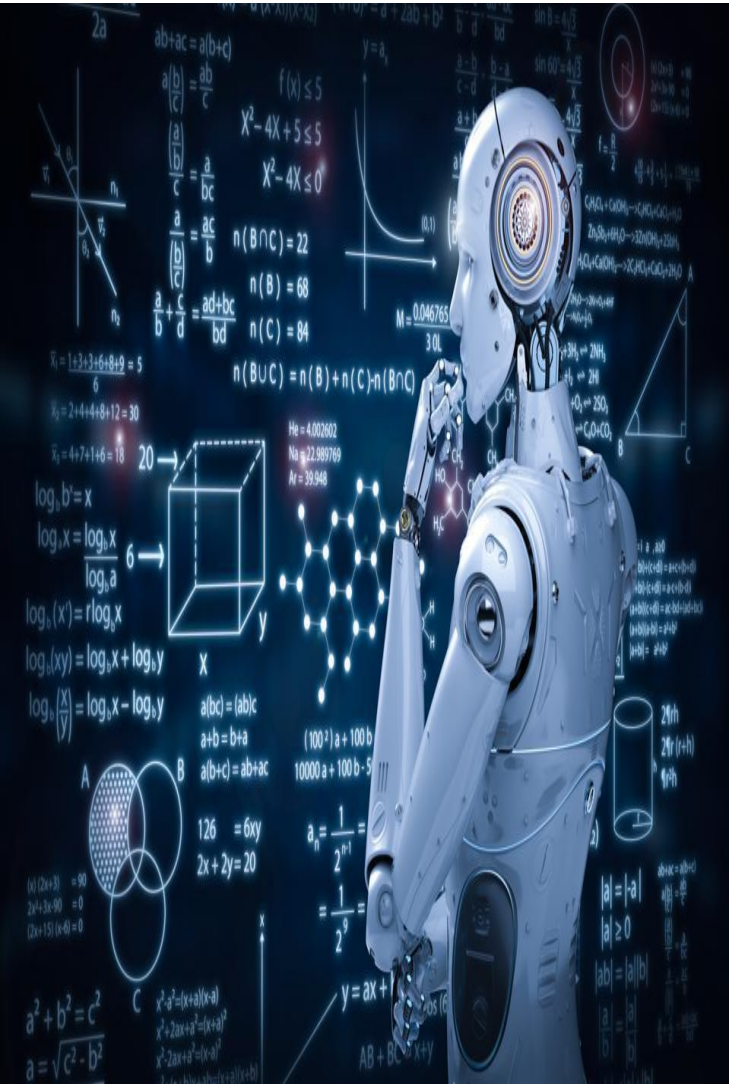


Practical Machine Learning

Day 13: SEP23 DBDA

Kiran Waghmare

Agenda

- FP Growth Algorithm
- Time Series Data Analysis

Pattern-Growth Approach: Mining Frequent Patterns Without Candidate Generation

- Bottlenecks of the Apriori approach
  - Breadth-first (i.e., level-wise) search
  - Candidate generation and test
    - Often generates a huge number of candidates
- The FPGrowth Approach (J. Han, J. Pei, and Y. Yin, SIGMOD' 00)
  - Depth-first search
  - Avoid explicit candidate generation
- Major philosophy: Grow long patterns from short ones using local frequent items only
  - "abc" is a frequent pattern
  - Get all transactions having "abc", i.e., project DB on abc: DB|abc
  - "d" is a local frequent item in DB|abc → abcd is a frequent pattern

Construct FP-tree from a Transaction Database

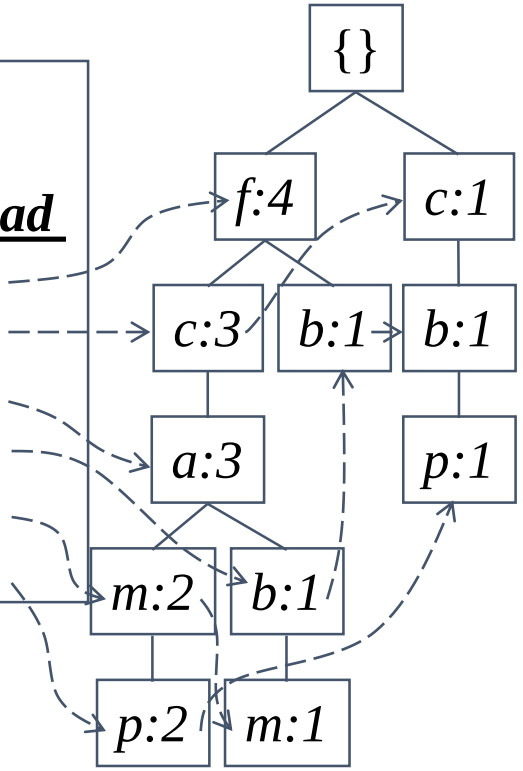
TID	Items bought	(ordered) frequent items
100	{f, a, c, d, g, i, m, p}	{f, c, a, m, p}
200	{a, b, c, f, l, m, o}	{f, c, a, b, m}
300	{b, f, h, j, o, w}	{f, b}
400	{b, c, k, s, p}	{c, b, p}
500	{a, f, c, e, l, p, m, n}	{f, c, a, m, p}

min_support = 3

1. Scan DB once, find frequent 1-itemset (single item pattern)
2. Sort frequent items in frequency descending order, f-list
3. Scan DB again, construct FP-tree

Header Table

Item	frequency	head
f	4	
c	4	
a	3	
b	3	
m	3	
p	3	



F-list = f-c-a-b-m-p

Time Series Analysis

- What is Time Series Analysis?
  - The analysis of data organized across units of time.
- Time series is a basic research design
  - Data for one or more variables is collected for many observations at different time periods
  - Usually regularly spaced
  - May be either
    - univariate - one variable description
    - multivariate - causal explanation

Importance

1. A very popular tool for Business Forecasting.
2. Basis for understanding past behavior.
3. Can forecast future activities/planning for future operations
4. Evaluate current accomplishments/evaluation of performance.
5. Facilitates comparison
6. Estimation of trade cycle

Time Series – Examples

- Stock price,
- Exchange rate, interest rate, inflation rate, national GDP
- Retail sales
- Electric power consumption
- Number of accident fatalities

Day	No. of Packets of milk sold
Monday	90
Tuesday	88
Wednesday	85
Thursday	75
Friday	72
Saturday	90
Sunday	102

Year	Population (in Million)
1921	251
1931	279
1941	319
1951	361
1961	439
1971	548
1981	685

- From example 1 it is clear that the sale of milk packets is decrease from Monday to Friday then again its start to increase.
- Same thing in example 2 the population is continuously increase.

Types of Components

1. Secular Trend or Trend
2. Seasonal Variations/Fluctuations
3. Cyclical Variations/Fluctuations
4. Irregular Variations/Movements

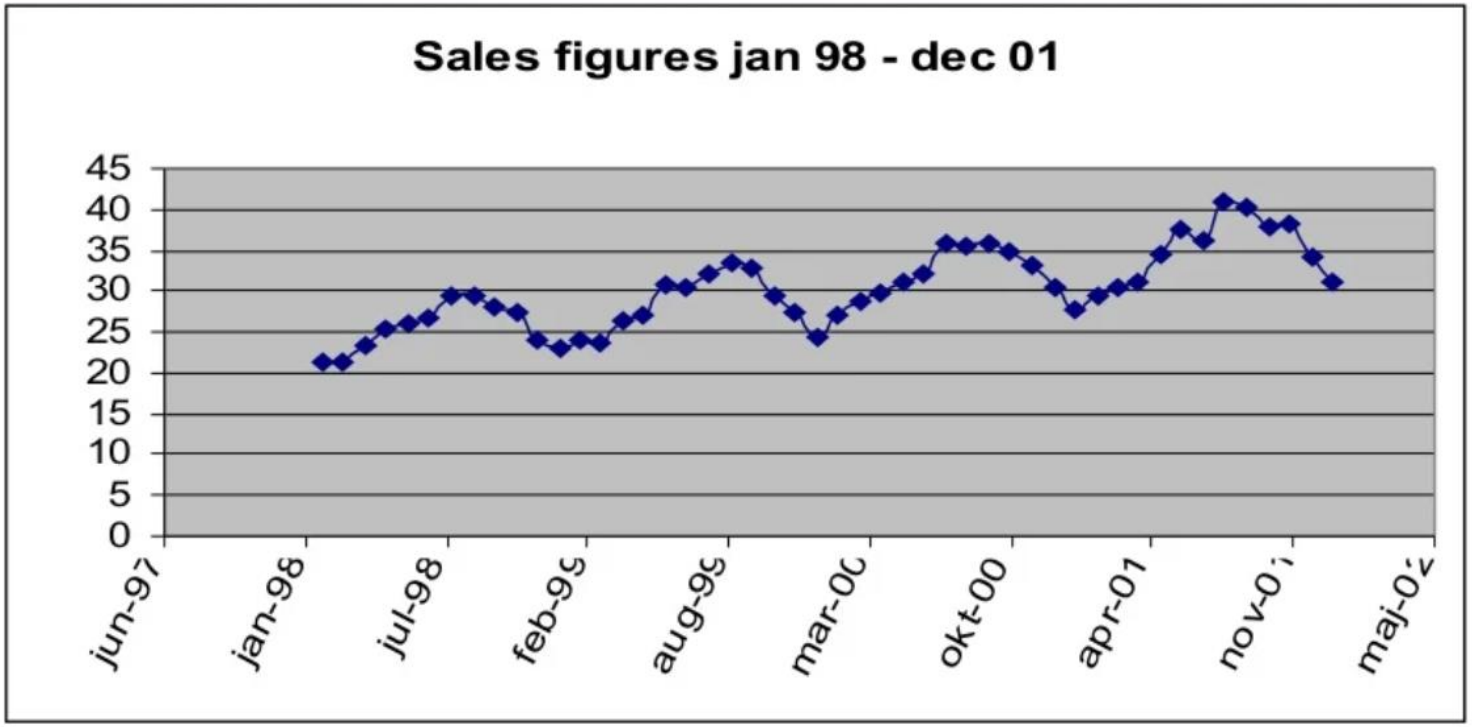
Time Series Analysis

A **Time Series** is a collection of observations made sequentially in time.

According to *Ya-lun Chou*, "A **Time Series** may be defined as a collection of readings belonging to different time periods, of some economic variables or composite of variables"

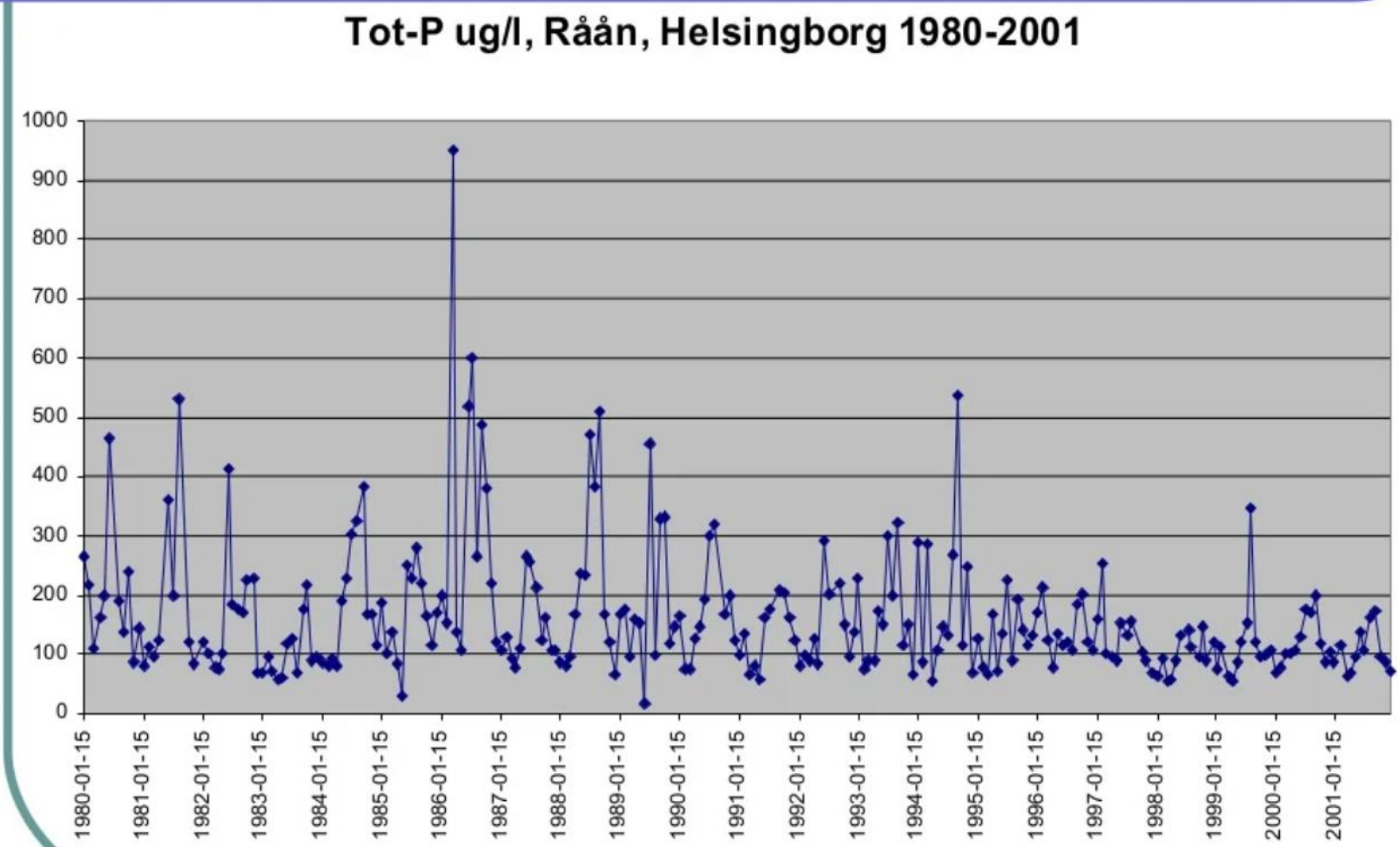
Examples: Financial time series, scientific time series, Demographic time series, Meteorological time series

Cont…



A study on random sample of 4000 graphics from 15 of the world's news papers published between 1974 and 1989 found that more than 75% of all graphics were time series.

Cont…

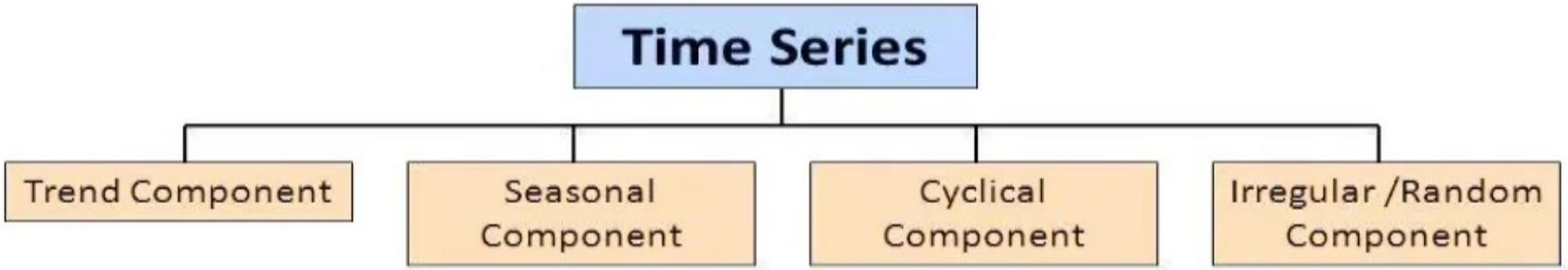

Tot-P ug/l, Råån, Helsingborg 1980-2001
1000
900
800
700
600
500
400
300
200
100
0
1980-01-15
1981-01-15
1982-01-15
1983-01-15
1984-01-15
1985-01-15
1986-01-15
1987-01-15
1988-01-15
1989-01-15
1990-01-15
1991-01-15
1992-01-15
1993-01-15
1994-01-15
1995-01-15
1996-01-15
1997-01-15
1998-01-15
1999-01-15
2000-01-15
2001-01-15

Time series components

Time series data can be broken into these four components:

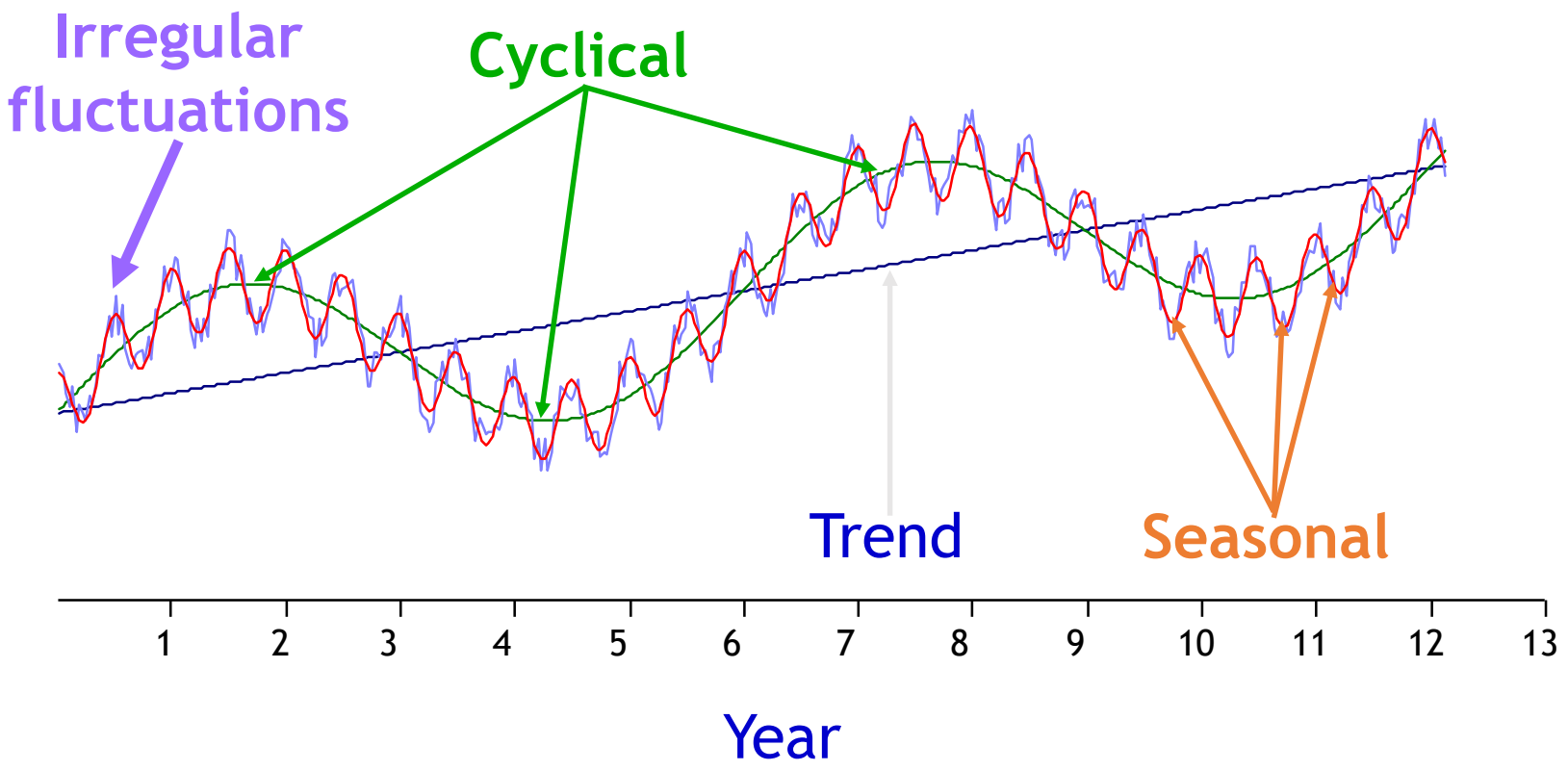
1. Secular trend
2. Seasonal variation
3. Cyclical variation
4. Irregular variation

Time-Series Components



Trend Component	Seasonal Component	Cyclical Component	Irregular /Random Component
Overall, persistent, long-term movement	Regular periodic fluctuations, usually within a 12-month period	Repeating swings or movements over more than one year	Erratic or residual fluctuations

Components of Time-Series Data



Predicting long term trends without smoothing?
What could go wrong?
Where do you commence your prediction from the bottom of a variation going up or the peak of a variation going down……..

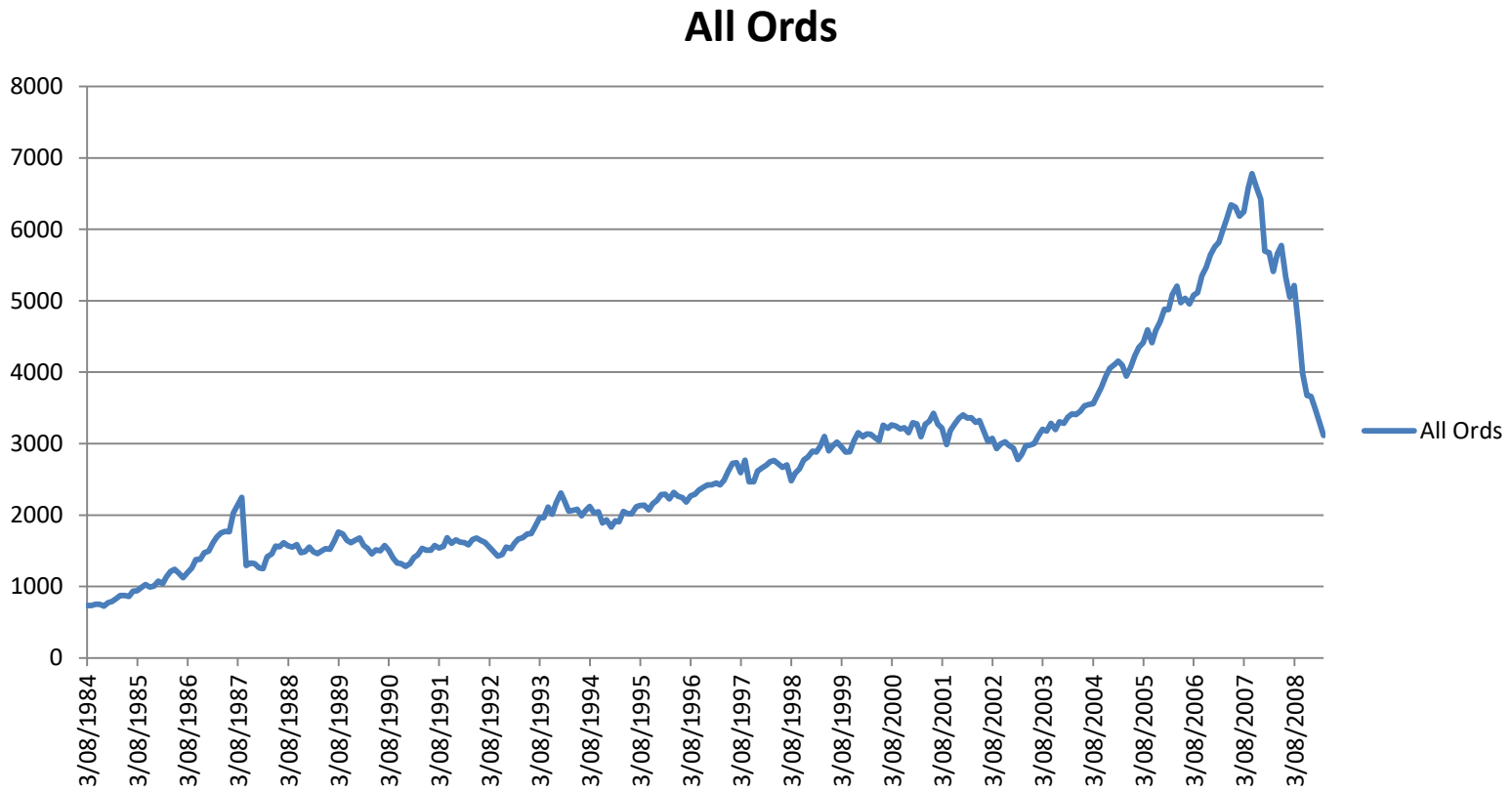
1. Secular Trend

This is the long term growth or decline of the series.

- In economic terms, long term may mean >10 years
- Describes the history of the time series
- Uses past trends to make prediction about the future
- Where the analyst can isolate the effect of a secular trend, changes due to other causes become clearer

Secular Trend

A secular trend identifies the underlying trend (direction) of the data: – increasing, decreasing or remaining constant. It is the long term direction of the data, usually described by the "line of best fit". And is deduced over a large number of periods. The following chart is a long term graph of the ASX200.



2. Seasonal Variation

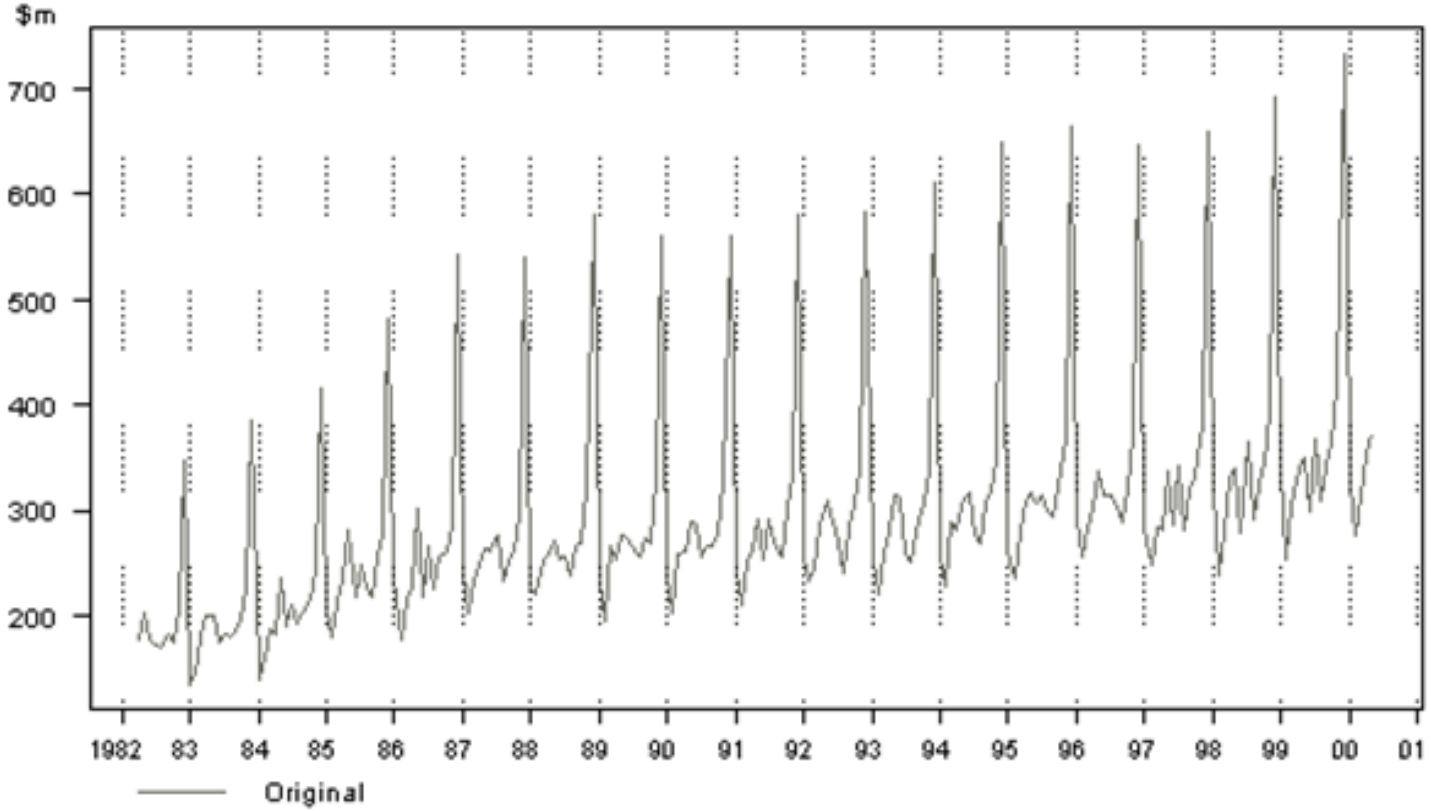
The seasonal variation of a time series is a pattern of change that **recurs** regularly over time.

Seasonal variations are usually due to the differences between seasons and to festive occasions such as Easter and Christmas.

Examples include:

- Air conditioner sales in Summer
- Heater sales in Winter
- Flu cases in Winter
- Airline tickets for flights during school vacations

Monthly Retail Sales in NSW Retail Department Stores



3. Cyclical variation

Cyclical variations also have recurring patterns but with a longer and more **erratic time scale** compared to Seasonal variations.

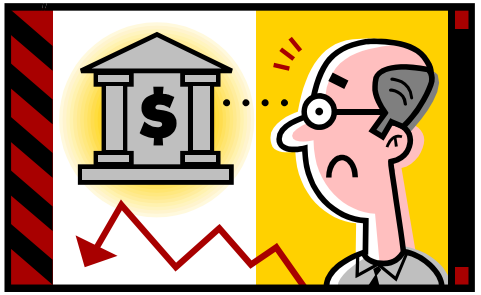
The name is quite misleading because these cycles can be far from regular and it is usually impossible to predict just how long periods of expansion or contraction will be.

There is no guarantee of a regularly returning pattern.

Cyclical variation

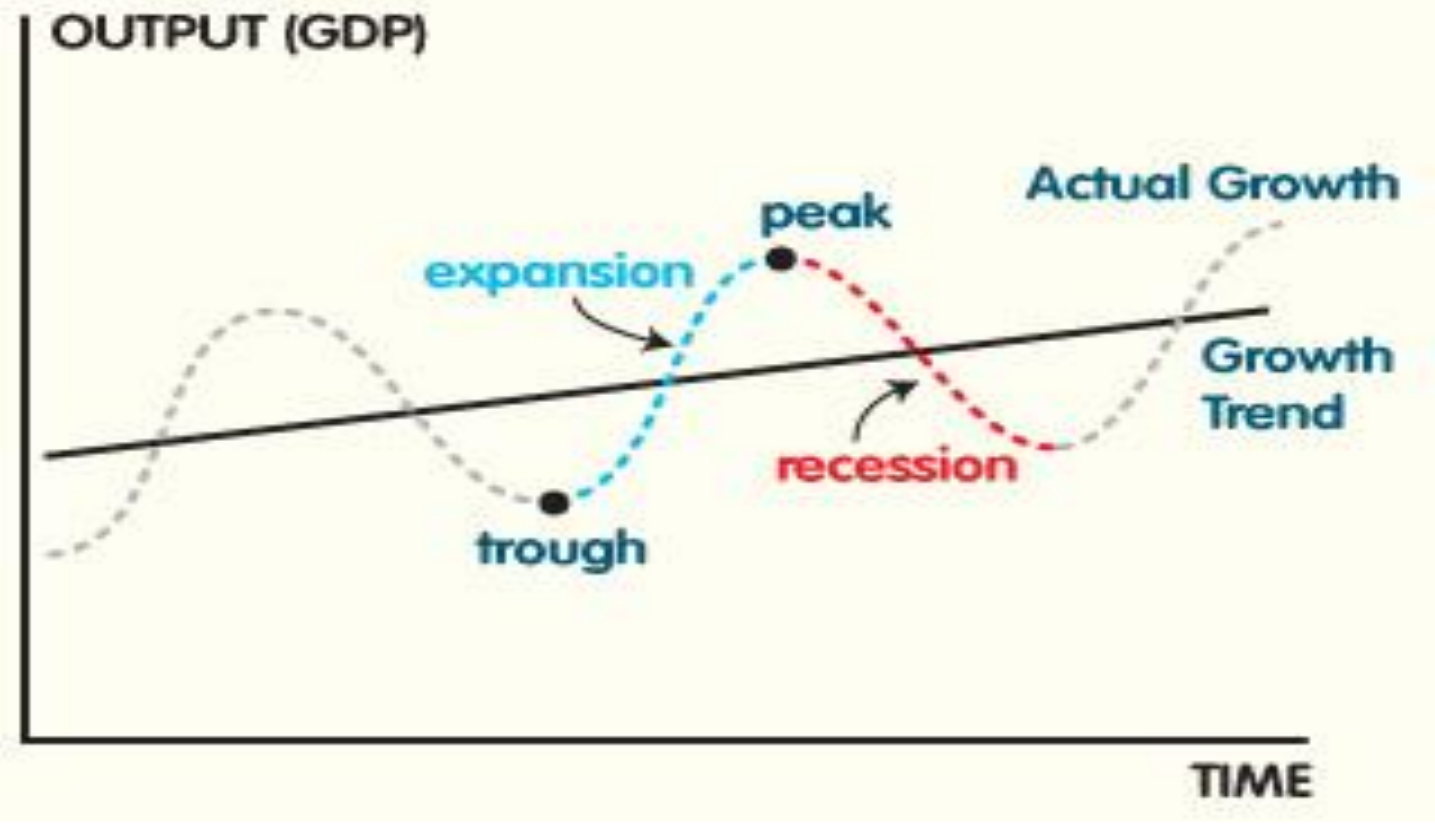
Example include:

- Floods
- Wars
- Changes in interest rates
- Economic depressions or recessions
- Changes in consumer spending



Cyclical variation

This chart represents an economic cycle, but we know it doesn't always go like this. The timing and length of each phase is not predictable.



4. Irregular variation

An irregular (or random) variation in a time series occurs over varying (usually short) periods.

It follows no pattern and is by nature unpredictable.

It usually occurs randomly and may be linked to events that also occur randomly.

Irregular variation cannot be explained mathematically.

Irregular variation

If the variation cannot be accounted for by secular trend, season or cyclical variation, then it is usually attributed to irregular variation. Example include:

- Sudden changes in interest rates
- Collapse of companies
- Natural disasters
- Sudden shift s in government policy
- Dramatic changes to the stock market
- Effect of Middle East unrest on petrol prices

Monthly Value of Building Approvals ACT)