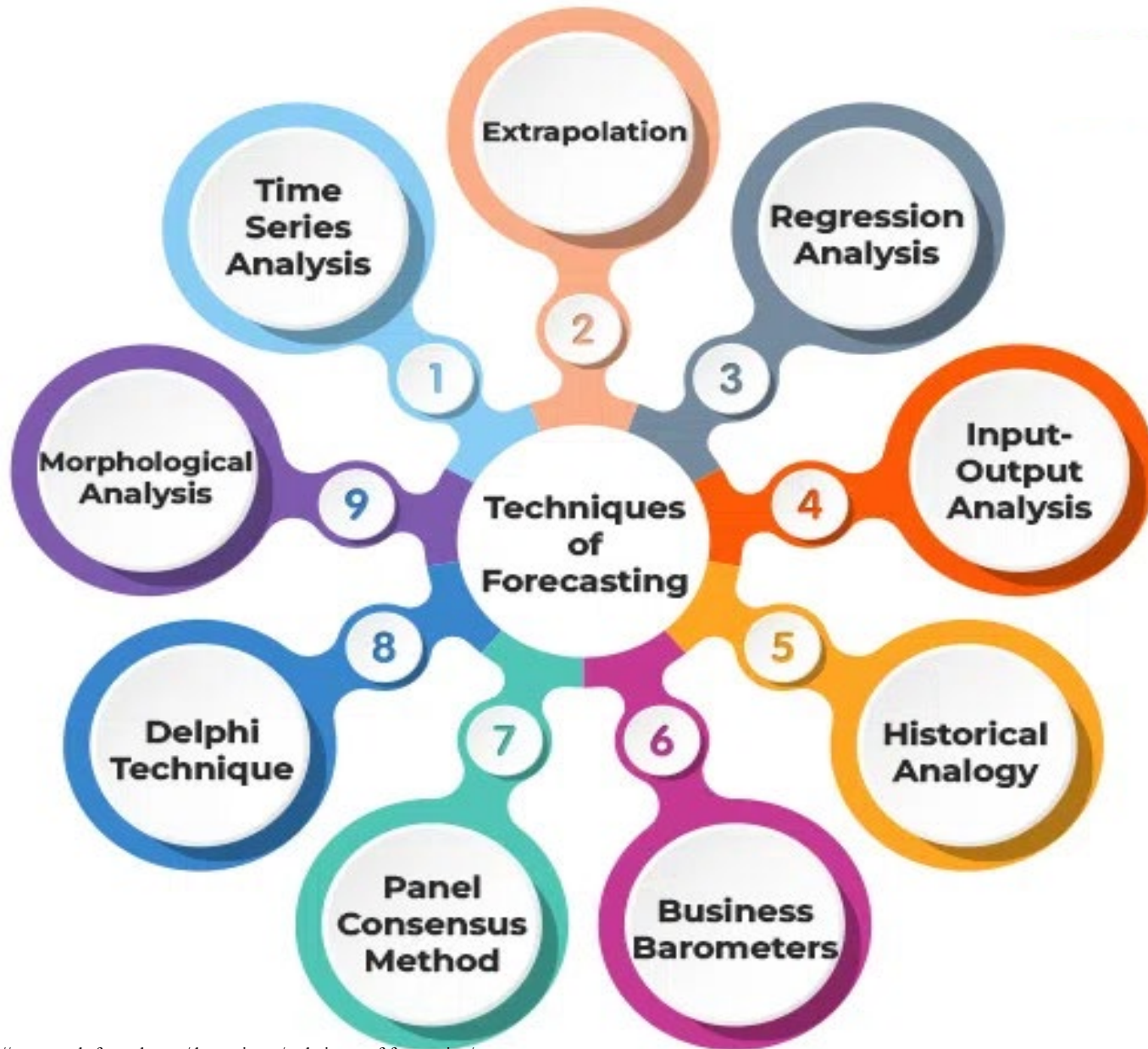
17.3 – Moving Averages and Exponential Smoothing

17.4 – Trend Projection

17.5 – Seasonality and Trend

17.6 – Time Series Decomposition

Forecasting Methods



Steps in the Process of Forecasting

Step 1 | Defining Objectives

Step 2 | Collecting and Analyzing Data

Step 3 | Selecting the Right Forecasting Method

Step 4 | Building and Testing Forecast Models

Step 5 | Monitoring and Adjusting



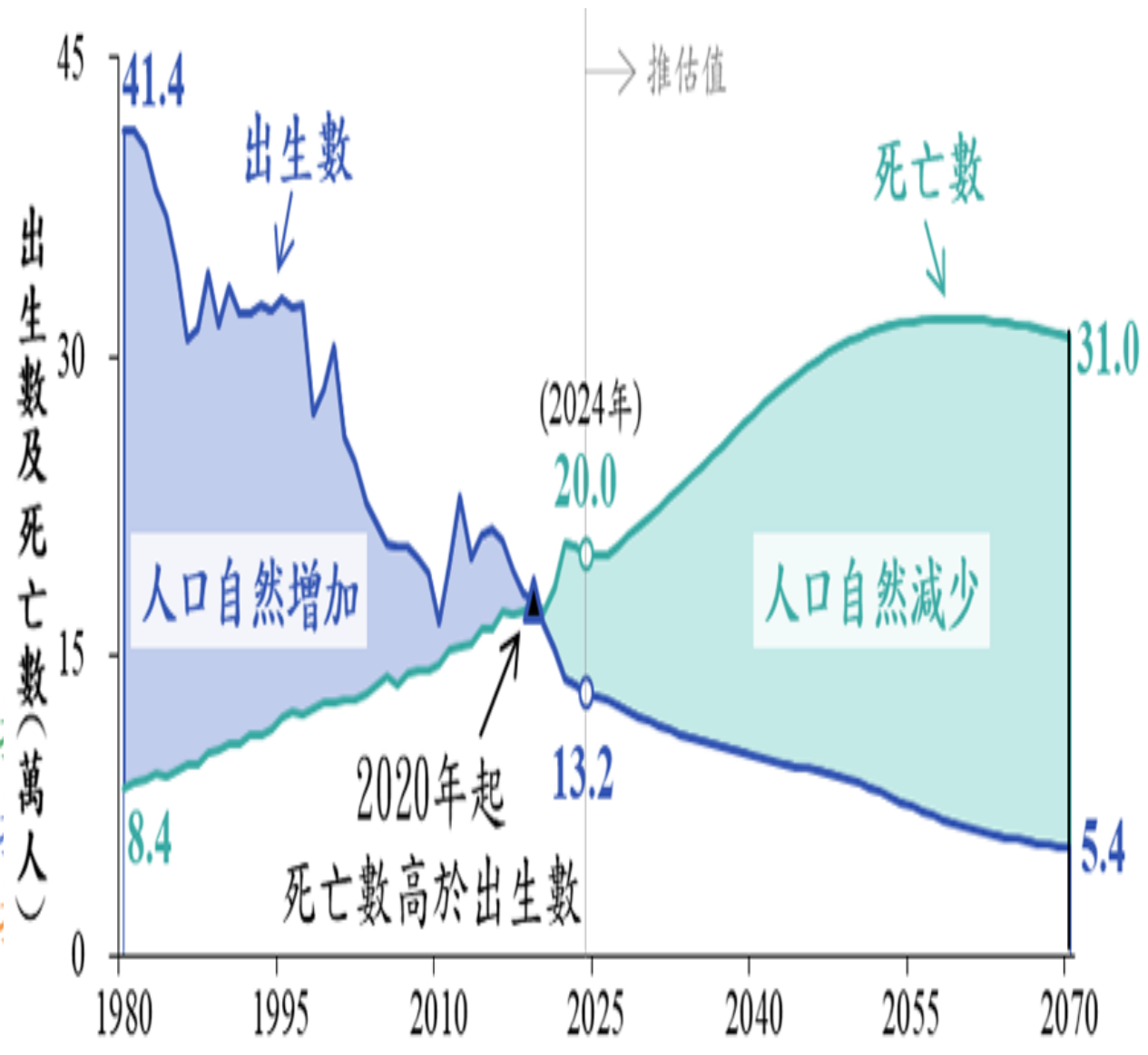
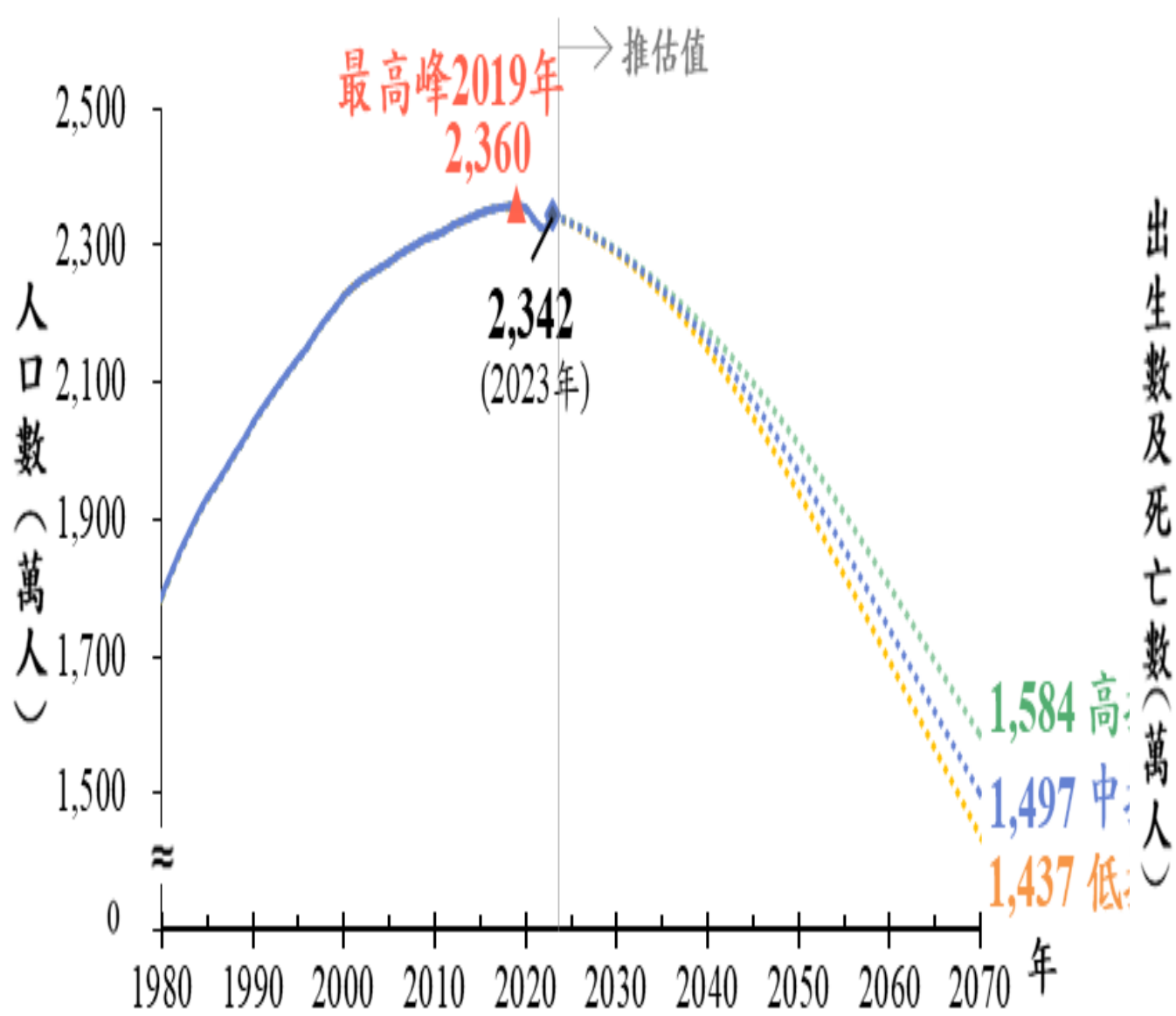
Forecasting Methods: Qualitative

- Forecasting methods can be classified as qualitative or quantitative.
 - Qualitative methods generally involve the use of expert judgment to develop forecasts.
 - Such methods are appropriate when historical data on the variable being forecast are either not applicable or unavailable.
 - We will focus exclusively on quantitative forecasting methods in this chapter.
- **Delphi Method** (德菲法)

中華民國人口推估（2024至2070年）

總人口數(萬人)

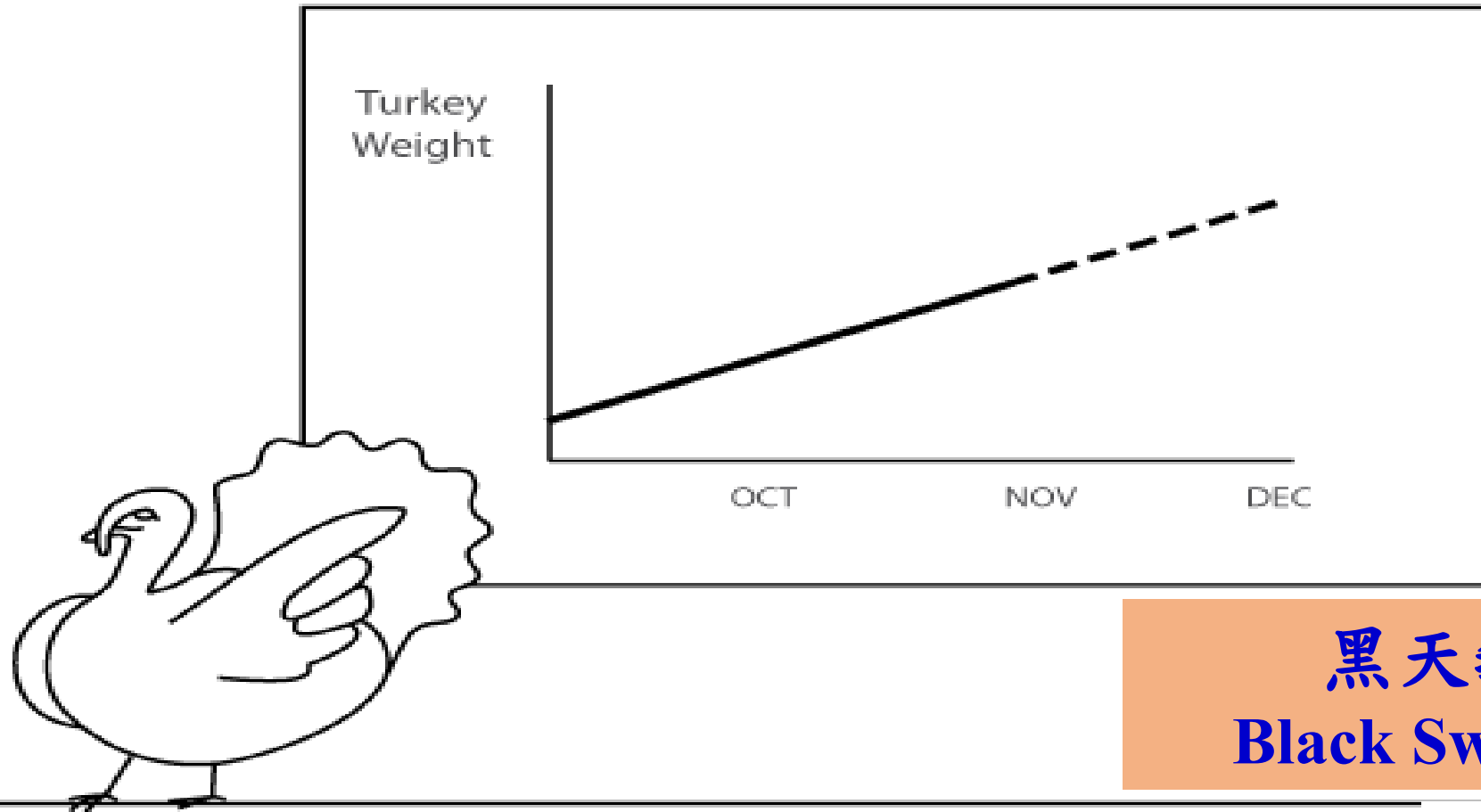
出生及死亡變動(萬人)



Forecasting Methods: Quantitative (1 of 5)

- Quantitative forecasting methods can be used when:
 - Past information about the variable being forecast is available,
 - The information can be quantified, and
 - It is reasonable to assume the pattern of the past will continue
- In such cases, a forecast can be developed using a time series method or a causal method.

THANKSGIVING PREDICTIVE ANALYTICS



黑天鵝理論
Black Swan Theory

"I see no reason why excellent growth shouldn't continue.."

<https://i0.wp.com/timoelliott.com/blog/wp-content/uploads/2013/07/turkey-predictive-analytics.png?ssl=1>

Q：灰犀牛理論？



<https://cdn.technasia.com/wp-content/uploads/2016/10/bs1.jpg>

Forecasting Methods: Quantitative (2 of 5)

- Quantitative methods are based on an analysis of historical data concerning one or more time series.
- A time series is a set of observations measured at successive points in time or over successive periods of time.
- If the historical data used are restricted to past values of the series that we are trying to forecast, the procedure is called a time series method.
- If the historical data used involve other time series that are believed to be related to the time series that we are trying to forecast, the procedure is called a causal method.

Forecasting Methods: Quantitative (3 of 5)

- Time Series Analysis
- The objective of time series analysis is to discover a pattern in the historical data or time series and then extrapolate the pattern into the future.
- The forecast is based solely on past values of the variable and/or past forecast errors.

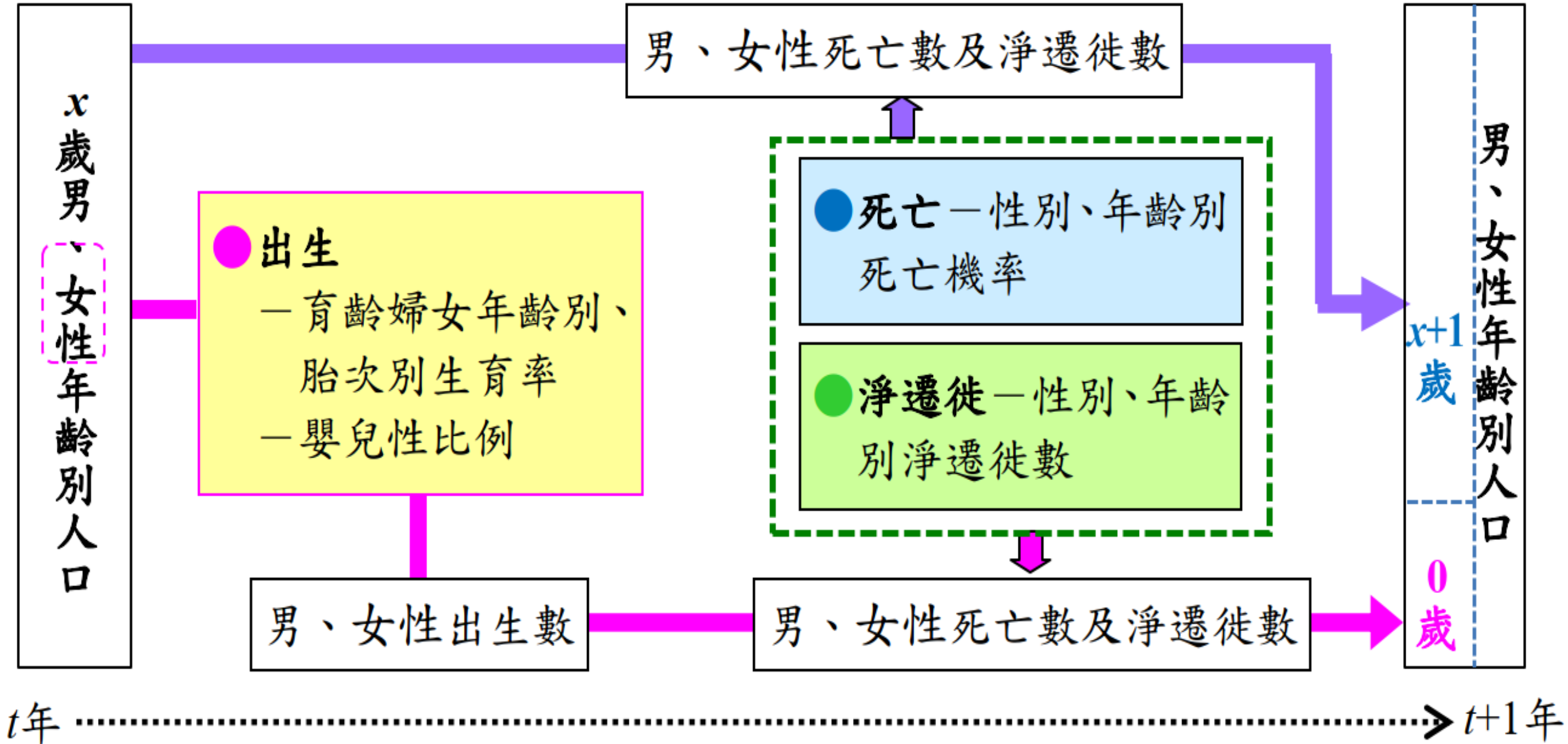
Forecasting Methods: Quantitative (4 of 5)

- Causal Methods
- Causal forecasting methods (因果預測法) are based on the assumption that the variable we are forecasting has a cause-effect relationship with one or more other variables.
- Looking at regression analysis as a forecasting tool, we can view the time series value that we want to forecast as the dependent variable.
- If we can identify a good set of related independent or explanatory variables, we may be able to develop an estimated regression equation for forecasting the time series.

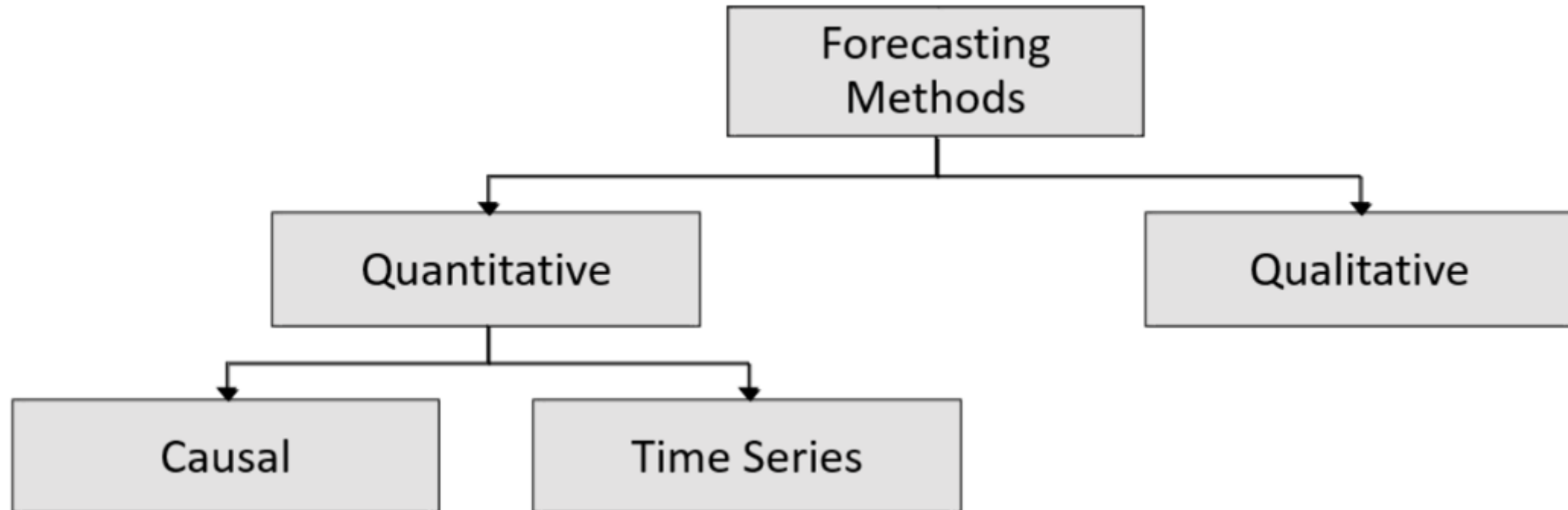
Forecasting Methods: Quantitative (5 of 5)

- Regression Analysis
- By treating time as the independent variable and the time series as a dependent variable, regression analysis can also be used as a time series method.
- Time-series regression refers to the use of regression analysis when the sole independent variable is time.
- Cross-sectional regression refers to the use of regression analysis when the independent variable(s) is(are) something other than time.

中華民國人口推估流程圖



Forecasting Methods



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.
- The pattern of the data is an important factor in understanding how the time series has behaved in the past.
- If such behavior can be expected to continue in the future, we can use it to guide us in selecting an appropriate forecasting method.

Time Series Plot (1 of 2)

- A useful first step in selecting an appropriate forecasting method is to construct a time series plot.
- A time series plot is a graphical presentation of the relationship between time and the time series variable.
- Time is on the horizontal axis, and the time series values are shown on the vertical axis.

Time Series Plot (2 of 2)

Example: Rosco Drugs

Sales of Comfort brand headache tonic (in bottles) for the past 10 weeks at Rosco Drugs are shown below. Rosco Drugs would like to identify the underlying pattern in the data to guide it in selecting an appropriate forecasting method.

Week	Sales
1	110
2	115
3	125
4	120
5	125
6	120
7	130
8	115
9	110
10	130

Time Series Patterns (1 of 6)

The common types of data patterns that can be identified when examining a time series plot include:

- Horizontal
- Trend (趨勢)
- Seasonal (季節性)
- Trend and Seasonal
- Cyclical (循環)

Time Series Patterns (2 of 6)

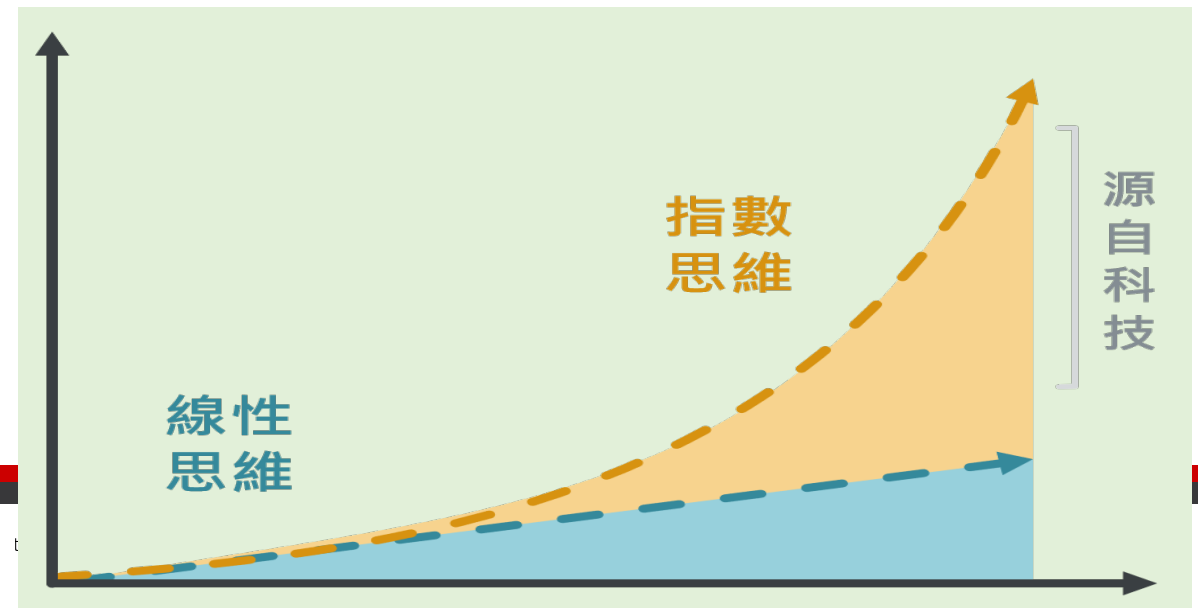
Horizontal Pattern

- A horizontal pattern exists when the data fluctuate around a constant mean.
- Changes in business conditions can often result in a time series that has a horizontal pattern shifting to a new level.
- A change in the level of the time series makes it more difficult to choose an appropriate forecasting method.

Time Series Patterns (3 of 6)

Trend Pattern

- A time series may show gradual shifts or movements to relatively higher or lower values over a longer period of time.
- Trend is usually the result of long-term factors such as changes in the population, demographics, technology, or consumer preferences.
- A systematic increase or decrease might be linear or nonlinear.
- A trend pattern can be identified by analyzing multiyear movements in historical data.



Time Series Patterns (4 of 6)

Seasonal Pattern

- Seasonal patterns are recognized by seeing the same repeating pattern of highs and lows over successive periods of time within a year.
- A seasonal pattern might occur within a day, week, month, quarter, year, or some other interval no greater than a year.
- A seasonal pattern does not necessarily refer to the four seasons of the year (spring, summer, fall, and winter).

Time Series Patterns (5 of 6)

Trend and Seasonal Pattern

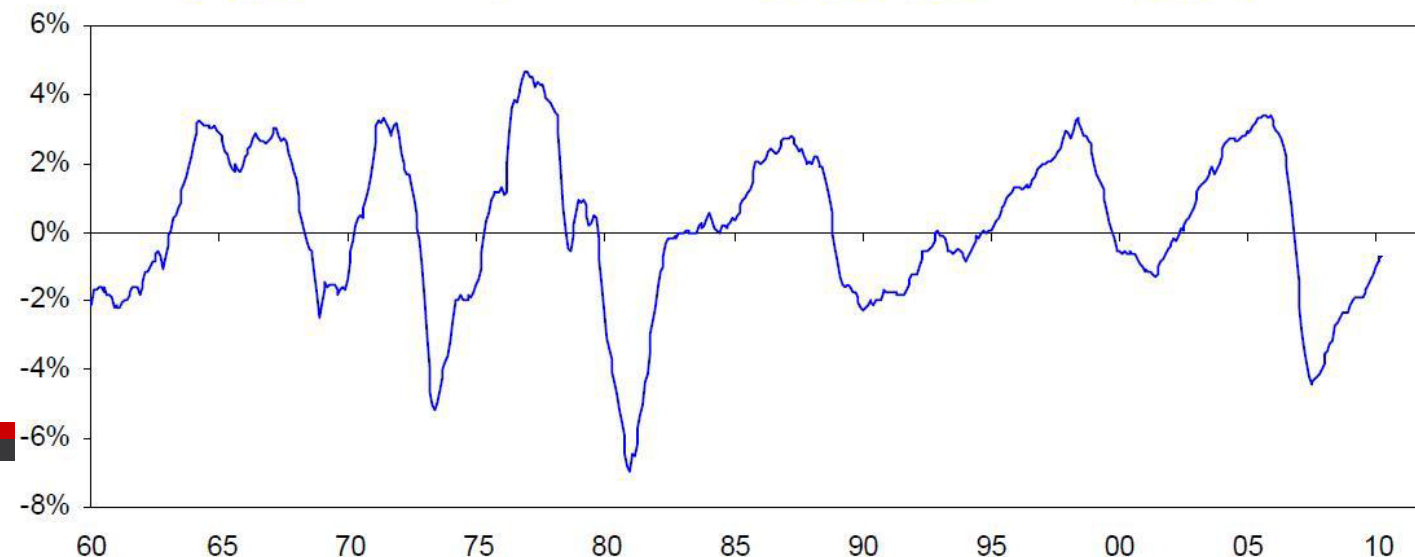
- Some time series include a combination of a trend and seasonal pattern.
- In such cases we need to use a forecasting method that has the capability to deal with both trend and seasonality.
- Time series decomposition can be used to separate or decompose a time series into trend and seasonal components.

Time Series Patterns (6 of 6)

Cyclical Pattern

- A cyclical pattern exists if the time series plot shows an alternating sequence of points below and above the trend line lasting more than one year.
- Often, the cyclical component of a time series is due to multiyear business cycles.
- Business cycles are extremely difficult, if not impossible, to forecast.
- In this chapter we do not deal with cyclical effects that may be present in the time series.

美國1960年~2010年期間的GDP缺口



Selecting a Forecasting Method

- The underlying pattern in the time series is an important factor in selecting a forecasting method.
- Thus, a time series plot should be one of the first things developed when trying to determine what forecasting method to use.
- If we see a horizontal pattern, then we need to select a method appropriate for this type of pattern.
- If we observe a trend in the data, then we need to use a method that has the capability to handle trend effectively.

Moving Averages and Exponential Smoothing

Now we discuss three forecasting methods that are appropriate for a time series with a **horizontal pattern**:

- Moving Averages
 - Weighted Moving Averages
 - Exponential Smoothing
-
- They are called smoothing methods because their objective is to smooth out the random fluctuations in the time series.
 - They are most appropriate for short-range forecasts.

Moving Averages (1 of 7)

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.

$$F_{t+1} = \frac{\sum(\text{most recent } k \text{ data values})}{k} = \frac{Y_t + Y_{t+1} + \cdots + Y_{t-k+1}}{k}$$

Where F_{t+1} = forecast of the time series for period $t + 1$

Y_t = actual value of the time series in period t

Each observation in the moving average calculation receives the same weight.

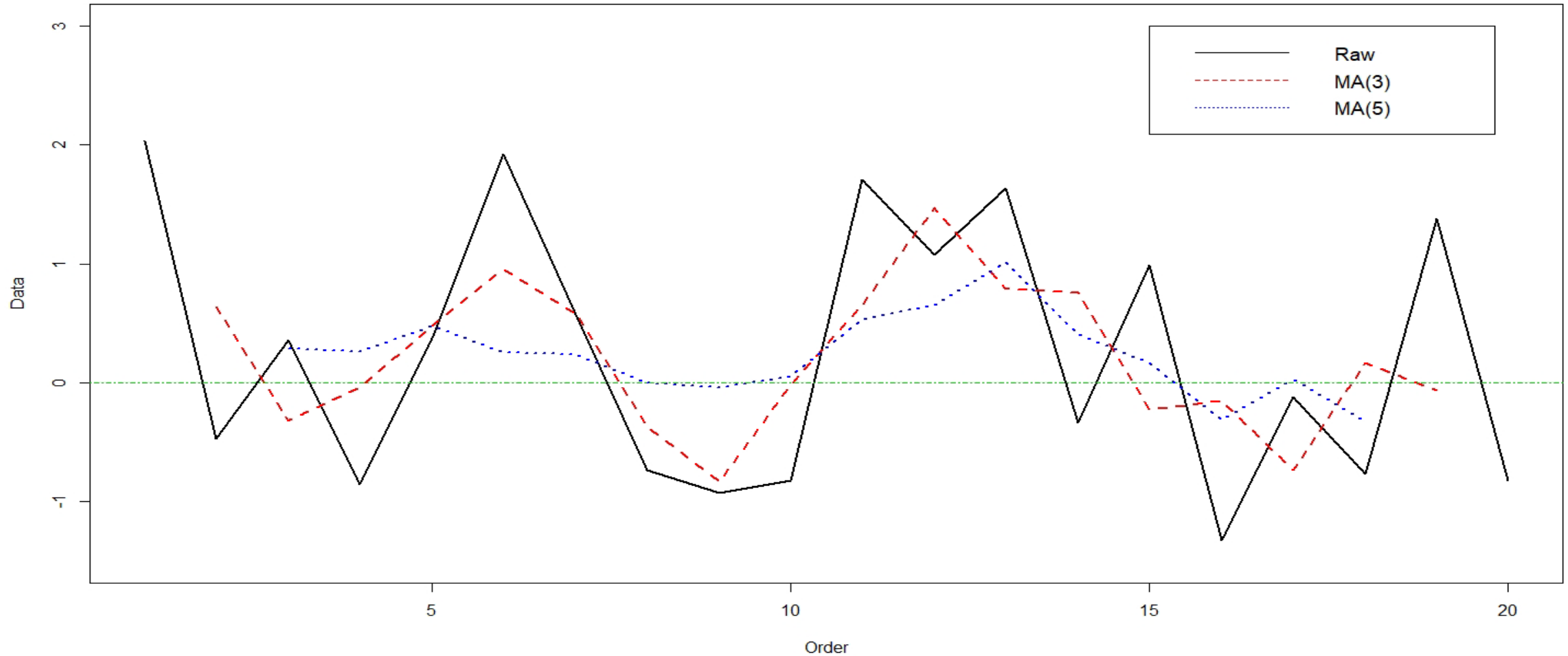
Moving Averages (2 of 7)

- 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.
- As a result, the average will change, or move, as new observations become available.

Moving Averages (3 of 7)

- To use moving averages to forecast, we must first select the order k , or number of time series values, to be included in the moving average.
- A smaller value of k will track shifts in a time series more quickly than a larger value of k .
- If more past observations are considered relevant, then a larger value of k is better.

An Example of Moving Average, $N(0,1)$ Random Numbers



Moving Averages (4 of 7)

Example: Rosco Drugs

If Rosco Drugs uses a 3-period moving average to forecast sales, what are the forecasts for weeks 4-11?

Week	Sales
1	110
2	115
3	125
4	120
5	125
6	120
7	130
8	115
9	110
10	130

Moving Averages (5 of 7)

<u>Week</u>	<u>Sales</u>	<u>3MA Forecast</u>
1	110	
2	115	
3	125	
4	120	116.7
5	125	120.0
6	120	123.3
7	130	121.7
8	115	125.0
9	110	121.7
10	130	118.3
11		118.3

Moving Averages (6 of 7)

<u>Week</u>	<u>Sales</u>	<u>3MA Forecast</u>	<u>Forecast Error</u>	<u>Absolute Error</u>	<u>Squared Error</u>	<u>Abs.% Error</u>
1	110					
2	115					
3	125					
4	120	116.7	3.3	3.3	10.89	2.75
5	125	120.0	5.0	5.0	25.00	4.00
6	120	123.3	-3.3	3.3	10.89	2.75
7	130	121.7	8.3	8.3	68.89	6.38
8	115	125.0	-10.0	10.0	100.00	8.70
9	110	121.7	-11.7	11.7	136.89	10.64
10	130	118.3	11.7	11.7	136.89	9.00
<u>Total</u>			3.33	53.3	489.45	44.22

Moving Averages (7 of 7)

3-MA Forecast Accuracy

$$MAE = \frac{53.3}{7} = 7.61$$

$$MSE = \frac{489.45}{7} = 69.92$$

$$MAPE = \frac{44.22}{7} = 6.32\%$$

The 3-week moving average approach provided more accurate forecasts than the naive approach.

Weighted Moving Averages (1 of 2)

To use this method we must first select the number of data values to be included in the average.

Next, we must choose the weight for each of the data values.

- The more recent observations are typically given more weight than older observations.
- For convenience, the weights should sum to 1.

Weighted Moving Averages (2 of 2)

An example of a 3-period weighted moving average (3WMA) is:

$$3WMA = 0.2(110) + 0.3(115) = 119$$

- In this example the weights are 0.2, 0.3, and 0.5 (which sum to 1).
- 125 is the most recent of the three observations.

Exponential Smoothing (1 of 6)

- This method is a special case of a weighted moving averages method; we select only the weight for the most recent observation.
- The weights for the other data values are computed automatically and become smaller as the observations grow older.
- The exponential smoothing forecast is a weighted average of all the observations in the time series.
- The term *exponential smoothing* comes from the exponential nature of the weighting scheme for the historical values.

Exponential Smoothing (2 of 6)

Exponential Smoothing Forecast

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

Where

F_{t+1} = forecast of the time series for period $t + 1$

Y_t = actual value of the time series in period t

F_t = forecast of the time series for period t

α = smoothing constant ($0 \leq \alpha \leq 1$)

And let $F_2 = Y_1$ to initiate the computations.

Exponential Smoothing (3 of 6)

Exponential Smoothing Forecast

- With algebraic manipulation, we can rewrite

$$F_{t+1} = \alpha Y_t + (1 - \alpha)F_t \text{ as:}$$

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

- We see that the new forecast F_{t+1} is equal to the previous forecast F_t plus an adjustment, which is α times the most recent forecast error, $Y_t - F_t$.

Exponential Smoothing (4 of 6)

Example: Rosco Drugs

If Rosco Drugs uses exponential smoothing to forecast sales, which value for the smoothing constant α , 0.1 or 0.8, gives better forecasts?

Week	Sales
1	110
2	115
3	125
4	120
5	125
6	120
7	130
8	115
9	110
10	130

Exponential Smoothing (5 of 6)

Using Smoothing Constant Value $\alpha = 0.1$

$$F_2 = Y_1 = 110$$

$$F_3 = .1Y_2 + .9F_2 = .1(115) + .9(110) = 110.50$$

$$F_4 = .1Y_3 + .9F_3 = .1(125) + .9(110.5) = 111.95$$

$$F_5 = .1Y_4 + .9F_4 = .1(120) + .9(111.95) = 112.76$$

$$F_6 = .1Y_5 + .9F_5 = .1(125) + .9(112.76) = 113.98$$

$$F_7 = .1Y_6 + .9F_6 = .1(120) + .9(113.98) = 114.58$$

$$F_8 = .1Y_7 + .9F_7 = .1(130) + .9(114.58) = 116.12$$

$$F_9 = .1Y_8 + .9F_8 = .1(115) + .9(116.12) = 116.01$$

$$F_{10} = .1Y_9 + .9F_9 = .1(110) + .9(116.01) = 115.41$$

Exponential Smoothing (6 of 6)

Using Smoothing Constant Value $\alpha = 0.8$

$$F_2 = \quad \quad \quad = 110$$

$$F_3 = .8(115) + .2(110) = 114.00$$

$$F_4 = .8(125) + .2(114) = 122.80$$

$$F_5 = .8(120) + .2(122.80) = 120.56$$

$$F_6 = .8(125) + .2(120.56) = 124.11$$

$$F_7 = .8(120) + .2(124.11) = 120.82$$

$$F_8 = .8(130) + .2(120.82) = 128.16$$

$$F_9 = .8(115) + .2(128.16) = 117.63$$

$$F_{10} = .8(110) + .2(117.63) = 111.53$$

Exponential Smoothing ($\alpha = 0.1$) (1 of 2)

<u>Week</u>	<u>Sales</u>	$\alpha = .1$ <u>Forecast</u>	<u>Forecast</u> <u>Error</u>	<u>Absolute</u> <u>Error</u>	<u>Squared</u> <u>Error</u>	<u>Abs.%</u> <u>Error</u>
1	110					
2	115	110.00	5.00	5.00	25.00	4.35
3	125	110.50	14.50	14.50	210.25	11.60
4	120	111.95	8.05	8.05	64.80	6.71
5	125	112.76	12.24	12.24	149.94	9.79
6	120	113.98	6.02	6.02	36.25	5.02
7	130	114.58	15.42	15.42	237.73	11.86
8	115	116.12	-1.12	1.12	1.26	0.97
9	110	116.01	-6.01	6.01	36.12	5.46
10	130	115.41	14.59	14.59	212.87	11.22
<u>Total</u>				82.95	974.22	66.98

Exponential Smoothing ($\alpha = 0.1$) (2 of 2)

Forecast Accuracy

$$\text{MAE} = \frac{82.95}{9} = 9.22$$

$$\text{MSE} = \frac{974.22}{9} = 108.25$$

$$\text{MAPE} = \frac{66.98}{9} = 7.44\%$$

Exponential smoothing (with $\alpha = 0.1$) provided less accurate forecasts than the 3-MA approach.

Exponential Smoothing ($\alpha = 0.8$) (1 of 2)

<u>Week</u>	<u>Sales</u>	$\alpha = .8$ <u>Forecast</u>	<u>Forecast</u> <u>Error</u>	<u>Absolute</u> <u>Error</u>	<u>Squared</u> <u>Error</u>	<u>Abs.%</u> <u>Error</u>
1	110					
2	115	110.00	5.00	5.00	25.00	4.35
3	125	114.00	11.00	11.00	121.00	8.80
4	120	122.80	-2.20	2.20	7.84	1.83
5	125	120.56	4.44	4.44	19.71	3.55
6	120	124.11	-4.11	4.11	16.91	3.43
7	130	120.82	9.18	9.18	84.23	7.06
8	115	128.16	-13.16	13.16	173.30	11.44
9	110	117.63	-7.63	7.63	58.26	6.94
10	130	111.53	18.47	18.47	341.27	14.21
			<u>Total</u>	75.19	847.52	61.61

Exponential Smoothing ($\alpha = 0.8$) (2 of 2)

Forecast Accuracy

$$\text{MAE} = \frac{75.19}{9} = 8.35$$

$$\text{MSE} = \frac{847.52}{9} = 94.17$$

$$\text{MAPE} = \frac{61.61}{9} = 6.85\%$$

Exponential smoothing (with $\alpha = 0.8$) provided more accurate forecasts than ES with $\alpha = 0.1$, but less accurate than the moving average (with $k = 3$).

Trend Projection

- If a time series exhibits a linear trend, the method of least squares may be used to determine a trend line (projection) for future forecasts.
- Least squares, also used in regression analysis, determines the unique trend line forecast which minimizes the mean square error between the trend line forecasts and the actual observed values for the time series.
- The independent variable is the time period and the dependent variable is the actual observed value in the time series.

Linear Trend Regression (1 of 5)

Using the method of least squares, the formula for the trend projection is:

$$T_t = b_0 + b_1t$$

where: T_t = linear trend forecast in period t
 b_0 = intercept of the linear trend line
 b_1 = slope of the linear trend line
 t = time period

Linear Trend Regression (2 of 5)

For the trend projection equation $T_t = b_0 + b_1 t$

$$b_1 = \frac{\sum_{t=1}^n (t - \bar{t})(Y_t - \bar{Y})}{\sum_{t=1}^n (t - \bar{t})^2} \quad b_0 = \bar{Y} - b_1 \bar{t}$$

where:

- Y_t = value of the time series in period t
- n = number of time periods (observations)
- \bar{Y} = average values of the time series
- \bar{t} = average value of t

Linear Trend Regression (3 of 5)

The number of plumbing repair jobs performed by Auger's Plumbing Service in the last nine months is listed on the right. Forecast the number of repair jobs Auger's will perform in December using the least squares method.

<u>Month</u>	<u>Jobs</u>
March	353
April	387
May	342
June	374
July	396
August	409
September	399
October	412
November	408

Linear Trend Regression (4 of 5)

(month)	t	$t - \bar{t}$	$(t - \bar{t})^2$	y_t	$(Y_t - \bar{Y})$	$(t - \bar{t})(Y_t - \bar{Y})$
(Mar.)	1	-4	16	353	-33.67	134.68
(Apr.)	2	-3	9	387	0.33	-0.99
(May)	3	-2	4	342	-44.67	89.34
(June)	4	-1	1	374	-12.67	12.67
(July)	5	0	0	396	9.33	0
(Aug.)	6	1	1	409	22.33	22.33
(Sep.)	7	2	4	399	12.33	24.66
(Oct.)	8	3	9	412	25.33	75.99
(Nov.)	9	4	16	408	21.33	85.32
Sum	45		60	3480		444.00

Linear Trend Regression (5 of 5)

$$\bar{t} = 45/9 = 5 \quad \text{and} \quad \bar{Y} = 3480/9 = 386.667$$

$$b_1 = \frac{\sum_{t=1}^n (t - \bar{t})(Y_t - \bar{Y})}{\sum_{t=1}^n (t - \bar{t})^2} = \frac{3480}{60} = 7.12$$

$$b_0 = \bar{Y} - b_1 \bar{t} = 386.667 - 7.12(5) = 351.07$$

$$T_{10} = 351.07 + (7.12)(10) = 422.27$$

Trend Projection (1 of 3)

Example: Auger's Plumbing Service

Forecast for December (Month 10) using a three-period ($k = 3$) weighted moving average with weights of 0.6, 0.3, and 0.1 for the newest to oldest data, respectively. Then, compare this Month 10 weighted moving average forecast with the Month 10 trend projection forecast.

<u>Month</u>	<u>Jobs</u>
March	353
April	387
May	342
June	374
July	396
August	409
September	399
October	412
November	408

Trend Projection (2 of 3)

Three-Month Weighted Moving Average

The forecast for December will be the weighted average of the preceding three months: September, October, and November.

$$\begin{aligned} F_{10} &= 0.1Y_{\text{Sep.}} + 0.3Y_{\text{Oct.}} + 0.6Y_{\text{Nov.}} \\ &= 0.1(399) + 0.3(412) + 0.6(408) \\ &= 408.3 \end{aligned}$$

Trend Projection:

$$F_{10} = 422.27 \text{ (from earlier slide)}$$

Trend Projection (3 of 3)

Conclusion:

Due to the positive trend component in the time series, the trend projection produced a forecast that is more in line with the trend that exists. The weighted moving average, even with heavy (0.6) weight placed on the current period, produced a forecast that is lagging behind the changing data.

Nonlinear Trend Regression (1 of 3)

- Sometimes time series have a curvilinear or nonlinear trend.
- A variety of nonlinear functions can be used to develop an estimate of the trend in a time series.
- One example is this quadratic trend equation:

$$T_t = b_0 + b_1t + b_2t^2$$

- Another example is this exponential trend equation:

$$T_t = b_0(b_1)^t$$

Nonlinear Trend Regression (2 of 3)

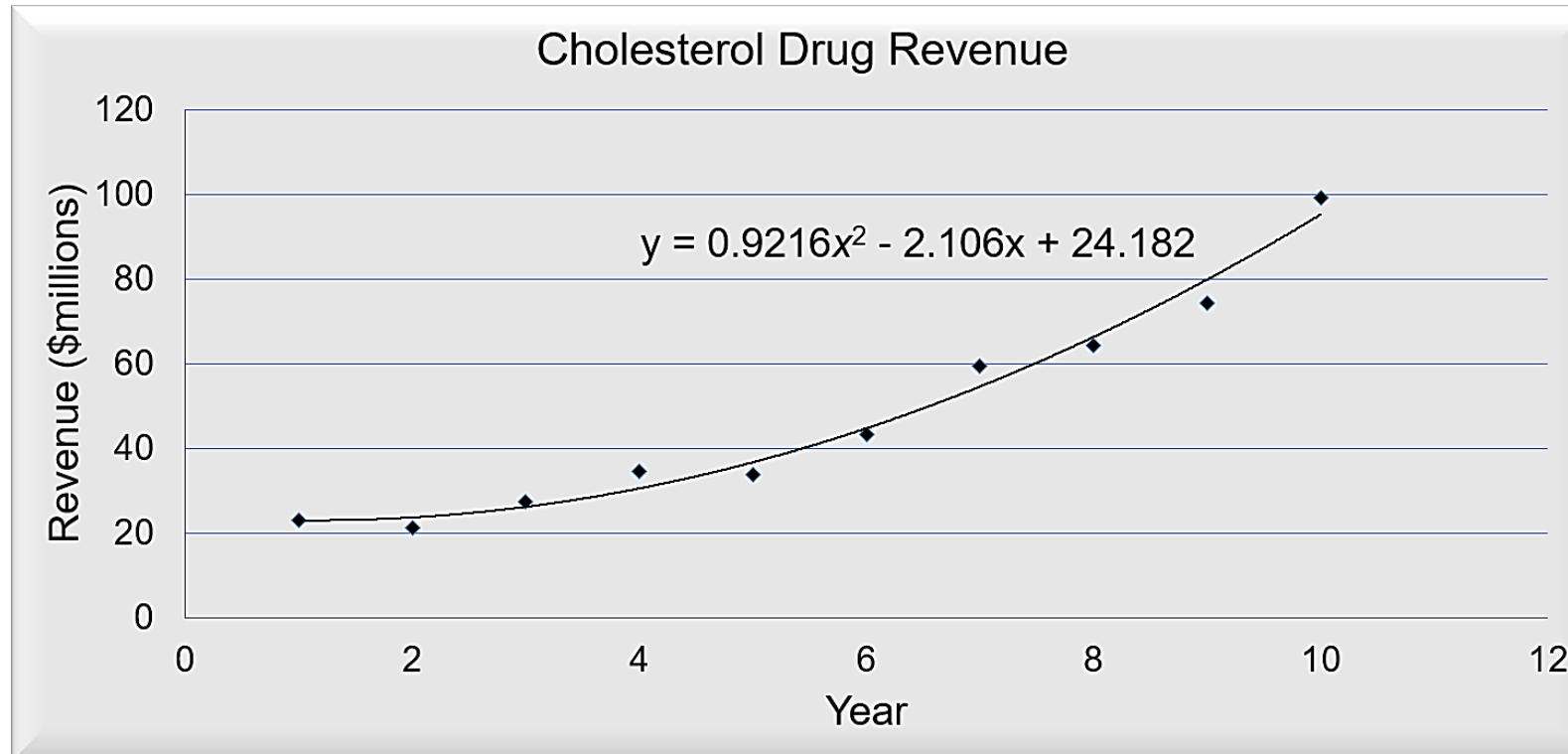
Example: Cholesterol Drug Revenue

The annual revenue in millions of dollars for a cholesterol drug for the first 10 years of sales is shown. A curvilinear function appears to be needed to model the long-term trend.

<u>Year</u>	<u>Revenue</u>
1	23.1
2	21.3
3	27.4
4	34.6
5	33.8
6	43.2
7	59.5
8	64.4
9	74.2
10	99.3

Nonlinear Trend Regression (3 of 3)

Example: Cholesterol Drug Revenue



Seasonality without Trend (1 of 6)

- To the extent that seasonality exists, we need to incorporate it into our forecasting models to ensure accurate forecasts.
- We will first look at the case of a seasonal time series with no trend and then discuss how to model seasonality with trend.

Seasonality without Trend (2 of 6)

Example: Umbrella Sales

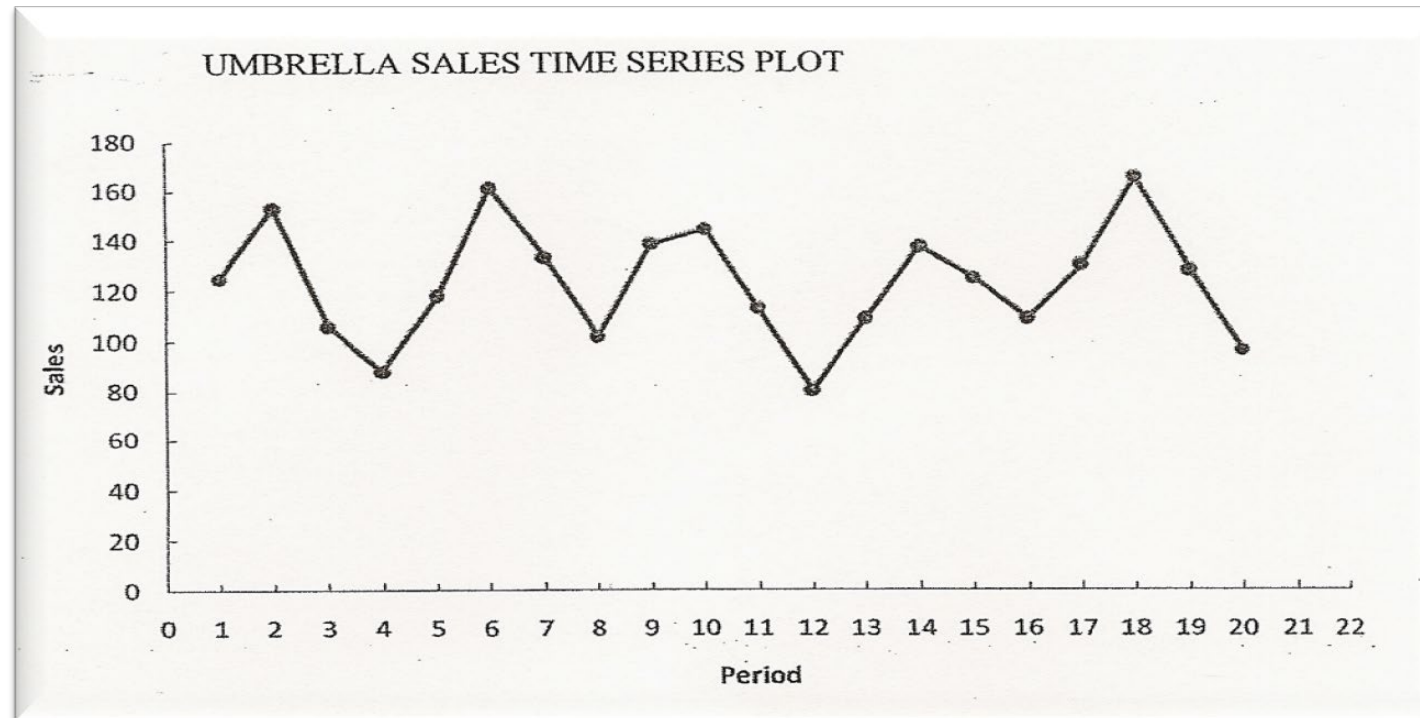
Sometimes it is difficult to identify patterns in a time series presented in a table.

Plotting the time series can be very informative.

<u>Year</u>	<u>Quarter 1</u>	<u>Quarter 2</u>	<u>Quarter 3</u>	<u>Quarter 4</u>
1	125	153	106	88
2	118	161	133	102
3	138	144	113	80
4	109	137	125	109
5	130	165	128	96

Seasonality without Trend (3 of 6)

Umbrella Sales Time Series Plot



Seasonality without Trend (4 of 6)

- The time series plot does not indicate any long-term trend in sales.
- However, close inspection of the plot does reveal a seasonal pattern.
 - The first and third quarters have moderate sales,
 - the second quarter the highest sales, and
 - the fourth quarter tends to be the lowest quarter in terms of sales.

Seasonality without Trend (5 of 6)

- Recall from an earlier chapter that dummy variables can be used to deal with categorical independent variables in a multiple regression model.
- We will treat the season as a categorical variable.
- Recall that when a categorical variable has k levels, $k - 1$ dummy variables are required.
- If there are four seasons, we need three dummy variables.
 - $Qtr1 = 1$ if Quarter 1, 0 otherwise
 - $Qtr2 = 1$ if Quarter 2, 0 otherwise
 - $Qtr3 = 1$ if Quarter 3, 0 otherwise

Seasonality without Trend (6 of 6)

- General Form of Estimated Regression Equation is:

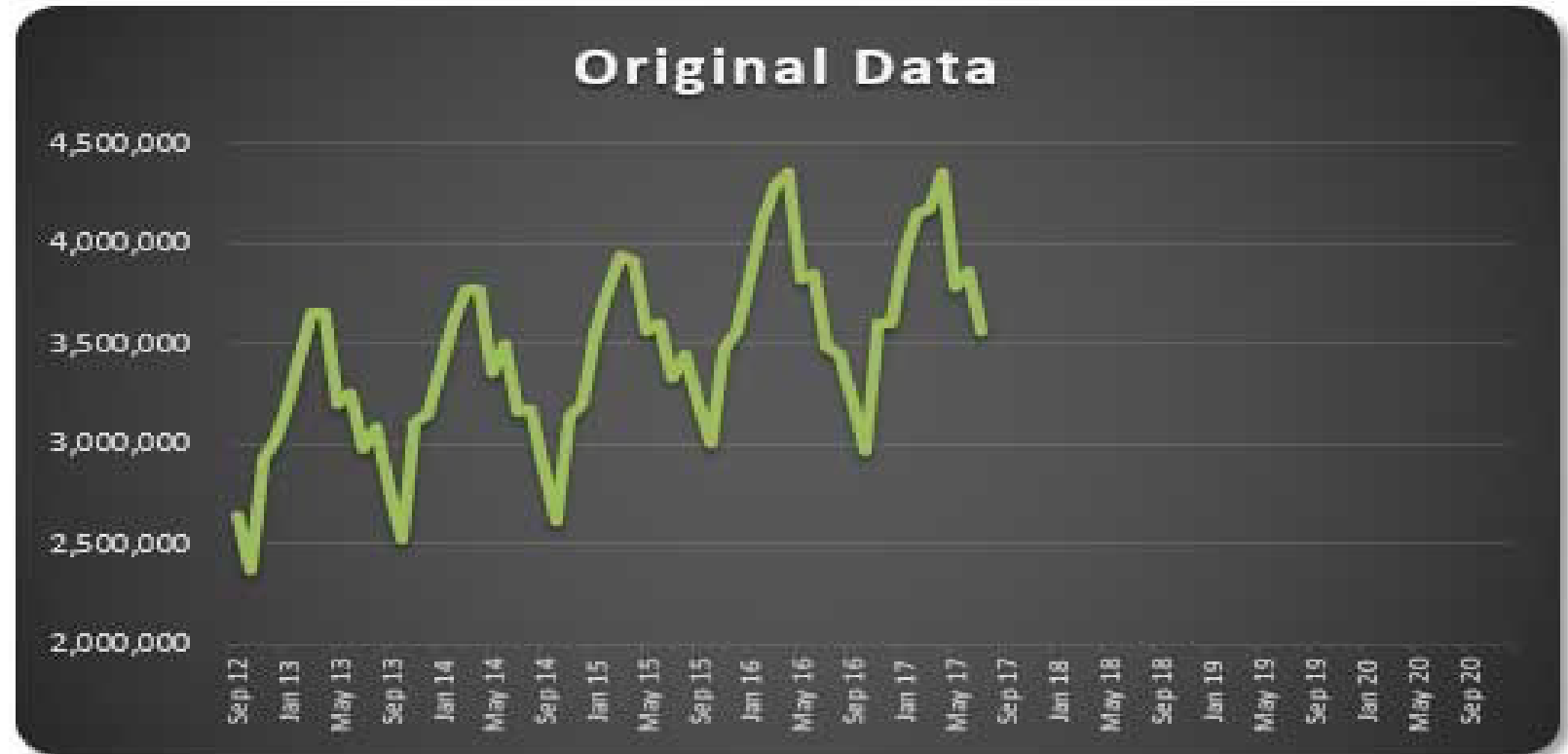
$$\hat{Y} = b_0 + b_1(Qtr1) + b_2(Qtr2) + b_3(Qtr3)$$

- Estimated Regression Equation is:

$$Sales = 95.0 + 29.0(Qtr\ 1) + 57.0(Qtr\ 2) + 26.0(Qtr\ 3)$$

- The forecasts of quarterly sales in year 6 are:
 - Quarter 1: Sales = $95 + 29(1) + 57(0) + 26(0) = 124$
 - Quarter 2: Sales = $95 + 29(0) + 57(1) + 26(0) = 152$
 - Quarter 3: Sales = $95 + 29(0) + 57(0) + 26(1) = 121$
 - Quarter 4: Sales = $95 + 29(0) + 57(0) + 26(0) = 95$

Date	Sales
30 Sep 12	2,644,539
31 Oct 12	2,359,800
30 Nov 12	2,925,918
31 Dec 12	3,024,973
31 Jan 13	3,177,100
28 Feb 13	3,419,595
31 Mar 13	3,649,702
30 Apr 13	3,650,668
31 May 13	3,191,526
30 Jun 13	3,249,428
31 Jul 13	2,971,484
31 Aug 13	3,074,209
30 Sep 13	2,785,466
31 Oct 13	2,515,361
30 Nov 13	3,105,958
31 Dec 13	3,139,059
31 Jan 14	3,380,355
28 Feb 14	3,612,886
31 Mar 14	3,765,824
30 Apr 14	3,771,842
31 May 14	3,356,365
30 Jun 14	3,490,100
31 Jul 14	3,163,659
31 Aug 14	3,167,124
30 Sep 14	2,883,810
31 Oct 14	2,610,667
30 Nov 14	3,129,205
31 Dec 14	3,200,527
31 Jan 15	3,547,804



Seasonality and Trend

Source: https://www.fm-magazine.com/content/fmm-home/news/2017/dec/how-to-use-excel-in-seasonal-forecasting-201718012/_jcr_content/article/articleparsys/image.img.jpg/1513353870032.jpg

Seasonality and Trend (1 of 4)

- We will now extend the regression approach to include situations where the time series contains both a seasonal effect and a linear trend.
- We will introduce an additional independent variable to represent time.

Example: Terry's Tie Shop

Business at Terry's Tie Shop can be viewed as falling into three distinct seasons: (1) Christmas (November and December); (2) Father's Day (late May to mid June); and (3) all other times.

Average weekly sales (\$) during each of the three seasons during the past four years are shown on the next slide.

Seasonality and Trend (2 of 4)

Example: Terry's Tie Shop

Determine a forecast for the average weekly sales in year 5 for each of the three seasons.

<u>Year</u>	<u>Season 1</u>	<u>Season 2</u>	<u>Season 3</u>
1	1856	2012	985
2	1995	2168	1072
3	2241	2306	1105
4	2280	2408	1120

Seasonality and Trend (3 of 4)

- There are three seasons, so we will need two dummy variables.
 - $Seas1 = 1$ if Season 1, 0 otherwise
 - $Seas2 = 1$ if Season 2, 0 otherwise
- General Form of Estimated Regression Equation is:

$$\hat{Y} = b_0 + b_1(Seas1) + b_2(Seas2) + b_3(t)$$

- Estimated Regression Equation is:

$$Sales = 797.0 + 1095.43(Seas1) + 1189.47(Seas2) + 36.47(t)$$

Seasonality and Trend (4 of 4)

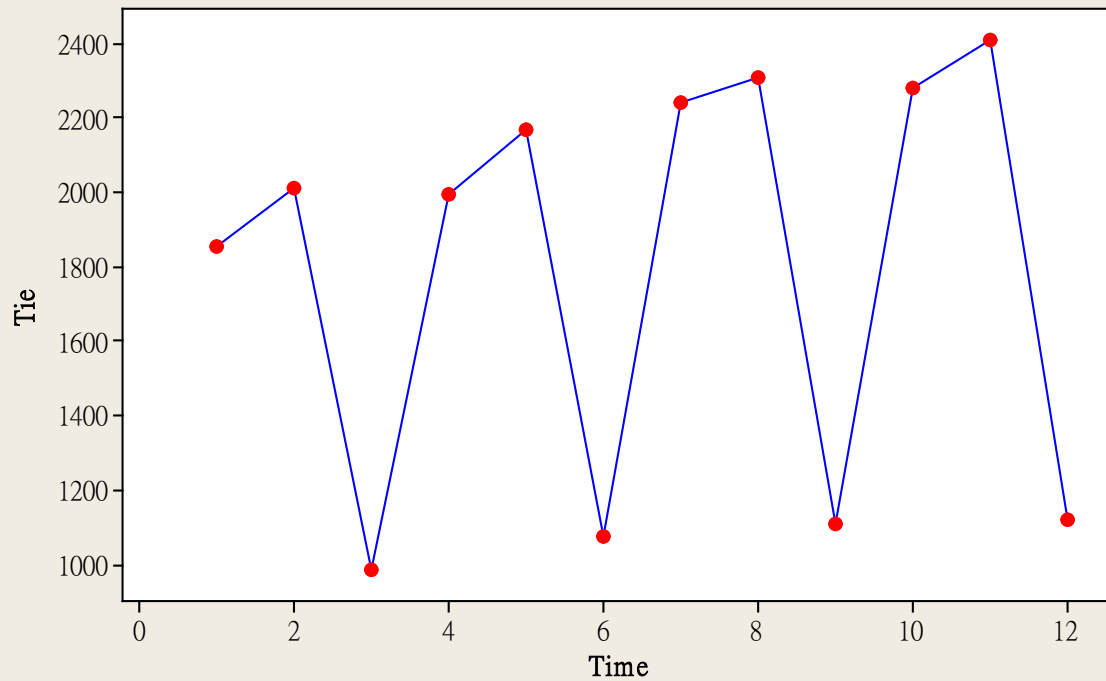
The forecasts of average weekly sales in the three seasons of year 5 are:

$$\begin{aligned}\text{Seas. 1: Sales} &= 797 + 1095.43(1) + 1189.47(0) + 36.47(13) \\ &= 2366.5\end{aligned}$$

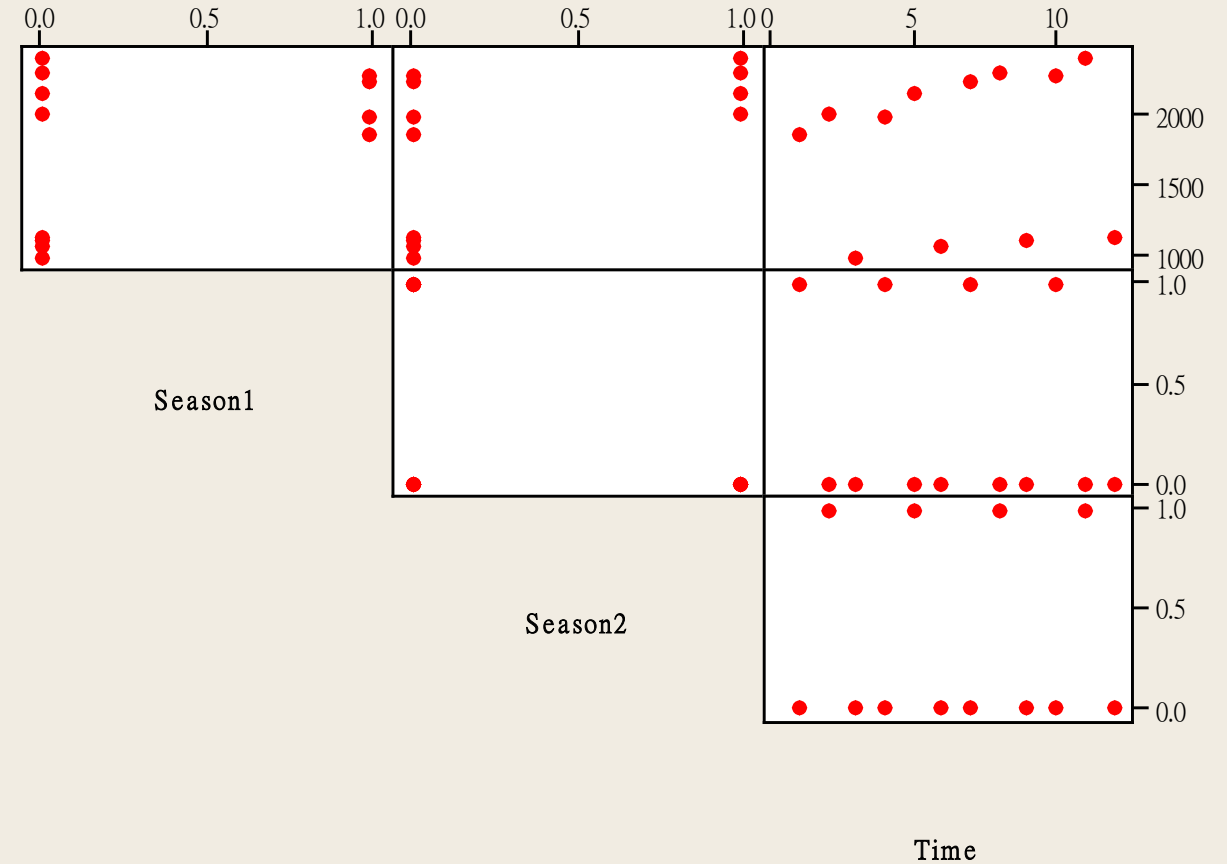
$$\begin{aligned}\text{Seas. 2: Sales} &= 797 + 1095.43(0) + 1189.47(1) + 36.47(14) \\ &= 2497.0\end{aligned}$$

$$\begin{aligned}\text{Seas. 3: Sales} &= 797 + 1095.43(0) + 1189.47(0) + 36.47(15) \\ &= 1344.0\end{aligned}$$

Scatterplot of Tie vs Time



Matrix Plot of Tie, Season1, Season2, Time



The regression equation is

$$\text{Tie} = 797 + 1095 \text{ Season1} + 1189 \text{ Season2} + 36.5 \text{ Time}$$

Predictor	Coef	SE Coef	T	P
Constant	797.00	59.67	13.36	0.000
Season1	1095.43	53.19	20.60	0.000
Season2	1189.47	52.06	22.85	0.000
Time	36.467	6.290	5.80	0.000

$$S = 73.0861 \quad R\text{-Sq} = 98.7\% \quad R\text{-Sq}(\text{adj}) = 98.3\%$$

Unusual Observations

Obs	Season1	Tie	Fit	SE Fit	Residual	St Resid
12	0.00	1120.0	1234.6	46.2	-114.6	-2.02R

Time Series Decomposition (1 of 5)

- Time series decomposition can be used to separate or decompose a time series into seasonal, trend, and irregular (error) components.
- While this method can be used for forecasting, its primary applicability is to get a better understanding of the time series.
- Understanding what is really going on with a time series often depends upon the use of deseasonalized data.

Time Series Decomposition (2 of 5)

- Decomposition methods assume that the actual time series value at period t is a function of three components: trend, seasonal, and irregular.
- How these three components are combined to give the observed values of the time series depends upon whether we assume the relationship is best described by an additive or a multiplicative model.

Time Series Decomposition (3 of 5)

An additive model follows the form:

$$Y_t = \text{Trend}_t + \text{Seasonal}_t + \text{Irregular}_t$$

where:

Trend_t = trend value at time period t

Seasonal_t = seasonal value at time period t

Irregular_t = irregular value at time period t

An additive model is appropriate in situations where the seasonal fluctuations do not depend upon the level of the time series.

Time Series Decomposition (4 of 5)

A multiplicative model follows the form:

$$Y_t = \text{Trend}_t \times \text{Seasonal}_t \times \text{Irregular}_t$$

where:

Trend_t = trend value at time period t

Seasonal_t = seasonal value at time period t

Irregular_t = irregular value at time period t

A multiplicative model is appropriate, for example, if the seasonal fluctuations grow larger as the sales volume increases because of a long-term linear trend.

Time Series Decomposition (5 of 5)

Example: Terry's Tie Shop

Determine a forecast for the average weekly sales in year 5 for each of the three seasons.

<u>Year</u>	<u>Season 1</u>	<u>Season 2</u>	<u>Season 3</u>
1	1856	2012	985
2	1995	2168	1072
3	2241	2306	1105
4	2280	2408	1120

Calculating the Seasonal Indexes (1 of 9)

Step 1. Calculate the centered moving averages.

There are three distinct seasons in each year. Hence, take a three-season moving average to eliminate seasonal and irregular factors.

For example:

$$1^{\text{st}} \text{ CMA} = (1856 + 2012 + 985)/3 = 1617.67$$

$$2^{\text{nd}} \text{ CMA} = (2012 + 985 + 1995)/3 = 1664.00$$

Etc.

Calculating the Seasonal Indexes (2 of 9)

Step 2. Center the CMAs on integer-valued periods.

The first centered moving average computed in step 1 (1617.67) will be centered on season 2 of year 1. Note that the moving averages from step 1 center themselves on integer-valued periods because n is an odd number.

Calculating the Seasonal Indexes (3 of 9)

Year	Season	Dollar Sales (Y_t)	Moving Average
1	1	1856	1617.67
	2	2012	
	3	985	
2	1	1995	1716.00
	2	2168	1745.00
	3	1072	1827.00
3	1	2241	1873.00
	2	2306	1884.00
	3	1105	1897.00
4	1	2280	1931.00
	2	2408	1936.00
	3	1120	

Calculating the Seasonal Indexes (4 of 9)

- The centered moving average values tend to “smooth out” both the seasonal and irregular fluctuations in the time series.
- The centered moving averages represent the trend in the data and any random variation that was not removed by using the moving averages to smooth the data.

Calculating the Seasonal Indexes (5 of 9)

Step 3. Determine the seasonal & irregular factors ($S_t l_t$).

By dividing each actual by the moving average for the same time period, we identify the combined seasonal-irregular effect in the time series.

$$S_t l_t = \frac{Y_t}{\text{Moving average for period } t}$$

Calculating the Seasonal Indexes (6 of 9)

Year	Season	Dollar Sales (Y_t)	Moving Average	$S_t l_t$
1	1	1856		
	2	2012	1617.67	1.244
	3	985	1664.00	.592
2	1	1995	1716.00	1.163
	2	2168	1745.00	1.242
	3	1072	1827.00	.587
3	1	2241	1873.00	1.196
	2	2306	1884.00	1.224
	3	1105	1897.00	.582
4	1	2280	1931.00	1.181
	2	2408	1936.00	1.244
	3	1120		

Calculating the Seasonal Indexes (7 of 9)

Step 4. Determine the average seasonal factors.

Averaging all $S_t l_t$ values corresponding to that season:

$$\begin{aligned} \text{Season 1: } (1.163 + 1.196 + 1.181)/3 &= 1.180 \\ \text{Season 2: } (1.244 + 1.242 + 1.224 + 1.244)/4 &= 1.238 \\ \text{Season 3: } (.592 + .587 + .582)/3 &= \underline{0.587} \\ &3.005 \end{aligned}$$

Calculating the Seasonal Indexes (8 of 9)

Step 5. Scale the seasonal factors (S_t).

Average the seasonal factors = $(1.180 + 1.238 + 0.587)/3 = 1.002$. Then, divide each seasonal factor by the average of the seasonal factors.

Season 1: $1.180/1.002 = 1.178$

Season 2: $1.238/1.002 = 1.236$

Season 3: $.587/1.002 = \underline{0.586}$
3.000

Calculating the Seasonal Indexes (9 of 9)

Year	Season	Dollar Sales (Y_t)	Moving Average	$S_t l_t$	Scaled S_t
1	1	1856			1.178
	2	2012	1617.67	1.244	1.236
	3	985	1664.00	.592	.586
2	1	1995	1716.00	1.163	1.178
	2	2168	1745.00	1.242	1.236
	3	1072	1827.00	.587	.586
3	1	2241	1873.00	1.196	1.178
	2	2306	1884.00	1.224	1.236
	3	1105	1897.00	.582	.586
4	1	2280	1931.00	1.181	1.178
	2	2408	1936.00	1.244	1.236
	3	1120			.586

Using the Deseasonalizing Time Series to Identify Trend (1 of 3)

Step 6. Determine the deseasonalized data. Divide the data point values, Y_t , by S_t .

Year	Season	Dollar Sales (Y_t)	Moving Average	$S_t l_t$	Scaled S_t	Y_t/S_t
1	1	1856			1.178	1576
	2	2012	1617.67	1.244	1.236	1628
	3	985	1664.00	.592	.586	1681
2	1	1995	1716.00	1.163	1.178	1694
	2	2168	1745.00	1.242	1.236	1754
	3	1072	1827.00	.587	.586	1829
3	1	2241	1873.00	1.196	1.178	1902
	2	2306	1884.00	1.224	1.236	1866
	3	1105	1897.00	.582	.586	1886
4	1	2280	1931.00	1.181	1.178	1935
	2	2408	1936.00	1.244	1.236	1948
	3	1120			.586	1911

Using the Deseasonalizing Time Series to Identify Trend (2 of 3)

Step 7. Determine a trend line of the deseasonalized data.

Using the least squares method for $t = 1, 2, \dots, 12$, gives:

$$T_t = b_0 + b_1 t$$

$$\text{Deseasonalized Sales}_t = 1580.11 + 33.96t$$

Using the Deseasonalizing Time Series to Identify Trend (3 of 3)

Step 8. Determine the deseasonalized predictions.

Substitute $t = 13, 14,$ and 15 into the least squares equation:

$$T_{13} = 1580.11 + (33.96)(13) = 2022$$

$$T_{14} = 1580.11 + (33.96)(14) = 2056$$

$$T_{15} = 1580.11 + (33.96)(15) = 2090$$

Seasonal Adjustments

Step 9. Take into account the seasonality.

Multiply each deseasonalized prediction by its seasonal factor to give the following forecasts year 5:

Season 1: $(1.178)(2022) = 2382$

Season 2: $(1.236)(2056) = 2541$

Season 3: $(0.586)(2090) = 1225$