Forecasting

Akanksha Bharadwaj
Asst. Professor, CS/IS

BITS Pilani
Pilani Campus

innovate achieve lead

BITS Pilani
Pilani Campus

SS ZG536, ADV STAT TECHNIQUES FOR ANALYTICS
Contact Session 10

Definition of Forecasting

- Forecasting is the process of making predictions of the future based on past and present data and most commonly by analysis of trends.

- A common place example might be estimation of some variable of interest at some specified future date.

Forecasting is required in many situations

- Deciding whether to build another power generation plant in the next five years requires forecasts of future demand
- Scheduling staff in a call center next week requires forecasts of call volumes
- Stocking an inventory requires forecasts of stock requirements.

Forecasting

- Some things are easier to forecast than others. The time of the sunrise tomorrow morning can be forecast precisely.

- On the other hand, tomorrow's lotto numbers cannot be forecast with any accuracy.

The predictability of an event or a quantity depends on several factors including:

- how well we understand the factors that contribute to it;
- how much data is available;
- whether the forecasts can affect the thing we are trying to forecast.

Example

- For example, forecasts of electricity demand can be highly accurate because all three conditions are usually satisfied.

- We have a good idea of the contributing factors: electricity demand is driven largely by temperatures, with smaller effects for calendar variation such as holidays, and economic conditions.

- Provided there is a sufficient history of data on electricity demand and weather conditions,

- and we have the skills to develop a good model linking electricity demand and the key driver variables, the forecasts can be remarkably accurate.

Example

- On the other hand, when forecasting currency exchange rates, only one of the conditions is satisfied: there is plenty of available data.

- However, we have a limited understanding of the factors that affect exchange rates, and forecasts of the exchange rate have a direct effect on the rates themselves.

- forecasting whether the exchange rate will rise or fall tomorrow is about as predictable as forecasting whether a tossed coin will come down as a head or a tail.

- In both situations, you will be correct about 50% of the time, whatever you forecast.

- In situations like this, forecasters need to be aware of their own limitations, and not claim more than is possible.

FORECASTING

- Virtually all areas of business, including production, sales, employment, transportation, distribution, and inventory, produce and maintain time-series data.
- Table provides an example of time-series data released by the Office of Market Finance,U.S. Department of the Treasury.
- The table contains the bond yield rates of three-month Treasury Bills for a 17-year period

TABLE 15.1

Bond Yields of Three-Month Treasury Bills

Year	Average Yield
1	14.03%
2	10.69
3	8.63
4	9.58
5	7.48
6	5.98
7	5.82
8	6.69
9	8.12
10	7.51
11	5.42
12	3.45
13	3.02
14	4.29
15	5.51
16	5.02
17	5.07

Principles of Forecasting

- There are many types of forecasting models.
- They differ in their degree of complexity, the amount of data they use, and the way they generate the forecast.
- However, some features are common to all forecasting models. They include the following:
- Forecasts are rarely perfect
- Forecasts are more accurate for groups or families of items rather than for individual items.
- Forecasts are more accurate for shorter than longer time horizons.

Time Series

- A time series is a collection of observations of well-defined data items obtained through repeated measurements over time.
- For example, measuring the value of retail sales each month of the year would comprise a time series.
- This is because sales revenue is well defined, and consistently measured at equally spaced intervals.
- Data collected irregularly or only once are not time series.

Time Series

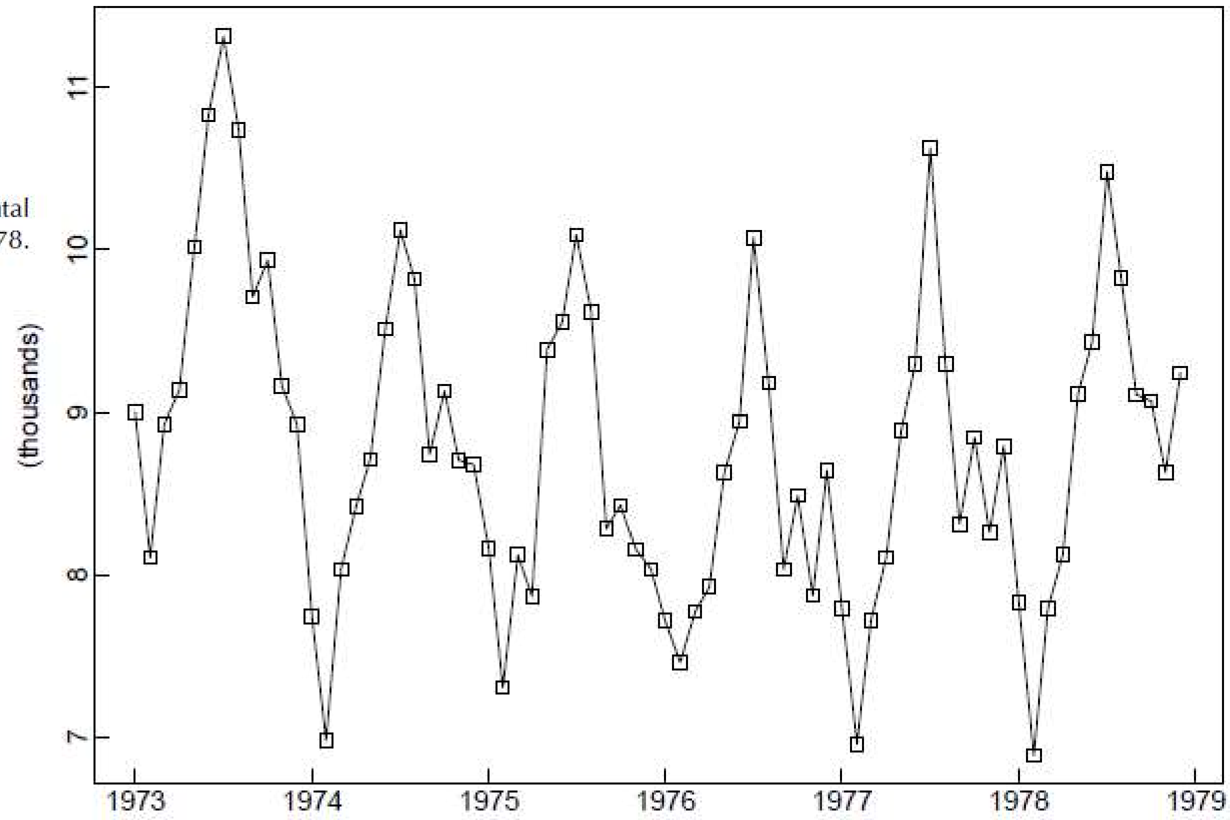
- A **time series** is a set of observations x_t, each one being recorded at a specific time t.

- A discrete-time time series is one in which the set T_0 of times at which observations are made is a discrete set, as is the case, for example, when observations are made at fixed time intervals.

- Continuous time series are obtained when observations are recorded continuously over some time interval, e.g., when $T_0 = [0, 1]$.

Example

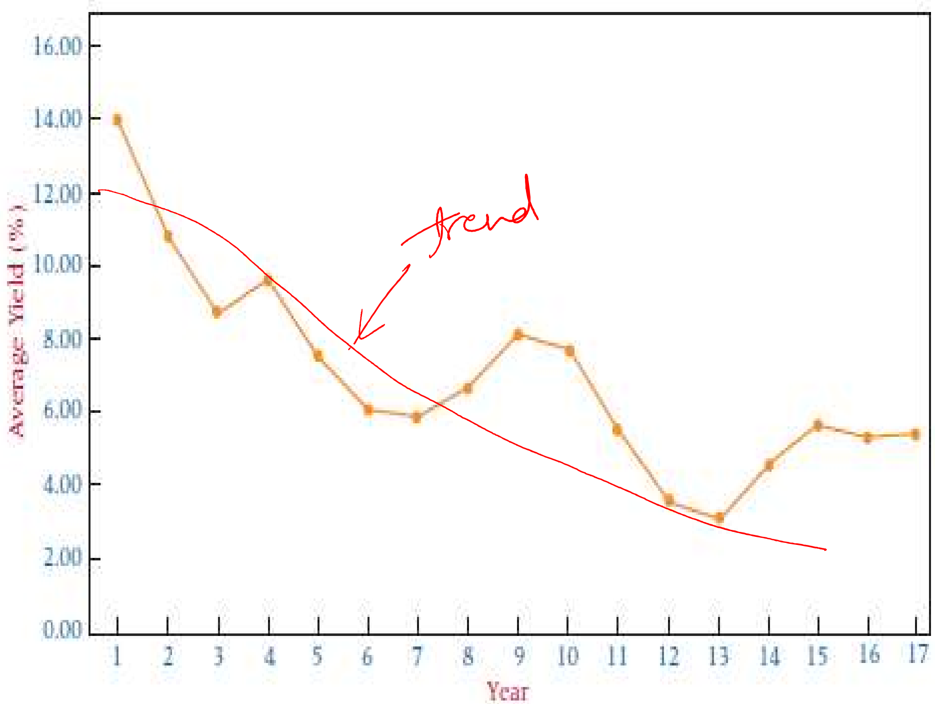
The monthly accidental deaths data, 1973–1978.



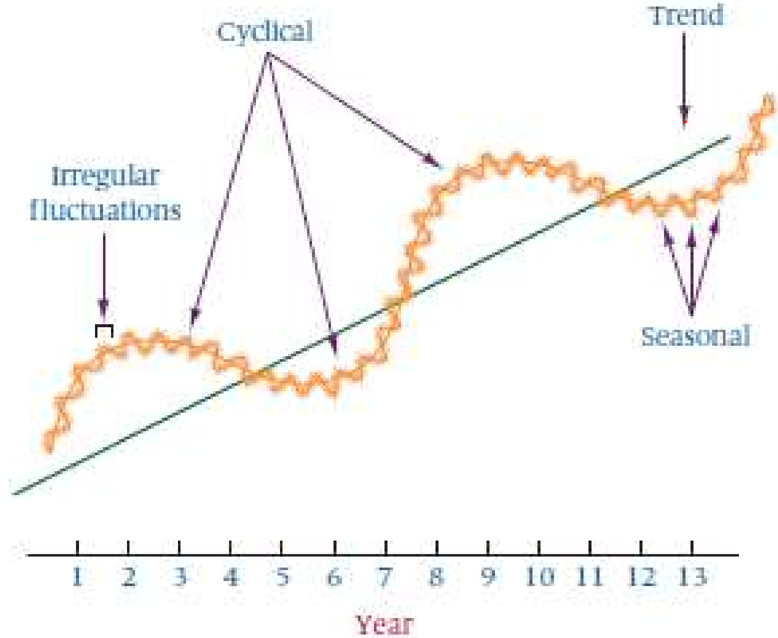
Why use time series data?

- To develop forecasting models
  - What will the rate of inflation be next year?

- To estimate dynamic causal effects
  - If the Fed increases the Federal Funds rate now, what will be the effect on the rates of inflation and unemployment in 3 months? in 12 months?



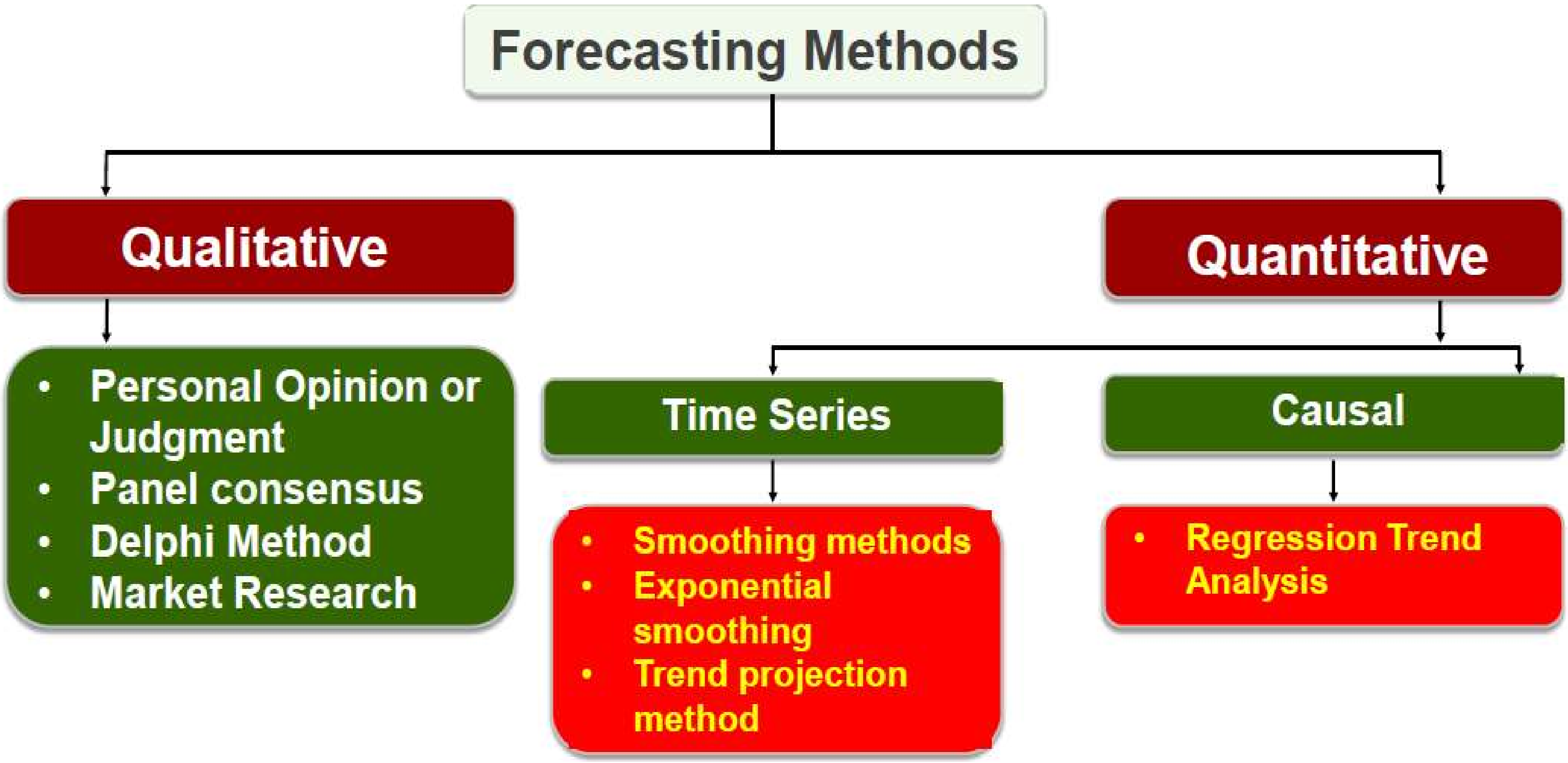
Excel Graph of Bond Yield
Time-Series Data (new)



Time-Series Components

- **Trend:** *The long-term general direction* of data
- **Cycles** are *patterns of highs and lows through which data move over time periods usually of more than a year.*
- **Seasonal effects**, on the other hand, are *shorter cycles, which usually occur in time periods of less than one year.*
- Often seasonal effects are measured by the month, but they may occur by quarter, or may be measured in as small a time frame as a week.
- **Irregular fluctuations** are *rapid changes or "bleeps" in the data, which occur in even shorter time frames than seasonal effects.*
- Irregular fluctuations can happen as often as day to day. They are subject to momentary change and are often unexplained.

Types of Forecasting methods



Qualitative methods

Personal Opinion

- Individuals forecasts future based on their own judgment or opinion without any formal model

Panel consensus

- Panel of individuals are encouraged to share information, opinions, and assumptions, if any, to predict future value of some variable under study

Qualitative method

Type	Characteristics	Strengths	Weaknesses
Executive opinion	A group of managers meet & come up with a forecast	Good for strategic or new-product forecasting	One person's opinion can dominate the forecast
Market research	Uses surveys & interviews to identify customer preferences	Good determinant of customer preferences	It can be difficult to develop a good questionnaire
Delphi method	Seeks to develop a consensus among a group of experts	Excellent for forecasting long-term product demand,	Time consuming to develop

Quantitative forecasting

Time Series Models:

- Assumes information needed to generate a forecast is contained in a time series of data
- Assumes the future will follow same patterns as the past

Causal Models or Associative Models

- Explores cause-and-effect relationships
- Uses leading indicators to predict the future
- Eg. House sales and appliance sales

The Measurement of Forecasting Error

How does a decision maker know which forecasting technique is doing the best job in predicting the future?

- One way is to compare forecast values with actual values and determine the amount of **forecasting error**
- An examination of individual errors gives some insight into the accuracy of the forecasts.
- However, this process can be tedious, especially for large data sets

ERROR OF AN INDIVIDUAL FORECAST

$$e_t = x_t - F_t$$

where

e_t = the error of the forecast
x_t = the actual value
F_t = the forecast value

Mean Absolute Deviation (MAD)

- The **mean absolute deviation (MAD)** is *the mean, or average, of the absolute values of the errors.*

MEAN ABSOLUTE DEVIATION

$$MAD = \frac{\sum |e_i|}{\text{Number of Forecasts}}$$

Actual value = 10

$P_1 = 9$

$P_2 = 12$

$e_1 = 10 - 9 = 1$

$e_2 = 10 - 12 = -2$

MAD example

Example

TABLE 15.2

Nonfarm Partnership Tax Returns

Year	Actual	Forecast	Error
1	1,402	—	—
2	1,458	1,402	56.0
3	1,553	1,441.2	111.8
4	1,613	1,519.5	93.5
5	1,676	1,585.0	91.0
6	1,755	1,648.7	106.3
7	1,807	1,723.1	83.9
8	1,824	1,781.8	42.2
9	1,826	1,811.3	14.7
10	1,780	1,821.6	−41.6
11	1,759	1,792.5	−33.5

The mean absolute error can be computed for the forecast errors in Table 15.2 as follows.

$$\text{MAD} = \frac{|56.0| + |111.8| + |93.5| + |91.0| + |106.3| + |83.9| + |42.2| + |14.7| + |-41.6| + |-33.5|}{10} = 67.45$$

Mean Square Error (MSE)

- The MSE is *computed by squaring each error (thus creating a positive number) and averaging the squared errors.*

MEAN SQUARE ERROR

$$MSE = \frac{\sum e_i^2}{\text{Number of Forecasts}}$$

MSE example

Example

TABLE 15.2

Nonfarm Partnership Tax Returns

Year	Actual	Forecast	Error
1	1,402	—	—
2	1,458	1,402	56.0
3	1,553	1,441.2	111.8
4	1,613	1,519.5	93.5
5	1,676	1,585.0	91.0
6	1,755	1,648.7	106.3
7	1,807	1,723.1	83.9
8	1,824	1,781.8	42.2
9	1,826	1,811.3	14.7
10	1,780	1,821.6	−41.6
11	1,759	1,792.5	−33.5

The mean square error can be computed for the errors shown in Table 15.2 as follows.

$$\text{MSE} = \frac{(56.0)^2 + (111.8)^2 + (93.5)^2 + (91.0)^2 + (106.3)^2 + (83.9)^2 + (42.2)^2 + (14.7)^2 + (-41.6)^2 + (-33.5)^2}{10} = 5{,}584.7$$

Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE)

$$= \frac{\sum \left| \frac{X_i - Fi}{X_i} \right|}{\textit{Number of forecasts}} \text{ x } 100$$

X is actual value and F is forecasted value

When to prefer MAD

- MAD is a better measure of error than MSE if forecast error does not have a symmetric distribution.

- It gives us the average difference between F and D, disregarding the accuracy of F.

When to prefer MAPE

- MAPE (mean absolute percentage error): good measure of forecast error when the underlying forecast has significant **seasonality** and demand varies considerably from one period to the next.

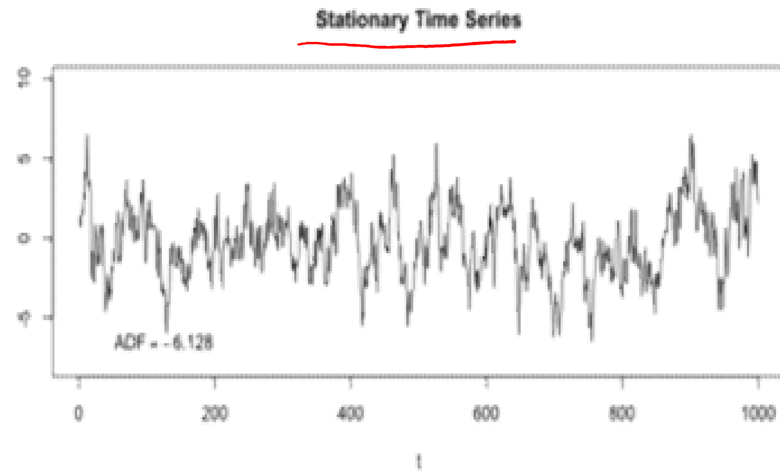
When to prefer MSE

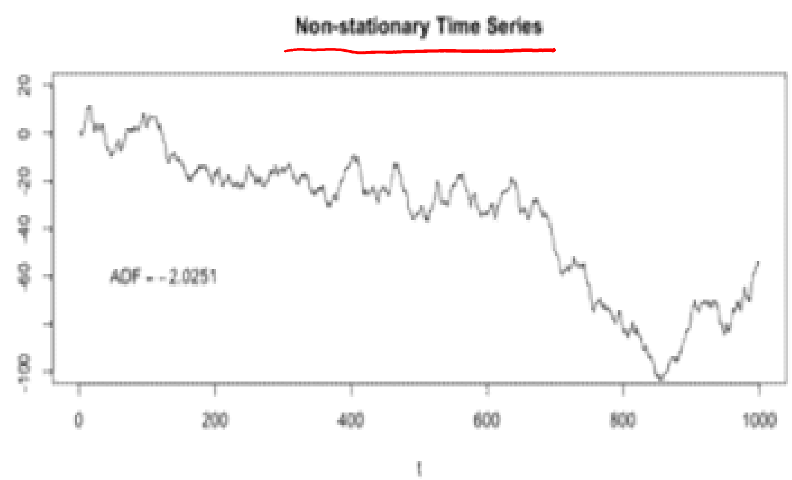
- MSE (mean squared error): MSE is related to the **variance** of the forecast error.
- We estimate the random component of demand has a mean of 0 and a **variance** of MSE.
- The MSE penalizes large errors much more significantly than small errors because all errors are squared.
- Because of this, it is a good idea to use the MSE to compare forecasting methods if **the cost of a large error is much larger than the gains from very accurate forecasts**.
- MSE is appropriate when forecast error has a distribution that is symmetric about zero.

Stationary time series

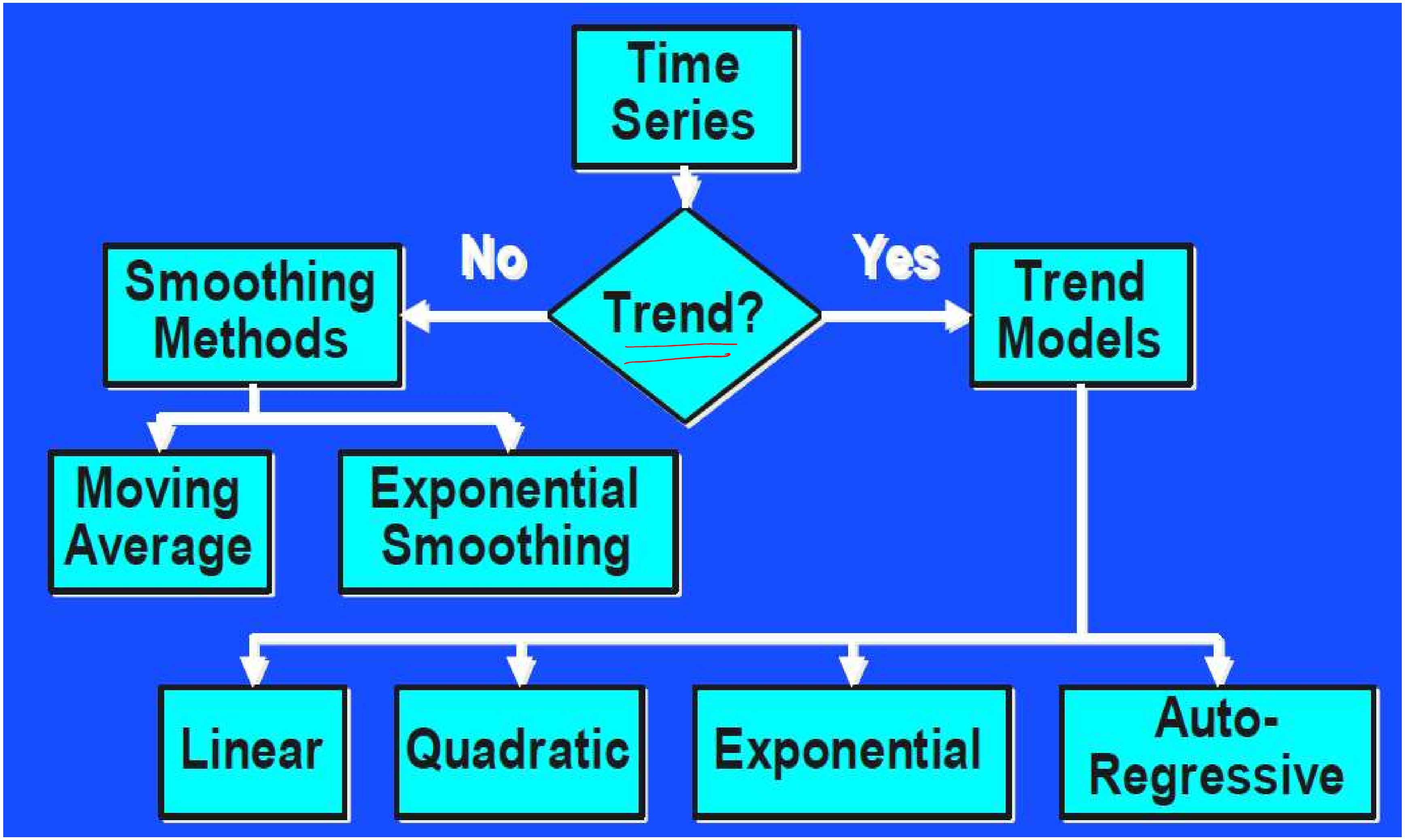
- *Time-series data that contain no trend, cyclical, or seasonal effects* are said to be **stationary**.

- Techniques used to forecast stationary data analyze only the irregular fluctuation effects.





Analysis



Time Series
Trend?
No
Yes
Smoothing Methods
Trend Models
Moving Average
Exponential Smoothing
Linear
Quadratic
Exponential
Auto-Regressive

SMOOTHING TECHNIQUES

- Several techniques are available to forecast time-series data that are **stationary** or that include no significant trend, cyclical, or seasonal effects.

- These techniques are often referred to as **smoothing techniques** because they *produce forecasts based on "smoothing out" the irregular fluctuation effects in the time-series data.*

  (1) naive forecasting models,

  (2) averaging models, and

  (3) exponential smoothing

Naïve Forecasting Models

- They are *simple models in which it is assumed that the more recent time periods of data represent the best predictions or forecasts for future outcomes.*
- Naïve models do not take into account data trend, cyclical effects, or seasonality.
- For this reason, naive models seem to work better with data that are reported on a daily or weekly basis or in situations that show no trend or seasonality.
- The simplest of the naïve forecasting methods is the model in which the forecast for a given time period is the value for the previous time period.

$$F_t = x_{t-1}$$

where

F_t = the forecast value for time period t

x_{t-1} = the value for time period $t - 1$

Example

- Table is representing the total reported domestic rail, truck, and air shipments of bell peppers in the United States for a given year reported by the U.S. Department of Agriculture.

Month	Shipments (millions of pounds)
January	336
February	308
March	582
April	771
May	935
June	808
July	663
August	380
September	333
October	412
November	458
December	412

- Prediction for Jan next year can be taken as 412 using naïve model

- Use 336 as the prediction for Jan next year

Averaging Models

- **Averaging models** are computed by *averaging data from several time periods and using the average as the forecast for the next time period.*

Moving Averages

- Suppose we were to attempt to forecast the heating oil cost for October of year 3 by using MA as the forecasting method.
- It would seem to make sense to use the 12 months prior to October of year 3 (i.e. October of year 2 through September of year 3) to average for the new forecast.
- Suppose in September of year 3 the cost of heating oil is 53.3 cents

$$= \frac{56.7 + 57.2 + 58.0 + 58.2 + 58.3 + 57.7 + 56.7 + 56.8 + 55.5 + 53.8 + 52.8 + 53.3}{12} = 56.25$$

Cost of Residential Heating Oil (cents per gallon)

Time Frame	Cost of Heating Oil
January (year 1)	66.1
February	66.1
March	66.4
April	64.3
May	63.2
June	61.6
July	59.3
August	58.1
September	58.9
October	60.9
November	60.7
December	59.4
January (year 2)	61.3
February	63.3
March	62.1
April	59.8
May	58.4
June	57.6
July	55.7
August	55.1
September	55.7
October	56.7
November	57.2
December	58.0
January (year 3)	58.2
February	58.3
March	57.7
April	56.7
May	56.8
June	55.5
July	53.8
August	52.8
Sep	53.3

- The Moving Average basically filters out rapid fluctuations .i.e. high frequency noise. Thus, it acts as a low-pass filter.

Disadvantages:

- It is difficult to choose the optimal length of time for which to compute the moving average, and
- moving averages do not usually adjust for such time-series effects as trend, cycles, or seasonality.
- To determine the more optimal lengths for which to compute the moving averages, we would need to forecast with several different average lengths and compare the errors produced by them.

Example

- Shown here are shipments (in millions of dollars) for electric lighting and wiring equipment over a 12-month period. Use these data to compute a 4-month moving average for all available months.

4-Month Moving Forecast

Month	Shipments	Average	Error
January	1056	—	—
February	1345	—	—
March	1381	—	—
April	1191	—	—
May	1259	1243.25	15.75
June	1361	1294.00	67.00
July	1110	1298.00	−188.00
August	1334	1230.25	103.75
September	1416	1266.00	150.00
October	1282	1305.25	−23.25
November	1341	1285.50	55.50
December	1382	1343.25	38.75

Month	Shipments
January	1056
February	1345
March	1381
April	1191
May	1259
June	1361
July	1110
August	1334
September	1416
October	1282
November	1341
December	1382

Weighted Moving Averages

- A forecaster may want to place more weight on certain periods of time than on others. For example, a forecaster might believe that the previous month's value is three times as important in forecasting as other months.
- *A moving average in which some time periods are weighted differently than others* is called a **weighted moving average**

As an example, suppose a 3-month weighted average is computed by weighting last month's value by 3, the value for the previous month by 2, and the value for the month before that by 1. This weighted average is computed as

$$\bar{x}_{weighted} = \frac{\boxed{3}(M_{t-1}) + 2(M_{t-2}) + 1(M_{t-3})}{6}$$

Example

- Compute a 4-month weighted moving average for the electric lighting and wiring data, using weights of 4 for last month's value, 2 for the previous month's value, and 1 for each of the values from the 2 months prior to that

Month	Shipments	4-Month Weighted Moving Average Forecast	Error
January	1056	—	—
February	1345	—	—
March	1381	—	—
April	1191	—	—
May	1259	1240.9	18.1
June	1361	1268.0	93.0
July	1110	1316.8	−206.8
August	1334	1201.5	132.5
September	1416	1272.0	144.0
October	1282	1350.4	−68.4
November	1341	1300.5	40.5
December	1382	1334.8	47.2

Month	Shipments
January	1056
February	1345
March	1381
April	1191
May	1259
June	1361
July	1110
August	1334
September	1416
October	1282
November	1341
December	1382

Exponential Smoothing

- Another forecasting technique, **exponential smoothing**, is *used to weight data from previous time periods with exponentially decreasing importance in the forecast.*
- Exponential smoothing is accomplished by multiplying the actual value for the present time period, X_t, by a value between 0 and 1

EXPONENTIAL SMOOTHING

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

where

F_{t+1} = the forecast for the next time period $(t + 1)$
F_t = the forecast for the present time period (t)
X_t = the actual value for the present time period
α = a value between 0 and 1 referred to as the exponential smoothing constant.

Example

- The U.S. Census Bureau reports the total units of new privately owned housing started over a 16-year recent period in the United States are given here. Use exponential smoothing to forecast the values for each ensuing time period. Work the problem using α = .2, .5, and .8.

Year	Total Units (1000)
1	1193
2	1014
3	1200
4	1288
5	1457
6	1354
7	1477
8	1474
9	1617
10	1641
11	1569
12	1603
13	1705
14	1848
15	1956
16	2068

Solution

Year	Total Units (1000)	α = .2 F	α = .2 e	α = .5 F	α = .5 e	α = .8 F	α = .8 e
1	1193	—	—	—	—	—	—
2	1014	1193.0	−179.0	1193.0	−179.0	1193.0	−179.0
3	1200	1157.2	42.8	1103.5	96.5	1049.8	150.2
4	1288	1165.8	122.2	1151.8	136.2	1170.0	118.0
5	1457	1190.2	266.8	1219.9	237.1	1264.4	192.6
6	1354	1243.6	110.4	1338.4	15.6	1418.5	−64.5
7	1477	1265.7	211.3	1346.2	130.8	1366.9	110.1
8	1474	1307.9	166.1	1411.6	62.4	1455.0	19.0
9	1617	1341.1	275.9	1442.8	174.2	1470.2	146.8
10	1641	1396.3	244.7	1529.9	111.1	1587.6	53.4
11	1569	1445.2	123.8	1585.5	−16.5	1630.3	−61.3
12	1603	1470.0	133.0	1577.2	25.8	1581.3	21.7
13	1705	1496.6	208.4	1590.1	114.9	1598.7	106.3
14	1848	1538.3	309.7	1647.6	200.4	1683.7	164.3
15	1956	1600.2	355.8	1747.8	208.2	1815.1	140.9
16	2068	1671.4	396.6	1851.9	216.1	1927.8	140.2

	α = .2	α = .5	α = .8
MAD:	209.8	128.3	111.2
MSE:	53,110.5	21,628.6	15,245.4

Exercise (HW)

- Following are time-series data for eight different periods. Use exponential smoothing to forecast the values for periods 3 through 8. Use the value for the first period as the forecast for the second period. Compute forecasts using two different values of alpha, α= 0.1 and α= 0.8. Compute the errors for each forecast and compare the errors produced by using the two different exponential smoothing constants

Time Period	Value		Time Period	Value
1	211		5	242
2	228	211	6	227
3	236	212.7	7	217
4	241		8	203

$F_3 = 0.1 \times (228) + 0.9 \times 211$

≈ 212.7

$F_4 = 0.1 \times (236) + 0.9 \times 212.7$

Solution

Period	Value	$F(\alpha = .1)$	Error	$F(\alpha = .8)$	Error	Difference
1	211					
2	228	211				
3	236	213	23	225	11	12
4	241	215	26	234	7	19
5	242	218	24	240	2	22
6	227	220	7	242	-15	22
7	217	221	-4	230	-13	9
8	203	220	-17	220	-17	0

Using alpha of .1 produced forecasting errors that were larger than those using alpha = .8 for the first three forecasts. For the next two forecasts (periods 6 and 7), the forecasts using alpha = .1 produced smaller errors. Each exponential smoothing model produced the same amount of error in forecasting the value for period 8. There is no strong argument in favor of either model.

SEASONAL EFFECTS

- **Seasonal effects** are *patterns of data behavior that occur in periods of time of less than one year.*
- How can we separate out the seasonal effects?
- One of the main techniques for isolating the effects of seasonality is **decomposition**

Time series decomposition

Objective is to estimate the overall time series as a combination of long term trend and seasonality

- Additive Model
- Multiplicative Model

- The trend and seasonality can be decomposed using smoothing and regression methods
- Exponential smoothing is suitable with constant variance and no seasonality. Recommended for short-term forecast.
- Another method for stationarizing the time series is by doing transformation .e.g. Differencing

Procedure for decomposition

- Assess the trend component by smoothing or curve fitting (regression with time)
- Assess/account for seasonality i.e. deseasonalize the data
- For **Additive Adjustment**: Look at all periods of a given type (eg. First Qtr periods where data is quarterly ,or all August period where the data is monthly) & compute an average deviation of the actual values from the smooth or fitted values in those periods. The average can then be added to the trend to adjust for seasonality.

Procedure for decomposition

- For **multiplicative adjustment**: Instead of calculating the average deviation, compute an average ratio, also called seasonal indices, of the actual values to the smooth or fitted values in those periods. The indices are then used as multiplier to adjust for seasonality.

- Forecast by projecting trend component in to the future and then adding or multiplying the seasonal component, as the case may be according to your chosen model

Example

- Let us consider the following quarterly sales data for J C Penney Sales company for the last 24 quarters
- Forecast sales for the 25th Quarter ?

Period	JC Penney Sales	Period	JC Penney Sales
1	4452	13	7339
2	4507	14	7104
3	5537	15	7639
4	8157	16	9661
5	6481	17	7528
6	6420	18	7207
7	7208	19	7538
8	9509	20	9573
9	6755	21	7522
10	6483	22	7211
11	7129	23	7729
12	9072	24	9542

Solution

Step-1: Fit a trend line. We can have a Linear or a polynomial fit using Regression with t

$$\hat{y}_t = 5903.2174 + 118.75261t$$

$$\hat{y}_t = 6354.9514 + 118.75261t - 9.4274932(t - 12.5)^2$$

The forecast value of trend component for the 25th Qtr will be

$$\begin{aligned} \hat{y}_{25} &= 5903.2174 + 118.75261(25) \\ &= 8872 \end{aligned}$$

Step-2: Compute additive/multiplicative seasonal adjustment factor for the first quarter

Period, t	y_t	$\hat{y}_t$	$y_t - \hat{y}_t$	$\frac{y_t}{\hat{y}_t}$
1	4452	6022	−1570	.7393
5	6481	6497	−16	.9975
9	6755	6972	−217	.9685
13	7339	7447	−108	.9855
17	7528	7922	−394	.9503
21	7522	8397	−875	.8958
			−3180	5.5369

Then note that the average $y_t - \hat{y}_t$ is

$$\frac{-3180}{6} = -530$$

and the average $y_t/\hat{y}_t$ is

$$\frac{5.5369}{6} = .9228$$

Step-3: So, forecast for the 25th Qtr period is obtained by adjusting the trend fit for seasonality as either

$$\hat{y}_{25} = 8872 + (-530) = 8342$$

(making use of an "additive" seasonality adjustment) or as

$$\hat{y}_{25} = 8872(.9228) = 8187$$

(making use of a "multiplicative" seasonality adjustment).

Exercise (HW)

- For the same example, what will be the sales forecast for the 28th Qtr period? (Hint: calculate seasonality factor for the 4th Qtr .i.e. 4, 8, 12,16, 20, 24)

Time series data raises new technical issues

- Time lags
- Correlation over time (serial correlation, a.k.a. autocorrelation)
- Forecasting models built on regression methods like autoregressive (AR) models
- Conditions under which dynamic effects can be estimated, and how to estimate them
- Calculation of standard errors when the errors are serially correlated

Stationarity

- A common assumption in many time series techniques is that the data are stationary.

- A stationary process has the property that the mean, variance and autocorrelation structure do not change over time.

- If past effects accumulate and the values increase toward infinity, then stationarity is not met.

Independent and identically distributed (iid) noise

- simplest model for a time series is one in which there is no trend or seasonal component and in which the observations are simply independent and identically distributed (iid) random variables with zero mean.
- We refer to such a sequence of random variables X_1, X_2, . . . as iid noise $\{X_t\} \sim \mathrm{IID}(0, \sigma^2)$
- Indicates that the random variables X_t are independent and identically distributed, each with mean 0 and variance σ^2 .
- Although iid noise is a rather uninteresting process for forecasting, it plays an important role as a building block for more complicated time series models.