STAT 115: Introductory Methods for Time Series Analysis and Forecasting

Concepts and Techniques

School of Statistics
University of the Philippines Diliman

FORECASTING

- ***Forecasting*** is an activity that calculates/predicts some future events or conditions, usually as a result of rational study or analysis of pertinent data.

Qualitative and Quantitative Forecasting

- Qualitative is an intuitive and educated guess.

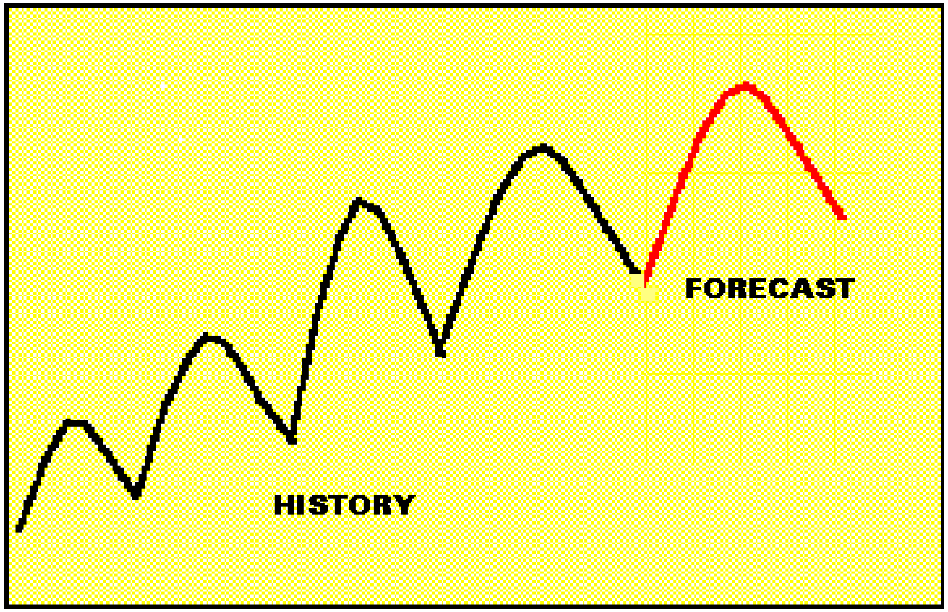
- Quantitative is based on some mathematical (deterministic) or statistical model.

Statistical Forecasting Techniques

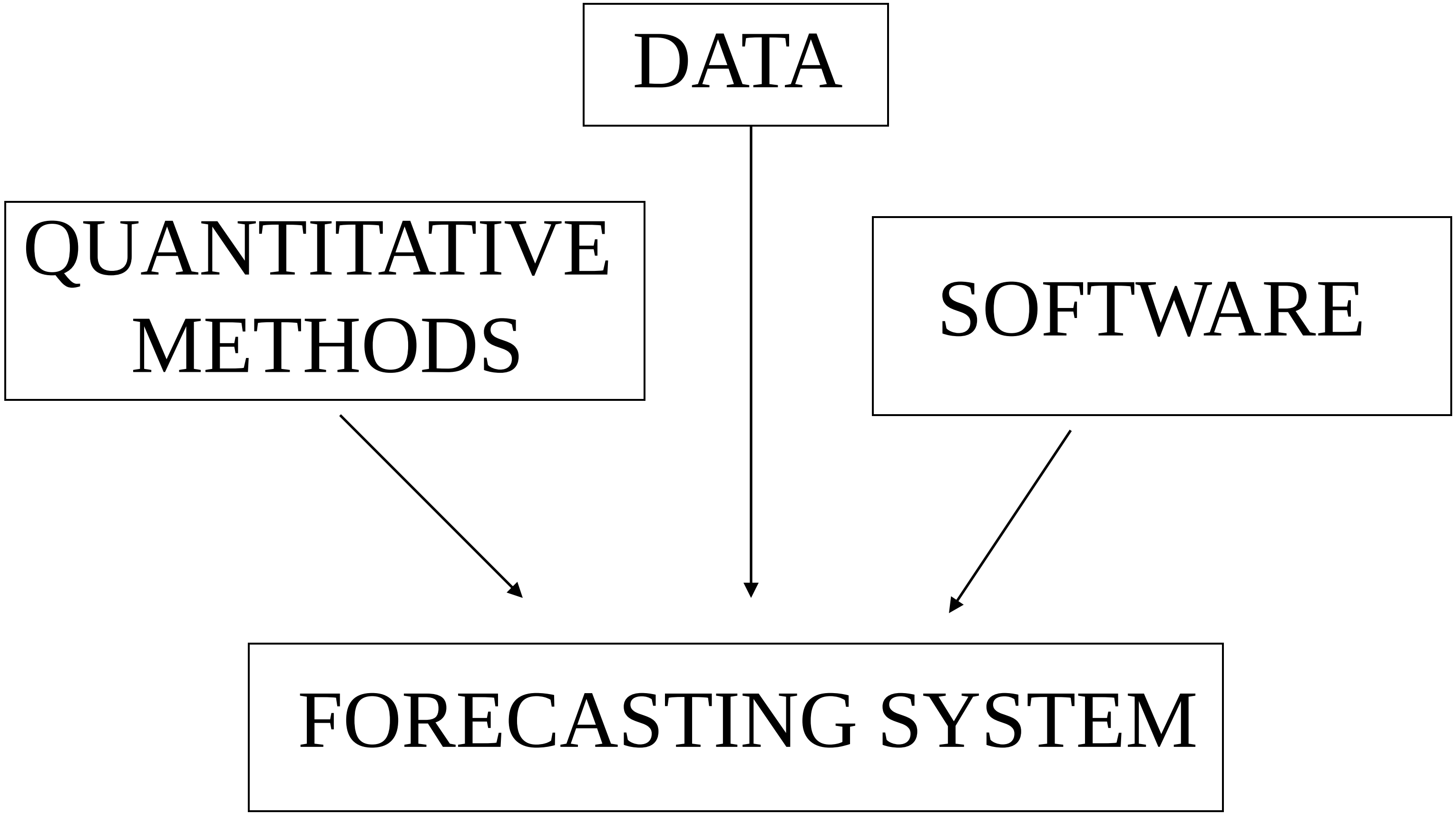
- A **Statistical Forecasting Technique** is one that generates forecasts by extrapolating patterns in historical data.
- **Forecasting** is linked to the building of **statistical models** in the sense that before one can forecast a variable of interest, one must build a model for it and estimate the model's parameter(s) using historical data.

Statistical Forecasting Techniques

- The estimated model summarizes the dynamic patterns in the data. It **provides a statistical characterization of the links between the past and future values.**

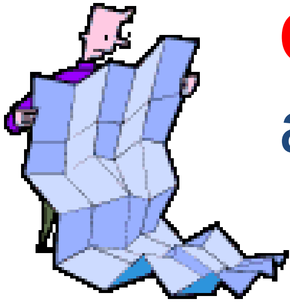


Forecasting is an interplay of the data, statistical techniques and the software



Types of Data

- Cross Section
- Time Series



Cross-section data refers to data taken at one point in time across a group.

Per Capita GRDP, 2001
In Current Prices, Pesos

Region	Per Capita GRDP
MM	127,820
CAR	54,503
Ilocos	23,669
C. Valley	24,982
C. Luzon	35,021
S. Tagalog	45,531
Bicol	19,853
W. Visayas	33,652
C. Visayas	42,836
E. Visayas	20,754
W. Mindanao	24,168
N. Mindanao	45,221
S. Mindanao	36,929
C. Mindanao	32,598
CARAGA	21,326
ARMM	14,156
PHILS	**45,453**

A **time series** refers to data gathered sequentially in time.

Year	Demand (million pesos)
1998	2.25
1999	2.85
2000	2.95
2001	3.65
2002	4.20

Anatomy of a Model

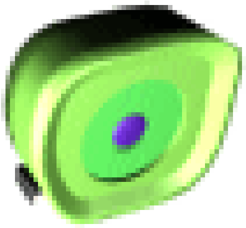
A fundamental assumption of most econometric methods is that actual value consists of a **pattern** and an **error.**

Actual Value = Pattern + Error

The pattern and error correspond to a **model**. A model is a set of assumptions that summarizes the system governing a time series.

When the pattern is clear, this behavior may be used to explain the behavior of the series & predict future values.

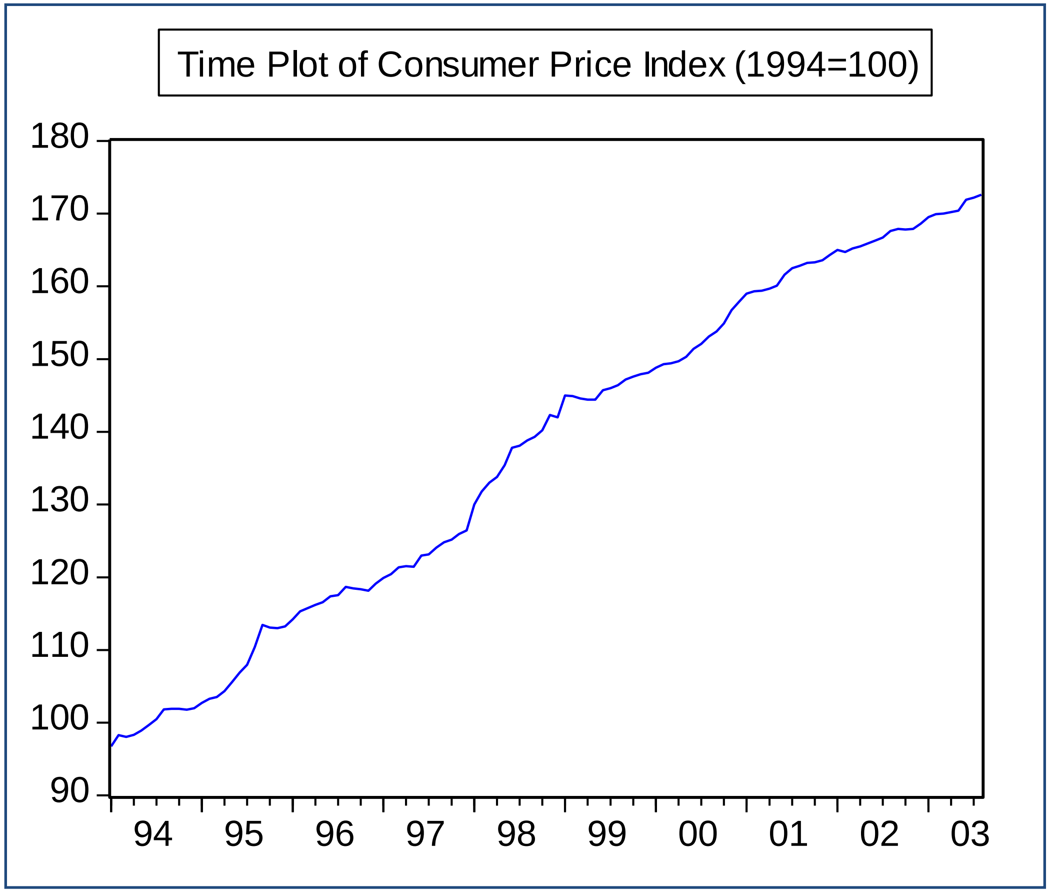
Note:



- Elements entering the model should be assessed using summary statistics.
- Aside from the usual measures of averages, standard deviations, and **autocorrelations** are important to justify their “predictive” ability in the model.

The Time Plot

A pattern in a series may be visually assessed through a **time plot**. A time plot is a line chart of a data series against time.



A time series is usually defined by a model.

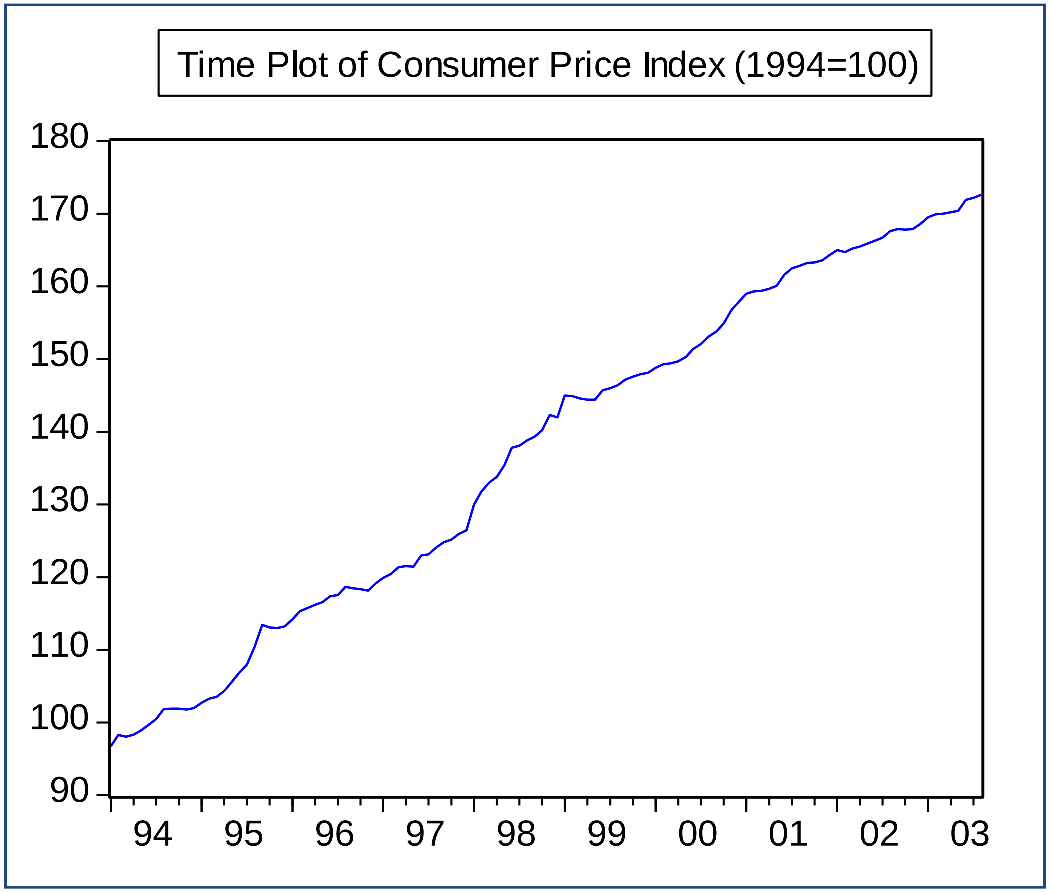
Value = function of (Trend, Cycle, Season, Random Shock)

For example, the gross domestic product in the first quarter of 2003 is expressed as

$$GDP_{2003.Q1} = f\,(\text{Trend Component}_{2003.Q1},\ \text{Cyclical Component}_{2003.Q1},\ \text{First Quarter Seasonal Index},\ \text{Random disturbance})$$

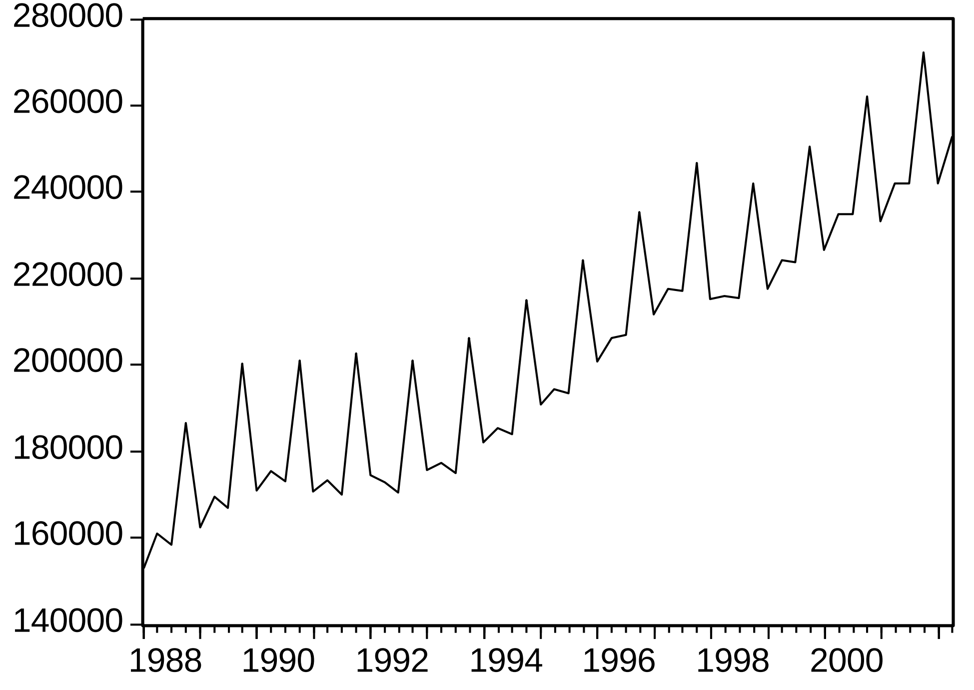
Components of a Series

The **trend** is the general ***long-run*** movement of the series.

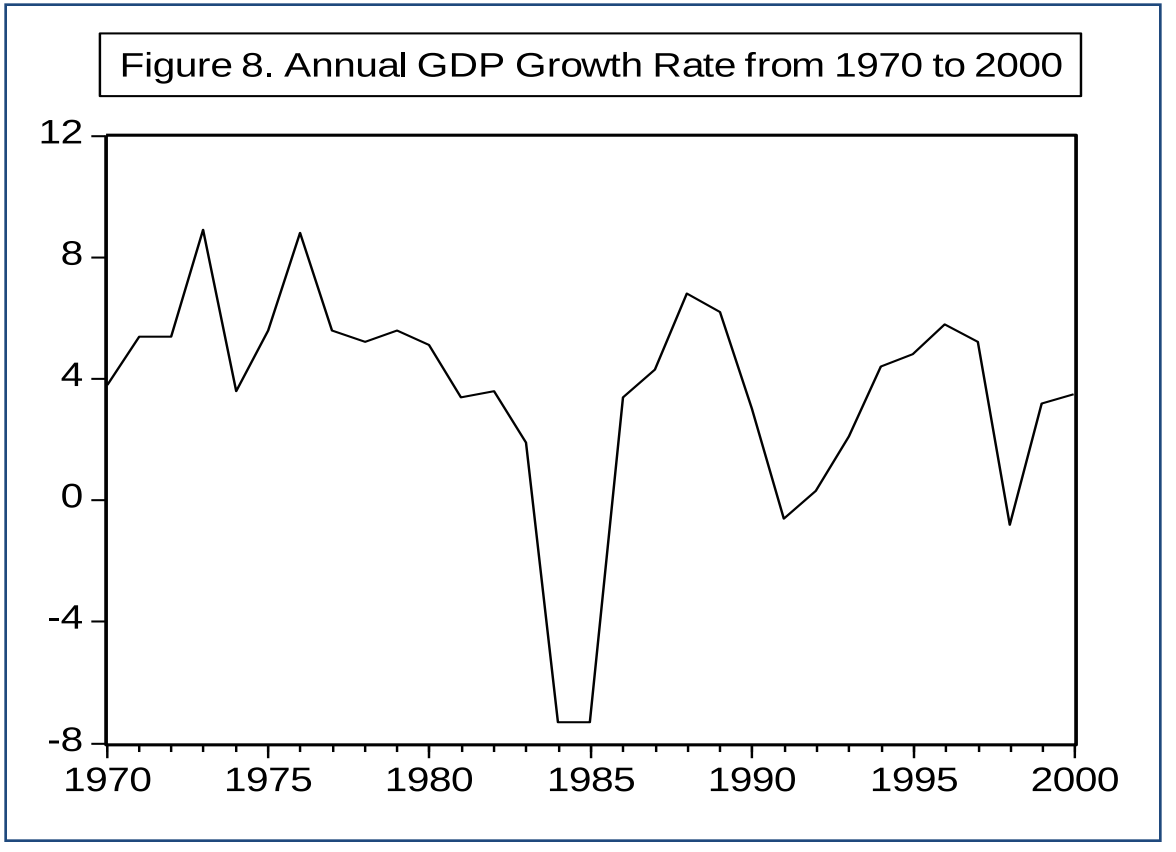


The **seasonality** is the regular rise and fall pattern in the series that occurs ***within a year*** and then repeated on a yearly basis.

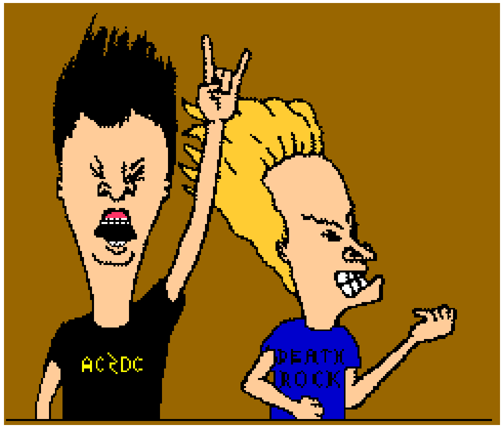
Figure 2. Real Gross Domestic Product (RGDP) from 1st Quarter 1998 to 2nd Quarter 2002



The **cycle** is the upward and downward change in the data pattern that occurs over a ***longer duration***.



The **random disturbance** is the collection of all other factors (shocks) affecting the series not due to trend, cycle or season.



Examples of random disturbances are day-to-day natural variation in the demand that is the result of a consumer's specific day-to-day consumption "mood".

Measuring Forecast Accuracy

Mean Absolute Deviation $\text{MAD} = \sum_{t=1}^{T} \frac{\left| Y_t - \hat{Y}_t \right|}{T}$

Mean Absolute Percentage Error $\text{MAPE} = \frac{100}{T} \sum_{t=1}^{T} \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|$

Mean Square Error $\text{MSE} = \sum_{t=1}^{T} \frac{\left(Y_t - \hat{Y}_t \right)^2}{T}$

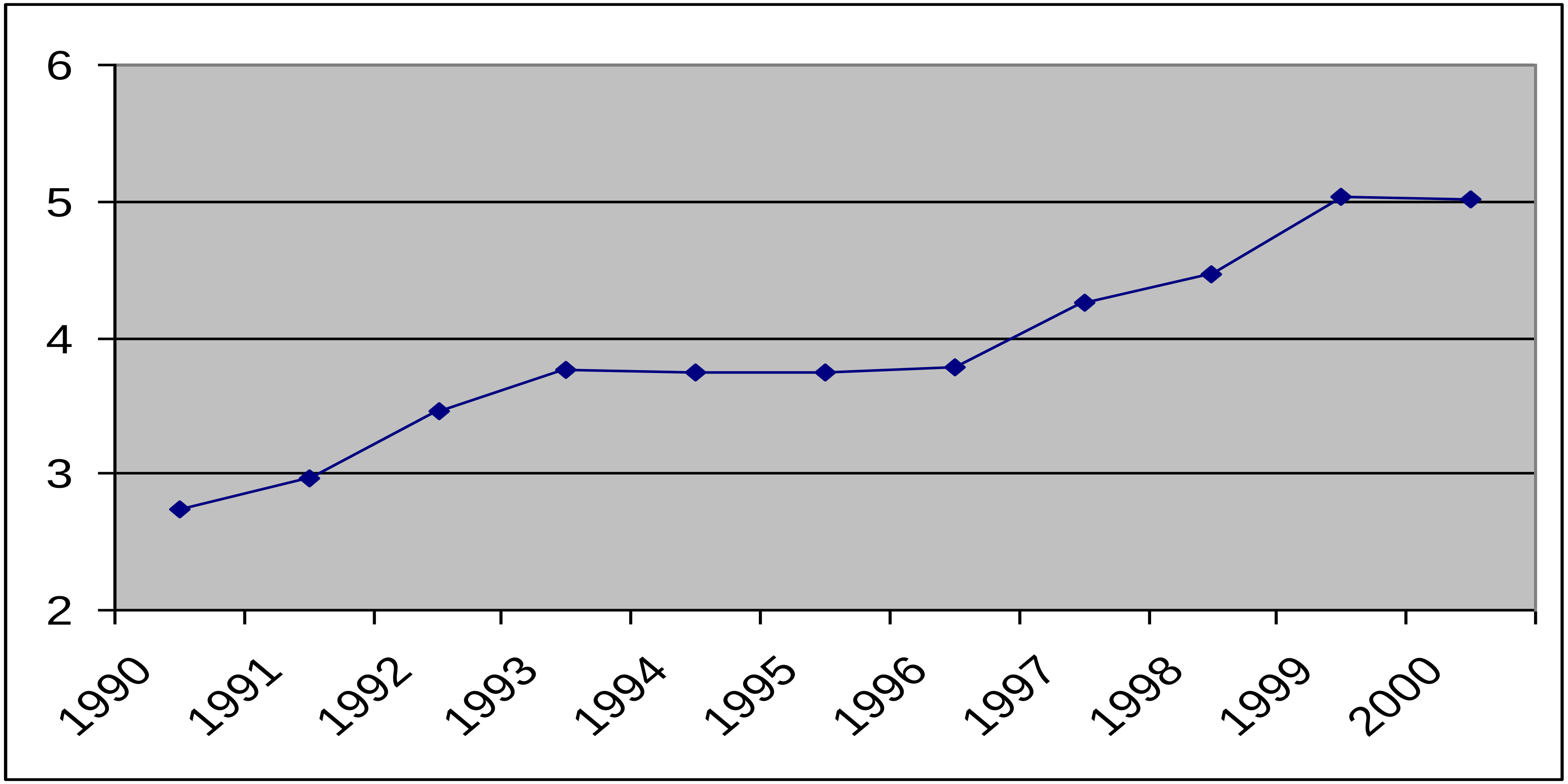
Root Mean Square Error $\text{RMSE} = \sqrt{MSE}$

Addressing Data Gaps

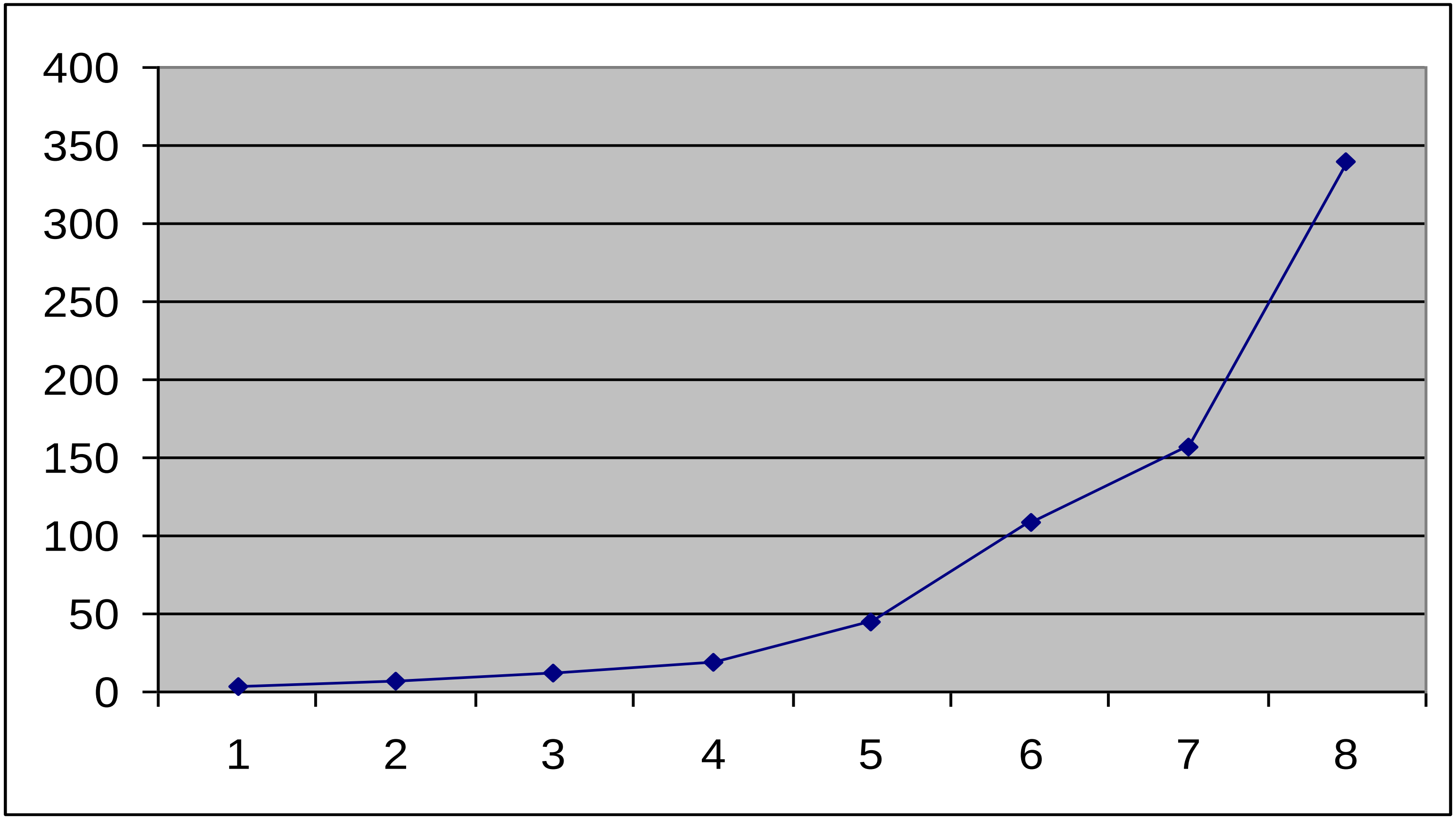
- An implicit assumption of time series analysis is that data are collected, observed or recorded at **regular intervals of time**.

- Forecasting methods lose some degree of efficiency when there are gaps in the time series. Thus, there is a need to estimate missing data.

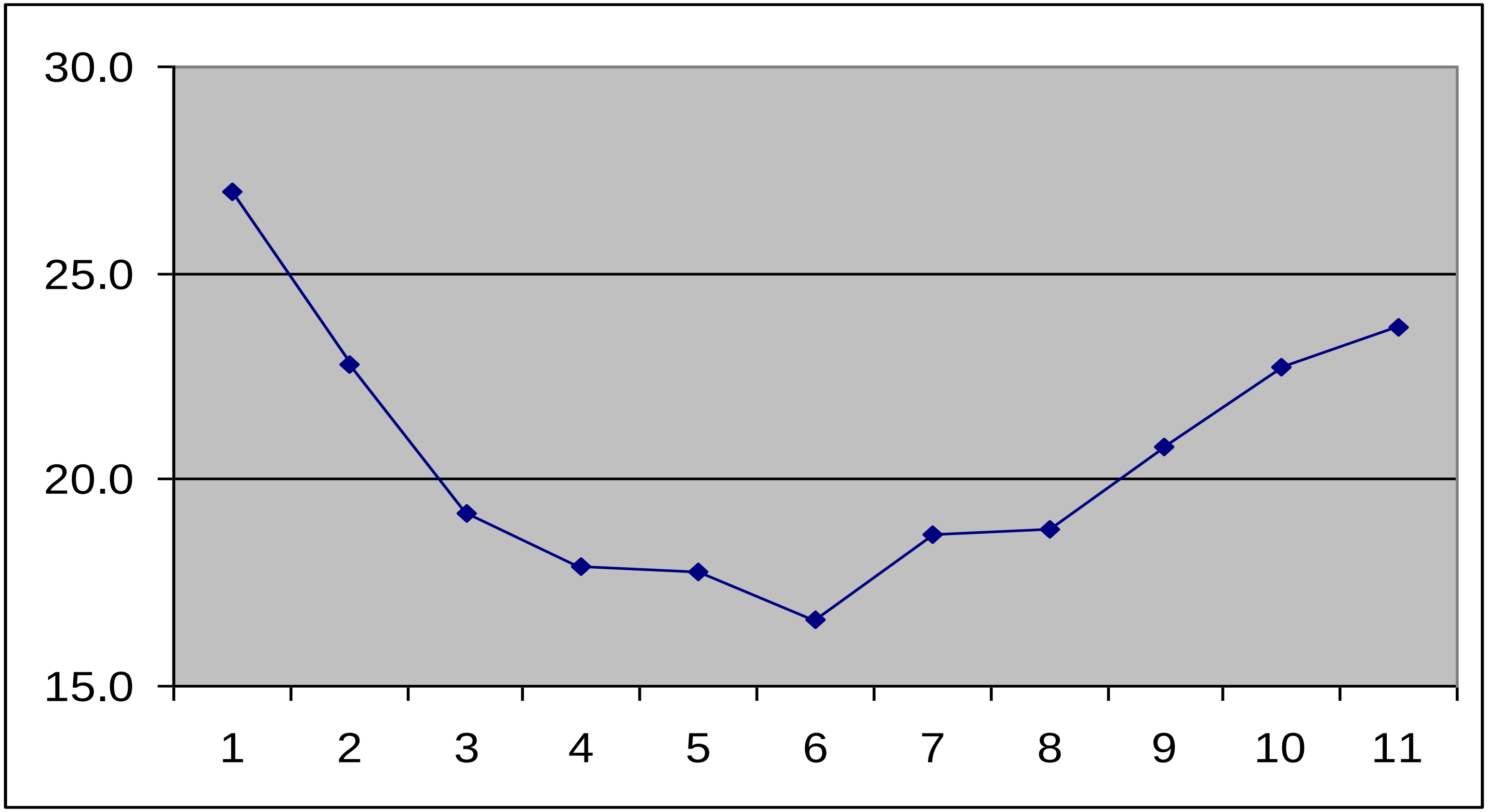
- Estimation of data gaps are usually based on some trend patterns evident in the series. Most common trend patterns are **linear**, **exponential** and **quadratic**.



A **linear** trend is exhibited by a line with minimal variation.



An **exponential** increase is exhibited by a graph curving upwards.



A **quadratic** trend is somewhat U-shaped, concave or convex.

Simple Moving Averages

Simple Moving Averages

The **Simple Moving Average** (SMA) is useful in modeling a ***series*** without trend nor seasonality but only a fluctuation about a common long-term level.

The method of simple moving averages assumes that a future value will equal an average of past values.

In general, an N-period moving average denotes that a new forecast moves one period by adding the newest actual and dropping the oldest actual.

For example, a 4-period SMA for May is SMA4(May) = (Jan+Feb+Mar+Apr)/4.

If the value of the series from Jan to Apr are:

Jan	Feb	Mar	Apr	May
120	124	122	123	?

SMA4(May) = (120+124+122+123)/4 = **122.25**

Now, a 4-period SMA for June is

SMA4(Jun) = (Feb+Mar+Apr+May)/4.

If the value of the series from Feb to May are:

Jan	Feb	Mar	Apr	May	June
120	124	122	123	125	?

SMA4(Jun) = (124+122+123+125)/4 = **123.5**

"H the optimal number of periods in a moving average?"

Here are some considerations

The optimal number of period is one that gives ***minimal forecast error.***

A long-period SMA yields low forecast error when the series is very random and erratic, i.e., it is not possessing high levels of autocorrelation.

A short-period SMA yields low forecast error when the series random yet moves smoothly up and down, i.e., it is highly autocorrelated.

The number of period serves as the "length of memory" of the SMA. A four-period SMA "remembers" the past four actuals while an eight-period SMA "recalls" the eight most recent realizations.

Disadvantages of the SMA

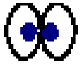
- SMA’s do not model trend or seasonality.
- All data needed to calculate the average must be stored and processed.
- It is difficult to determine the optimal number of periods without the judgment of the researcher who knows or has institutional memory of the series.

Exponential Smoothing

Single Exponential Smoothing

Single exponential smoothing (SES) is another forecasting tool for data with no trend or seasonality. SES can be used for short series.

Why "exponential"?



SES uses forecasting equations based on past observations that are given exponentially decaying weights.

Why is smoothing important?

Smoothing procedures allow estimation of future values by "learning" the historical behavior of a series. When a forecast is needed as input in some prediction equation (like regression models), smoothing methods "supply" such forecasts.

To forecast using SES, we model

Forecast now = Constant x Past Actual + (1-constant) x Past Forecast

The forecast equation for a series Y_t is

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

- α is called the **smoothing constant**. It is between 0 and 1.
- When a great amount of smoothing is desired, a "small" α must be used.
- α is chosen such that a chosen measure of forecast performance (e.g., MAD) is minimized

Example: Suppose a company desires to forecast (monthly) demand of a product using SES with α=0.3. Last month's demand was 1,000 and the forecast was 900. Thus, the forecast for this month is

$$F_t = 0.3(1{,}000) + (1-0.3)(900) = \mathbf{930\ units}$$

Now suppose the actual demand for this month is 980 units, the forecast for next month is

$$\mathbf{F_{t+1} = 0.3(980) + (1-0.3)(930) = 945\ units}$$

Further, if the actual demand next month is 950 units, then the forecast two months hence is

$$F_{t+2} = 0.3(950) + (1-0.3)(945) = \mathbf{946.5\ units}$$

Double Exponential Smoothing (Brown's One-Parameter Method)

- This is applicable to data with trend but no seasonality.
- The forecast equation for Y_{t+L} is

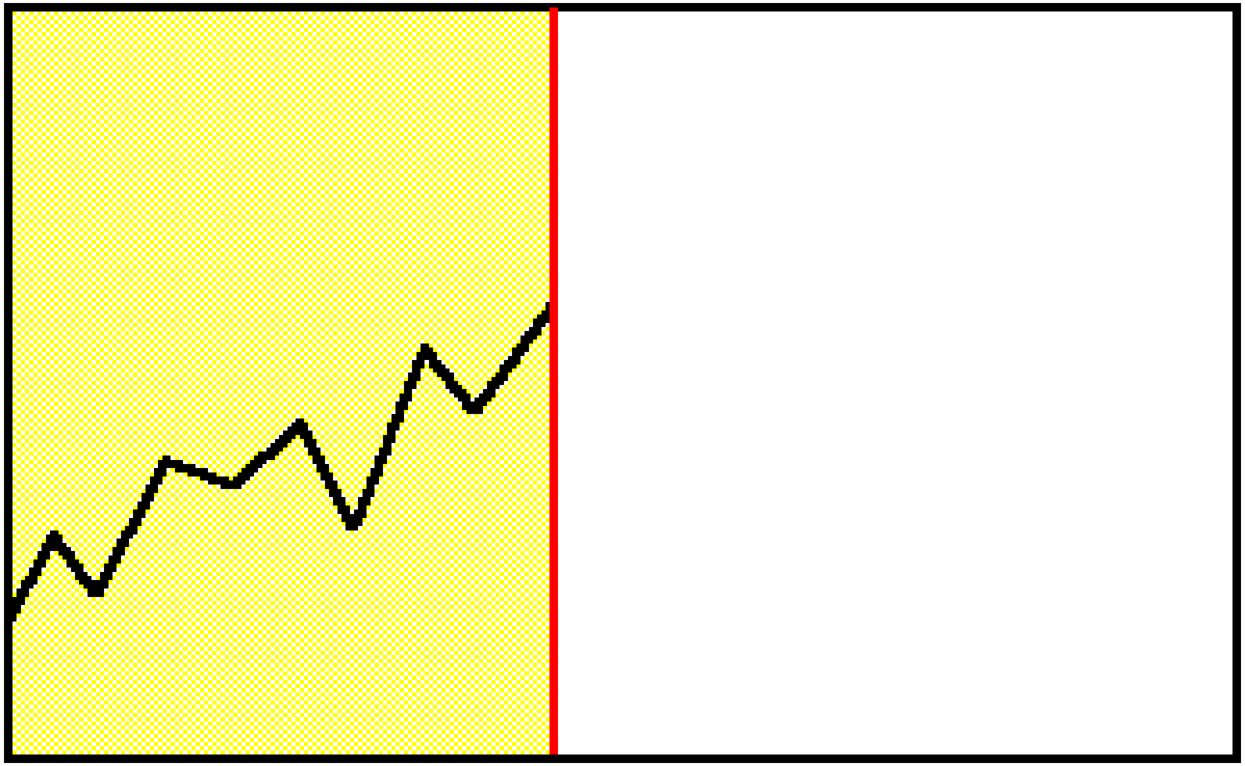
$$F_{t+L} = \text{MEAN}_t + \text{TREND}_t \times L$$

where MEAN and TREND are the "intercept" and "slope", respectively, of the forecast equation.

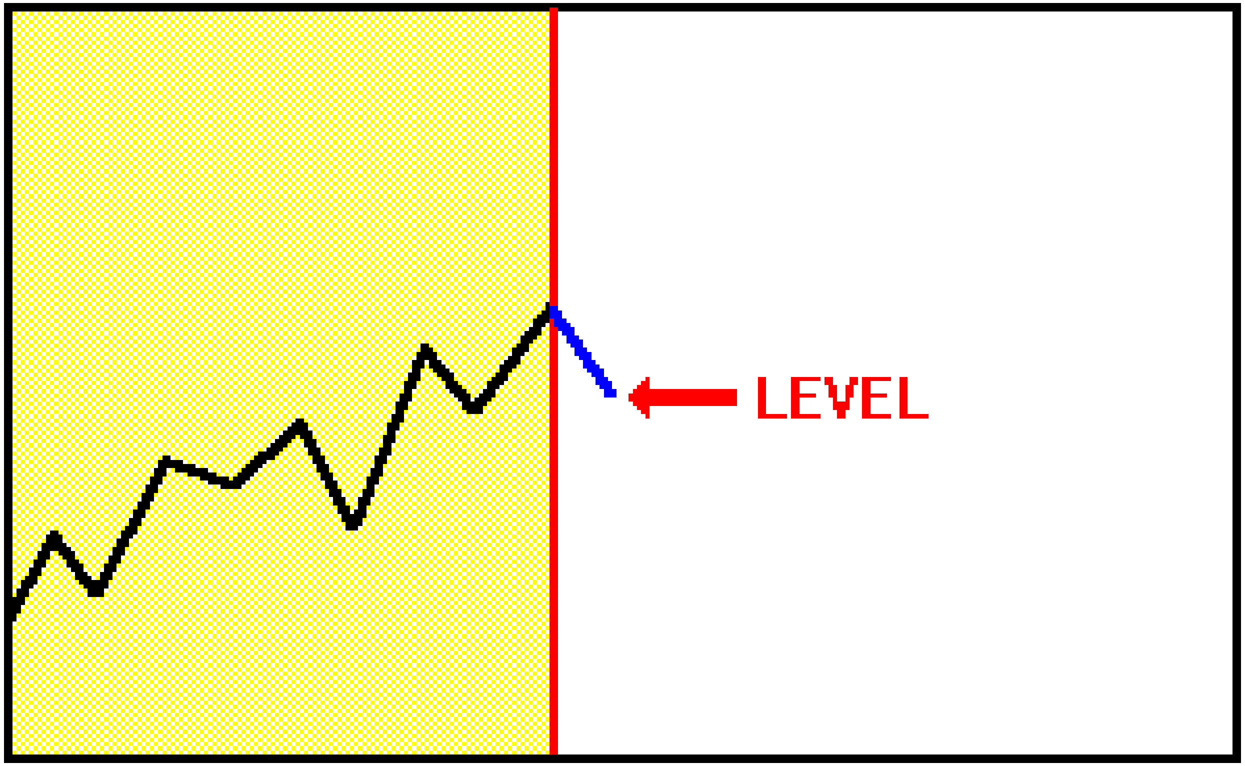
Holt-Winters Method

- Assumes that a reasonably consistent seasonal pattern exists, and establishes this
- Combines current level, trend, and seasonalities to forecast future values
- Uses exponential smoothing to estimate each of these parameters

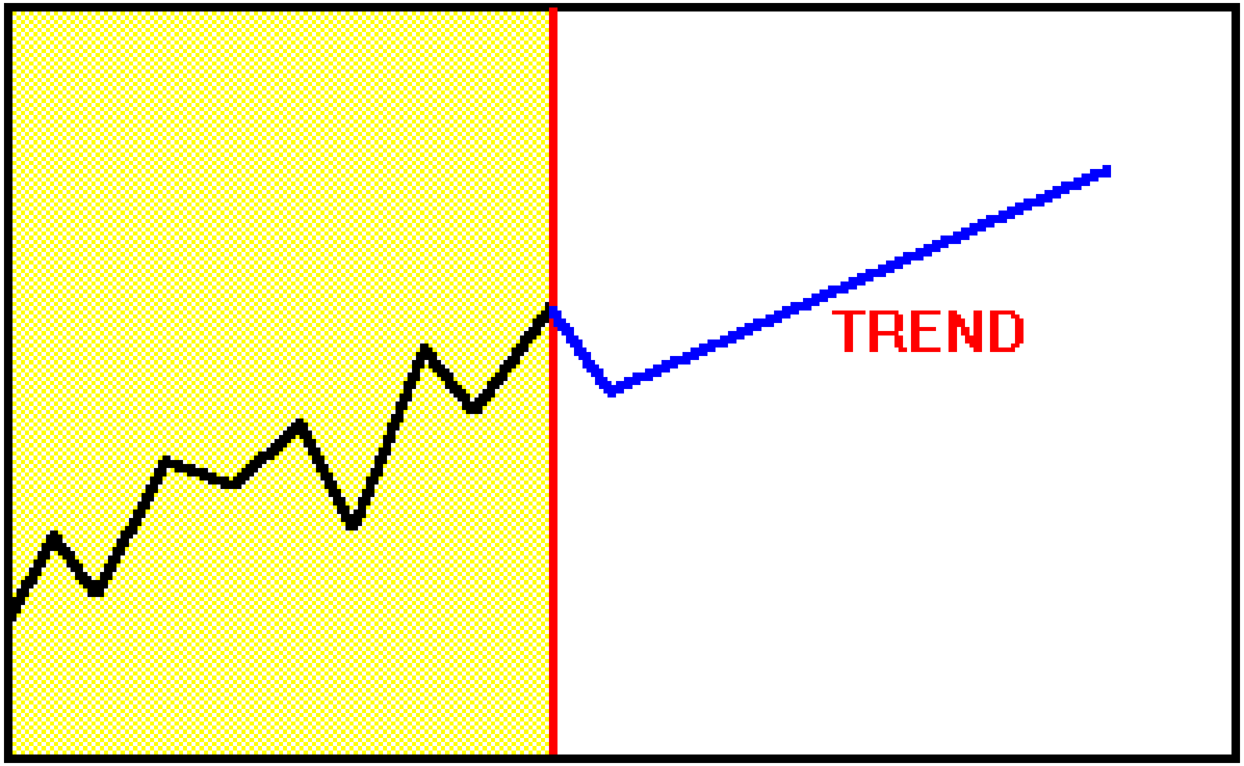
Holt-Winters Method



Holt-Winters Method

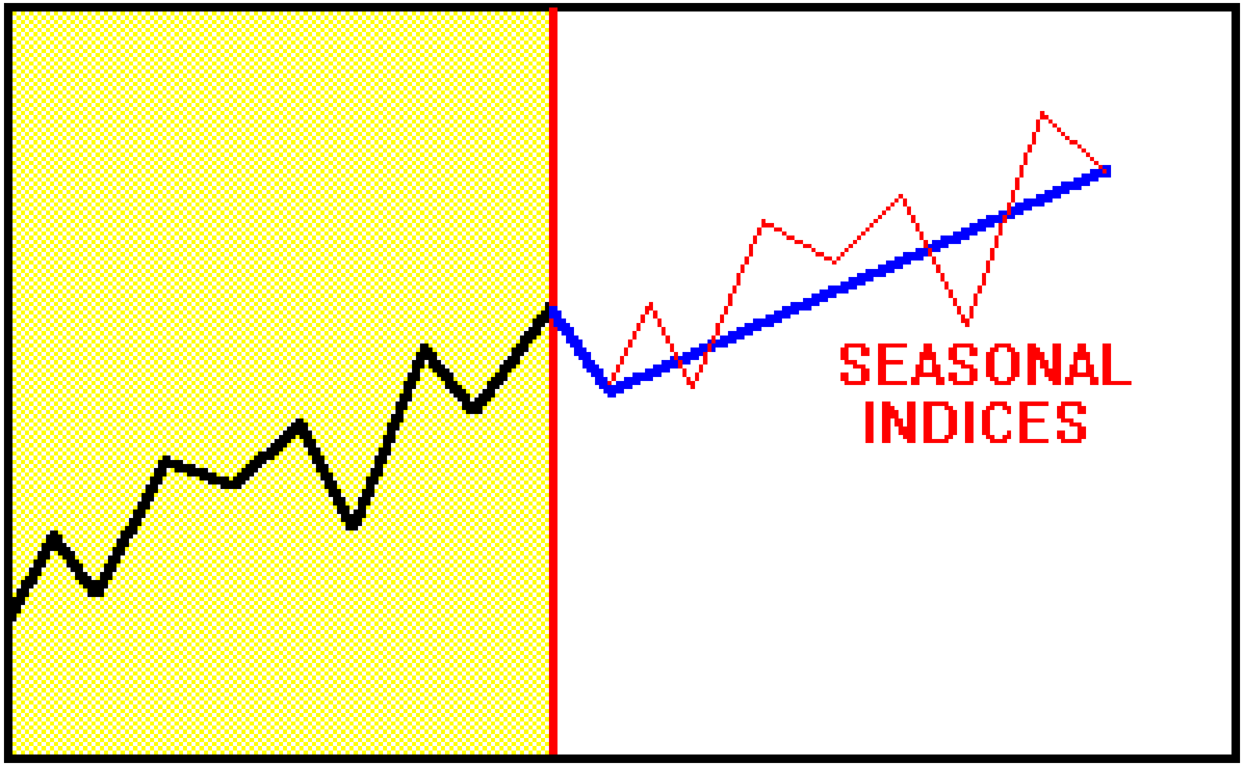


Holt-Winters Method



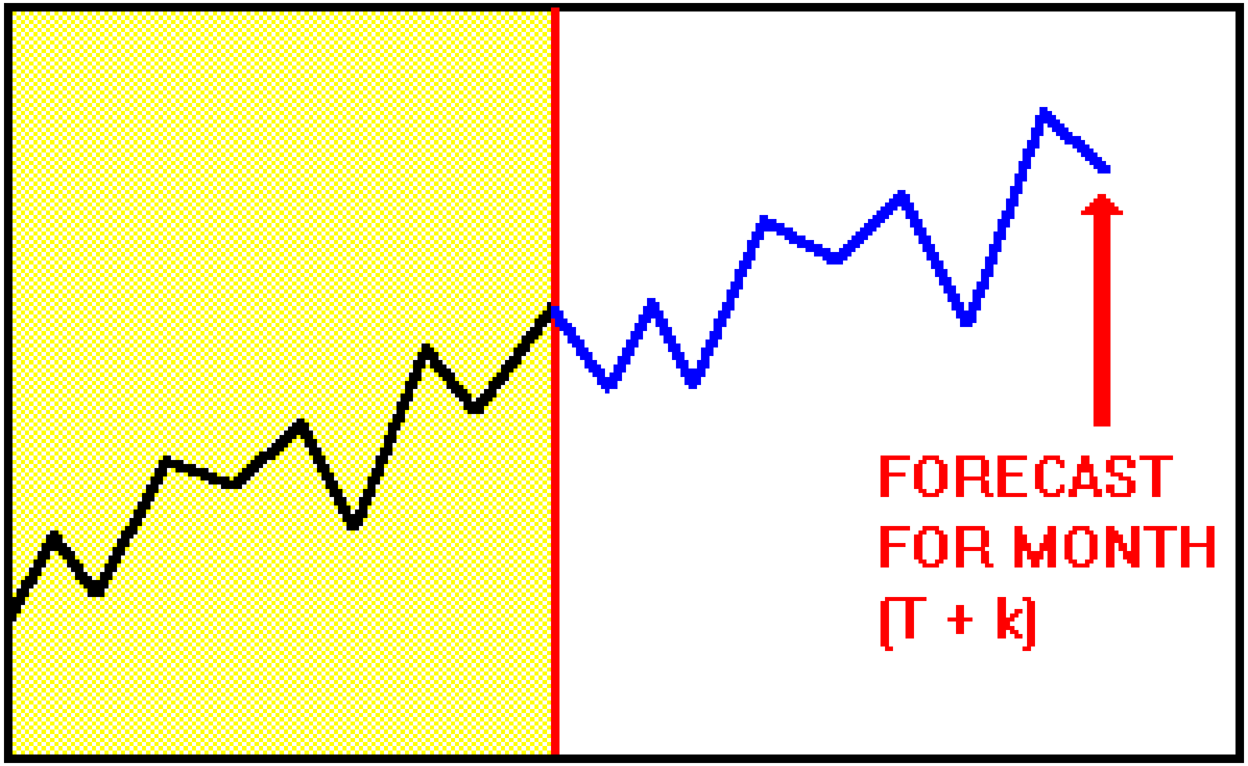
Exponential Smoothing

Holt-Winters Method



Exponential Smoothing

Holt-Winters Method



Double Exponential Smoothing (Holt's Two-Parameter Method)

- This is applicable to data with trend but no seasonality.
- This procedure uses two smoothing constants, one each for the mean and the trend; thus allowing for greater flexibility in the values of MEAN and TREND.

Double Exponential Smoothing (Holt's Two-Parameter Method)

- The forecast equation for Y_{t+L} is

$$F_{t+L} = \text{MEAN}_t + \text{TREND}_t \times L$$

where MEAN and TREND are the "intercept" and "slope", respectively, of the forecast equation.

- L is the forecast horizon from time t.

Triple Exponential Smoothing (Additive Model)

- This method is applicable to data with trend and additive seasonality.

- The forecast equation for Y_{t+L} is

  $$F_{t+L} = \text{MEAN}_t + \text{TREND}_t \times L + \text{SEASON}_{t-s+L}$$

  where SEASON is the seasonality part and *s* is the length of seasonality.

Triple Exponential Smoothing (Multiplicative Model)

- This is applicable to data with trend and multiplicative seasonality.
- The forecast equation for Y_{t+L} is

  $$F_{t+L} = (\text{MEAN}_t + \text{TREND}_t \times L) \times \text{SEASON}_{t-s+L}$$

  where SEASON is the seasonality index and s is the length of seasonality.

"I know of no way of judging the future but by the past."

Patrick Henry