BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Introduction to Statistical Methods

ISM Team

Sessions 13, 14 & 15
Time Series Analysis
(26th August, 2nd & 9th Sep 2023)

Time Series

Time Series ⇨ Definition

A time series is defined as a set of observations on a variable generated sequentially in time. The measurement of the variable may be made continuously or at discrete (equally spaced) intervals.

Often a variable continuous in time is measured at regular intervals and this produces a discrete series of data.

Time Series → Definition

Time series may be represented as a set of data (Y_1, Y_2, Y_3, …, Y_n), Y_t denoting the value of the variable Y at time t.

Example of continuous variable:

-Temperature on a chemical reactor

- Level of tide at a particular site

- Amplitude of an electrical signal

Time Series → Definition

A discrete time series data may be generated from the accumulation of data over a period of time.

Example of discrete variable:

- Monthly sales
- Daily rainfall
- Production of a crop over different years in a country

innovate achieve lead

Time Series ➡ Graphs

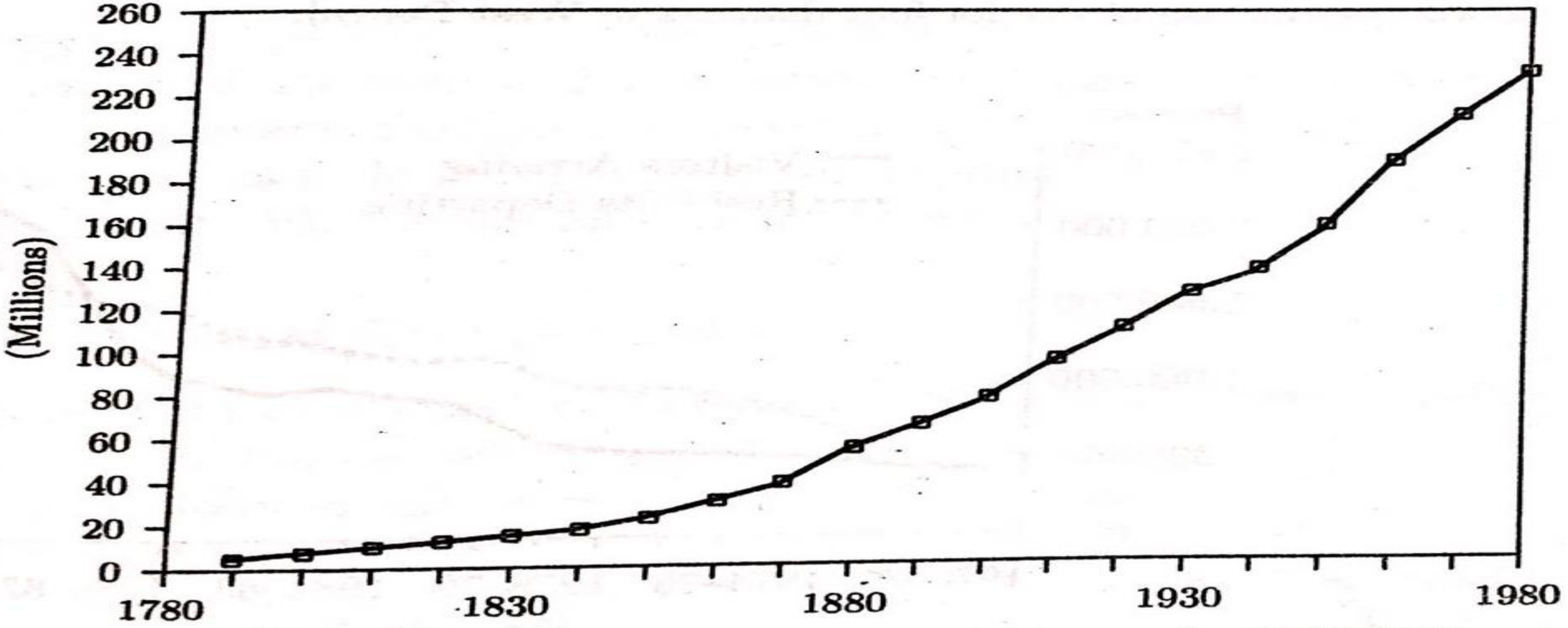


Fig 14.1 U.S. Population at ten year intervals, 1790-1980.
[*Source*. Brockwell and Davis, 1990.]

Time Series ⇨ Graphs

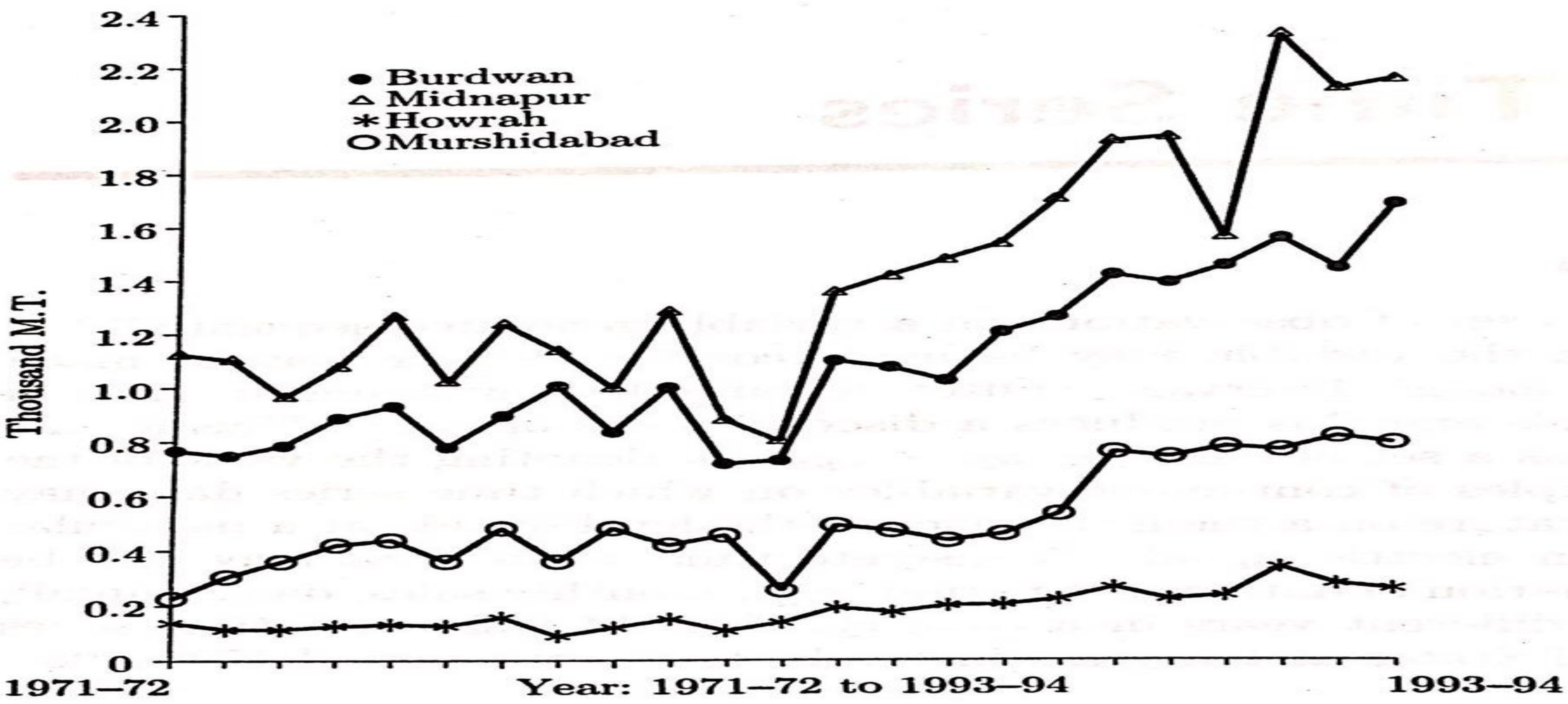


Fig 14.2 District-wise production of rice for four districts in West Bengal.

Time Series → Graphs

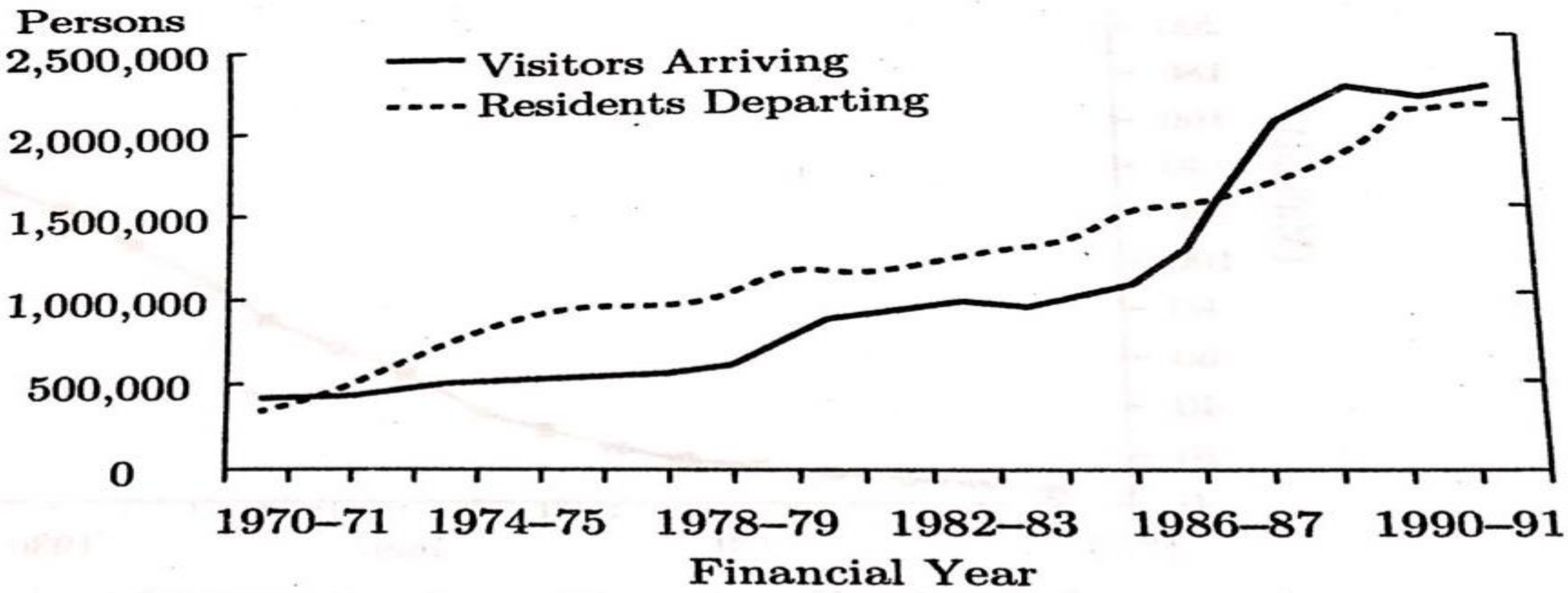


Fig 14.3 Short-term movement to and from Australia.
[*Source*: Immigration update, December, Quarter 1991, Bureau of Immigration Research, Statistics Section, Australia.]

Statistical Methods for Data Science

Time Series → Graphs

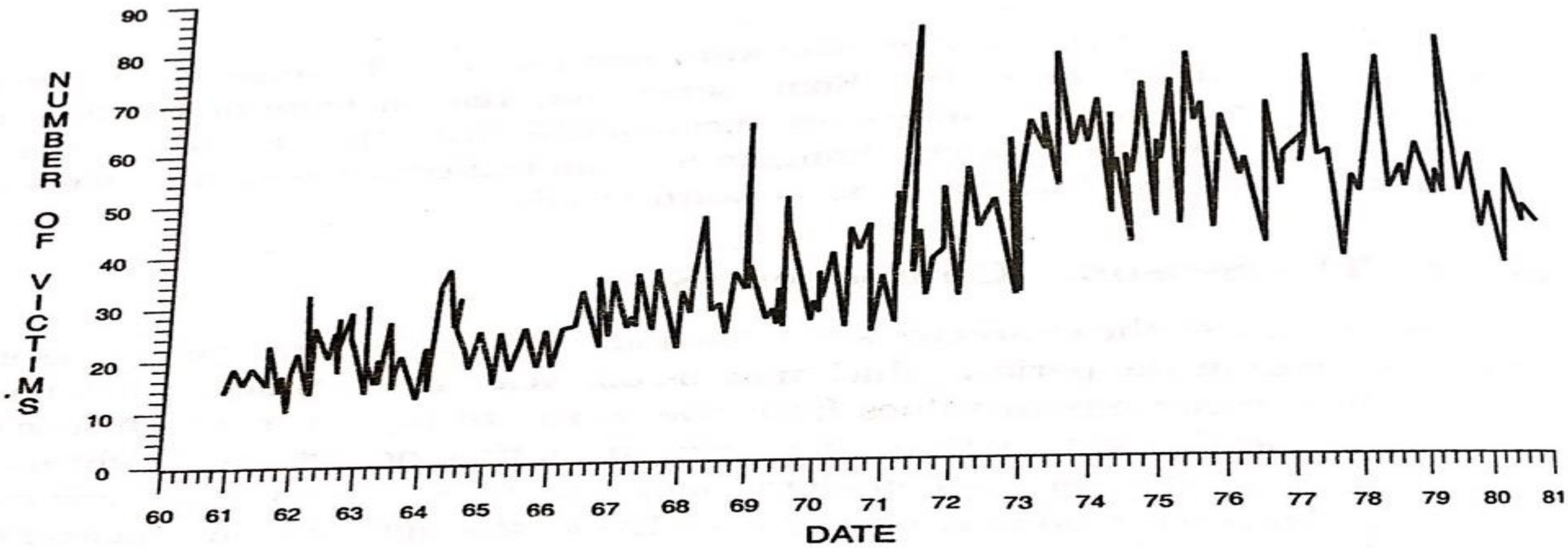


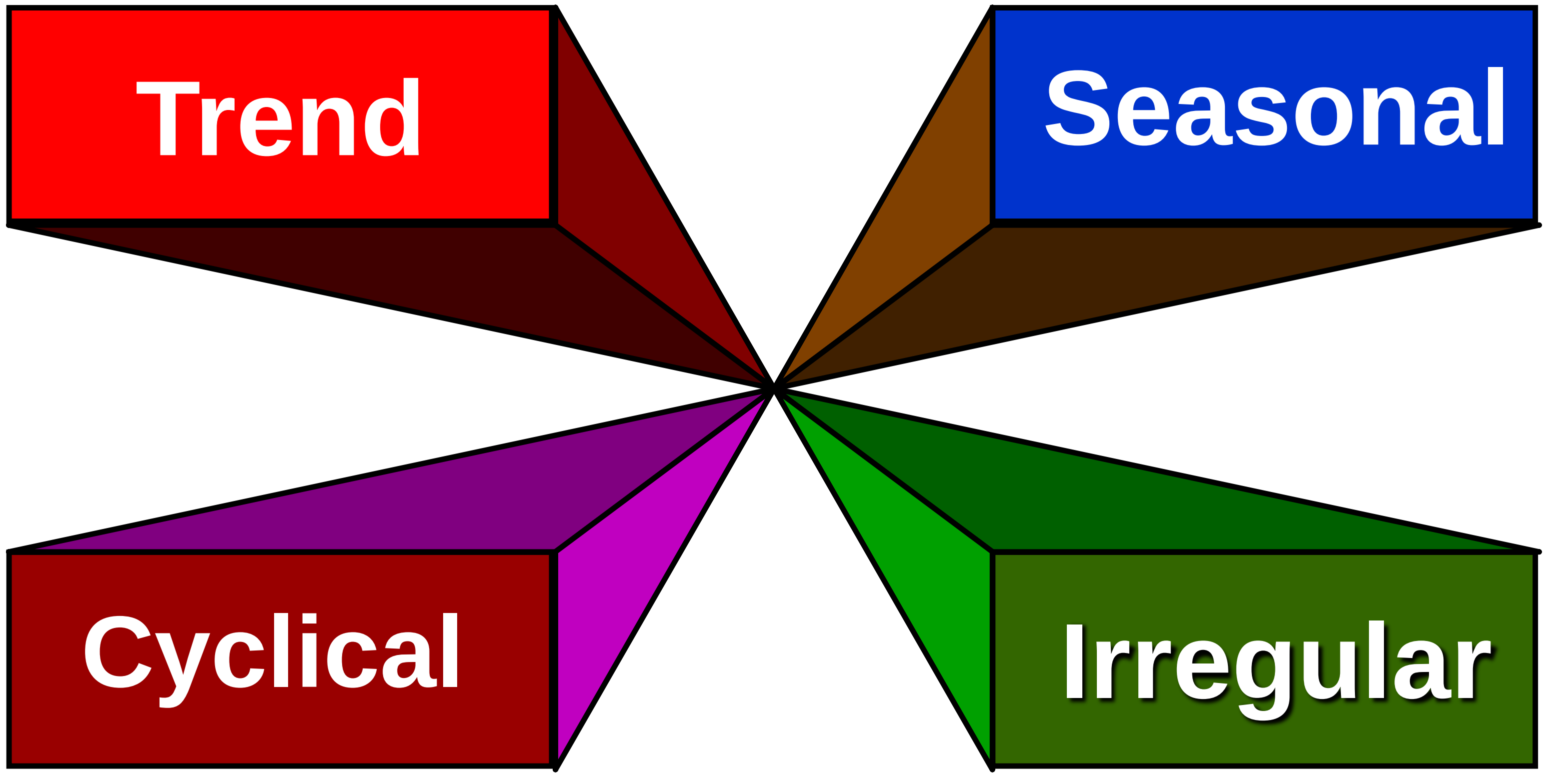
Fig 14.5 Frequency of homicide victims by month, 1961-1980 (excluding manslaughters and infanticides).
[*Source*: McLeod, MacNeil and Bhattacharyya (1985).]

- **Time Series Models:**
  - A **time series** is a set of observations Y_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*** *time series* are obtained when observations are recorded continuously over some time interval, e.g., when T_0 [0, 1].

Time Series ➡ Decomposition Models

This is an important technique for all types of time series analysis, especially for seasonal adjustment. It seeks to construct, from an observed time series, a number of component series (that could be used to reconstruct the original by additions or multiplications) where each of these has a certain characteristic or type of behaviour. For example, time series are usually decomposed into:

Time Series → Decomposition of Time Series as Components



Time Series ➡ Decomposition view

- 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 a 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.
    - 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 into the future and then adding or multiplying the seasonal component, as the case may be according to your chosen model

T_t: the trend component at time *t*, which reflects the long-term progression of the series (secular variation). A trend exists when there is a persistent increasing or decreasing direction in the data. **The trend component does not have to be linear**

Trend:

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

Time Series ➡ Seasonal variation

S_t: the seasonal component at time *t*, reflecting seasonality (seasonal variation). A seasonal pattern exists when a time series is influenced by seasonal factors. Seasonality occurs over a fixed and known period (e.g., the quarter of the year, the month, or day of the week)

Time Series → 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 Hindu, Christian or Muslim festivals.

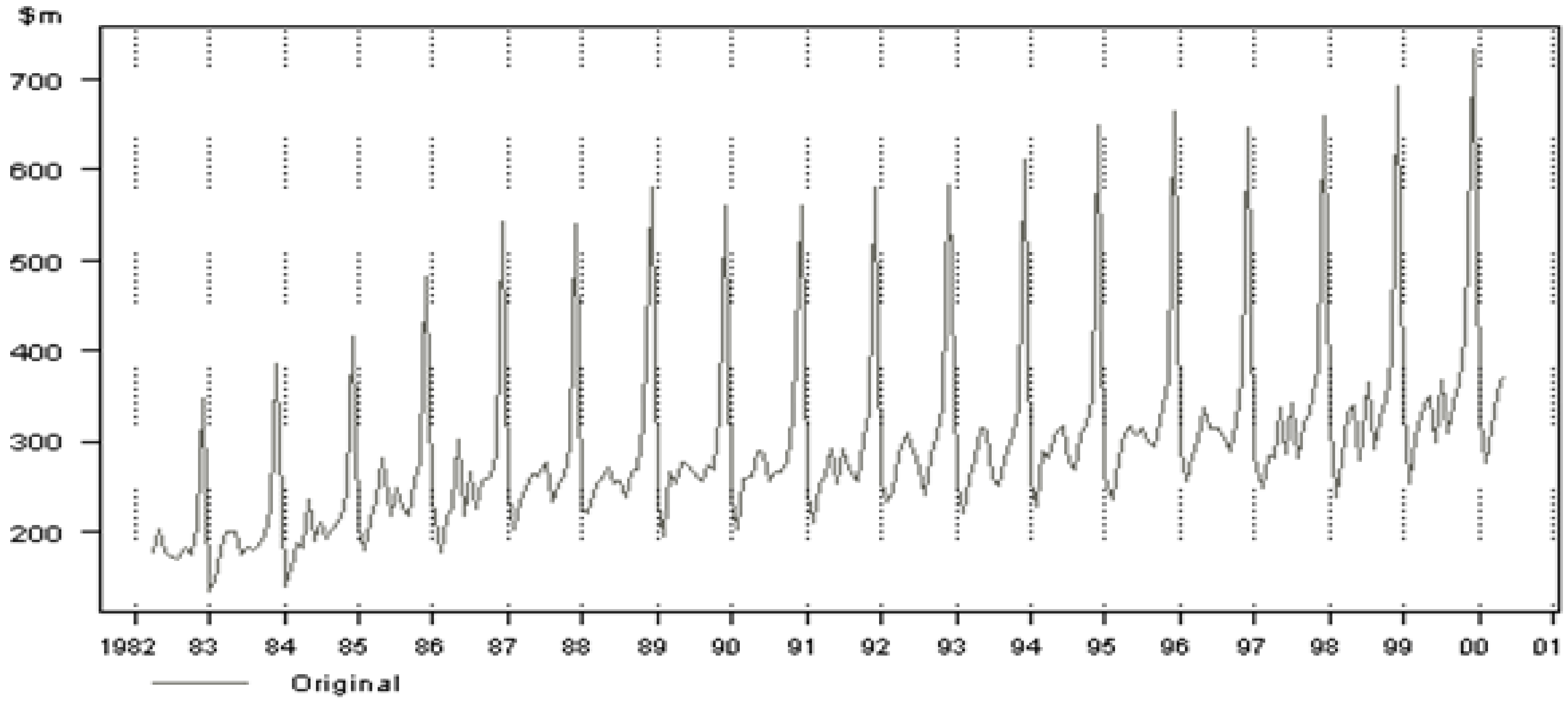
Examples include:

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

Time Series ➡ Seasonal variation

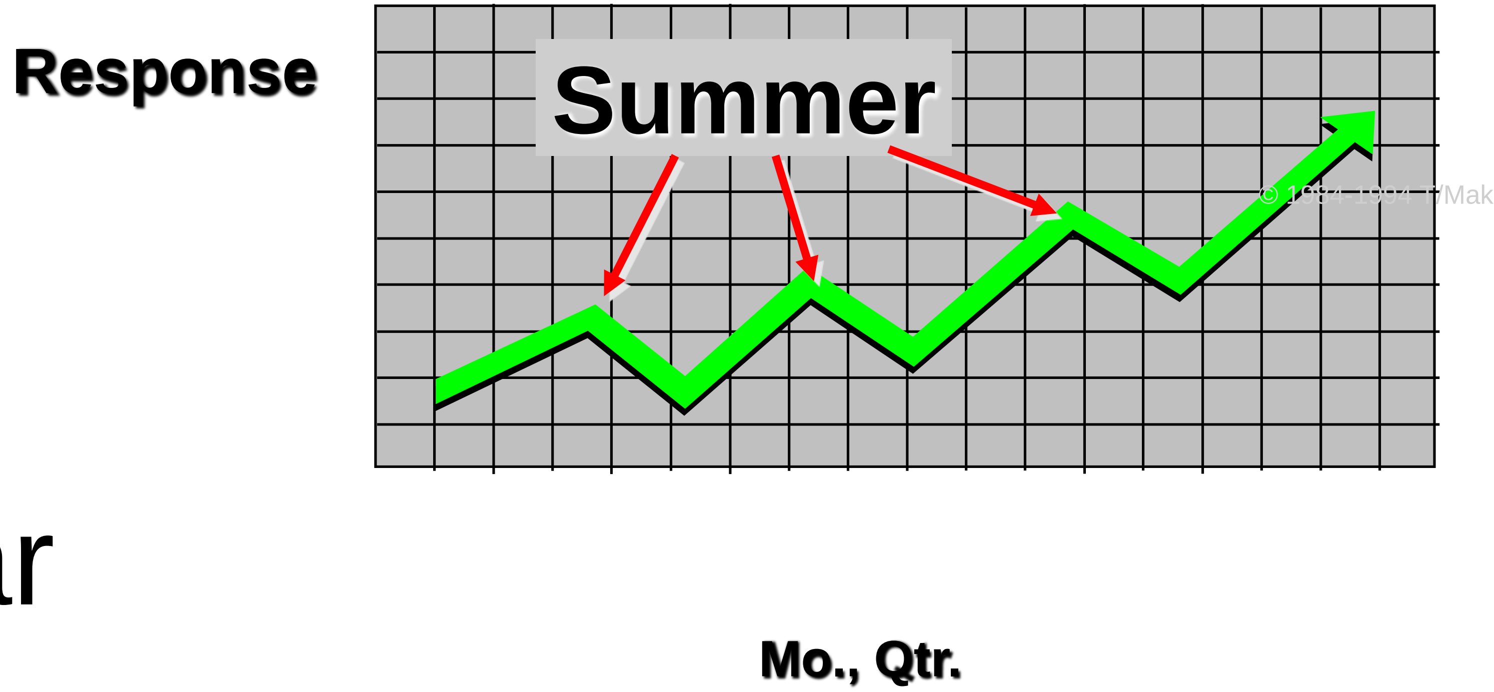
Seasonality in a time series can be identified by **regularly spaced peaks and troughs** which have a consistent direction and approximately the same magnitude every year, relative to the trend. The following diagram depicts a strongly seasonal series. There is an obvious large seasonal increase in December retail sales in New South Wales due to Christmas shopping. In this example, the magnitude of the seasonal component increases over time, as does the trend.

Time Series ➡ Seasonal variation



Time Series ⇨ Time Series Components

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



innovate achieve lead

Time Series ➡ Seasonal variation

Seasonal Movement

Seasonal movement refers to regular periodic fluctuations that occur in each time period – yearly, monthly, daily. Some examples are speciality cards for Valentine's Day, monthly travel passes and off-peak heating.

Seasonal variations greatly impact on the outcomes of recorded data and often belie the underlying trend. Businesses need to identify the seasonal impact:

- So that a measurement (index) can be used to adjust the expected outcome.

- In order to recognise the direction of the underlying trend.

Time Series → Cyclical variation

C_t: the cyclical component at time *t*, which reflects repeated but non-periodic fluctuations. The duration of these fluctuations is usually of at least two years

Time Series → 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.

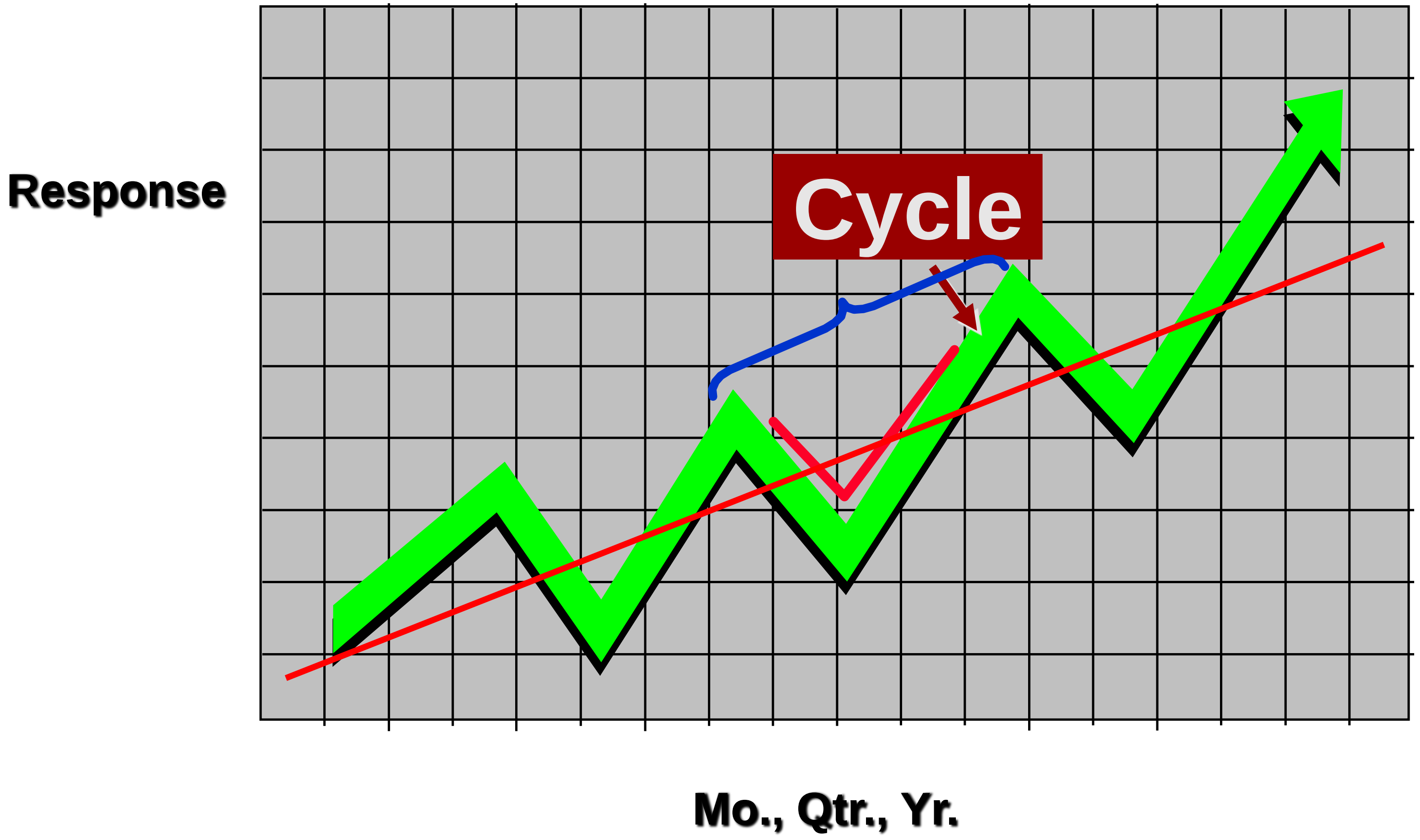
Time Series → Cyclical variation

Example include:

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

Cyclical:

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



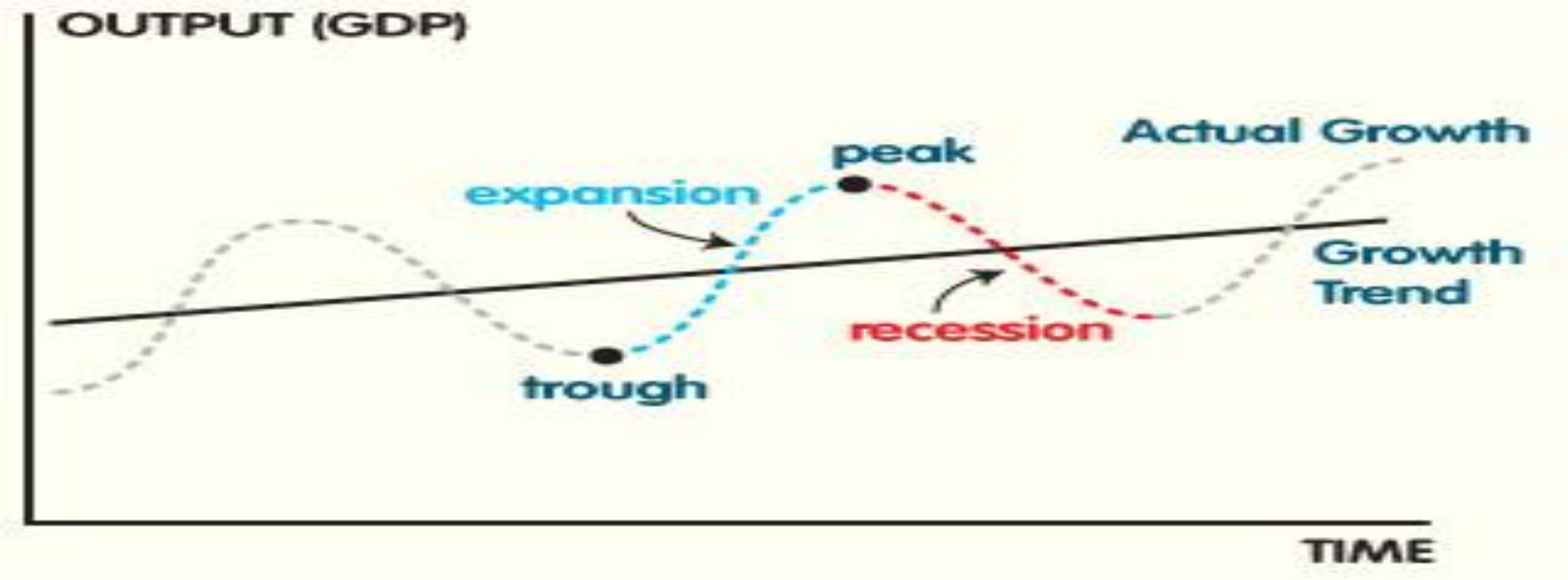
Time Series → Cyclical variation

Cyclical Movement

This reflects the level of business activity and economic movement over time by fluctuating patterns, known as the economic cycle. These variations measure periods of expansion and contraction in industry and the economy. Their regularity and intensity are not predictable, however certain economic indicators contribute to their existence – level of investment, confidence in the economy, GDP, trade indexes and government policy.

Time Series ⇨ 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.



Time Series → Irregular variation

I_t: the irregular component (or "noise") at time *t*, which describes random, irregular influences. It represents the residuals or remainder of the time series after the other components have been removed.

Time Series ➡ 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.

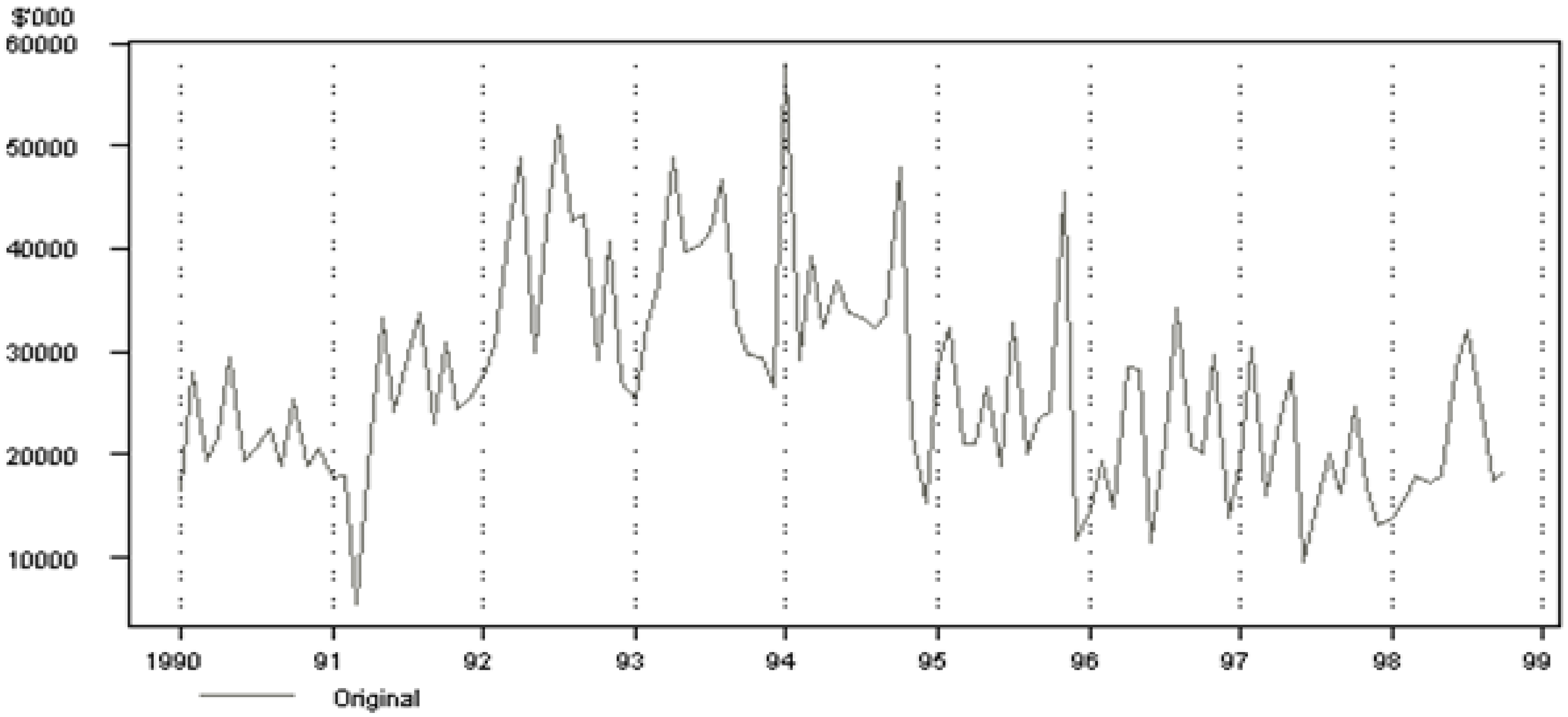
Irregular variation cannot be explained mathematically.

Time Series ⇨ Irregular variation

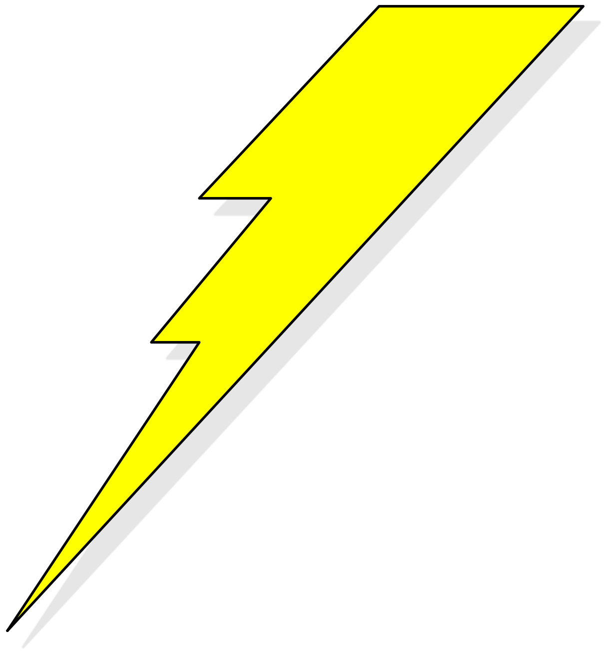
If the variation cannot be accounted for by secular trend, season or cyclical variation, then it is usually attributed to irregular variation. Example include:

- –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.

Time Series ➡ Irregular variation



Irregular:



- Erratic, unsystematic, ‘residual’ fluctuations
- Due to random variation or unforeseen events like Union strike, War
- Short duration & nonrepeating
- Rapid changes or bleeps in the data caused by short term unanticipated and non-recurring factors. Irregular fluctuations can happen as often as day to day

Time Series ➡ Irregular variation

The irregular component results from short term fluctuations in the series which are neither systematic nor predictable. In a highly irregular series, these fluctuations can dominate movements, which will mask the trend and seasonality. The graph above is of a highly irregular time series.

http://www.youtube.com/watch?v=ca0rDWo7IpI

good you tube video

Time Series ➡ Irregular variation

Irregular Movements

These patterns refer to random variations that impact greatly on the level of business activity, often called natural variation. Some examples are extreme weather patterns (flood, fire, cyclone), extreme business variation (stock market crash, drop in $A), political climate (sudden elections, wars, death of a leader), and industry changes (pilot strikes, waterside strikes.). The resulting patterns will exert a great pressure on the predicted underlying trends and for this reason must be accounted for when planning for the future. However, the irregular movements are unpredictable.

Hence a time series using an additive model can be thought of as:

$$y_t = T_t + C_t + S_t + I_t$$

Where as the multiplicative model can be thought of as:

$$y_t = T_t \times C_t \times S_t \times I_t$$

Time Series ⇨ Decomposition Models

An additive model would be used when the variations around the trend do not vary with the level of the time series.

Multiplicative model would be appropriate if the trend is proportional to the level of the time series.

Time Series ➡ Decomposition of Time Series Components

Sometimes the trend and cyclical components are grouped into one, called the trend-cycle component. The trend-cycle component can just be referred to as the "trend" component, even though it may contain cyclical behaviour. For example, a seasonal decomposition of time series plot decomposes a time series into seasonal, trend and irregular components using loess and plots the components separately, whereby the cyclical component (if present in the data) is included in the "trend" component plot.

Time Series ➔ 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 commonplace example might be estimation of some variable of interest at some specified future date.

Time Series → Applications of Forecasting

Forecasting is required in many situations:

(a) deciding whether to build another power generation plant in the next five years requires forecasts of future demand

(b) scheduling staff in a call centre next week requires forecasts of call volumes

(c) stocking an inventory requires forecasts of stock requirements.

Time Series ➡ Applications of Forecasting

- Forecasts can be required several years in advance

Example – in case of capital investments

- only a few minutes beforehand

  Example - telecommunication routing

Whatever the circumstances or time horizons involved, forecasting is an important aid to effective and efficient planning.

Time Series ⇨ Definition of 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 iPhone sale cannot be forecast with any accuracy.

Time Series ➡ Definition of Forecasting

- 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.

Time Series ⇨ Applications of Forecasting

- Forecasts of electricity demand can be highly accurate because all three conditions are usually satisfied.
- 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.

Forecasting:

is about predicting the future as accurately as possible, given all of the information available, including historical data and knowledge of any future events that might impact the forecasts

Time Series ➡ Forecasting, Planning and Goals

Planning:

is a response to forecasts and goals. Planning involves determining the appropriate actions that are required to make your forecasts match your goals.

Goals:

are what you would like to have happen. Goals should be linked to forecasts and plans, but this does not always occur. Too often, goals are set without any plan for how to achieve them, and no forecasts for whether they are realistic.

Short-term Forecasts

Short-term forecasts generally involve some form of scheduling which may include for example the seasons of the year for planning purposes.

The cyclical and seasonal factors are more important in these situations.

Short-term Forecasts

Such forecasts are usually prepared every 6 months or on a more frequent basis.

These forecasts are needed for the scheduling of personnel, production and transportation. As part of the scheduling process, forecasts of demand are often also required.

Medium-term Forecasts

Medium-term forecasts are generally prepared for planning, scheduling, budgeting and resource requirements purposes.

Medium-term forecasts are needed to determine future resource requirements, in order to purchase raw materials, hire personnel, or buy machinery and equipment.

Medium-term Forecasts

The trend factor, as well as the cyclical component, plays a key role in the medium-term forecast as the year to year variations in traffic growth are an important element in the planning process

Long-term Forecasts

Long-term forecasts are used mostly in connection with strategic planning to determine the level and direction of capital expenditures and to decide on ways in which goals can be accomplished.

The trend element generally dominates long term situations and must be considered in the determination of any long-run decisions.

Long-term Forecasts

The methods generally found to be most appropriate in long-term situations are econometric analysis and life-cycle analysis. Long-term forecasts are used in strategic planning. Such decisions must take account of market opportunities, environmental factors and internal resources.

Long-term Forecasts

It is also important that since the time span of the forecast horizon is long, forecasts should be calibrated and revised at periodic intervals (every two or three years depending on the situation).

Time Series ➡ The basic steps in a forecasting task

Step 1: Problem definition

Step 2: Gathering information

Step 3: Preliminary (exploratory) analysis

Step 4: Choosing and fitting models

Step 5: Using and evaluating a forecasting model

Time Series ➔ Types of forecasting methods

Forecasting Methods

- **Qualitative**
  - Personal Opinion or Judgment
  - Panel consensus
  - Delphi Method
  - Market Research
- **Quantitative**
  - **Time Series**
    - Smoothing methods
    - Exponential smoothing
    - Trend projection method
  - **Causal**
    - Regression Trend Analysis

Time Series ⇨ Qualitative forecasting

- **Personal Opinion**
  - Individuals forecasts future based on their own judgment or opinion without any formal model
  - Such assessments are relatively reliable and accurate

Time Series ⇨ Qualitative forecasting

- Panel consensus
  - Reduces the prejudice and ignorance that may arise in the individual judgment
  - Possible to develop consensus among group of individuals
  - Panel of individuals are encouraged to share information, opinions, and assumptions, if any, to predict future value of some variable under study

Delphi Method

The Delphi method was invented by Olaf Helmer and Norman Dalkey of the Rand Corporation in the 1950s for the purpose of addressing a specific military problem. The method relies on the key assumption that forecasts from a group are generally more accurate than those from individuals.

Delphi Method

The aim of the Delphi method is to construct consensus forecasts from a group of experts in a structured iterative manner. A facilitator is appointed in order to implement and manage the process. The Delphi method generally involves the following stages:

Delphi Method

- A panel of experts is assembled.
- Forecasting tasks/challenges are set and distributed to the experts.
- Experts return initial forecasts and justifications. These are compiled and summarised in order to provide feedback.

Delphi Method

- Feedback is provided to the experts, who now review their forecasts in light of the feedback. This step may be iterated until a satisfactory level of consensus is reached.
- Final forecasts are constructed by aggregating the experts' forecasts.

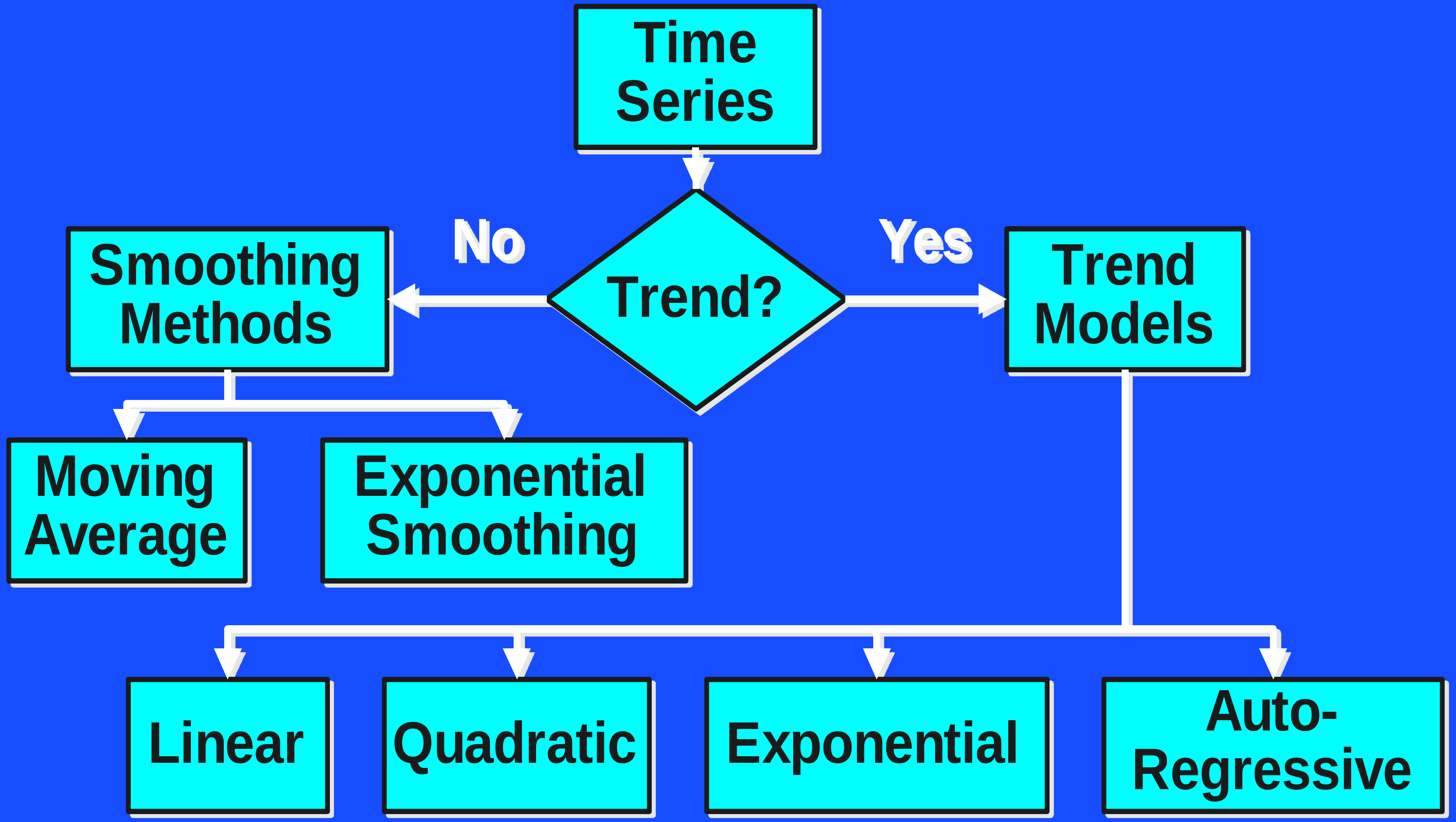
Delphi Method

- Each stage of the Delphi method comes with its own challenges. In what follows, we provide some suggestions and discussions about each one of these.

- **Market Research Method**
  - This method is used to collect based on well-defined objectives and assumptions about future value of variable.
  - Questionnaire is used to gather data and prepared summary of responses.
  - This method produces a narrow range of forecasts rather than a single view of future.

Time Series ➡ Qualitative forecasting

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, technological	Time consuming to develop



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

Modeling Time Series

Methods for forecasting a time series fall into two general classes: *smoothing methods* and *regression-based modeling methods*.

Although the smoothing methods do not explicitly use the time series components, it is a good idea to keep them in mind. The regression models explicitly estimate the components as a basis for building models.

Time Series ➡ Fitting a trend to Time Series

A linear trend between time series values Y_t and the time t is given by the equation $Y_t = a + bt + e$.

For the pair of values of (Y_t, t), a & b as estimated by the principles of least squares are

$$\hat{b} = \frac{\sum xY_t}{\sum x^2}$$

and

$$\hat{a} = \frac{\sum Y_t}{n}$$

$$where\ for\ odd\ numbers$$

$$x = \frac{t - middle\ value\ of\ t}{interval\ (h)}$$

and for even numbers

$$x = \frac{t - (mean\ of\ middle\ two\ values\ of\ t)}{\frac{1}{2}\, interval\ (h)}$$

Time Series → Fitting a trend to Time Series

The following data relates to production of crops (in tons) in a place from 2016 -2020. Fit a linear trend $Y_t = a + bt + e$ for the data given below:

	Year (t)	Production (tons) (Y_t)	x	x^2	xY_t
	2016	35	-2	4	-70
	2017	55	-1	1	-55
	2018	79	0	0	0
	2019	80	1	1	80
	2020	40	2	4	80
Sum	10090	289	0	10	35

$$\boldsymbol{x = \frac{t - 2018}{1} = t - 2018}$$

$$\hat{b} = \frac{\sum xY_t}{\sum x^2} = \frac{35}{10} = 3.5$$

$$\hat{a} = \frac{\sum Y_t}{n} = \frac{289}{5} = 57.8$$

$$\widehat{Y_t} = \hat{a} + \hat{b}\,x = 57.8 + 3.5x$$

$$The\ forecasted\ value Y_t$$
$$for\ the\ year\ 2021\ is$$

$$\widehat{Y_t} = \hat{a} + \hat{b}\,x = 57.8 + 3.5\,(3) = 68.3$$

Time Series ➡ Fitting a trend to Time Series

The following data relates to Sale (in '000) in a Mall from 2016 -2021. Fit a linear trend Y_t = a + bt +e for the data given below:

	Year (t)	Sales ('000) (Y_t)	x	x^2	xY_t
	2016	23	-5	25	-115
	2017	27	-3	9	-81
	2018	31	-1	1	-31
	2019	34	1	1	34
	2020	22	3	9	66
	2021	29	5	25	145
Sum	**12111**	**166**	**0**	**70**	**18**

$$\hat{b} = \frac{\sum xY_t}{\sum x^2} = \frac{18}{70} = 0.2571$$

$$\hat{a} = \frac{\sum Y_t}{n} = \frac{12111}{6} = 27.6667$$

$$\widehat{Y_t} = \hat{a} + \hat{b}\, x = 27.6667 + 0.2571x$$

$$The\ forecasted\ value Y_t$$
$$for\ the\ year\ 2022\ is$$

$$\boldsymbol{x = \frac{t - (2018 + 2019)/2}{1/2} = 2(t - 2018.5)}$$

$$\boldsymbol{\widehat{Y_t} = \hat{a} + \hat{b}\, x = 27.6667 + 0.2571\,(7) = 29.4664}$$

innovate achieve lead

Time Series ➡ Fitting a trend to Time Series

A trend need not be a linear, it may even be quadratic. Let the quadratic trend between time series values Y_t and the time t is given by the equation

$Y_t = a + bt + ct^2 + e.$

For the pair of values of (Y_t, t), a, b and c as estimated by the principles of least squares is given by the system of three normal equations as follows:

$$\sum Y_t = na + b\sum x + c\sum x^2$$

$$\sum xY_t = a\sum x + b\sum x^2 + c\sum x^3$$

$$\sum x^2Y_t = a\sum x^2 + b\sum x^3 + c\sum x^4$$

As there are three equations in three unknown quantities, any of the numerical techniques for solving the system of equations can be made use like Gauss – Elimination method.

Note: x is defined as in case of fitting linear trend

Time Series → Smoothing Techniques – Moving Averages

- Appropriate for a time series with a horizontal pattern ie., the data that are stationary.
  - Moving Average (the average of the most recent k data values forms the forecast for the next period)

$$F_{t+1} = \frac{\sum \text{most recent k data values}}{k}$$

$$F_{t+1} = \frac{Y_t + Y_{t-1} + \ldots + Y_{t-k+1}}{k}$$

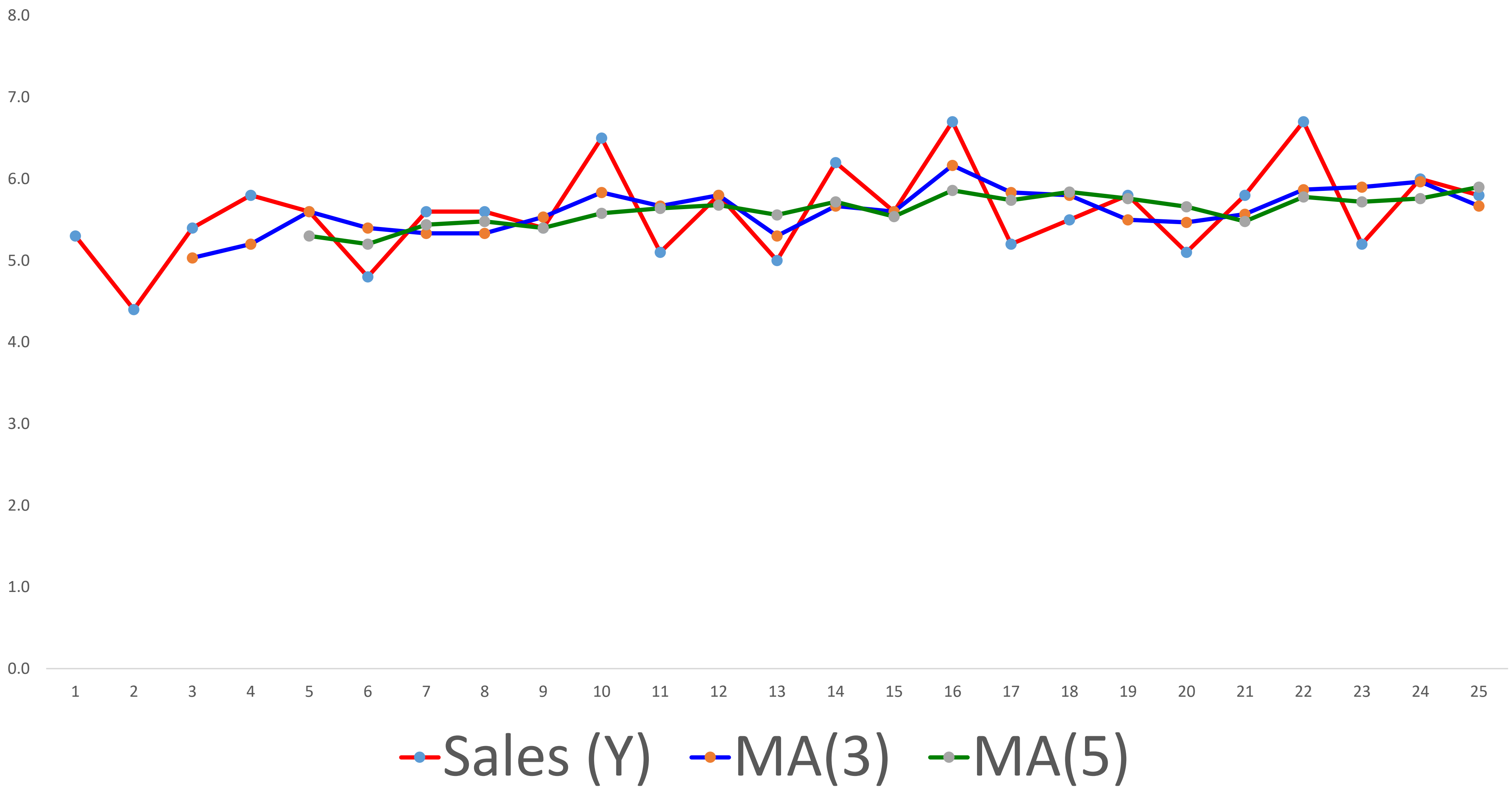
Statistical Methods for Data Science

innovate achieve lead

Time Series ➡ Simple Moving Average

Following data shows production volume (in ‘000 tones). Compute 3 and 5-year moving average

Period (t)	Sales (Y_t)	3yr MA sum	3yr MA sum	5yr MA sum	MA sum
1	5.3				
2	4.4	15.1	5.03		
3	5.4	15.6	5.20	26.5	5.30
4	5.8	16.8	5.60	26	5.20
5	5.6	16.2	5.40	27.2	5.44
6	4.8	16	5.33	27.4	5.48
7	5.6	16	5.33	27	5.40
8	5.6	16.6	5.53	27.9	5.58
9	5.4	17.5	5.83	28.2	5.64
10	6.5	17	5.67	28.4	5.68
11	5.1	17.4	5.80	27.8	5.56
12	5.8	15.9	5.30	28.6	5.72
13	5	17	5.67	27.7	5.54
14	6.2	16.8	5.60	29.3	5.86
15	5.6	18.5	6.17	28.7	5.74
16	6.7	17.5	5.83	29.2	5.84
17	5.2	17.4	5.80	28.8	5.76
18	5.5	16.5	5.50	28.3	5.66
19	5.8	16.4	5.47	27.4	5.48
20	5.1	16.7	5.57	28.9	5.78
21	5.8	17.6	5.87	28.6	5.72
22	6.7	17.7	5.90	28.8	5.76
23	5.2	17.9	5.97	29.5	5.90
24	6	17	5.67		
25	5.8				



Time Series ➡ Simple Moving Average

- The idea behind the moving averages is that observations which are nearby in time are also likely to be close in value.
- The average of the points near an observation will provide a reasonable estimate of the trend-cycle at that observation.
- The average eliminate some of the randomness in the data, and leaves a smooth trend-cycle component.

Time Series → Simple Moving Average

- Appropriate for a time series with a horizontal pattern ie., the data that are stationary.
  - Moving Average (the average of the most recent k data values forms the forecast for the next period)

$$F_{t+1} = \frac{\sum \text{most recent } k \text{ data values}}{k}$$

$$F_{t+1} = \frac{Y_t + Y_{t-1} + \ldots + Y_{t-k+1}}{k}$$

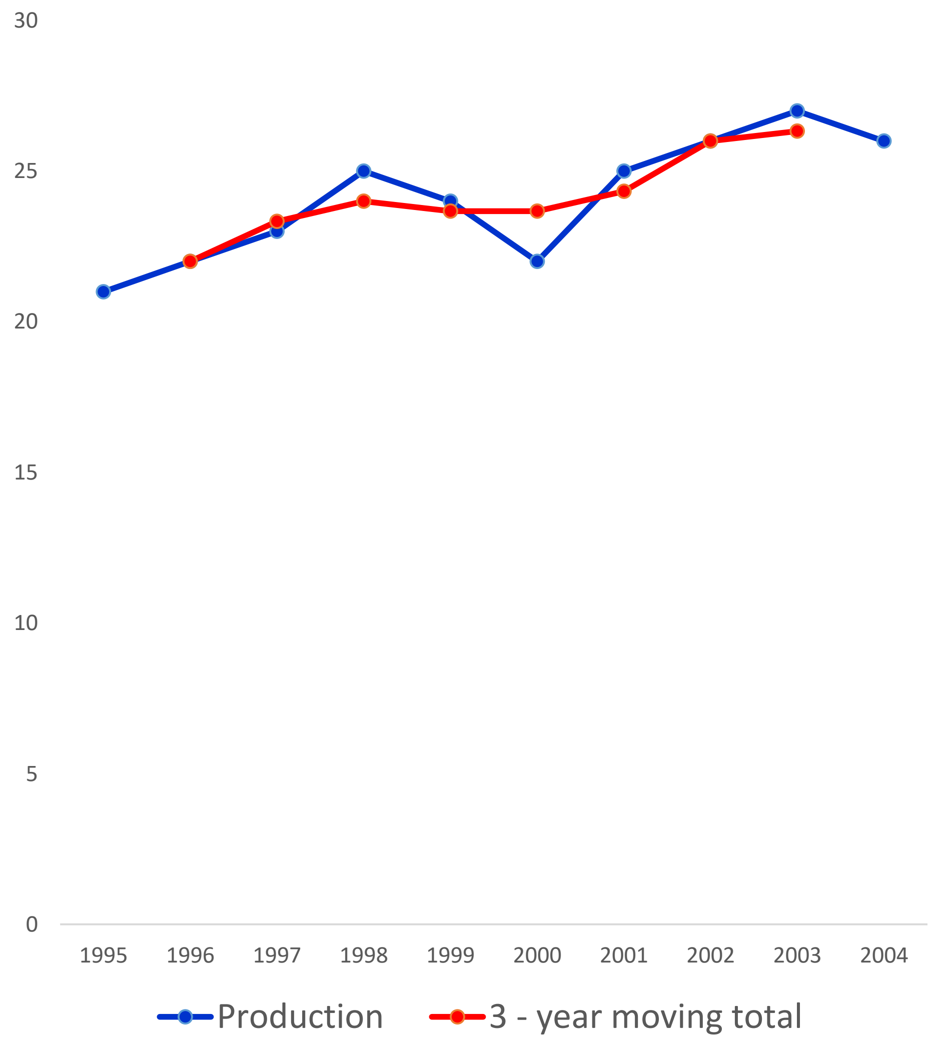
Statistical Methods for Data Science

innovate achieve lead

Time Series → Centered – Moving Averages

Following data shows production volume (in '000 tones). Compute 3-year moving average for all available years

Year	Production	3 - year moving total	3 - year moving average
1995	21		
1996	22	66	22.00
1997	23	70	23.33
1998	25	72	24.00
1999	24	71	23.67
2000	22	71	23.67
2001	25	73	24.33
2002	26	78	26.00
2003	27	79	26.33
2004	26		



Time Series → Centered – Moving Averages

Following data shows production volume (in ‘000 tones). Compute 4-year centered moving average

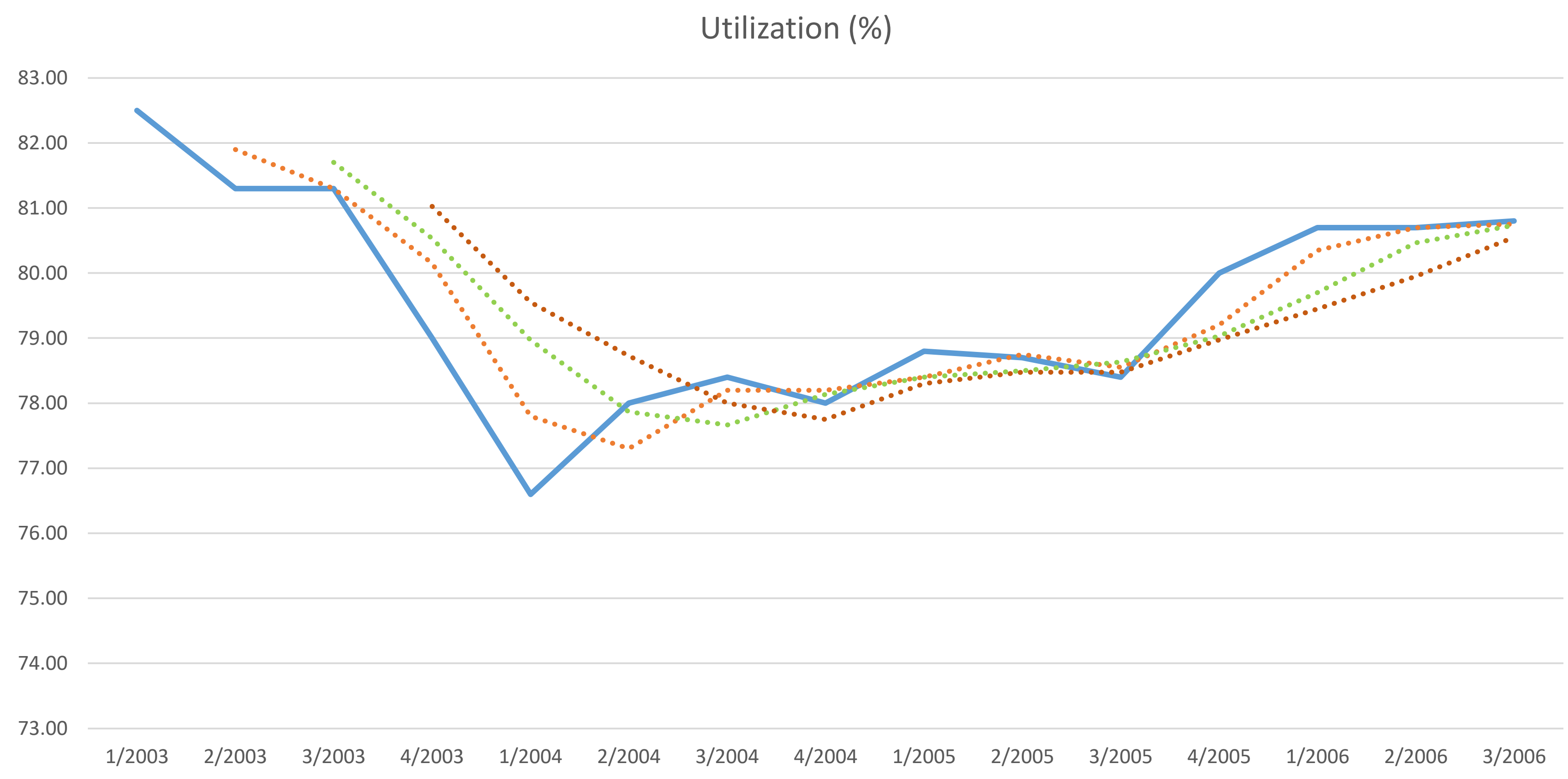
Year	Production(Y)	4-yearly moving average	4-year centered moving average
1995	21.0		
1996	22.0		
		22.75	
1997	23.0		23.13
		23.50	
1998	25.0		23.50
		23.50	
1999	24.0		23.75
		24.00	
2000	22.0		24.13
		24.25	
2001	25.0		24.63
		25.00	
2002	26.0		25.50
		26.00	
2003	27.0		26.38
		26.75	
2004	26.0		26.75
2005	28.0		

Time Series → Smoothing Techniques – Moving Averages

The following data represent 15 quarters of manufacturing capacity utilization

Quarter/Year	Utilization (%)
1/2003	82.50
2/2003	81.30
3/2003	81.30
4/2003	79.00
1/2004	76.60
2/2004	78.00
3/2004	78.40
4/2004	78.00
1/2005	78.80
2/2005	78.70
3/2005	78.40
4/2005	80.00
1/2006	80.70
2/2006	80.70
3/2006	80.80

Observe the 2,3 & 4 quarter moving average (drawn using Excel)



Thank you