Time Series Analysis

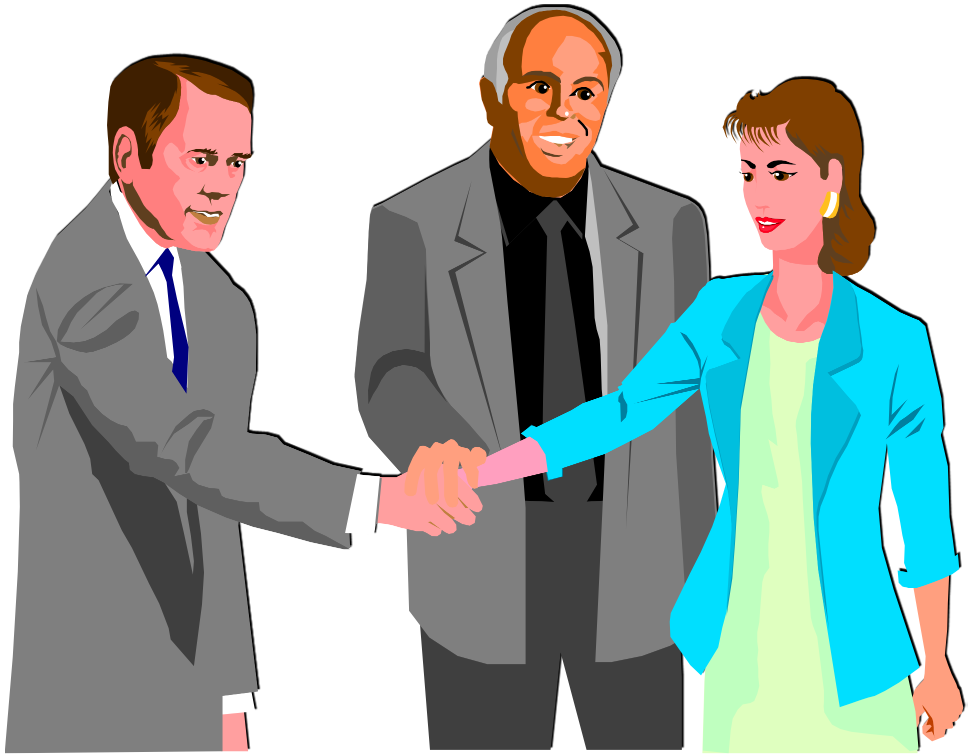
Anshu Pandey

Learning Objectives

- Describe what forecasting is
- Explain time series & its components
- Smooth a data series
  - Moving average
  - Exponential smoothing
- Forecast using trend models
  - Simple Linear Regression
  - Auto-regressive

What Is Forecasting?

- Process of predicting a future event
- Underlying basis of all business decisions
  - Production
  - Inventory
  - Personnel
  - Facilities



Forecasting Approaches

Qualitative Methods

- Used when situation is vague & little data exist
  - New products
  - New technology
- Involve intuition, experience
- e.g., forecasting sales on Internet

Quantitative Methods

Forecasting Approaches

Qualitative Methods

- Used when situation is vague & little data exist
  - New products
  - New technology
- Involve intuition, experience
- e.g., forecasting sales on Internet

Quantitative Methods

- Used when situation is ‘stable’ & historical data exist
  - Existing products
  - Current technology
- Involve mathematical techniques
- e.g., forecasting sales of color televisions

Quantitative Forecasting

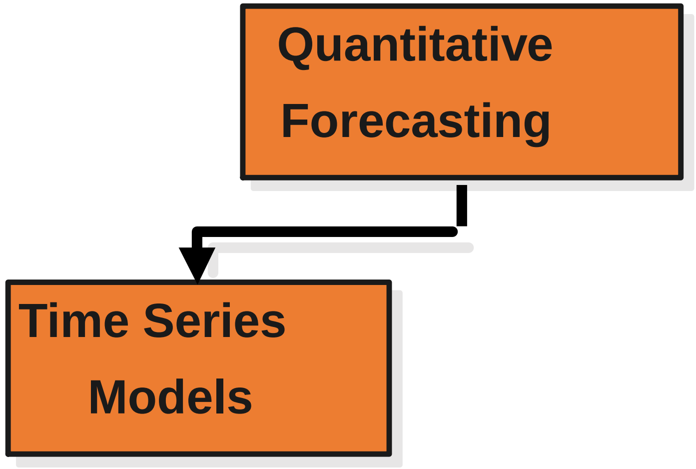
- Select several forecasting methods
- ‘Forecast’ the past
- Evaluate forecasts
- Select best method
- Forecast the future
- Monitor continuously forecast accuracy

Quantitative Forecasting Methods

Quantitative Forecasting Methods

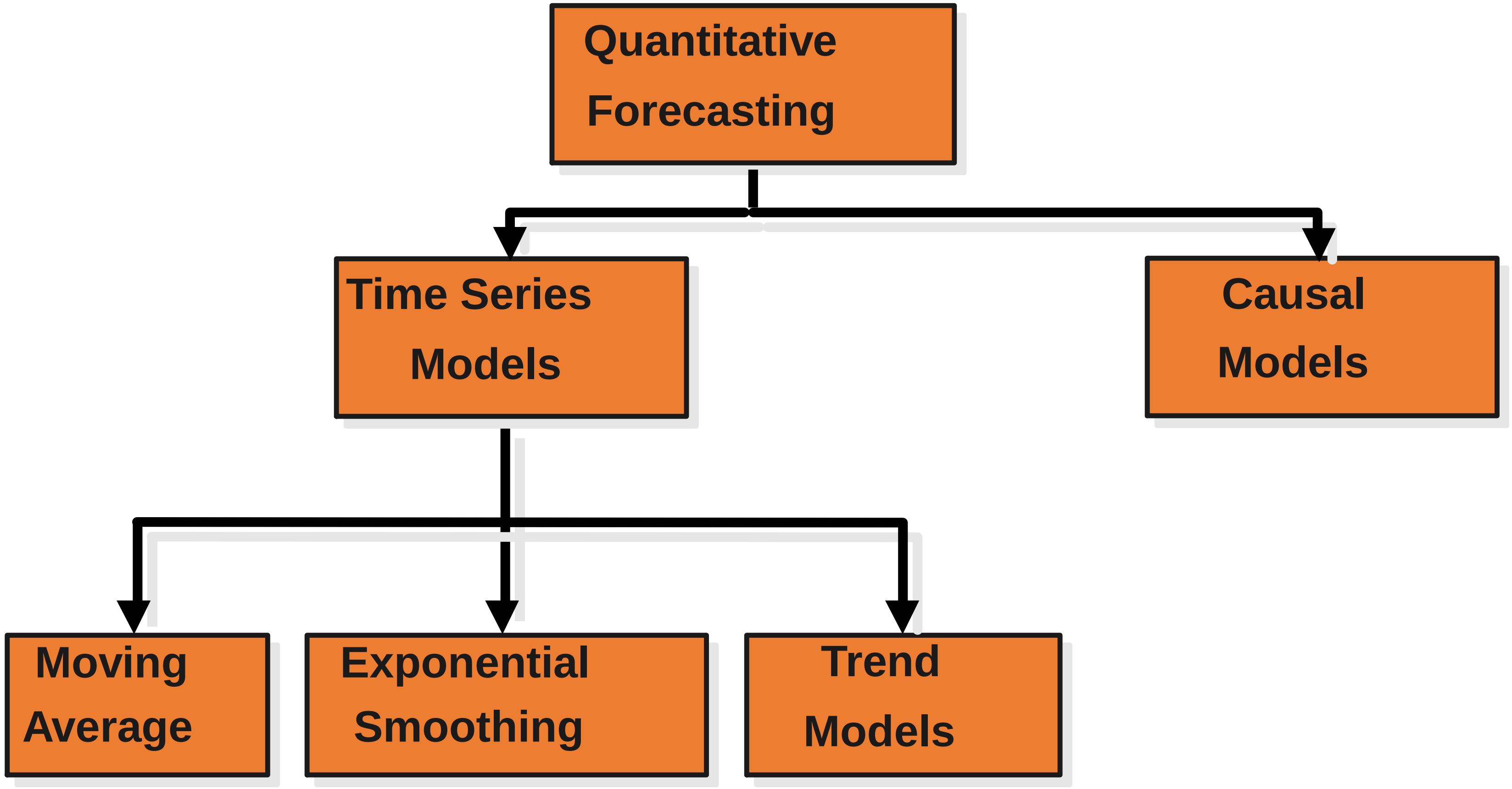
Quantitative
Forecasting

Quantitative Forecasting Methods

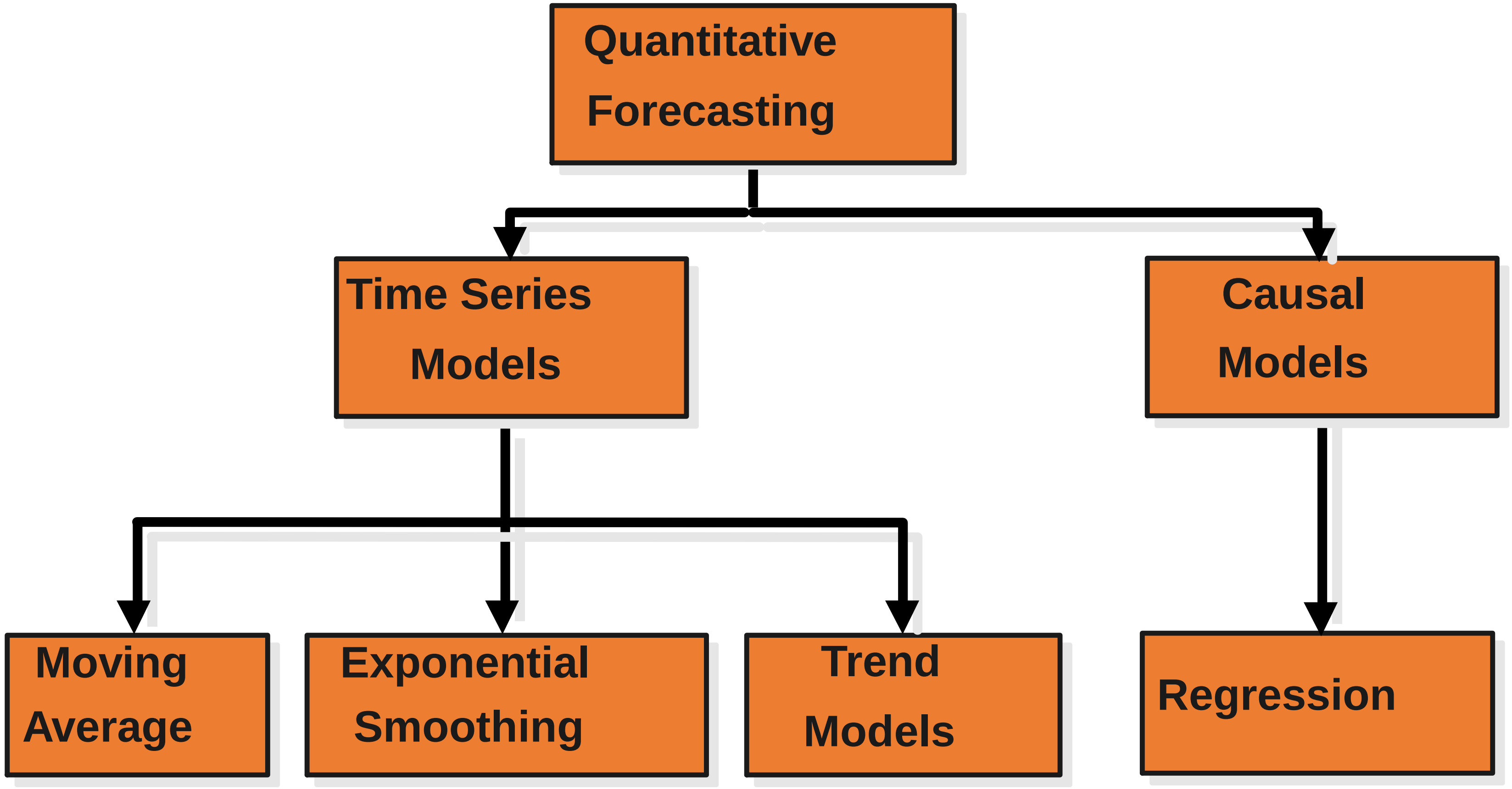


Quantitative Forecasting Methods

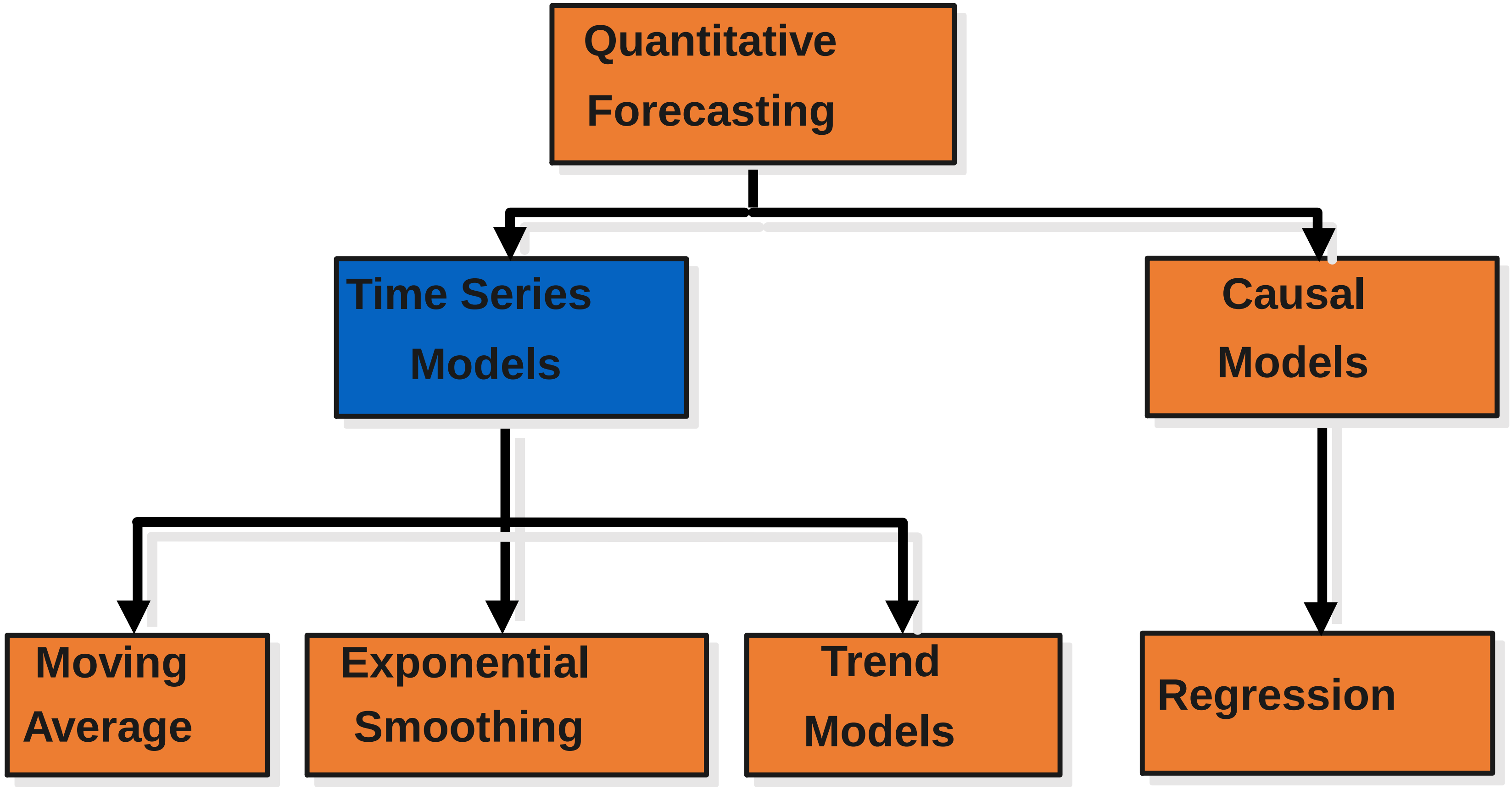
Quantitative Forecasting Methods



Quantitative Forecasting Methods



Quantitative Forecasting Methods



What is a Time Series?

- Set of evenly spaced numerical data
  - Obtained by observing response variable at regular time periods
- Forecast based only on past values
  - Assumes that factors influencing past, present, & future will continue
- Example

• Year:	1995	1996	1997	1998	1999
• Sales:	78.7	63.5	89.7	93.2	92.1

Time Series vs.
Cross Sectional Data

Time series data is a sequence of observations

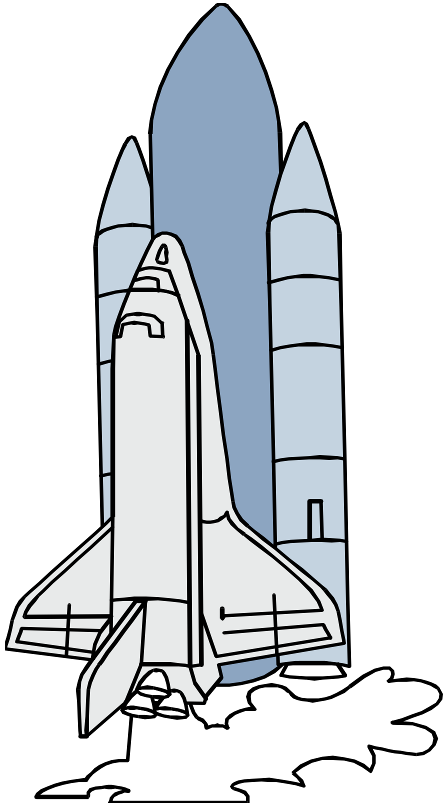
- **collected from a process**
- **with equally spaced periods of time.**

Time Series vs. Cross Sectional Data

Contrary to restrictions placed on cross-sectional data, the major purpose of forecasting with time series is to extrapolate beyond the range of the explanatory variables.

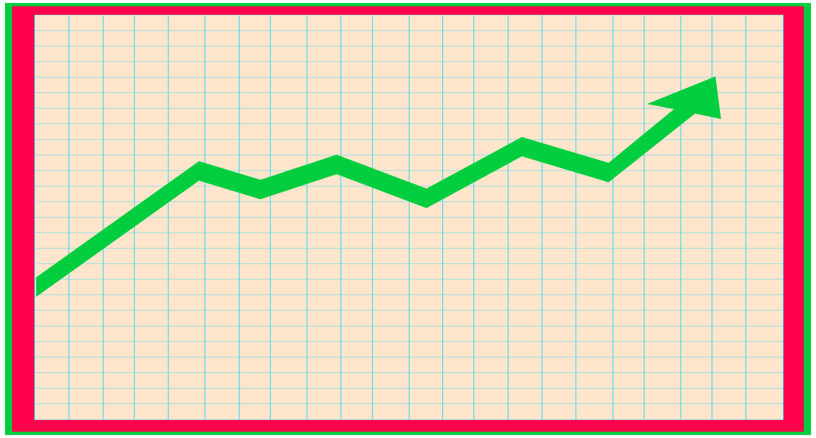
Time Series vs. Cross Sectional Data

Time series is dynamic, it does change over time.



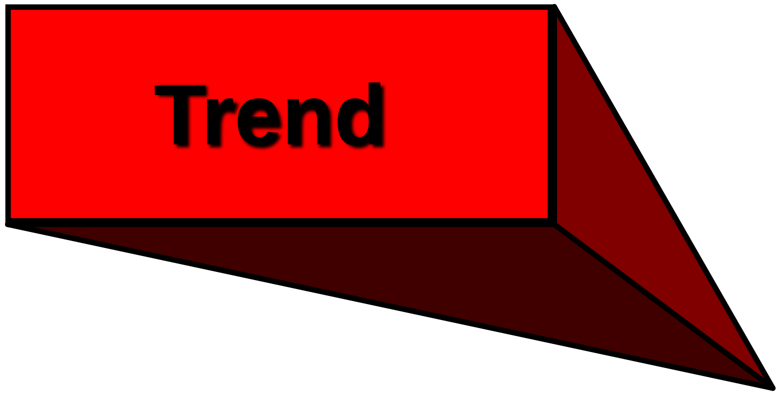
Time Series vs. Cross Sectional Data

When working with time series data, it is paramount that the data is plotted so the researcher can view the data.

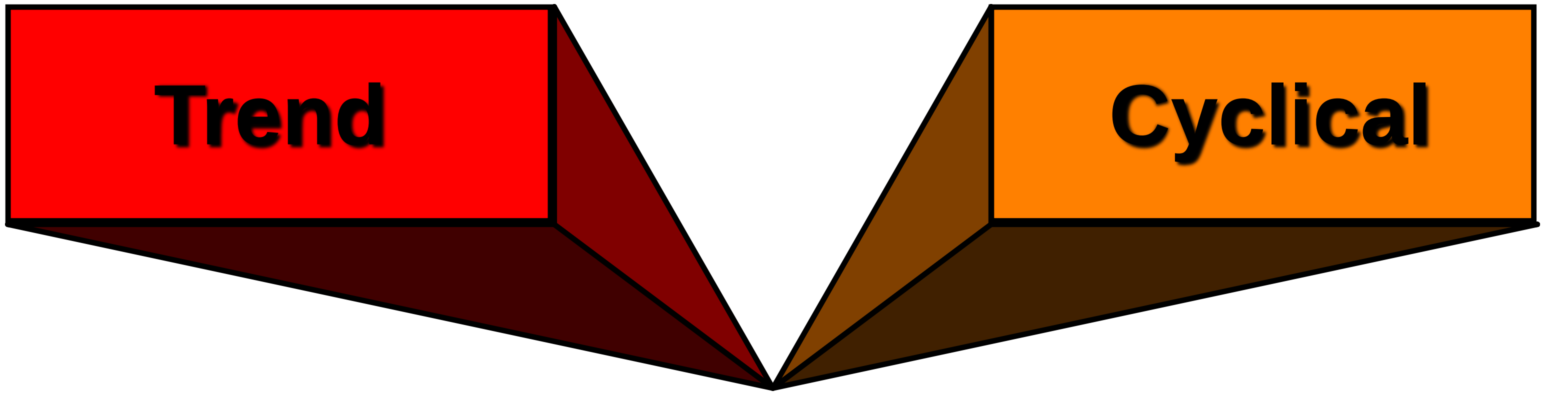


Time Series Components

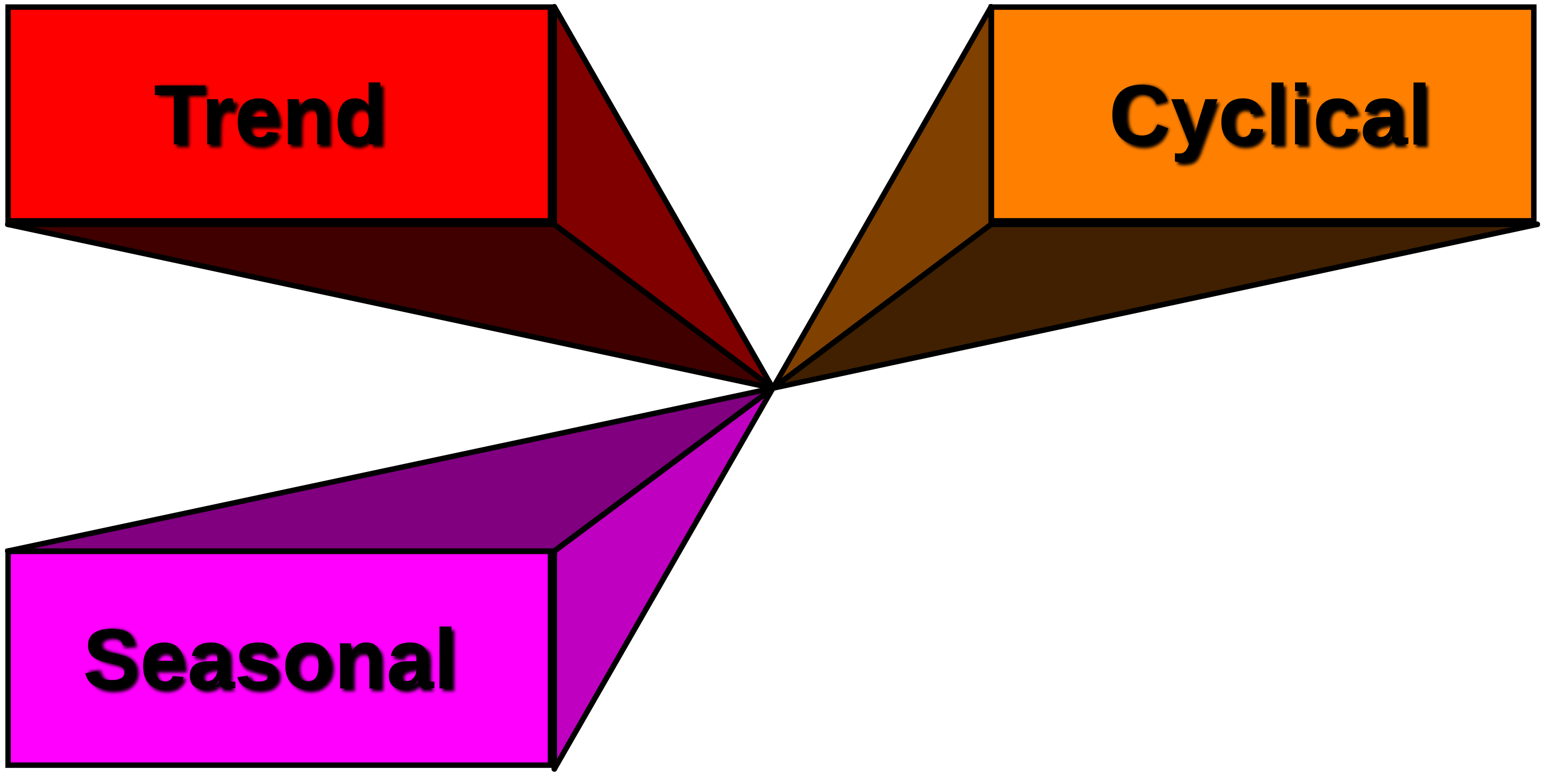
Time Series Components



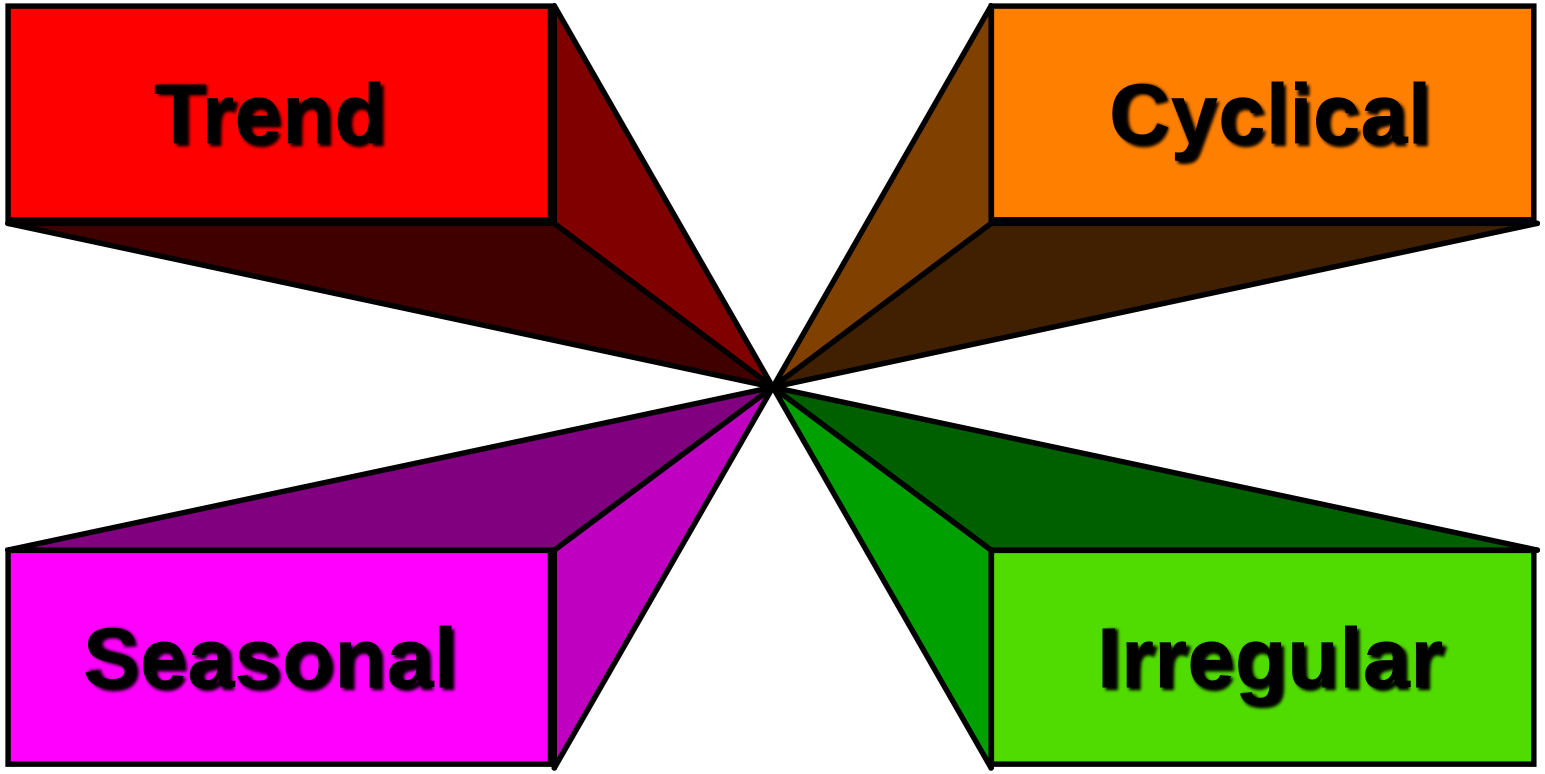
Time Series Components



Time Series Components

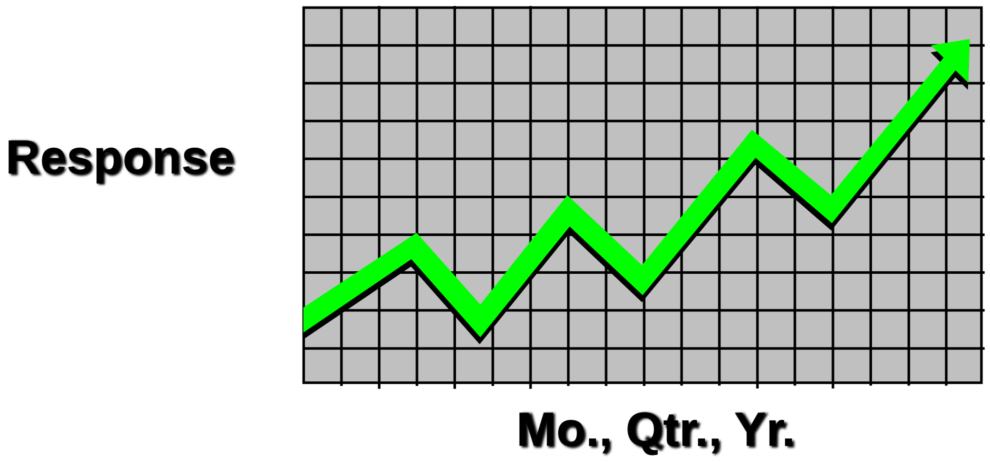


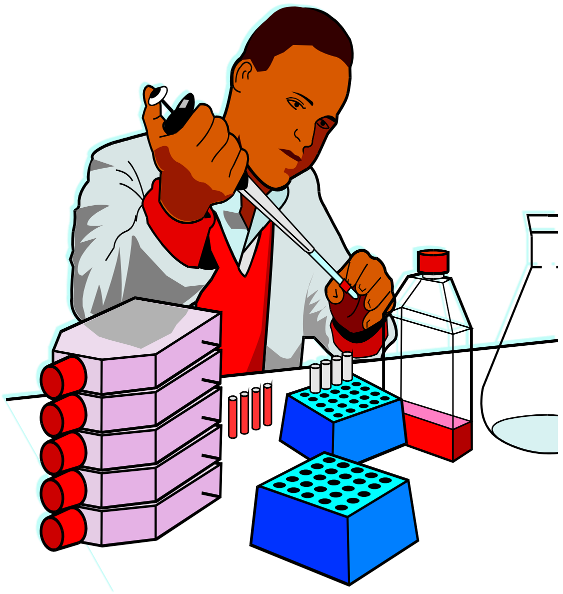
Time Series Components



Trend Component

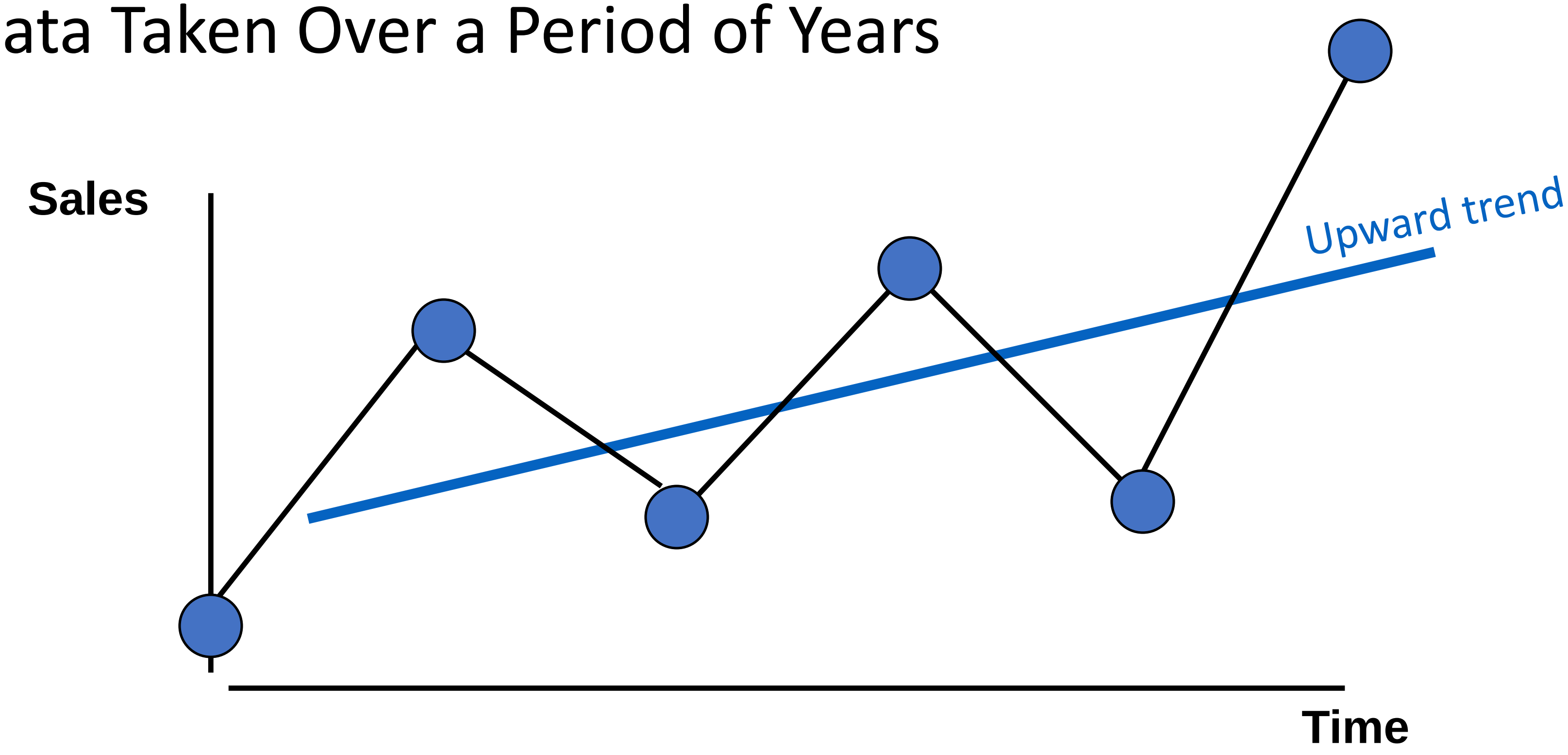
- Persistent, overall upward or downward pattern
- Due to population, technology etc.
- Several years duration





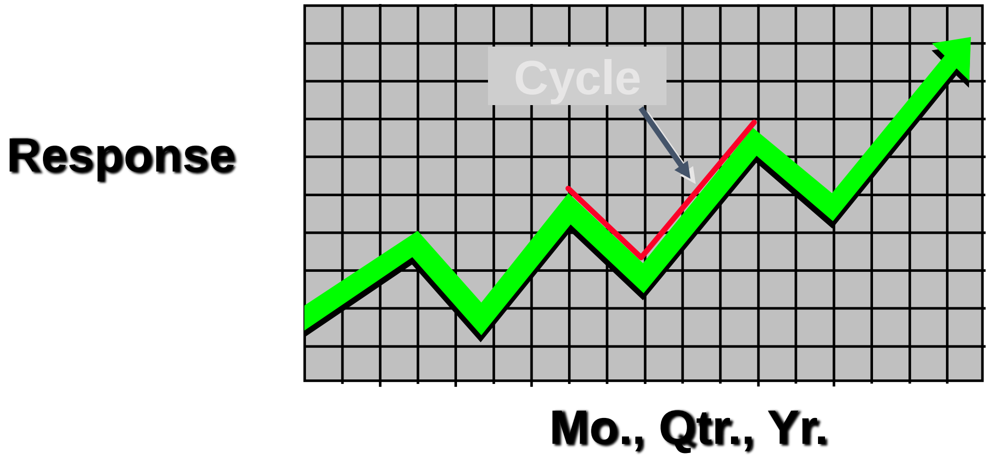
Trend Component

- Overall Upward or Downward Movement
- Data Taken Over a Period of Years



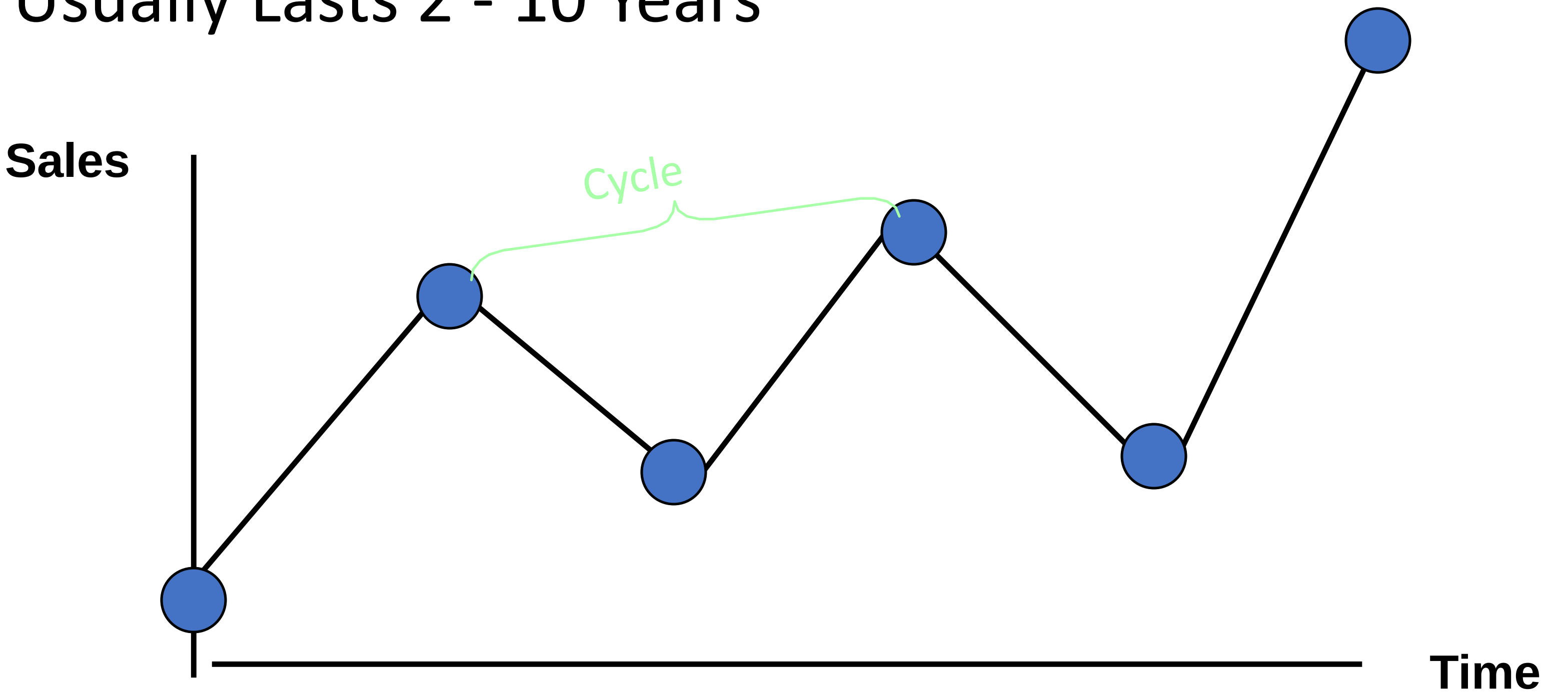
Cyclical Component

- Repeating up & down movements
- Due to interactions of factors influencing economy
- Usually 2-10 years duration



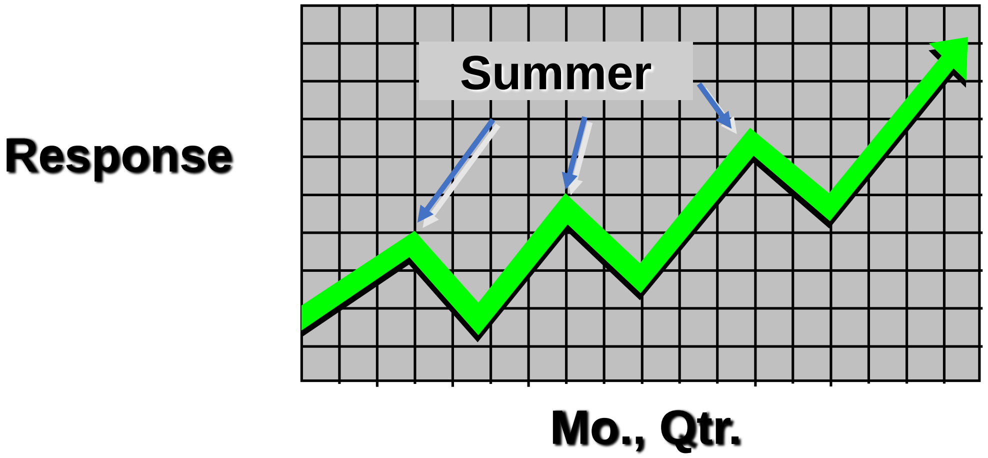
Cyclical Component

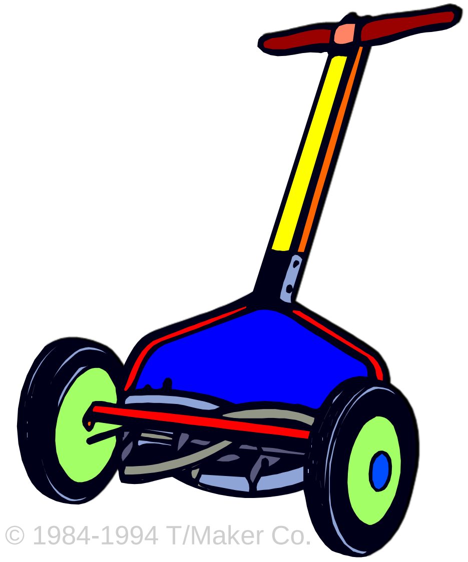
- Upward or Downward Swings
- May Vary in Length
- Usually Lasts 2 - 10 Years



Seasonal Component

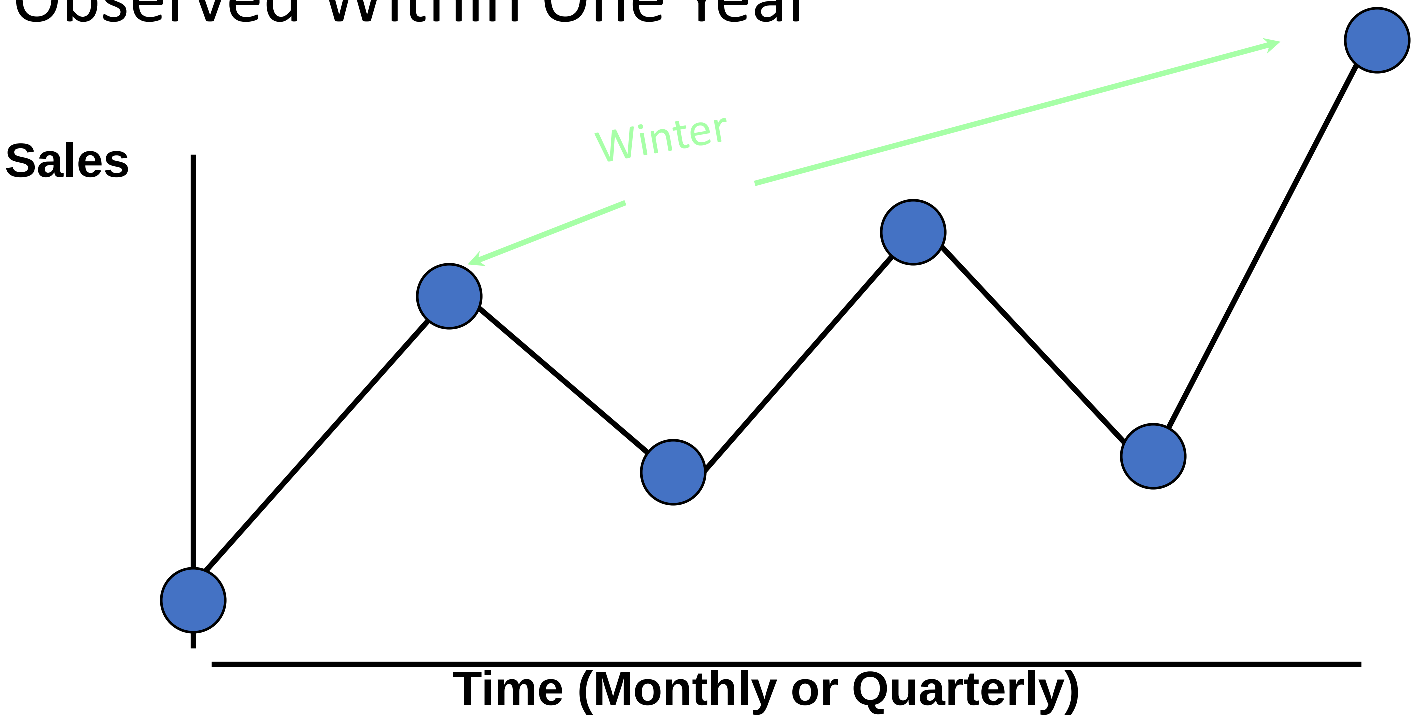
- Regular pattern of up & down fluctuations
- Due to weather, customs etc.
- Occurs within one year





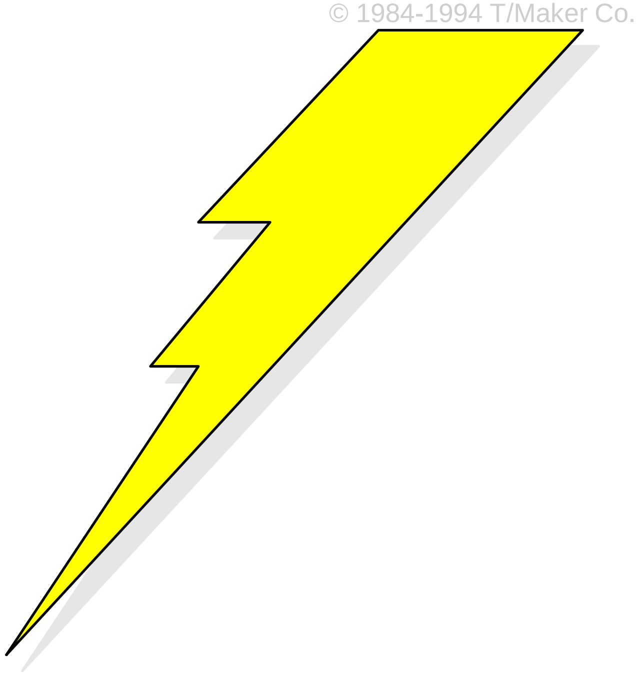
Seasonal Component

- Upward or Downward Swings
- Regular Patterns
- Observed Within One Year



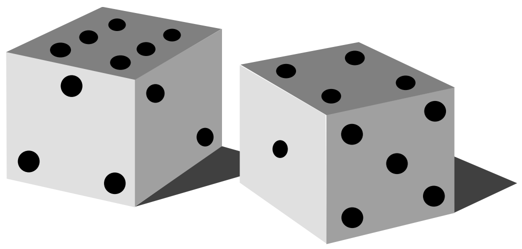
Irregular Component

- Erratic, unsystematic, ‘residual’ fluctuations
- Due to random variation or unforeseen events
  - Union strike
  - War
- Short duration & nonrepeating



Random or Irregular Component

- Erratic, Nonsystematic, Random, ‘Residual’ Fluctuations
- Due to Random Variations of
  - Nature
  - Accidents
- Short Duration and Non-repeating

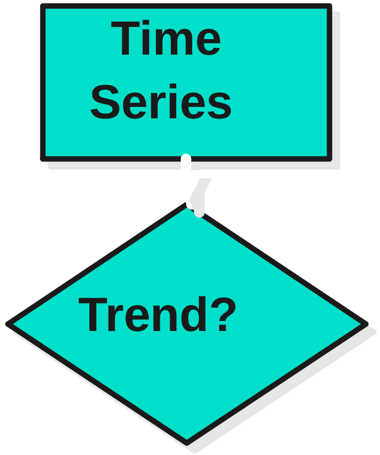


Time Series Forecasting

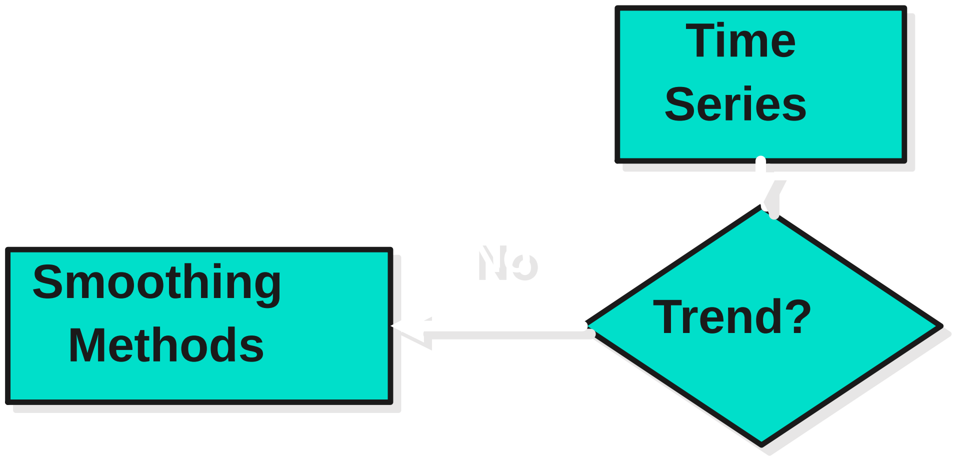
Time Series Forecasting

Time
Series

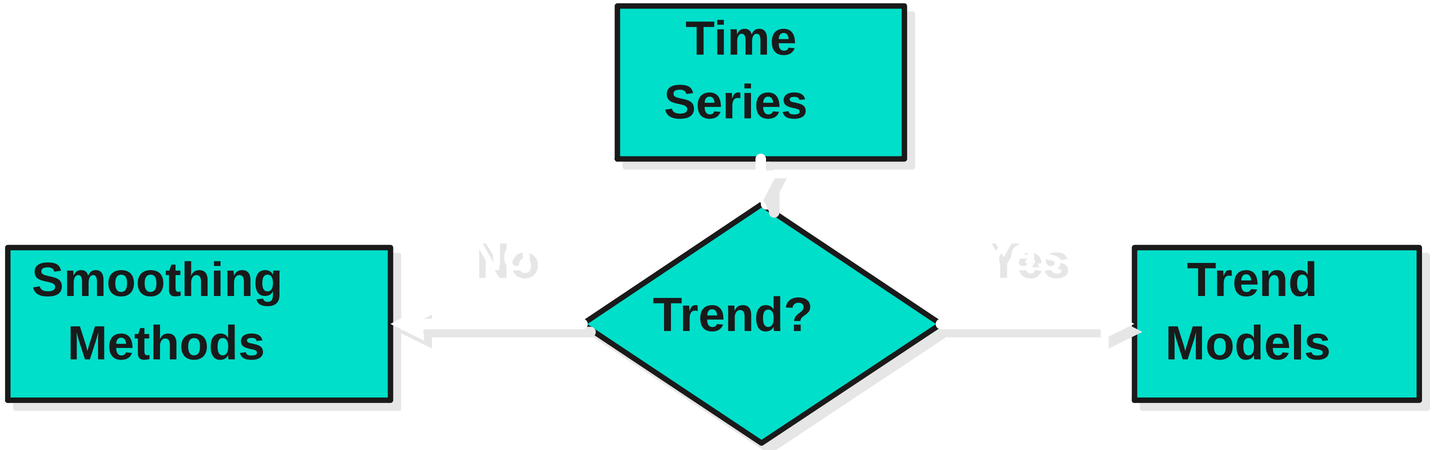
Time Series Forecasting



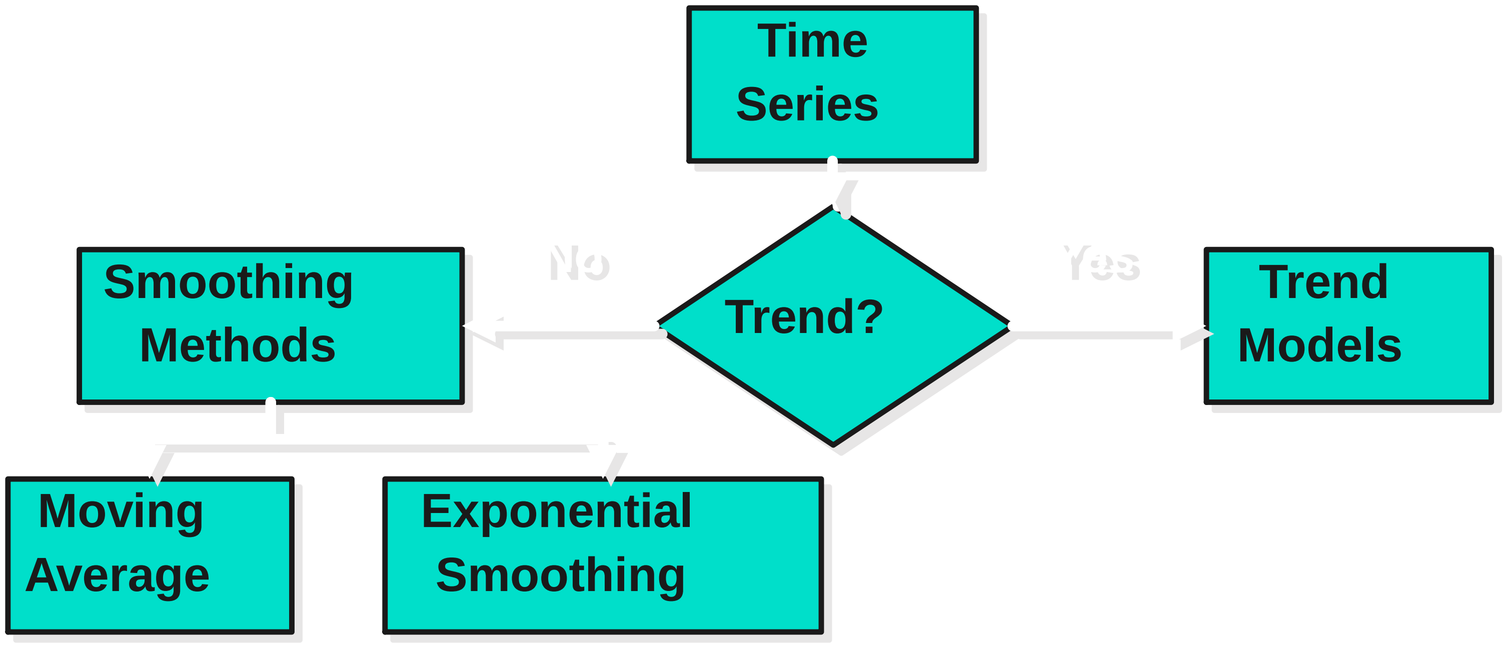
Time Series Forecasting



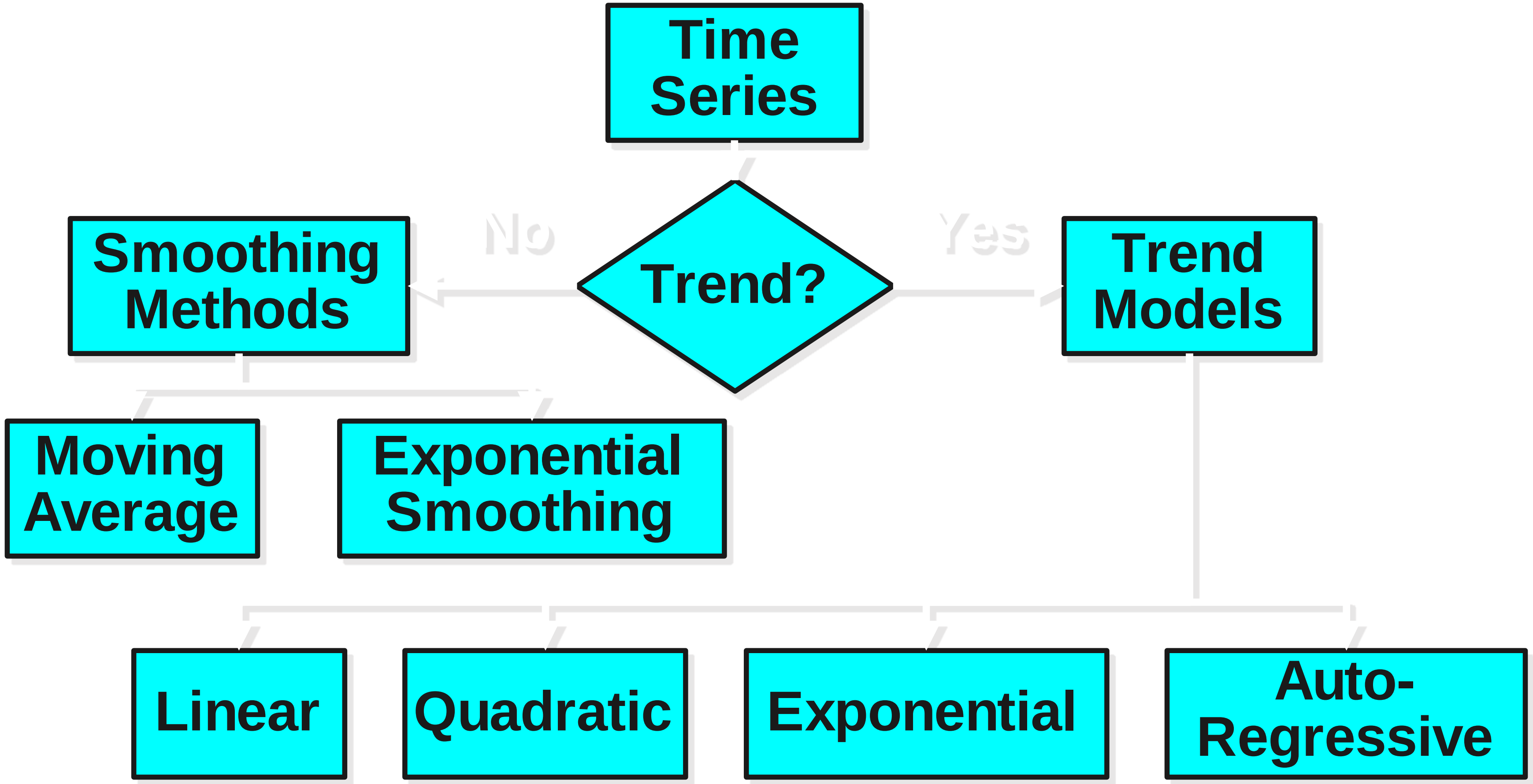
Time Series Forecasting



Time Series Forecasting



Time Series Forecasting



Time Series Analysis

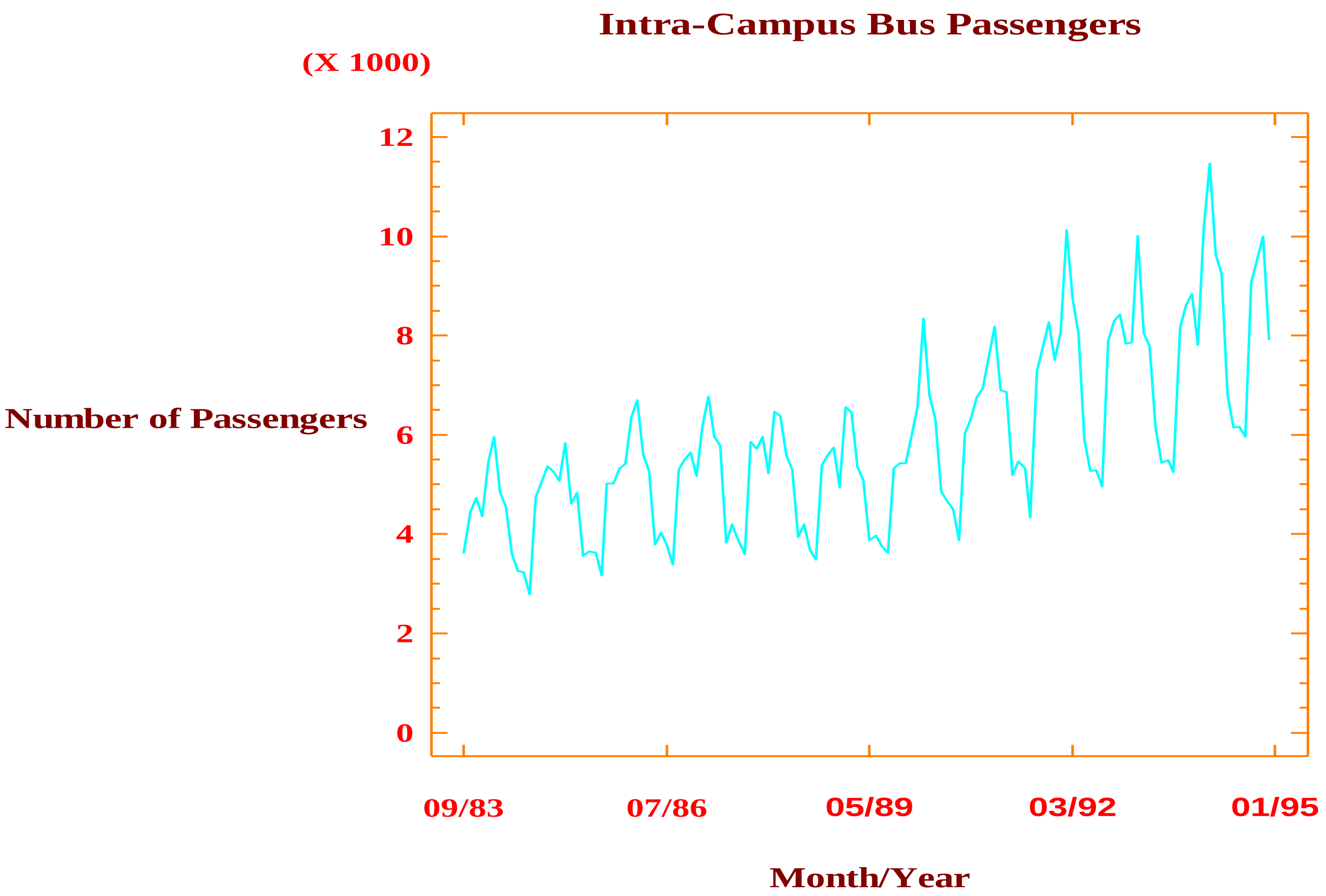
Plot time series data

$\Downarrow$

No $\Leftarrow$ trend? $\Rightarrow$ ***Yes***

$\Downarrow$			$\Downarrow$
Smoothing	Forecasting	$\Leftrightarrow$	Decomposition
$\Downarrow$	$\Downarrow$		$\Downarrow$
moving averages exponential smoothing	linear quadratic exponential autoregressive		cyclical analysis

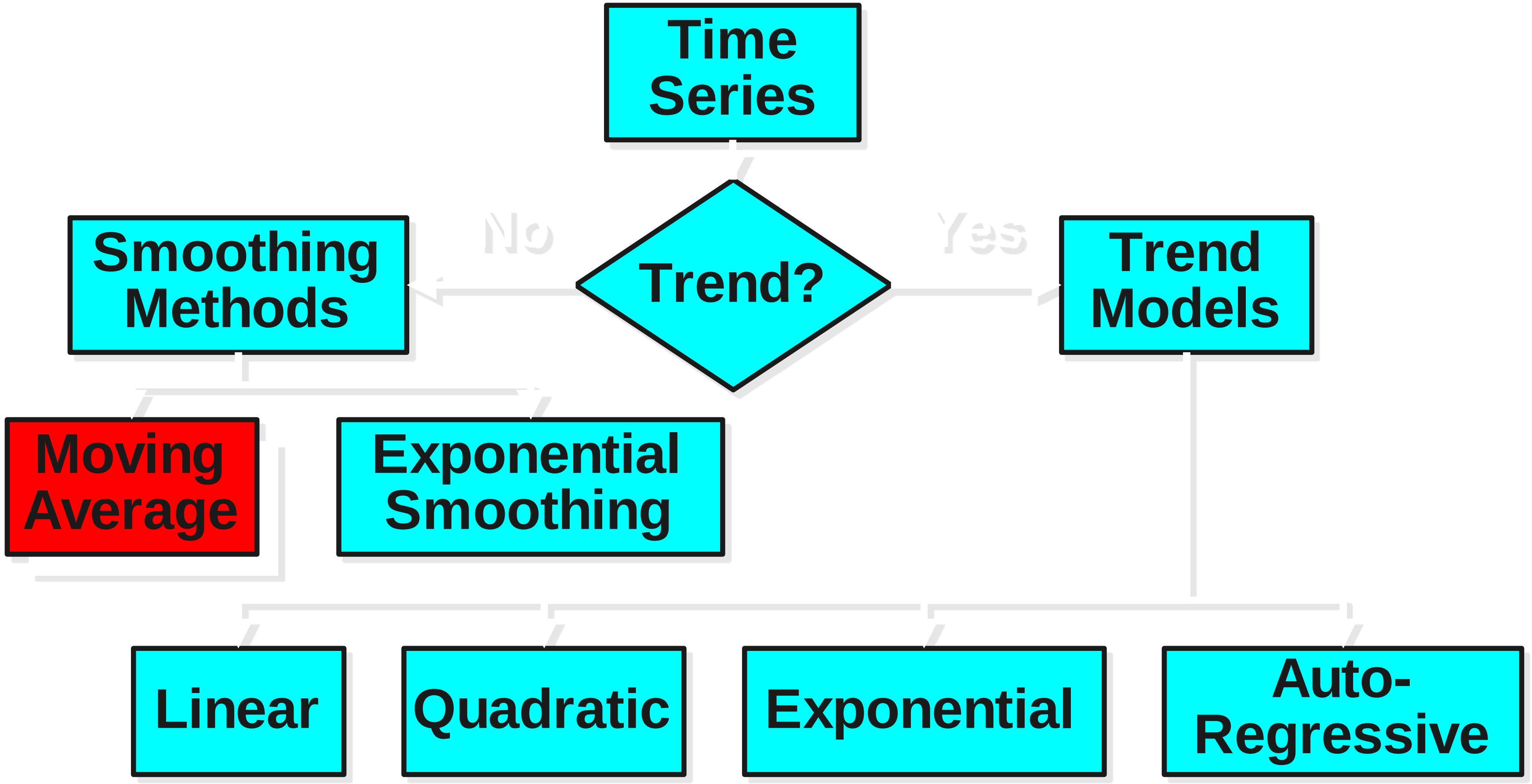
Plotting Time Series Data



Data collected by Coop Student (10/6/95)

Moving Average Method

Time Series Forecasting



Moving Average Method

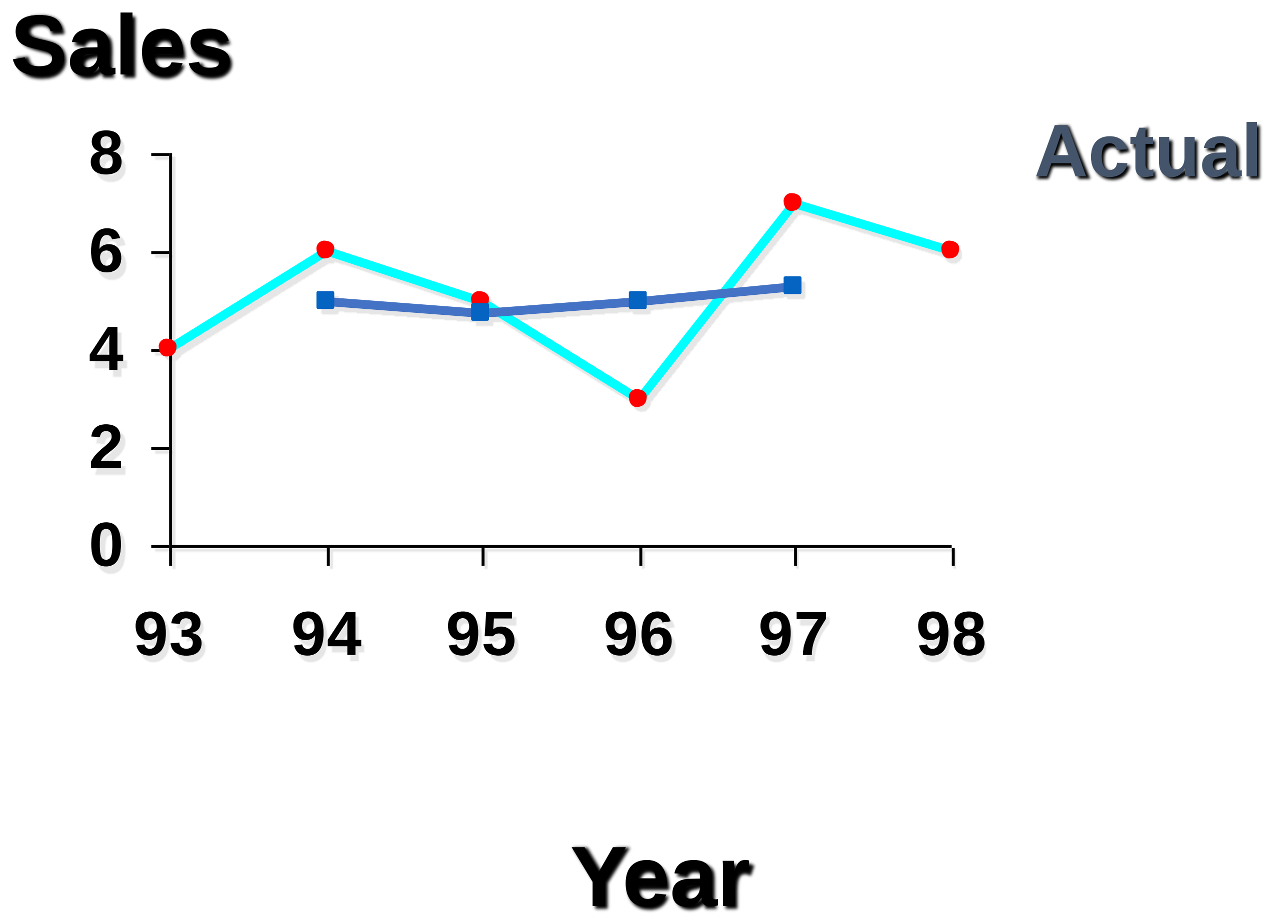
- Series of arithmetic means
- Used only for smoothing
  - Provides overall impression of data over time

Moving Average Method

- Series of arithmetic means
- Used only for smoothing
  - Provides overall impression of data over time

Used for elementary forecasting

Moving Average Graph



Moving Average
[An Example]

You work for Firestone Tire. You want to smooth random fluctuations using a 3-period moving average.

1995	20,000
1996	24,000
1997	22,000
1998	26,000
1999	25,000

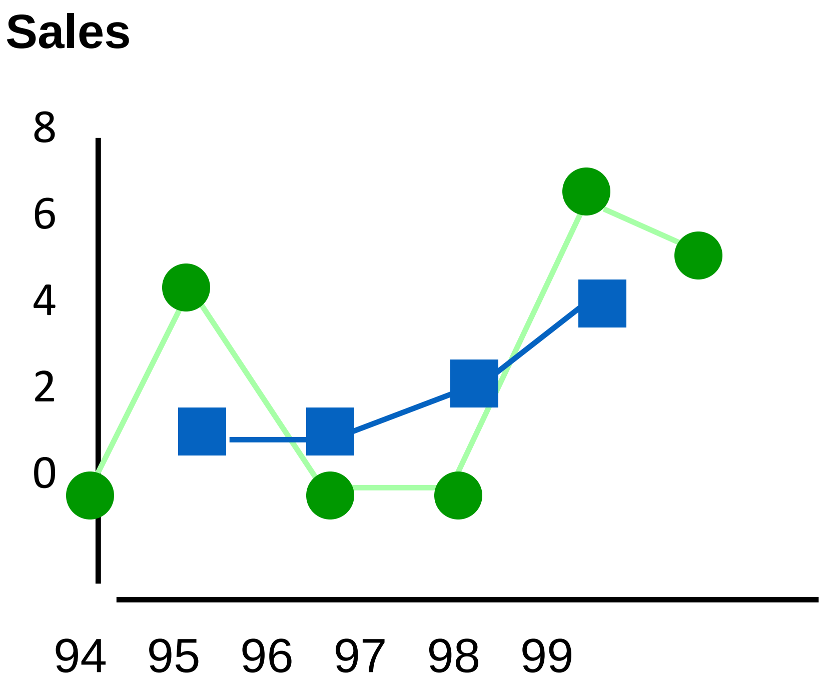
Moving Average

[Solution]

Year	Sales	MA(3) in 1,000
1995	20,000	NA
1996	24,000	(20+24+22)/3 = 22
1997	22,000	(24+22+26)/3 = 24
1998	26,000	(22+26+25)/3 = 24
1999	25,000	NA

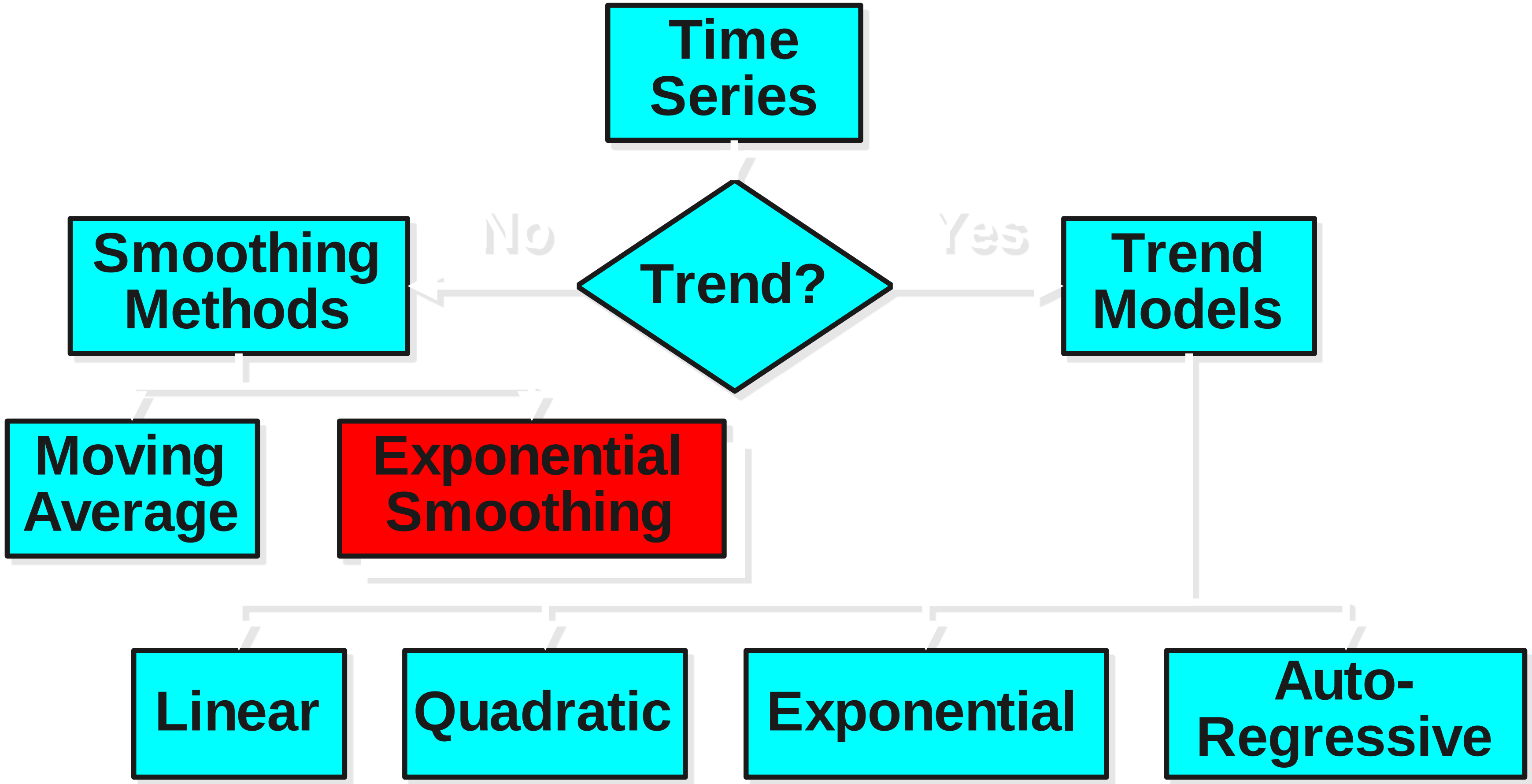
Moving Average

Year	Response ●	Moving Ave ■
1994	2	NA
1995	5	3
1996	2	3
1997	2	3.67
1998	7	5
1999	6	NA



Exponential Smoothing Method

Time Series Forecasting



Exponential Smoothing Method

- Form of weighted moving average
  - Weights decline exponentially
  - Most recent data weighted most
- Requires smoothing constant (*W*)
  - Ranges from 0 to 1
  - Subjectively chosen
- Involves little record keeping of past data

Exponential Smoothing

[An Example]

You're organizing a Kwanza meeting. You want to forecast attendance for 1998 using exponential smoothing (α = .20). Past attendance (00) is:

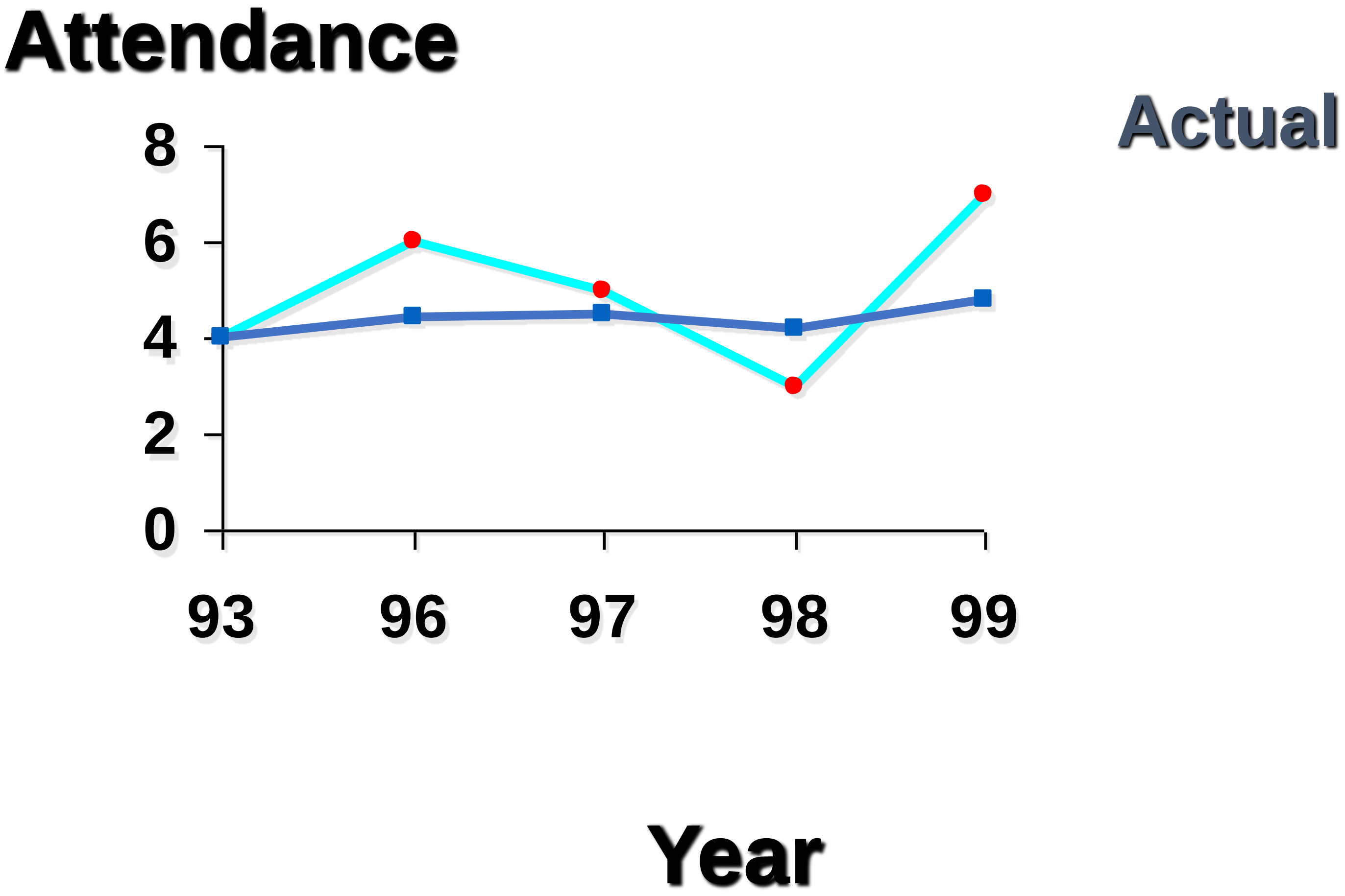
1995	4
1996	6
1997	5
1998	3
1999	7

Exponential Smoothing

$$E_i = W \cdot Y_i + (1 - W) \cdot E_{i-1}$$

Time	Y_i	Smoothed Value, E_i ($W = .2$)	Forecast $\hat{Y}_{i+1}$
1995	4	4.0	NA
1996	6	(.2)(6) + (1-.2)(4.0) = 4.4	4.0
1997	5	(.2)(5) + (1-.2)(4.4) = 4.5	4.4
1998	3	(.2)(3) + (1-.2)(4.5) = 4.2	4.5
1999	7	(.2)(7) + (1-.2)(4.2) = 4.8	4.2
2000	NA	NA	4.8

Exponential Smoothing [Graph]



Forecast Effect of Smoothing Coefficient (*W*)

$$\hat{Y}_{i+1} = W \cdot Y_i + W \cdot (1-W) \cdot Y_{i-1} + W \cdot (1-W)^2 \cdot Y_{i-2} + \ldots$$

W is...	Weight		
	Prior Period	2 Periods Ago	3 Periods Ago
	W	$W(1-W)$	$W(1-W)^2$
0.10	10%	9%	8.1%
0.90	90%	9%	0.9%

Simple Exponential Smoothing

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots$$

Simple Exponential Smoothing

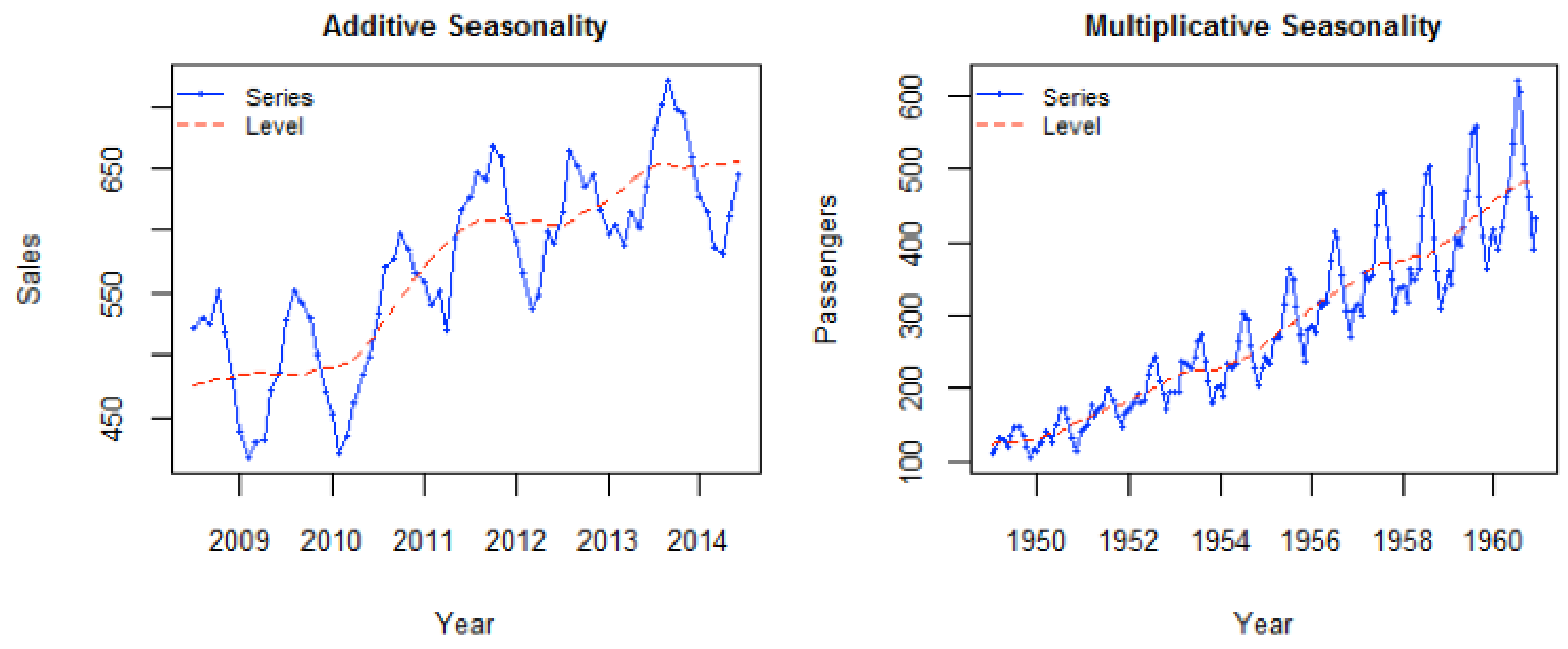
```
from statsmodels.tsa.api import ExponentialSmoothing, SimpleExpSmoothing, Holt
y_hat_avg = test.copy()
fit2 = SimpleExpSmoothing(np.asarray(train['Count'])).fit(smoothing_level=0.6,optimized=False)
y_hat_avg['SES'] = fit2.forecast(len(test))
plt.figure(figsize=(16,8))
plt.plot(train['Count'], label='Train')
plt.plot(test['Count'], label='Test')
plt.plot(y_hat_avg['SES'], label='SES')
plt.legend(loc='best')
plt.show()
```

Forecast equation : $\hat{y}_{t+h|t} = \ell_t + h\, b_t$

Level equation : $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1}+b_{t-1})$

Trend equation : $b_t = \beta*(\ell_t - \ell_{t-1})+(1-\beta)b_{t-1}$

Holt’s Linear Trend method


Additive Seasonality
Series
Level
Sales
450
550
650
2009
2010
2011
2012
2013
2014
Year
Multiplicative Seasonality
Series
Level
Passengers
100
200
300
400
500
600
1950
1952
1954
1956
1958
1960
Year

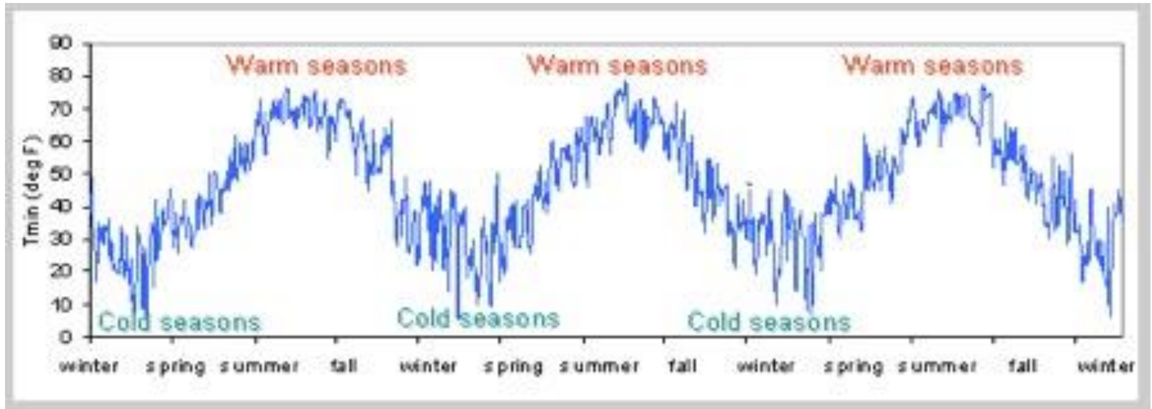
Holt’s Linear Trend method

```
y_hat_avg = test.copy()

fit1 = Holt(np.asarray(train['Count'])).fit(smoothing_level = 0.3,smoothing_slope = 0.1)
y_hat_avg['Holt_linear'] = fit1.forecast(len(test))

plt.figure(figsize=(16,8))
plt.plot(train['Count'], label='Train')
plt.plot(test['Count'], label='Test')
plt.plot(y_hat_avg['Holt_linear'], label='Holt_linear')
plt.legend(loc='best')
plt.show()

rms = sqrt(mean_squared_error(test.Count, y_hat_avg.Holt_linear))
print(rms)
```



$$
\begin{aligned}
\text{level} \quad L_t &= \alpha(y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1}); \\
\text{trend} \quad b_t &= \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}, \\
\text{seasonal} \quad S_t &= \gamma(y_t - L_t) + (1-\gamma)S_{t-s} \\
\text{forecast} \; F_{t+k} &= L_t + kb_t + S_{t+k-s},
\end{aligned}
$$

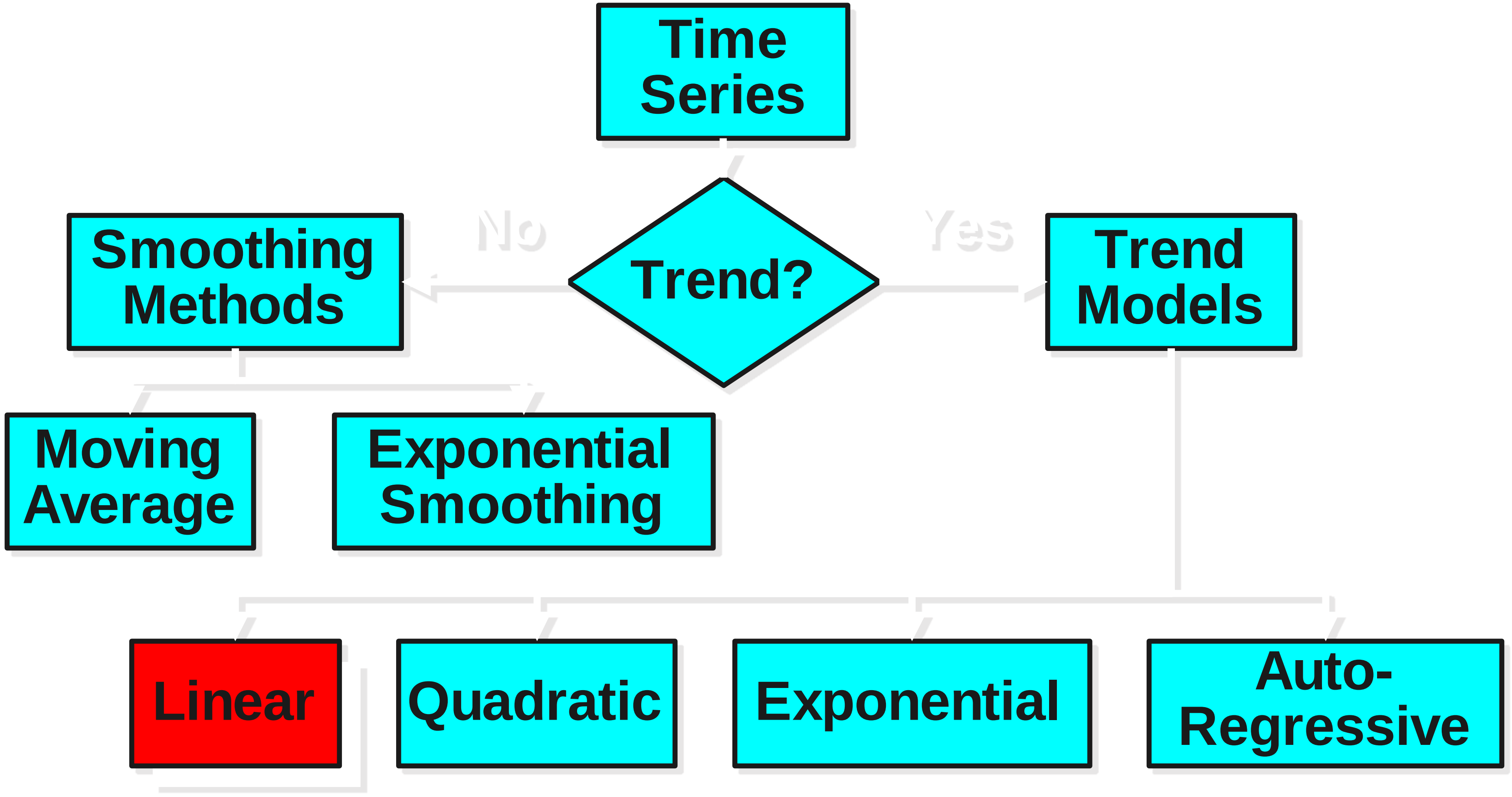
Holt-Winters Method

Holt-Winters Method

```
y_hat_avg = test.copy()
fit1 = ExponentialSmoothing(np.asarray(train['Count']) ,seasonal_periods=7
,trend='add', seasonal='add',).fit()
y_hat_avg['Holt_Winter'] = fit1.forecast(len(test))
plt.figure(figsize=(16,8))
plt.plot( train['Count'], label='Train')
plt.plot(test['Count'], label='Test')
plt.plot(y_hat_avg['Holt_Winter'], label='Holt_Winter')
plt.legend(loc='best')
plt.show()
```

Linear Time-Series Forecasting Model

Time Series Forecasting



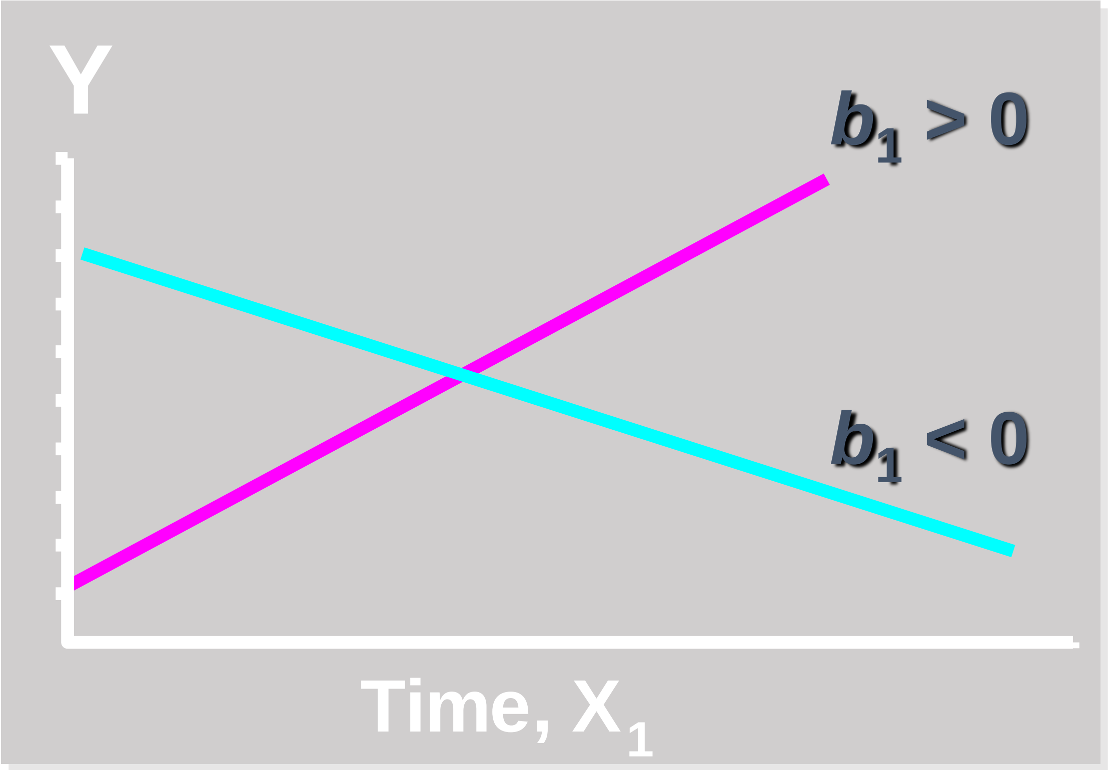
Linear Time-Series Forecasting Model

- Used for forecasting trend
- Relationship between response variable *Y* & time *X* is a linear function
- Coded *X* values used often

Year *X*:	1995	1996	1997	1998	1999
Coded year:	0	1	2	3	4
Sales *Y*:	78.7	63.5	89.7	93.2	92.1

Linear Time-Series Model

$$\hat{Y}_i = b_0 + b_1 X_{1i}$$



Linear Time-Series Model [An Example]

You're a marketing analyst for Hasbro Toys. Using coded years, you find $\hat{Y}_i = .6 + .7X_i$.

1995	1
1996	1
1997	2
1998	2
1999	4

Forecast 2000 sales.

Linear Time-Series [Example]

Year	Coded Year	Sales (Units)
1995	0	1
1996	1	1
1997	2	2
1998	3	2
1999	4	4
2000	5	?

2000 forecast sales: $\hat{Y}_i = .6 + .7\cdot(5) = 4.1$

The equation would be different if 'Year' used.

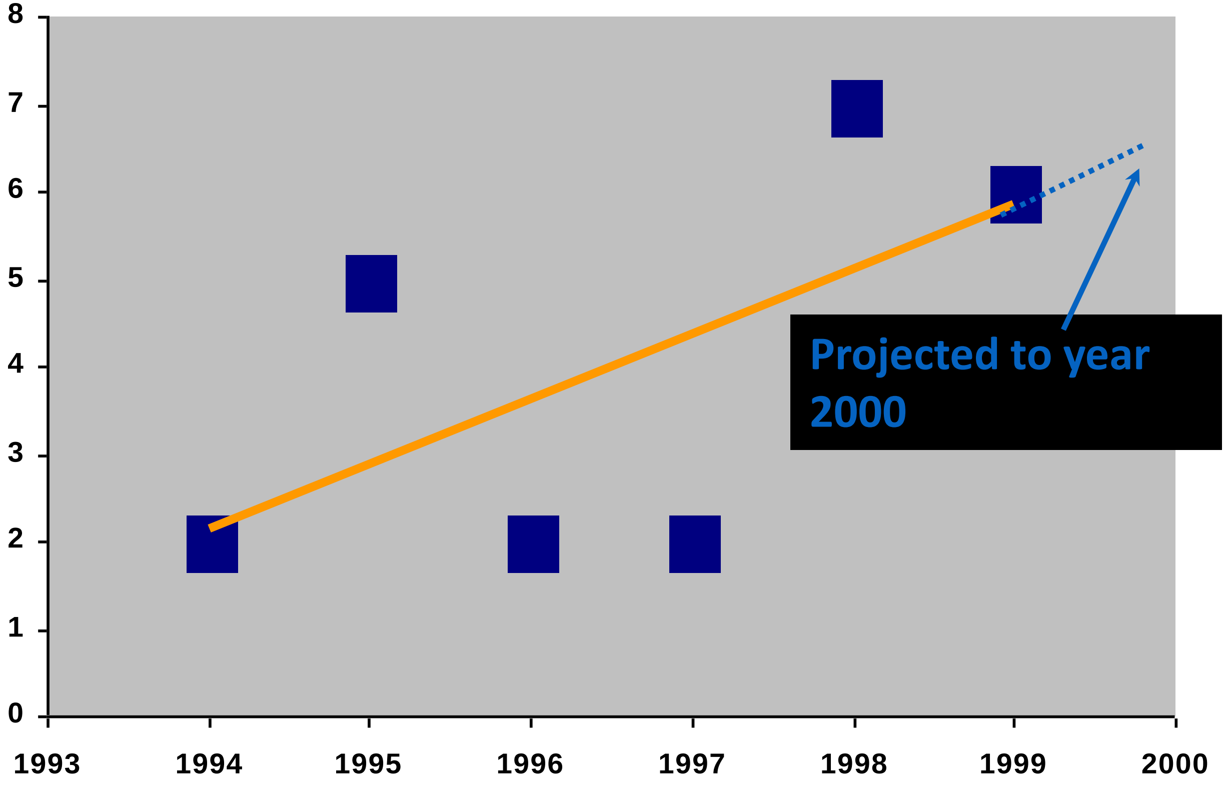
The Linear Trend Model

Year	Coded	Sales
94	0	2
95	1	5
96	2	2
97	3	2
98	4	7
99	5	6

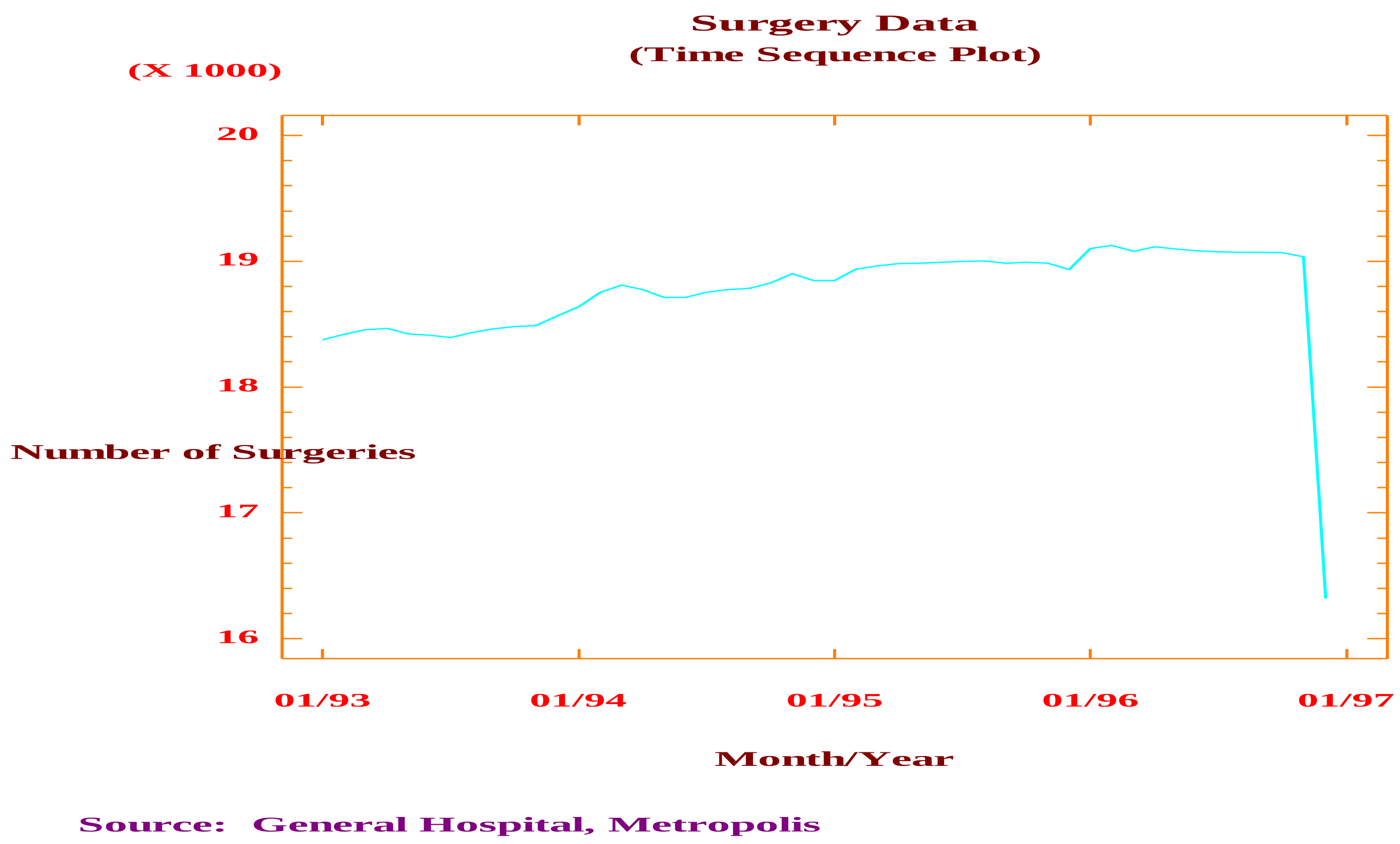
$$\hat{Y}_i = b_0 + b_1 X_i = 2.143 + .743 X_i$$

Excel Output

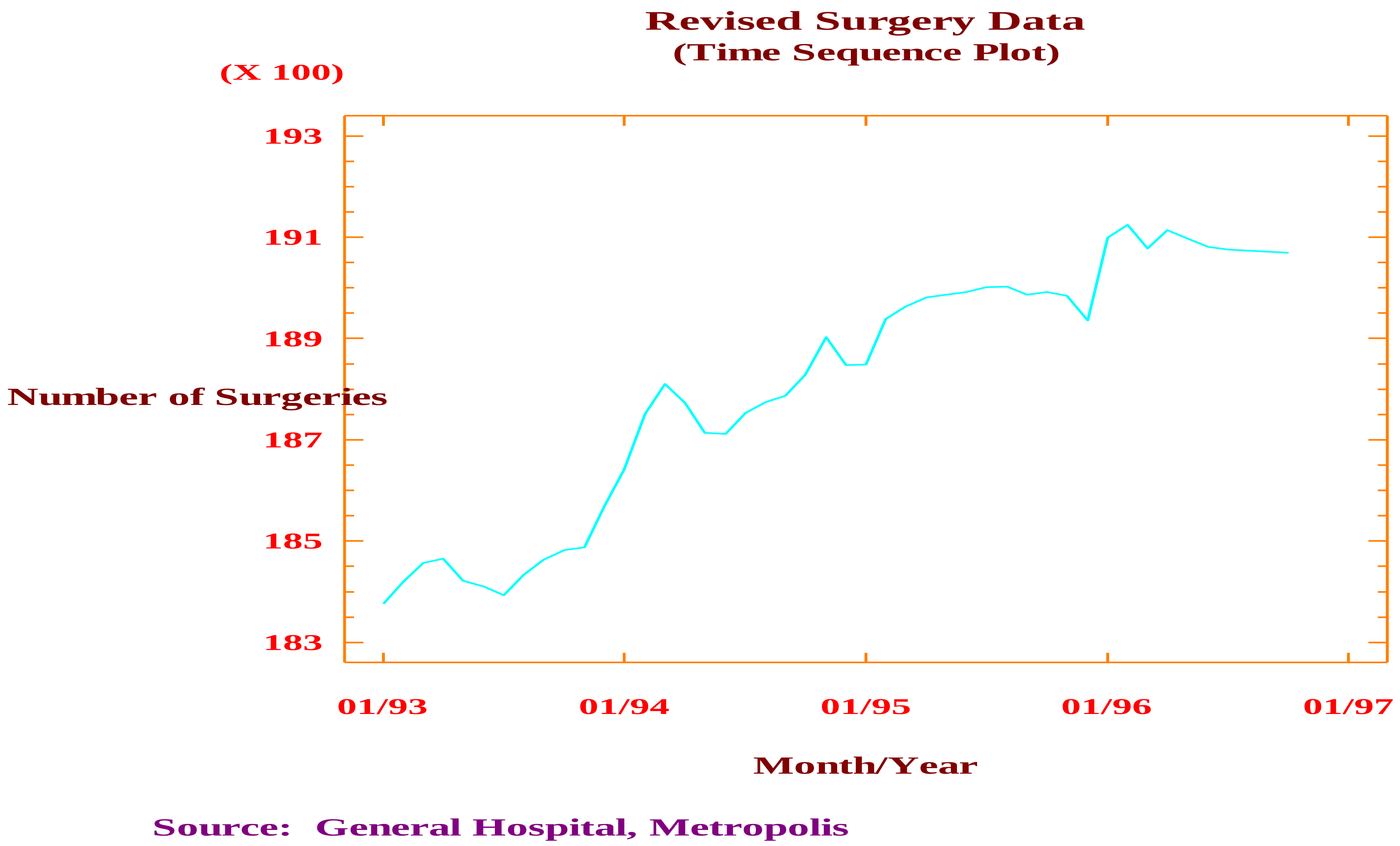
	Coefficients
Intercept	2.14285714
X Variable	0.74285714



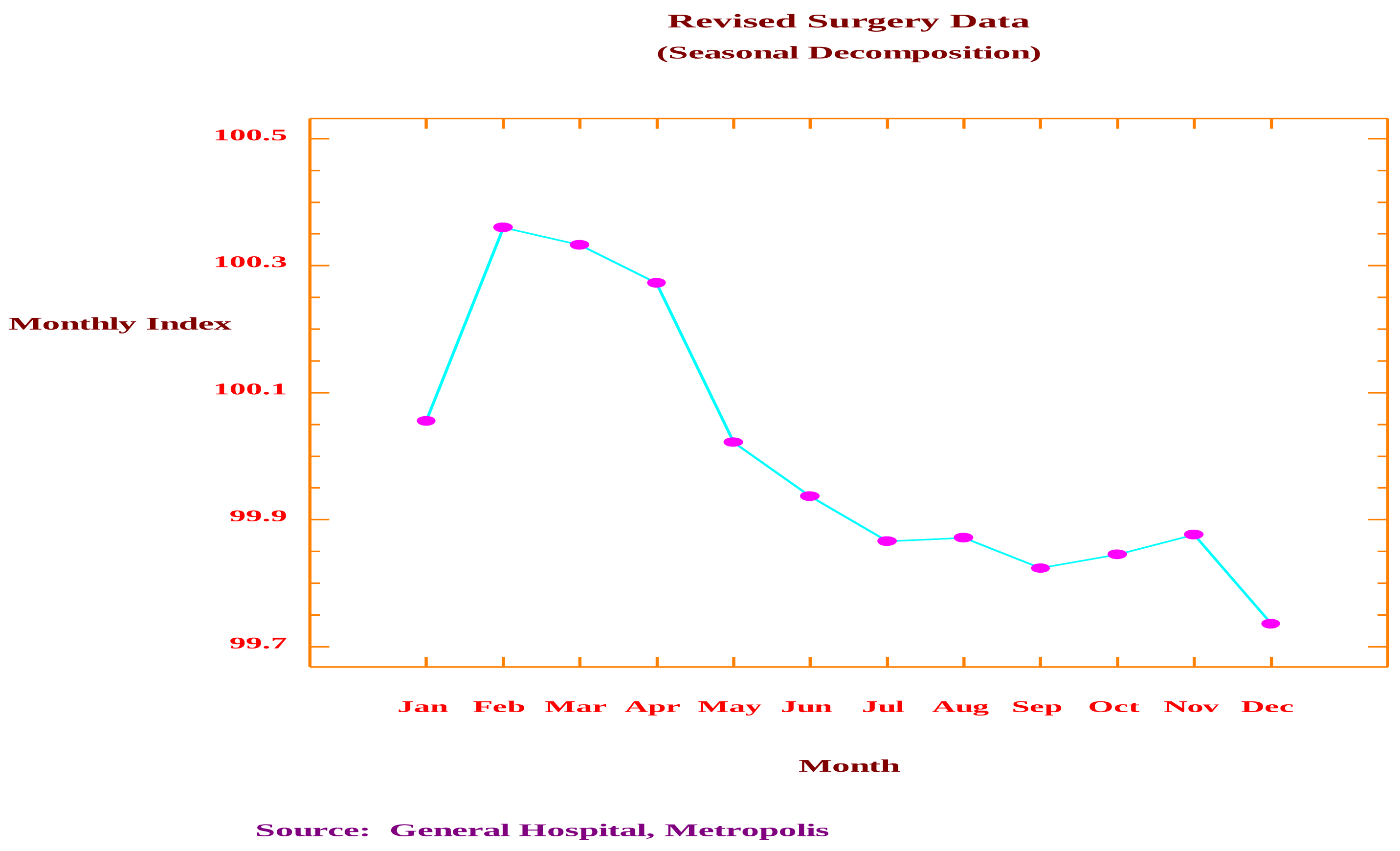
Time Series Plot



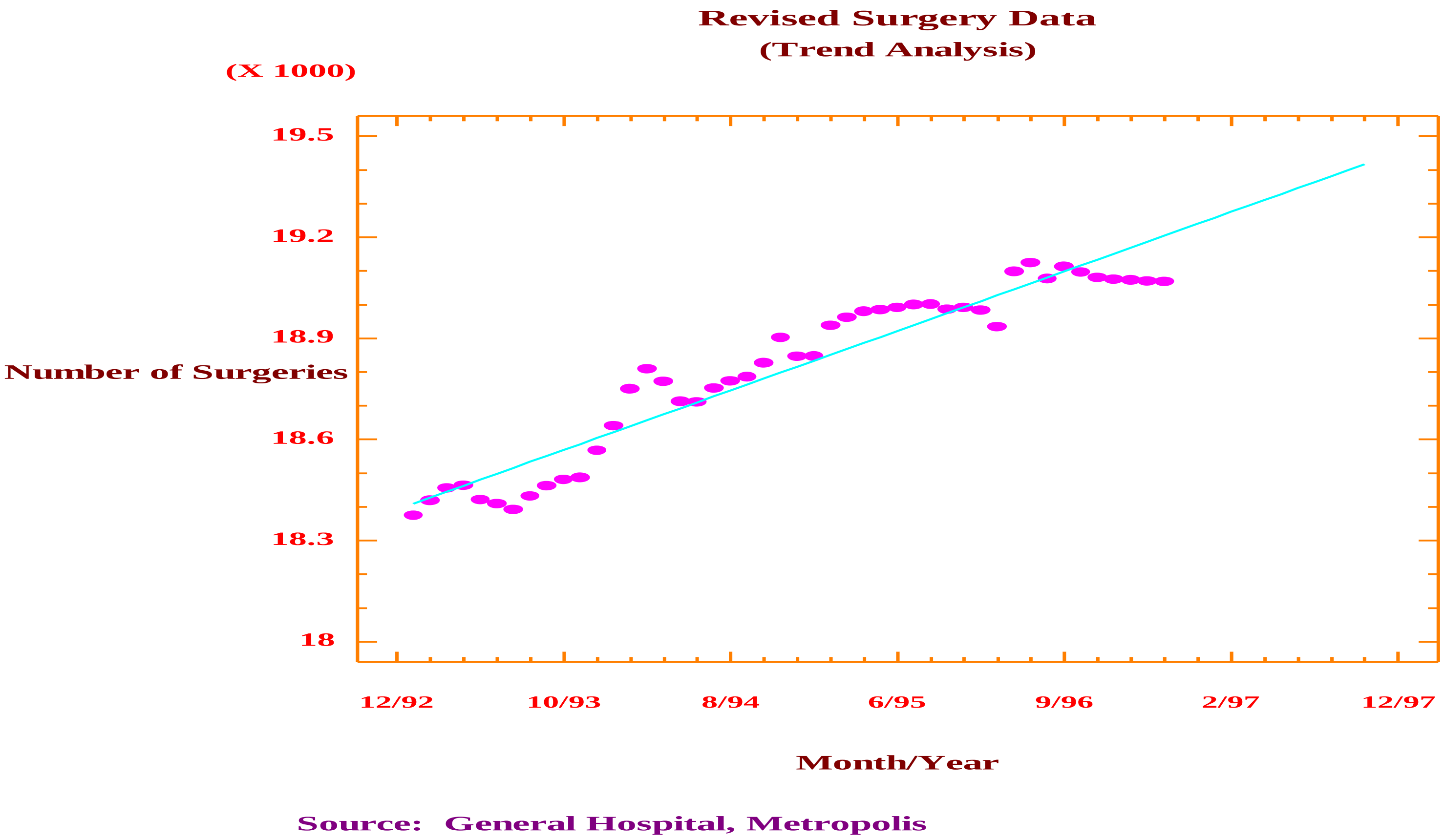
Time Series Plot [Revised]



Seasonality Plot

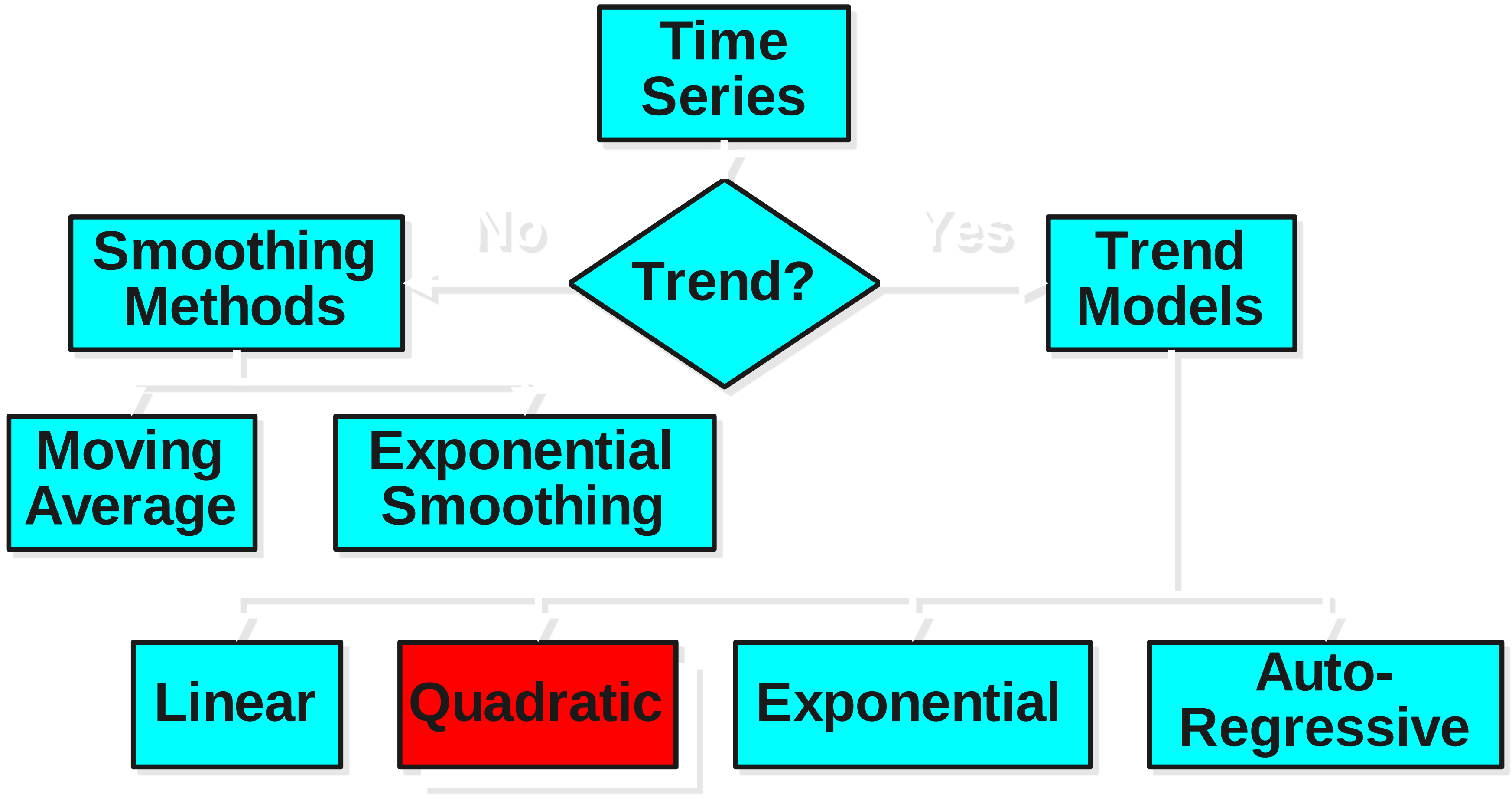


Trend Analysis



Quadratic Time-Series Forecasting Model

Time Series Forecasting



Quadratic Time-Series Forecasting Model

- Used for forecasting trend
- Relationship between response variable *Y* & time *X* is a quadratic function
- Coded years used

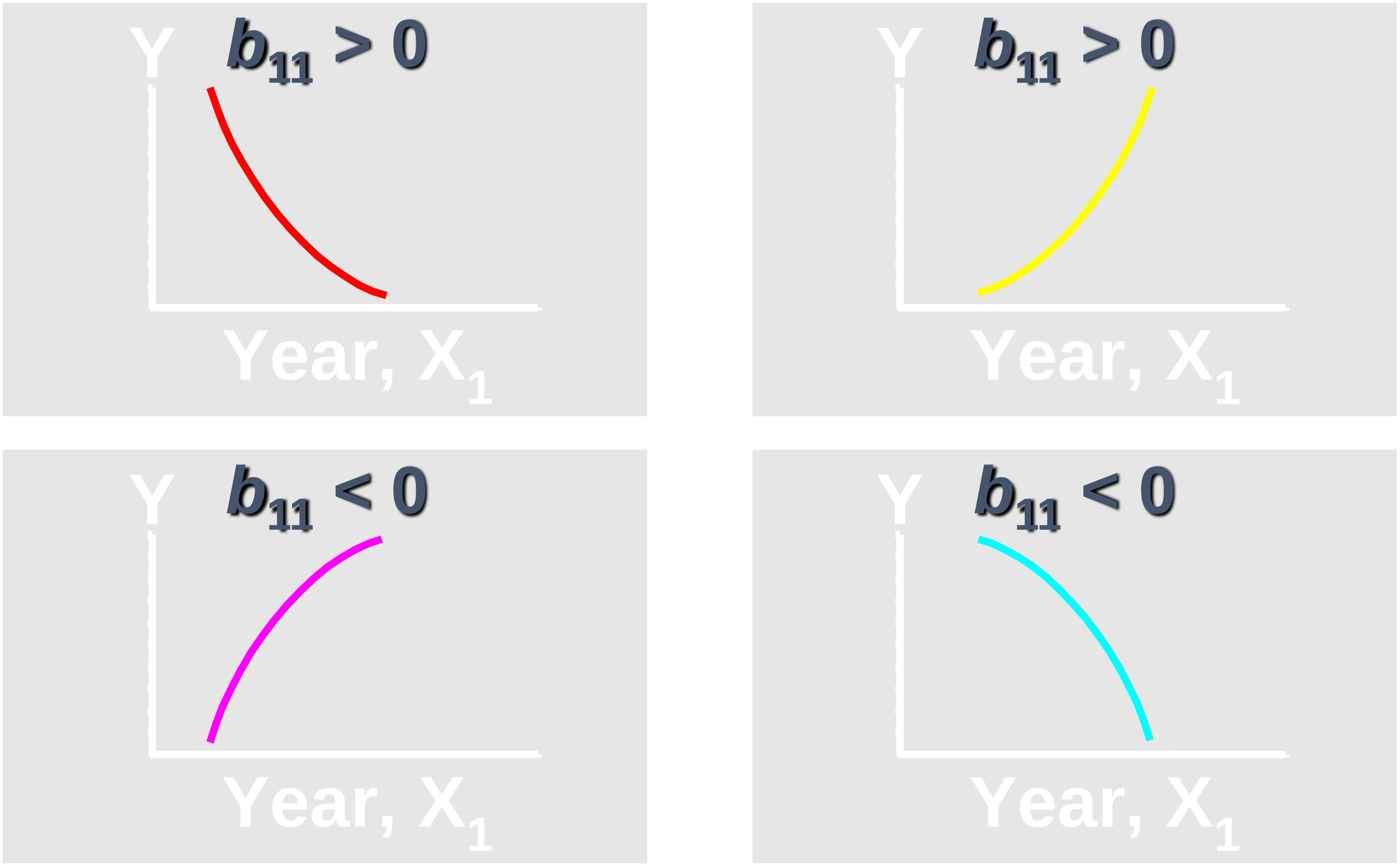
Quadratic Time-Series Forecasting Model

- Used for forecasting trend
- Relationship between response variable *Y* & time *X* is a quadratic function
- Coded years used
- Quadratic model

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_{11} X_{1i}^2$$

Quadratic Time-Series Model

Relationships



Quadratic Trend Model

Year	Coded	Sales
94	0	2
95	1	5
96	2	2
97	3	2
98	4	7
99	5	6

$$\hat{Y}_i = b_0 + b_1 X_i + b_2 X_i^2$$

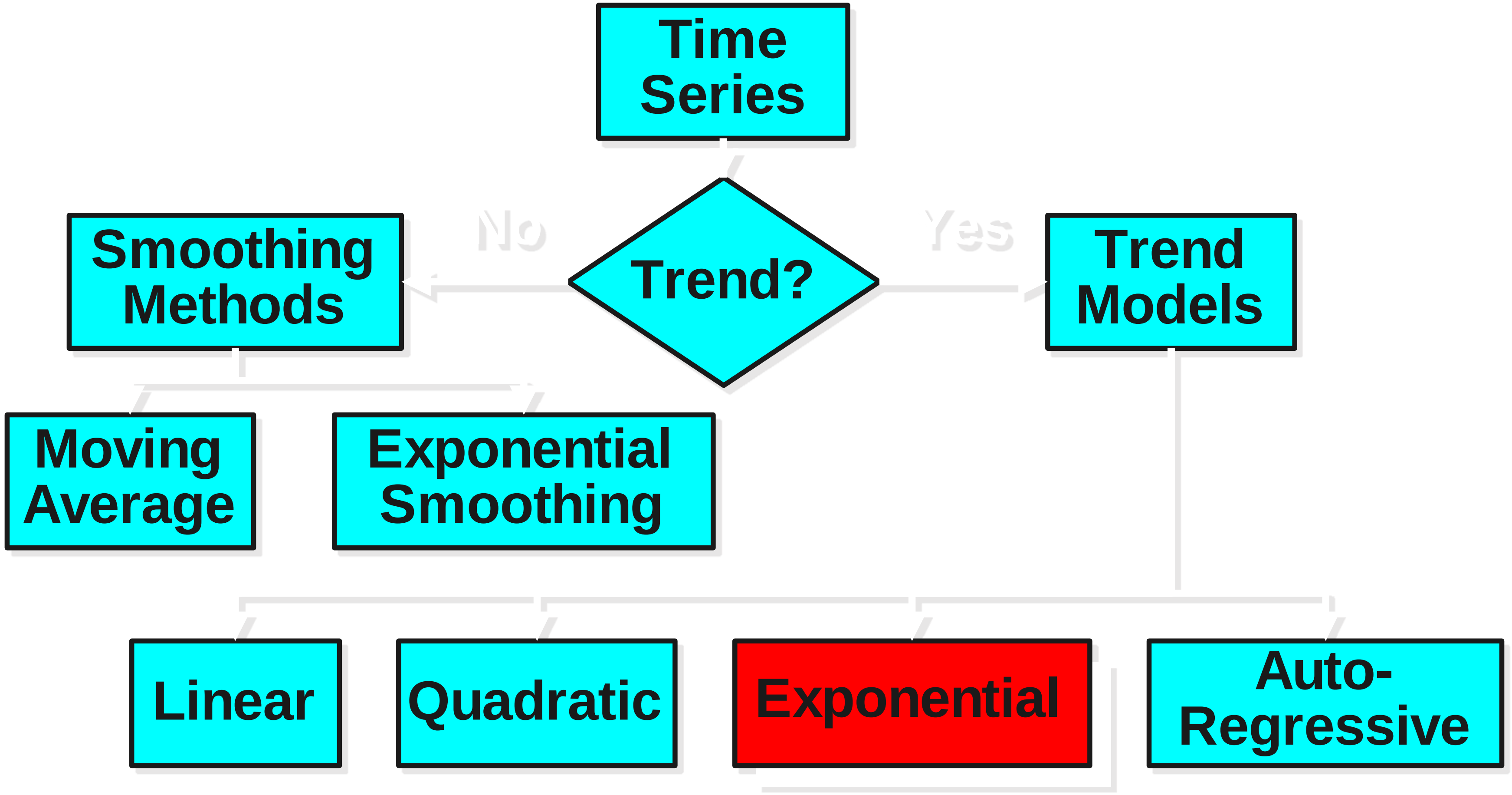
	Coefficients
Intercept	2.85714286
X Variable 1	-0.3285714
X Variable 2	0.21428571

Excel Output

$$\hat{Y}_i = 2.857 - 0.33 X_i + .214 X_i^2$$

Exponential Time-Series Model

Time Series Forecasting



Exponential Time-Series Forecasting Model

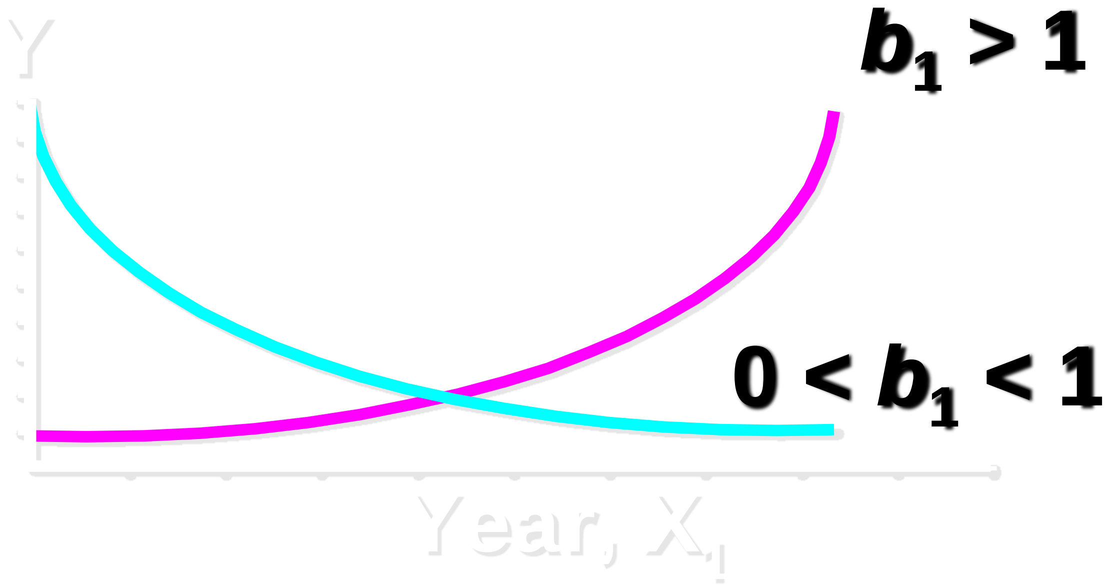
- Used for forecasting trend
- Relationship is an exponential function
- Series increases (decreases) at increasing (decreasing) rate

Exponential Time-Series Forecasting Model

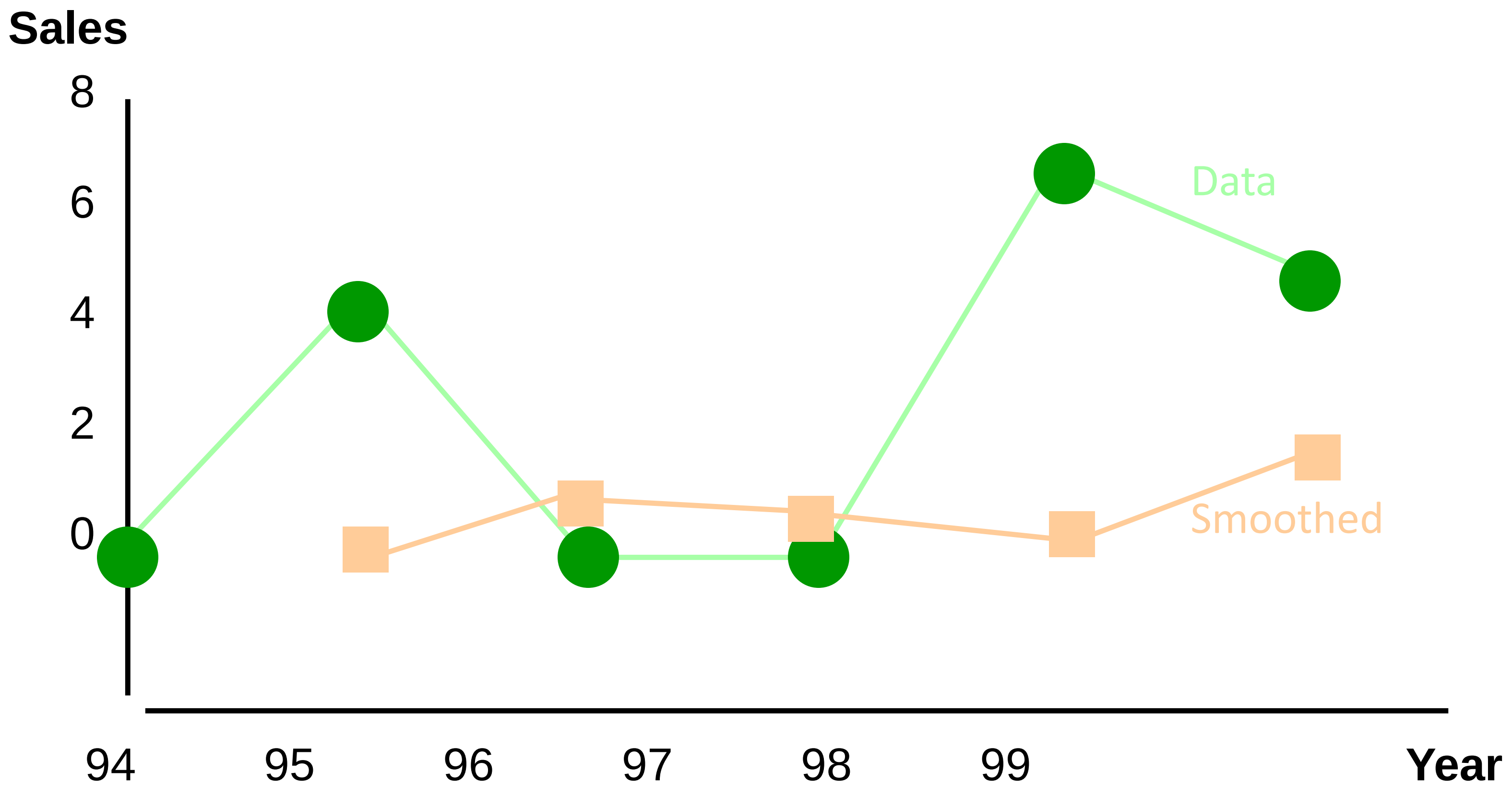
- Used for forecasting trend
- Relationship is an exponential function
- Series increases (decreases) at increasing (decreasing) rate

Exponential Time-Series Model

Relationships



Exponential Weight [Example Graph]



Exponential Trend Model

$$\hat{Y}_i = b_0 b_1^{X_i} \quad \text{or} \quad \log \hat{Y}_i = \log b_0 + X_1 \log b_1$$

Year	Coded	Sales
94	0	2
95	1	5
96	2	2
97	3	2
98	4	7
99	5	6

	Coefficients
Intercept	0.33583795
X Variable	0.08068544

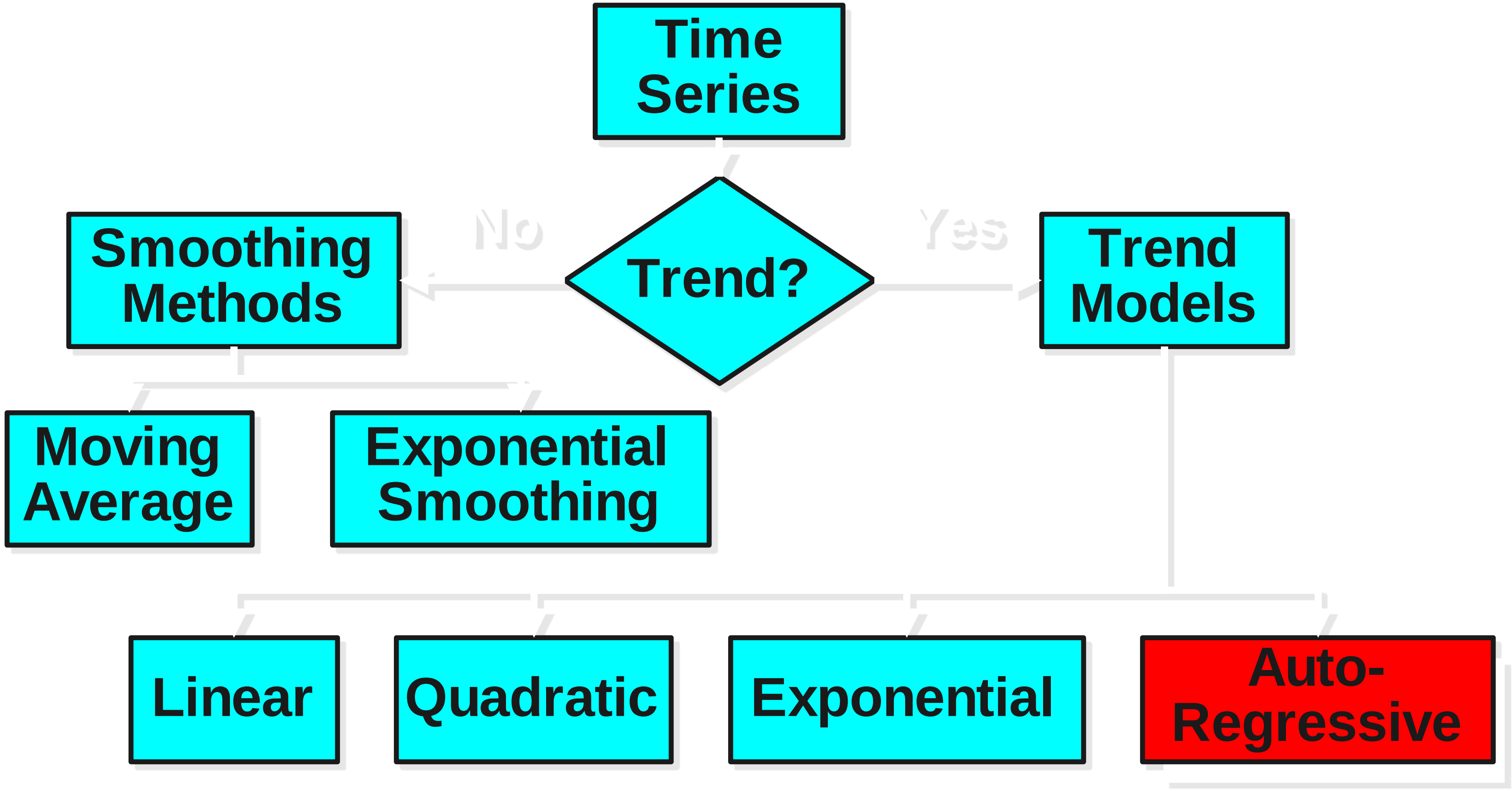
Excel Output of Values in logs

antilog(.33583795) =	2.17
antilog(.08068544) =	1.2

$$\hat{Y}_i = (2.17)(1.2)^{X_i}$$

Autoregressive Modeling

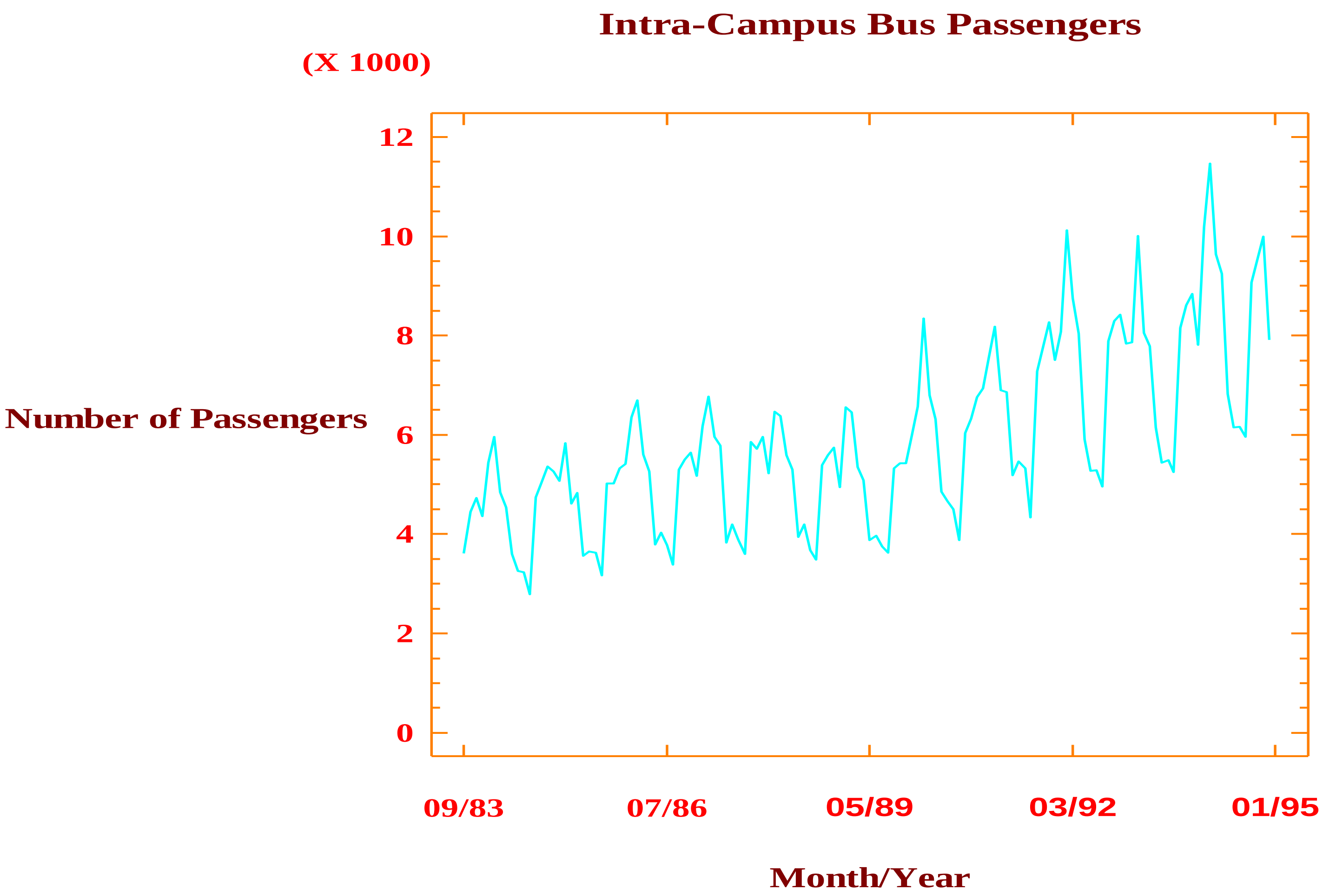
Time Series Forecasting



Autoregressive Modeling

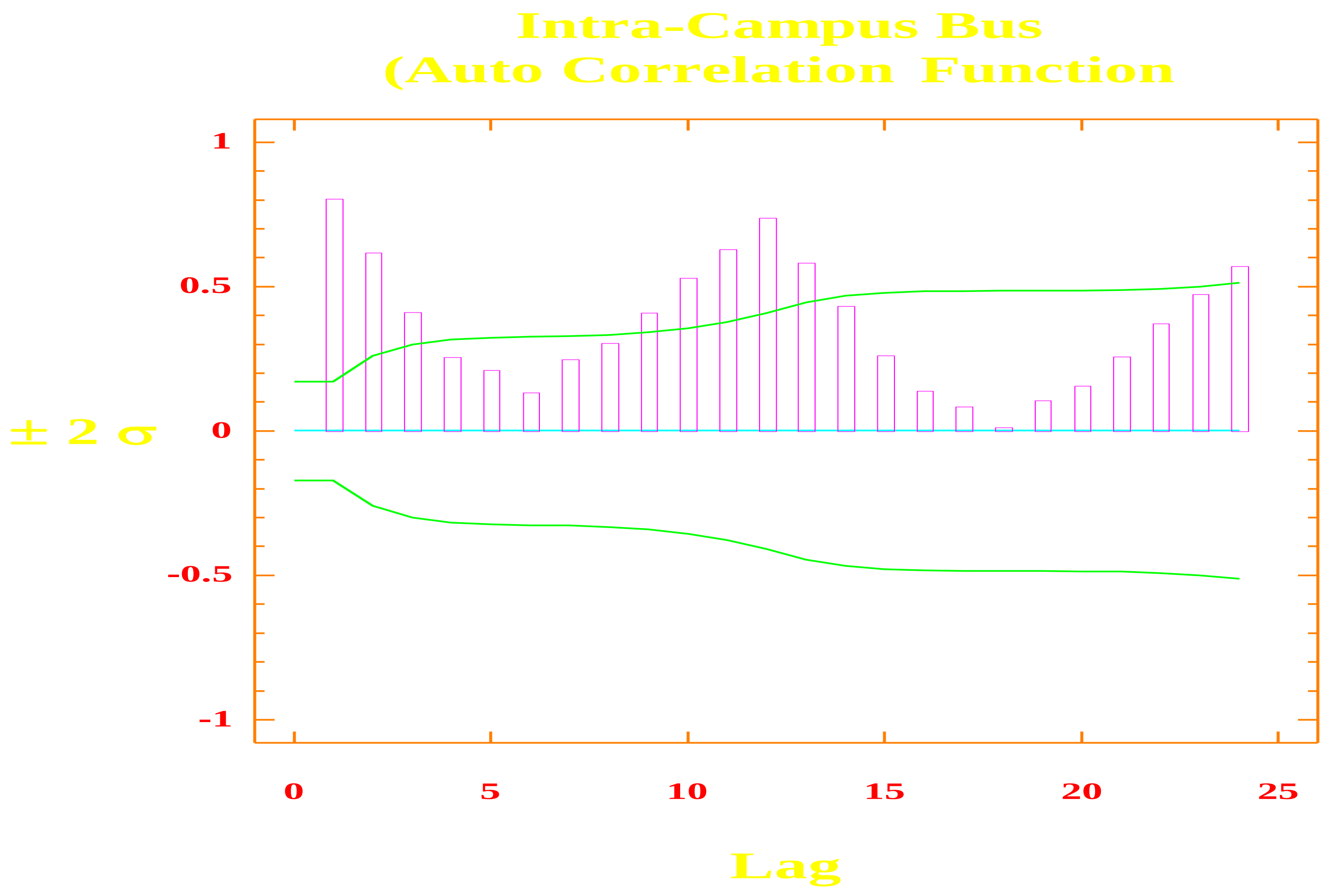
- Used for forecasting trend
- Like regression model
  - Independent variables are lagged response variables Y_{i-1}, Y_{i-2}, Y_{i-3} etc.
- Assumes data are correlated with past data values
  - 1st Order: Correlated with prior period
- Estimate with ordinary least squares

Time Series Data Plot



Data collected by Coop Student (10/6/95)

Auto-correlation Plot



Autoregressive Model [An Example]

The Office Concept Corp. has acquired a number of office units (in thousands of square feet) over the last 8 years. Develop the 2nd order Autoregressive models.

Year	Units
92	4
93	3
94	2
95	3
96	2
97	2
98	4
99	6

Autoregressive Model [Example Solution]

- Develop the 2nd order table
- Use Excel to run a regression model

Year	Y_i	Y_{i-1}	Y_{i-2}
92	4	---	---
93	3	4	---
94	2	3	4
95	3	2	3
96	2	3	2
97	2	2	3
98	4	2	2
99	6	4	2

Excel Output

	Coefficients
Intercept	3.5
X Variable 1	0.8125
X Variable 2	-0.9375

$$Y_i = 3.5 + .8125Y_{i-1} - .9375Y_{i-2}$$

Evaluating Forecasts

Quantitative Forecasting Steps

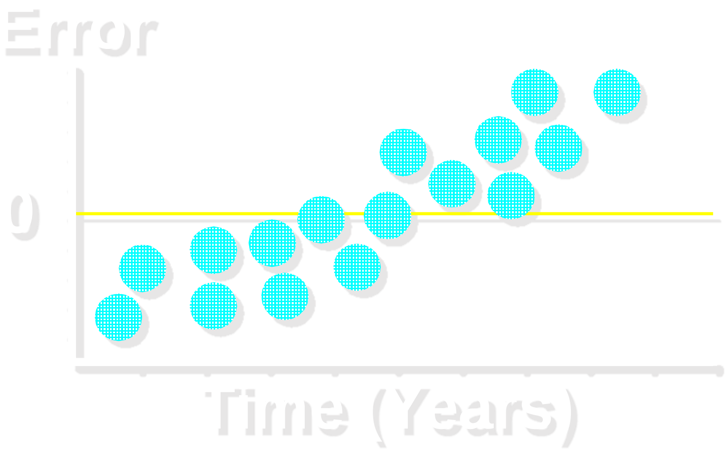
- Select several forecasting methods
- ‘Forecast’ the past
- Evaluate forecasts
- Select best method
- Forecast the future
- Monitor continuously forecast accuracy

Forecasting Guidelines

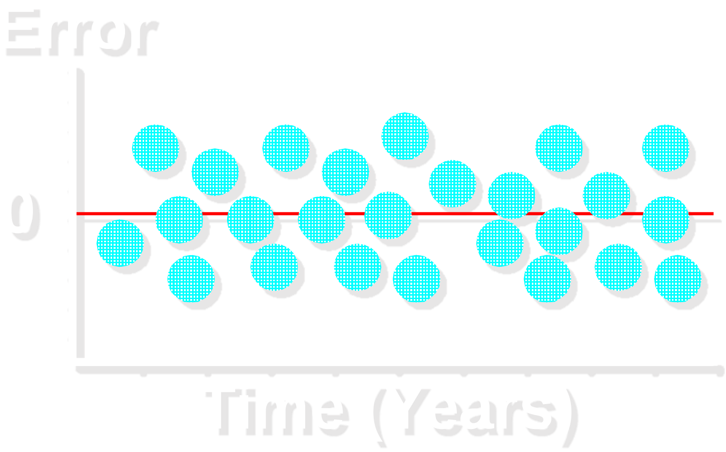
- No pattern or direction in forecast error
  - e_i = (Actual Y_i - Forecast Y_i)
  - Seen in plots of errors over time
- Smallest forecast error
  - Measured by mean absolute deviation
- Simplest model
  - Called principle of parsimony

Pattern of Forecast Error

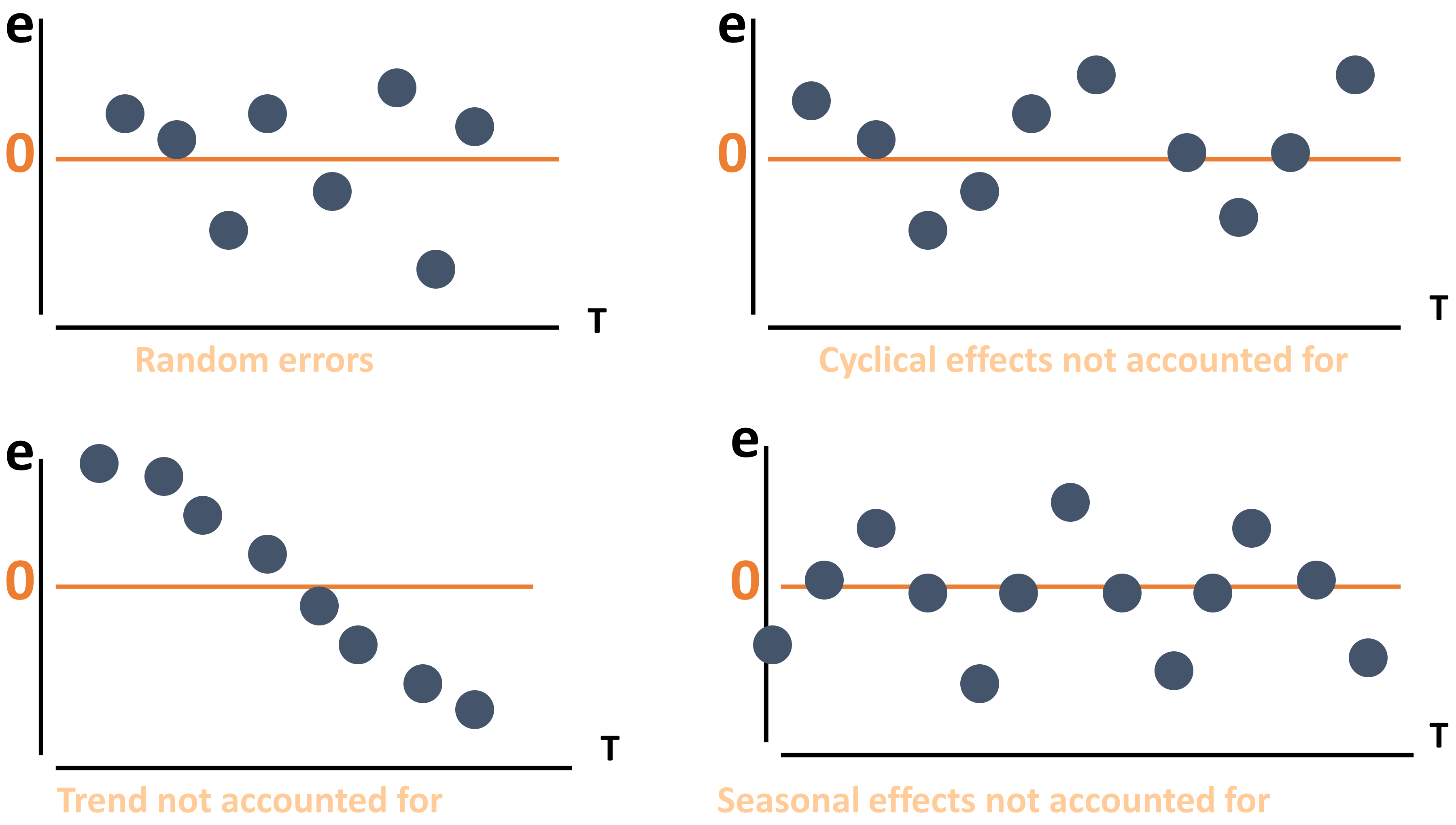
Trend Not Fully Accounted for



Desired Pattern



Residual Analysis



Principal of Parsimony

- Suppose two or more models provide good fit for data
- Select the Simplest Model
  - Simplest model types:
    - least-squares linear
    - least-square quadratic
    - 1st order autoregressive
  - More complex types:
    - 2nd and 3rd order autoregressive
    - least-squares exponential

Summary

- Described what forecasting is
- Explained time series & its components
- Smoothed a data series
  - Moving average
  - Exponential smoothing
- Forecasted using trend models

Questions?