Statistical Methods for Data Science

ISM Team

BITS Pilani

Pilani | Dubai | Goa | Hyderabad

innovate achieve lead

BITS Pilani

Forecasting and Time series

Time Series

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.

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 ➡ Definition of Forecasting

Forecasts can be required several years in advance (for the case of capital investments), or only a few minutes beforehand (for 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 lotto numbers cannot be forecast with any accuracy. The predictability of an event or a quantity depends on several factors including:

Time Series ➡ Definition of Forecasting

- 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 ⇨ Definition of Forecasting

- 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.

Time Series ⇨ Definition of Forecasting

- 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.

- 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.

Time Series ⇨ Definition of Forecasting

- If there are well-publicised forecasts that the exchange rate will increase, then people will immediately adjust the price they are willing to pay and so the forecasts are self-fulfilling. In a sense, the exchange rates become their own forecasts. This is an example of the “efficient market hypothesis”.

- Consequently, 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:

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

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.

Time Series ⇨ 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.

Forecasting the future involves uncertainty. Therefore, it is almost impossible to make a perfect prediction. Forecasters know that they have to live with a certain amount of error, which is the difference between what is forecast and what actually happens. The goal of forecasting is to generate good forecasts *on the average* over time and to keep forecast errors as low as possible

Forecasts are more accurate for groups or families of items rather than for individual items.

When items are grouped together, their individual high and low values can cancel each other out. The data for a group of items can be stable even when individual items in the group are very unstable.

Consequently, one can obtain a higher degree of accuracy when forecasting for a group of items rather than for individual items.

For example, you cannot expect the same degree of accuracy if you are forecasting sales of long-sleeved hunter green polo shirts that you can expect when forecasting sales of all polo shirts.

Forecasts are more accurate for shorter than longer time horizons.

The shorter the time horizon of the forecast, the lower the degree of uncertainty.

Time Series ⇨ Forecasting frame

Forecasting should be an integral part of the decision-making activities of management, as it can play an important role in many areas of a company. Modern organisations require short-term, medium-term and long-term forecasts, depending on the specific application.

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.

Short-term Forecasts

Some airport operators undertake ‘ultra short term’ forecasts for (e.g.) the next month in order to provide for specific requirement such as adequate staffing in the peaks.

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 ➡ Case studies

Case 1

The client was a large company manufacturing disposable tableware such as napkins and paper plates. They needed forecasts of each of hundreds of items every month. The time series data showed a range of patterns, some with trends, some seasonal, and some with neither. At the time, they were using their own software, written in-house, but it often produced forecasts that did not seem sensible. The methods that were being used were the following:

average of the last 12 months data;

average of the last 6 months data;

prediction from a straight line regression over the last 12 months;

prediction from a straight line regression over the last 6 months;

prediction obtained by a straight line through the last observation with slope equal to the average slope of the lines connecting last year's and this year's values;

prediction obtained by a straight line through the last observation with slope equal to the average slope of the lines connecting last year's and this year's values, where the average is taken only over the last 6 months.

They required us to tell them what was going wrong and to modify the software to provide more accurate forecasts. The software was written in COBOL, making it difficult to do any sophisticated numerical computation.

Time Series ➡ Case studies

Case 2

In this case, the client was the Australian federal government, who needed to forecast the annual budget for the Pharmaceutical Benefit Scheme (PBS). The PBS provides a subsidy for many pharmaceutical products sold in Australia, and the expenditure depends on what people purchase during the year. The total expenditure was around A$7 billion in 2009, and had been underestimated by nearly $1 billion in each of the two years before we were asked to assist in developing a more accurate forecasting approach.

In order to forecast the total expenditure, it is necessary to forecast the sales volumes of hundreds of groups of pharmaceutical products using monthly data. Almost all of the groups have trends and seasonal patterns. The sales volumes for many groups have sudden jumps up or down due to changes in what drugs are subsidised. The expenditures for many groups also have sudden changes due to cheaper competitor drugs becoming available.

Thus we needed to find a forecasting method that allowed for trend and seasonality if they were present, and at the same time was robust to sudden changes in the underlying patterns. It also needed to be able to be applied automatically to a large number of time series.

Time Series ➡ Case studies

Case 3

A large car fleet company asked us to help them forecast vehicle re-sale values. They purchase new vehicles, lease them out for three years, and then sell them. Better forecasts of vehicle sales values would mean better control of profits; understanding what affects resale values may allow leasing and sales policies to be developed in order to maximise profits.

At the time, the resale values were being forecast by a group of specialists. Unfortunately, they saw any statistical model as a threat to their jobs, and were uncooperative in providing information. Nevertheless, the company provided a large amount of data on previous vehicles and their eventual resale values.

Case 4

In this project, we needed to develop a model for forecasting weekly air passenger traffic on major domestic routes for one of Australia's leading airlines. The company required forecasts of passenger numbers for each major domestic route and for each class of passenger (economy class, business class and first class). The company provided weekly traffic data from the previous six years.

Air passenger numbers are affected by school holidays, major sporting events, advertising campaigns, competition behaviour, etc. School holidays often do not coincide in different Australian cities, and sporting events sometimes move from one city to another. During the period of the historical data, there was a major pilots' strike during which there was no traffic for several months. A new cut-price airline also launched and folded. Towards the end of the historical data, the airline had trialled a redistribution of some economy class seats to business class, and some business class seats to first class. After several months, however, the seat classifications reverted to the original distribution.

Step 1: Problem definition

Often this is the most difficult part of forecasting. Defining the problem carefully requires an understanding of the way the forecasts will be used, who requires the forecasts, and how the forecasting function fits within the organisation requiring the forecasts. A forecaster needs to spend time talking to everyone who will be involved in collecting data, maintaining databases, and using the forecasts for future planning.

Time Series → The basic steps in a forecasting task

Step 2: Gathering information

There are always at least two kinds of information required: (a) statistical data, and (b) the accumulated expertise of the people who collect the data and use the forecasts. Often, it will be difficult to obtain enough historical data to be able to fit a good statistical model. In that case, the judgmental forecasting methods of can be used. Occasionally, old data will be less useful due to structural changes in the system being forecast; then we may choose to use only the most recent data. However, remember that good statistical models will handle evolutionary changes in the system; don't throw away good data unnecessarily.

Time Series → The basic steps in a forecasting task

Step 3: Preliminary (exploratory) analysis

Always start by graphing the data. Are there consistent patterns? Is there a significant trend? Is seasonality important? Is there evidence of the presence of business cycles? Are there any outliers in the data that need to be explained by those with expert knowledge? How strong are the relationships among the variables available for analysis? Various tools have been developed to help with this analysis.

Step 4: Choosing and fitting models

The best model to use depends on the availability of historical data, the strength of relationships between the forecast variable and any explanatory variables, and the way in which the forecasts are to be used. It is common to compare two or three potential models. Each model is itself an artificial construct that is based on a set of assumptions (explicit and implicit) and usually involves one or more parameters which must be estimated using the known historical data.

Step 4: Choosing and fitting models

Some of the common models used for forecast are regression models, exponential smoothing methods, Box-Jenkins ARIMA models, Dynamic regression models, Hierarchical forecasting, and several advanced methods like neural networks and vector auto-regression.

Step 5: Using and evaluating a forecasting model

Once a model has been selected and its parameters estimated, the model is used to make forecasts. The performance of the model can only be properly evaluated after the data for the forecast period have become available. A number of methods have been developed to help in assessing the accuracy of forecasts.

Step 5: Using and evaluating a forecasting model

There are also organisational issues in using and acting on the forecasts. When using a forecasting model in practice, numerous practical issues arise such as how to handle missing values and outliers, or how to deal with short time series.

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.

Time Series ⇨ Qualitative forecasting

- **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

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

Thus a 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 also 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 county

Time Series → Definition

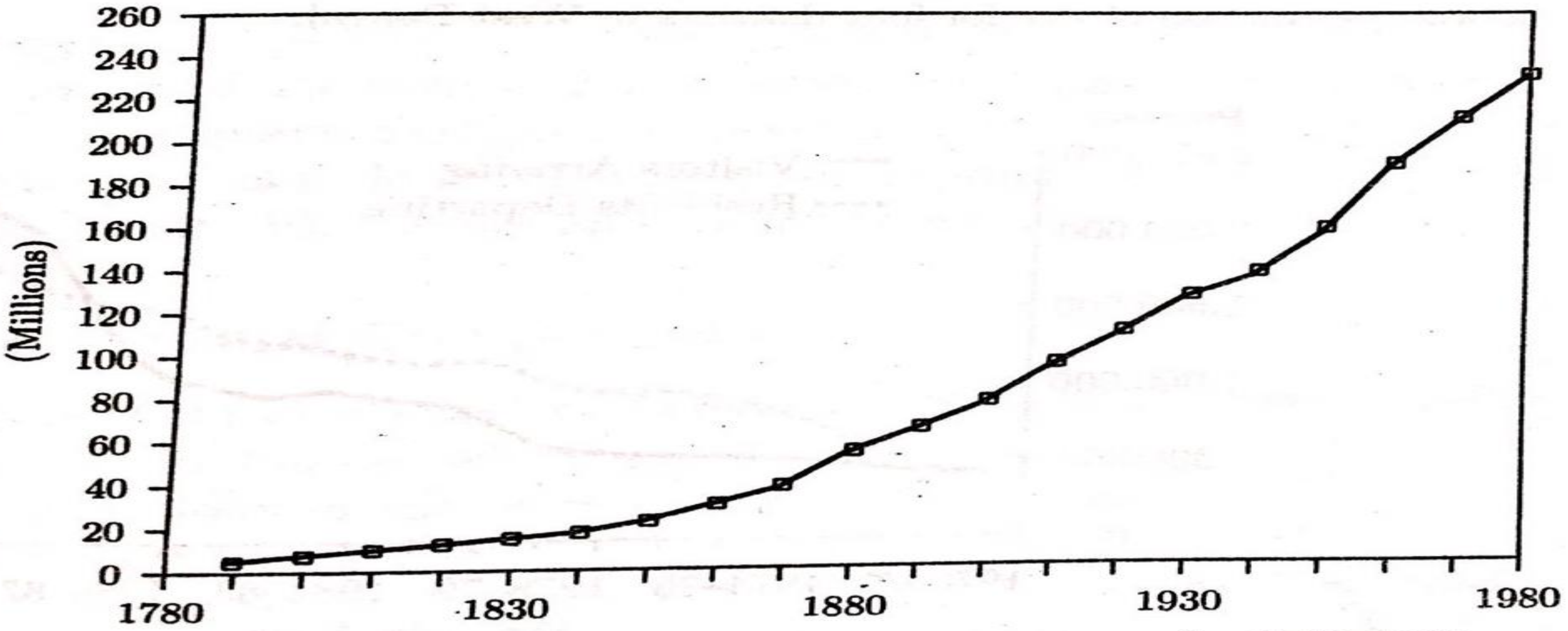


Fig 14.1 U.S. Population at ten year intervals, 1790–1980.
[*Source*. Brockwell and Davis, 1990.]

Time Series ➡ Definition

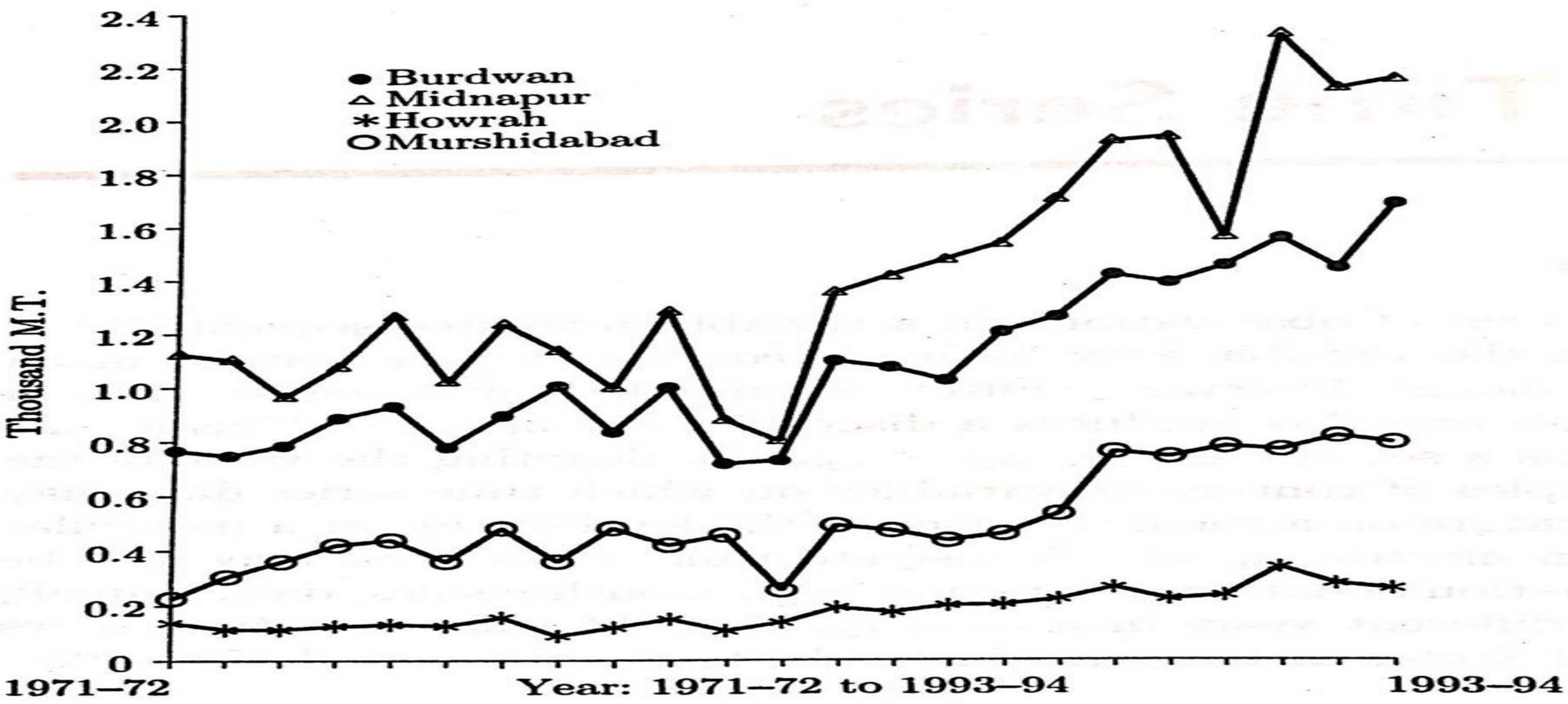


Fig 14.2 District-wise production of rice for four districts in West Bengal.

Time Series ➡ Definition

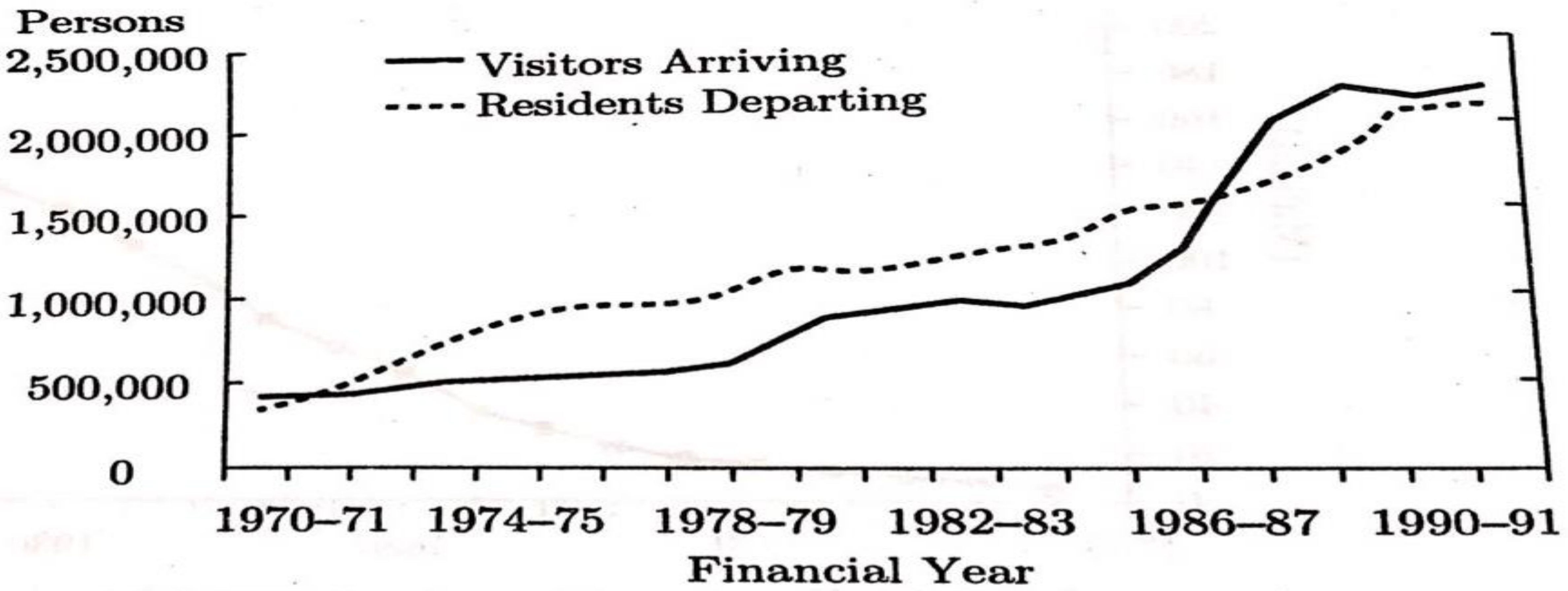


Fig 14.3 Short-term movement to and from Australia.
[*Source*: Immigration update, December, Quarter 1991, Bureau of Immigration Research, Statistics Section, Australia.]

Time Series ➡ Definition

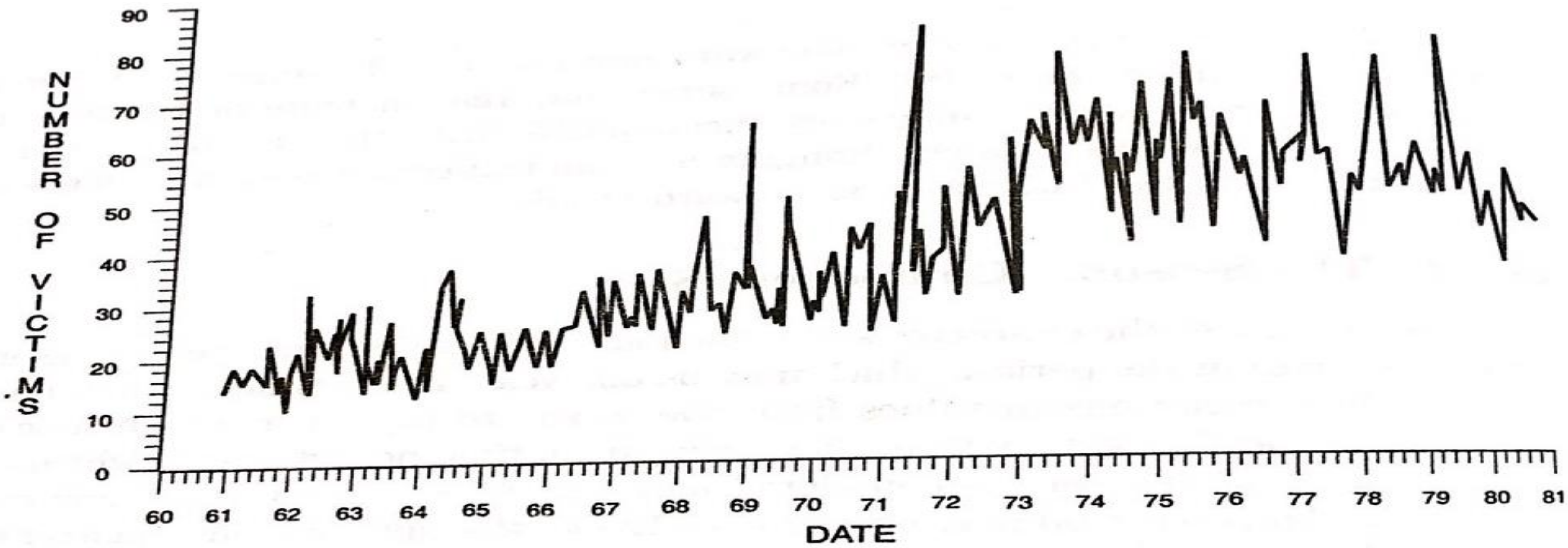


Fig 14.5 Frequency of homicide victims by month, 1961-1980 (excluding manslaughters and infanticides).
[*Source*: McLeod, MacNeil and Bhattacharyya (1985).]

Time Series ⇨ Qualitative Forecasting Methods

- Time Series Models:
- Causal Models or Associative Models

Time Series ⇨ Qualitative Forecasting Methods

- **Time Series Models:**
  - 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*** *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:

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

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 → **Decomposition Models**

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)

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.

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:

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

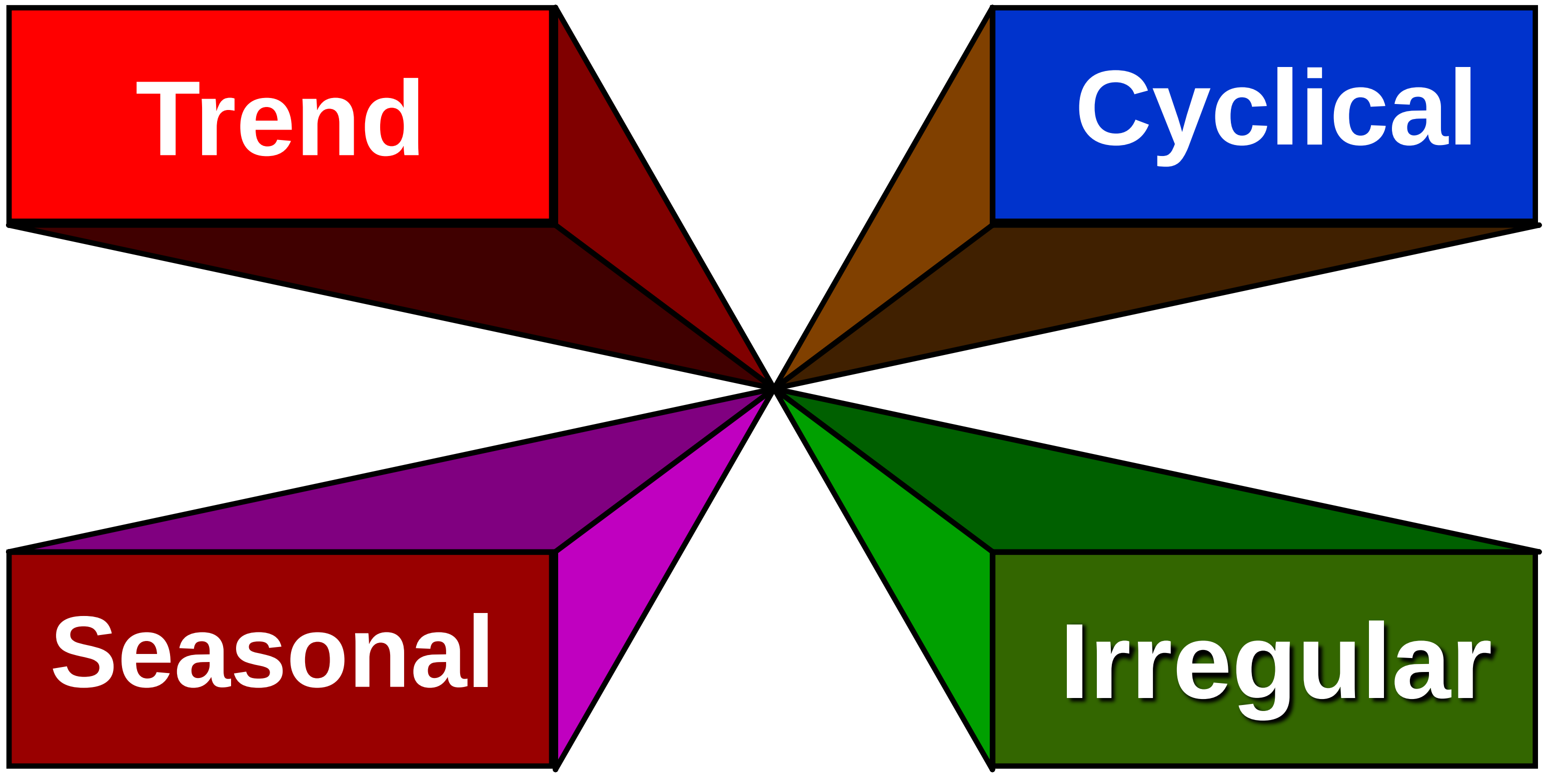
Time Series ⇨ Decomposition Models

An additive model would be used when the variations around the trend do not vary with the level of the time series whereas a multiplicative model would be appropriate if the trend is proportional to the level of the time series.

Time Series ➡ Decomposition Models

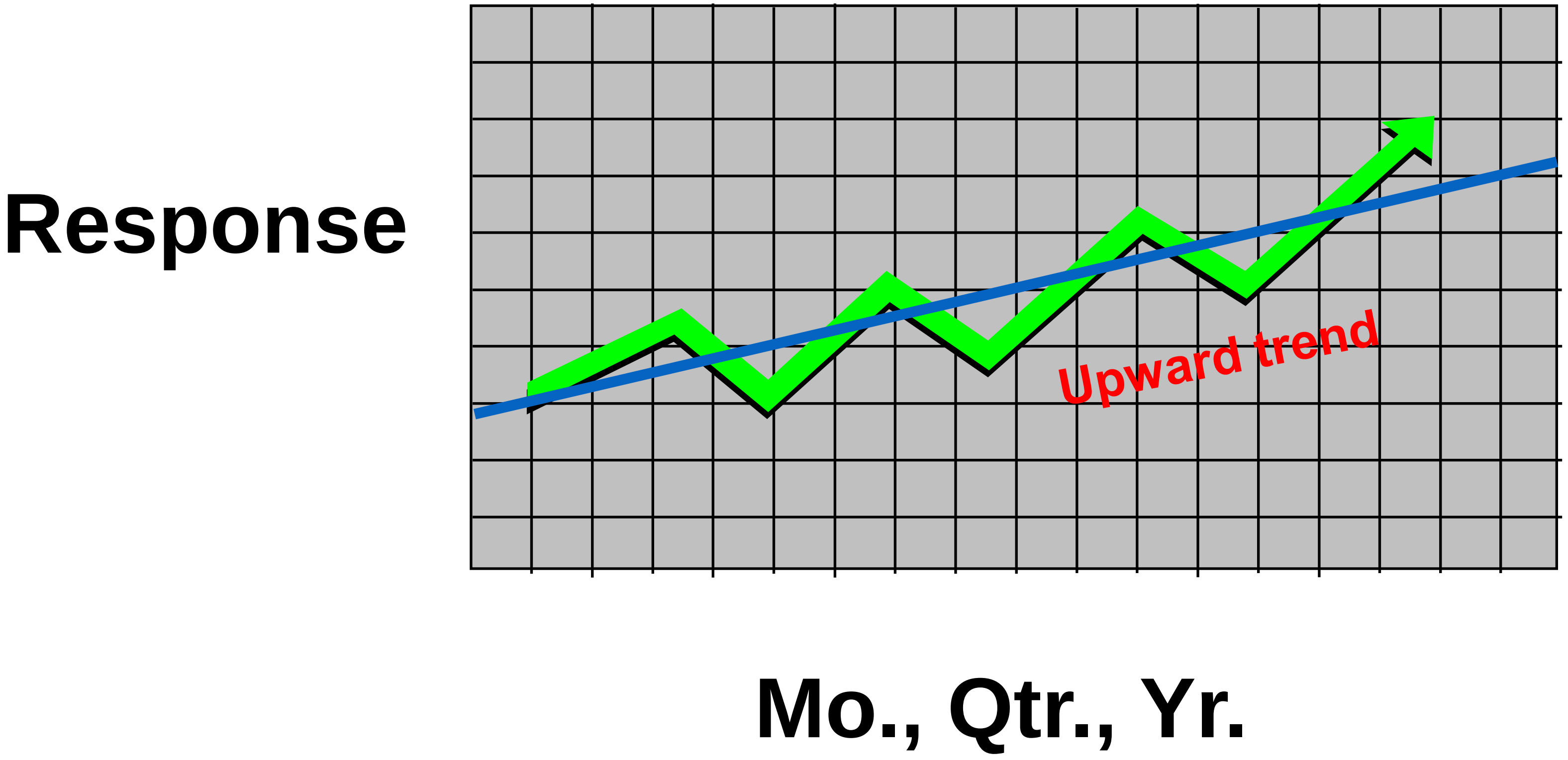
Sometimes the trend and cyclical components are grouped into one, called the trend-cycle component. The trend-cycle component can just be referred to as the "trend" component, even though it may contain cyclical behaviour. For example, a seasonal decomposition of time series plot decomposes a time series 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 ⇨ Time Series Components



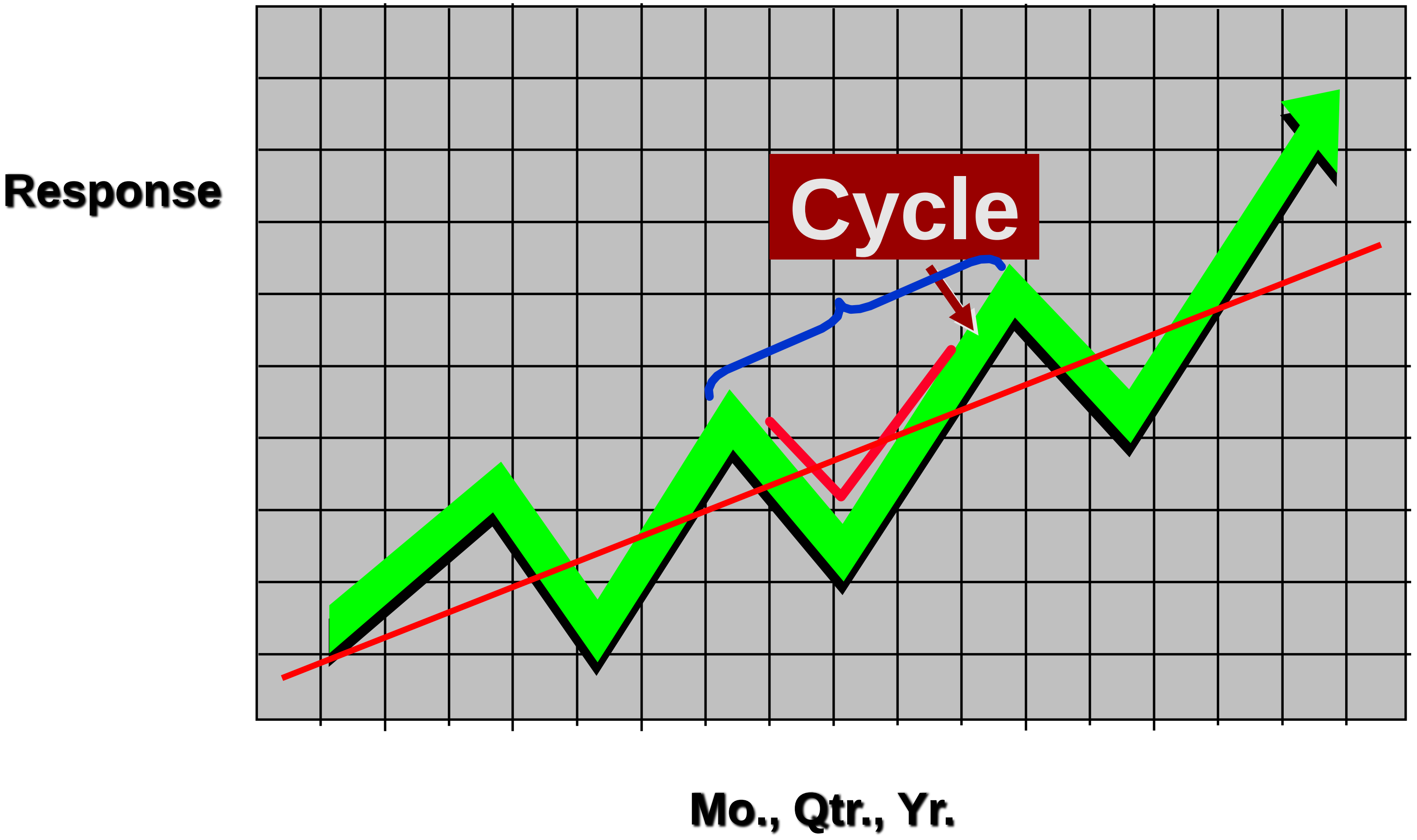
Trend:

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



Cyclical:

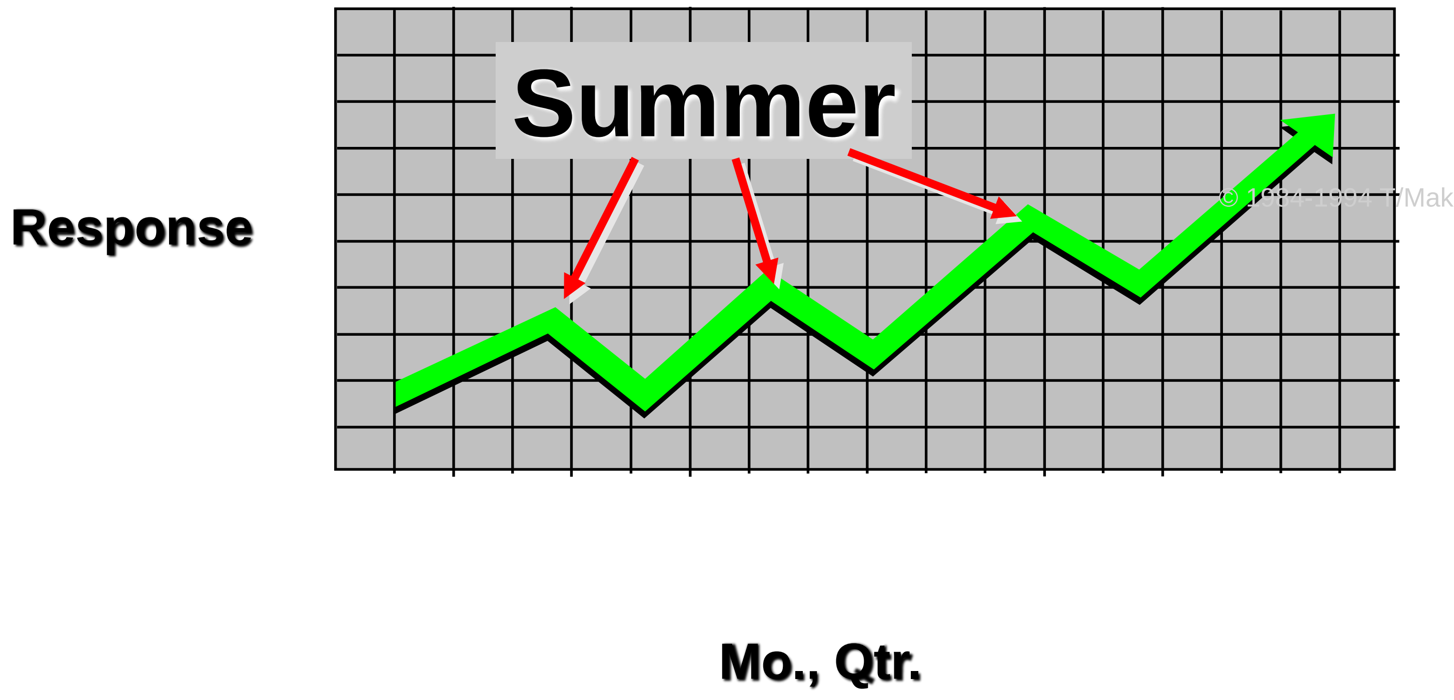
- Repeating up & down movements above trend line
- Due to interactions of factors influencing economy
- Usually 2-10 years duration



innovate achieve lead

Time Series ⇨ Time Series Components

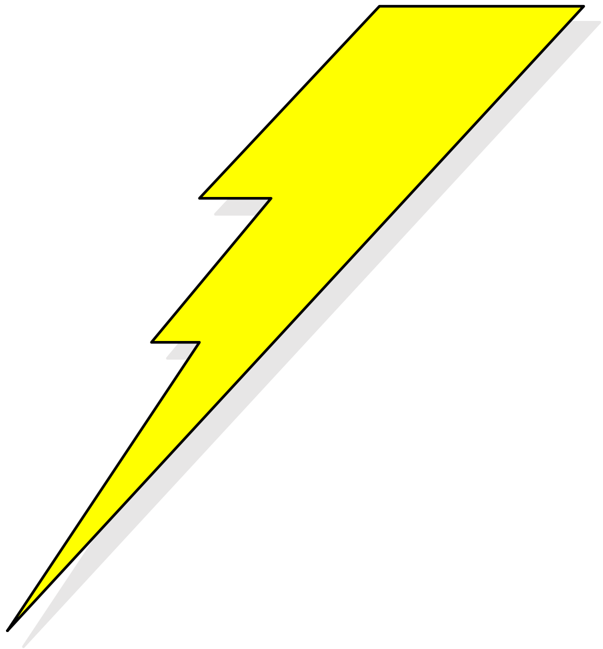
- Regular pattern of up & down fluctuations
- Due to weather, customs etc.
- Occurs within one year



Time Series ➡ Time Series Components

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



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 ⇨ 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)

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

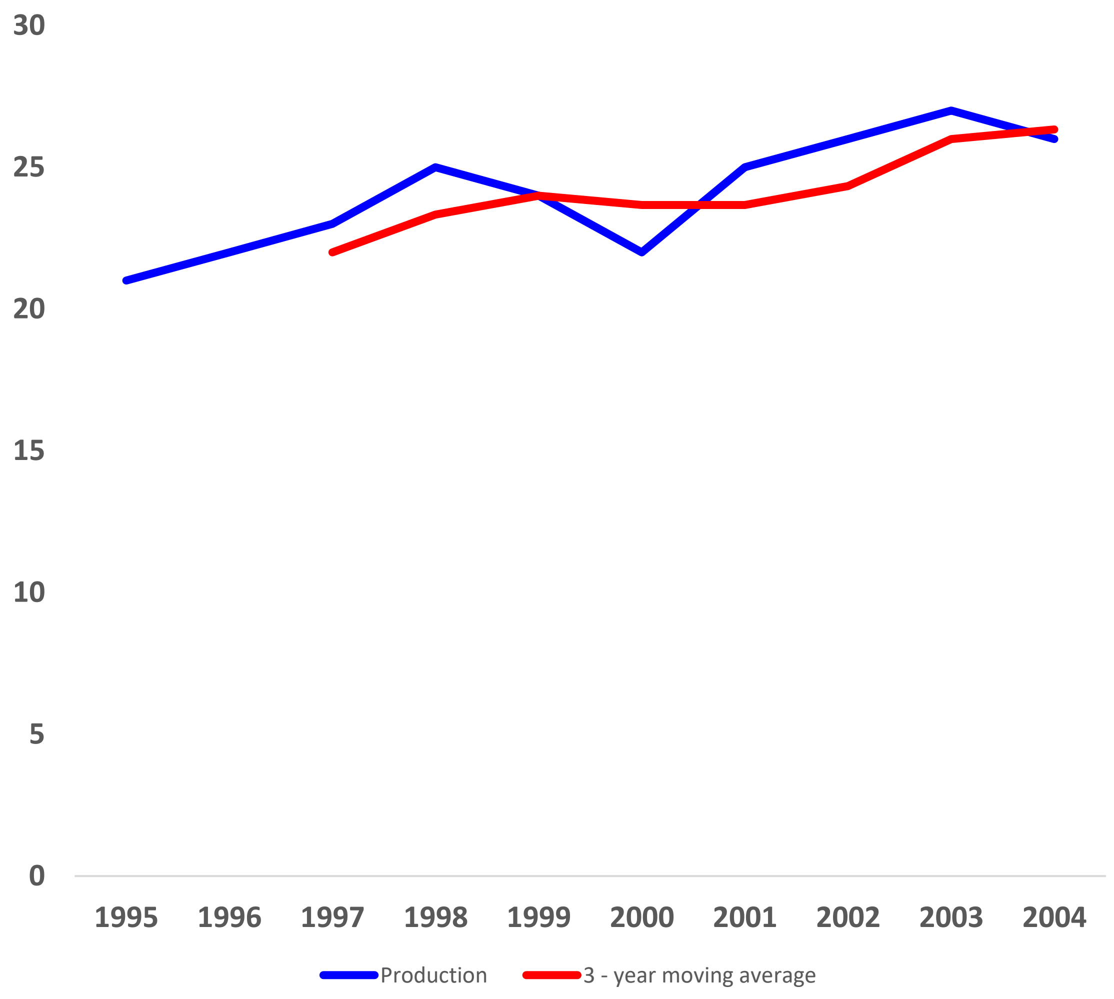
$$F_{t+1} = \frac{Y_t + Y_{t-1} + \ldots + Y_{t-k+1}}{k}$$

innovate achieve lead

Time Series → Smoothing Techniques – 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		
1997	23	66	22.00
1998	25	70	23.33
1999	24	72	24.00
2000	22	71	23.67
2001	25	71	23.67
2002	26	73	24.33
2003	27	78	26.00
2004	26	79	26.33

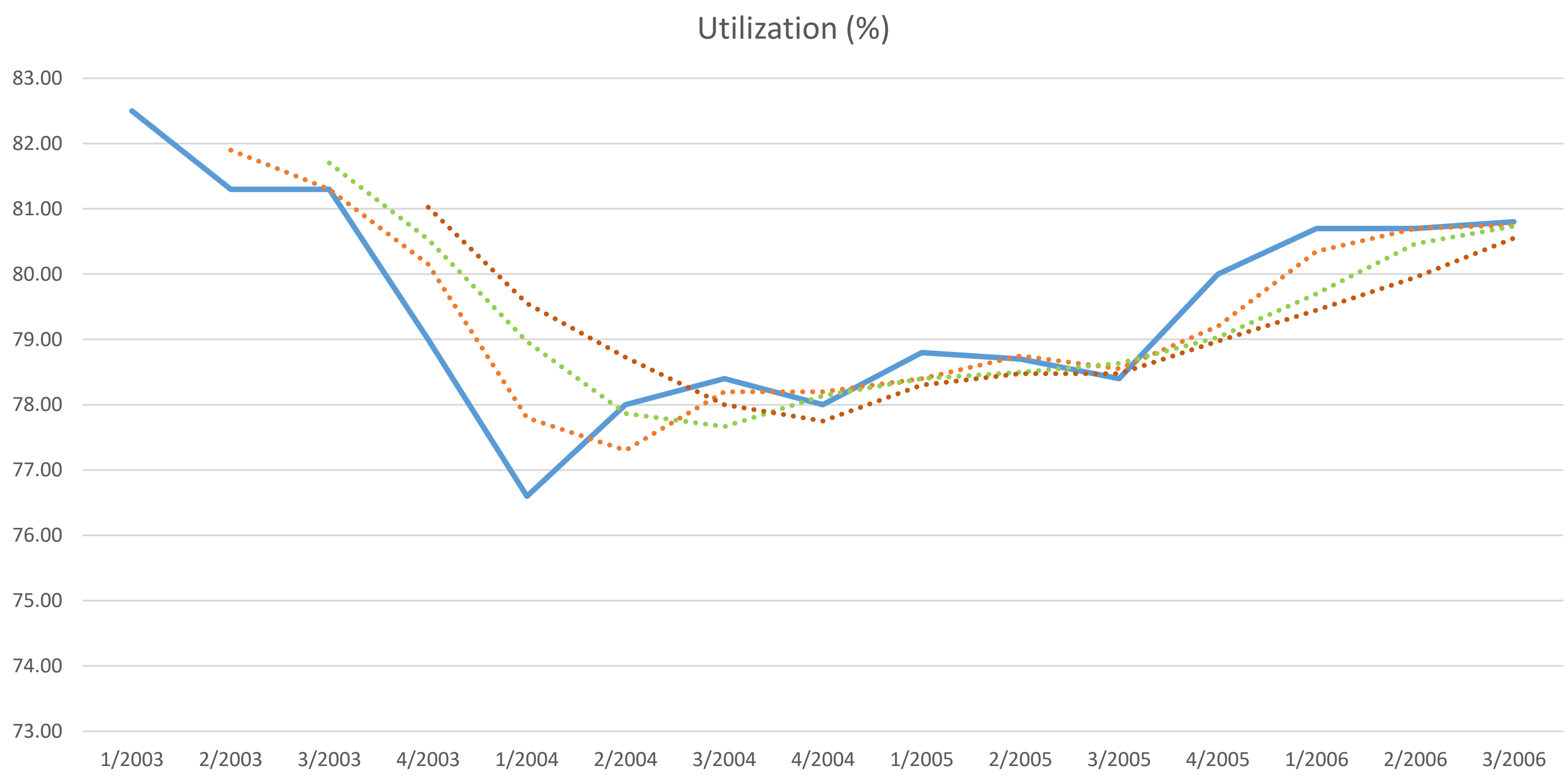


Time Series ➡ Smoothing Techniques – Moving Averages

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)



Time Series ⇨ Smoothing Techniques – Exponential

- Exponential (Weighted average of all the past time series values with exponentially decreasing importance in the forecast)

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

- α=1⇒ Forecast value is just the previous actual value. A naïve forecast.
- α=0⇒ Current Forecast value is just the previous forecast value.
- How to choose value of k and α?

innovate achieve lead

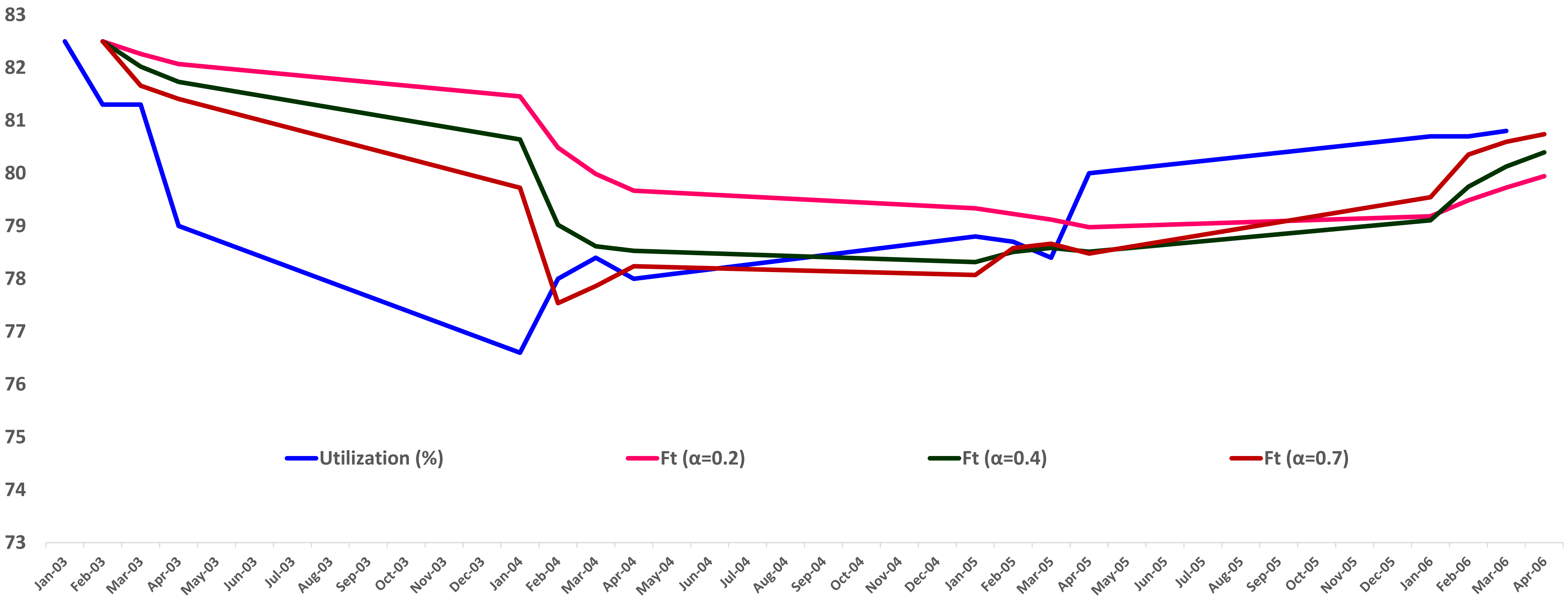
Time Series ➡ Smoothing Techniques – Exponential

- Done the analysis on the previous problem for 3 different values of α
- For α=0.7, it is better
- The exponential smoothing does not have a significant initial delay like the Moving Av. where the initial k values are not smoothed
- Formula can be rearranged as

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

Quarter/ Year	Utilization (%)	F_t (α=0.2)	F_t (α=0.4)	F_t (α=0.7)
1/2003	82.50			
2/2003	81.30	82.50	82.50	82.50
3/2003	81.30	82.26	82.02	81.66
4/2003	79.00	82.07	81.73	81.41
1/2004	76.60	81.45	80.64	79.72
2/2004	78.00	80.48	79.02	77.54
3/2004	78.40	79.99	78.61	77.86
4/2004	78.00	79.67	78.53	78.24
1/2005	78.80	79.34	78.32	78.07
2/2005	78.70	79.23	78.51	78.58
3/2005	78.40	79.12	78.59	78.66
4/2005	80.00	78.98	78.51	78.48
1/2006	80.70	79.18	79.11	79.54
2/2006	80.70	79.49	79.74	80.35
3/2006	80.80	79.73	80.13	80.60
3/2006	-	**79.94**	**80.40**	**80.74**

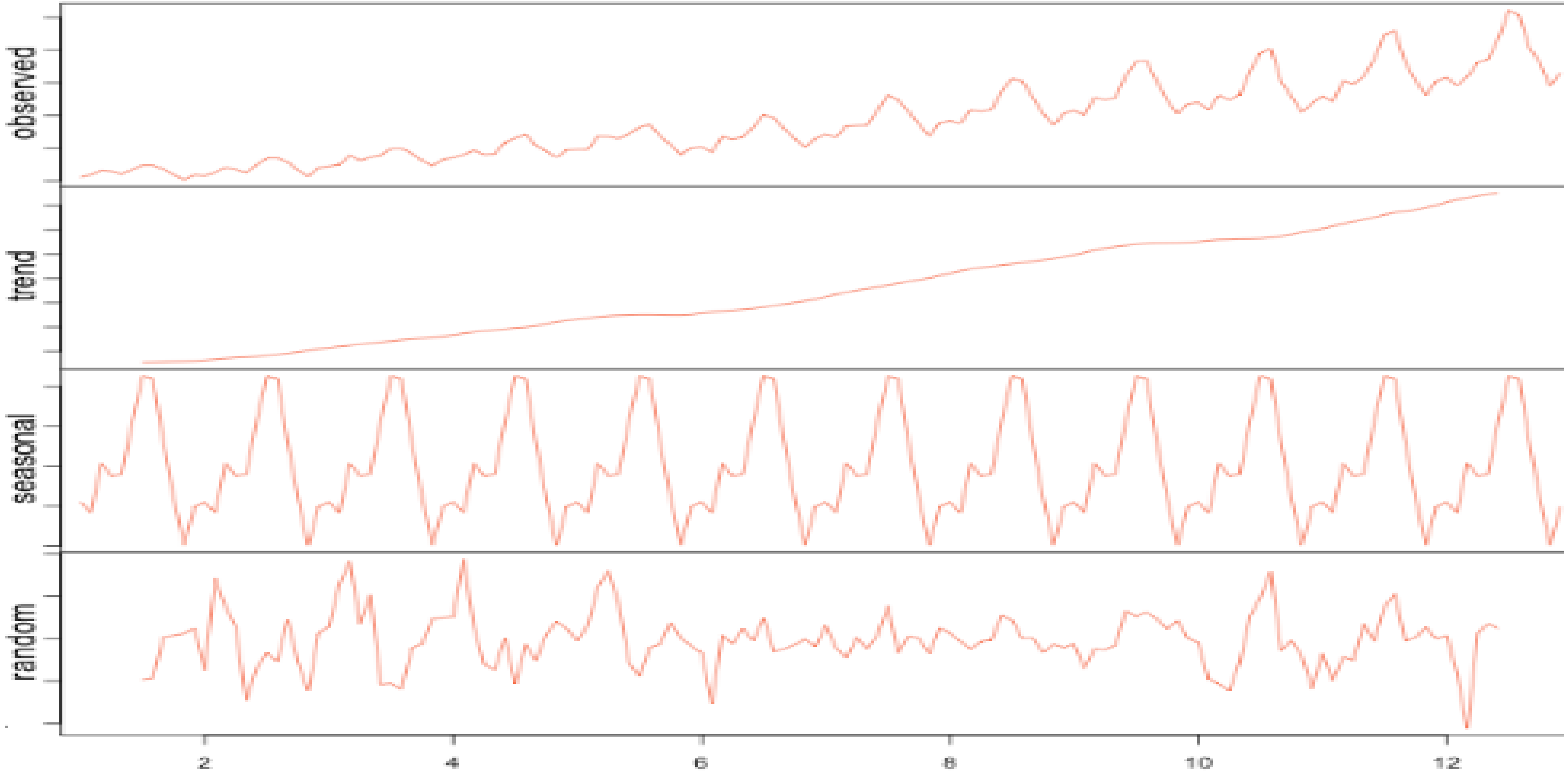
Time Series ➔ Smoothing Techniques – Exponential



Time Series ➡ Smoothing Techniques – Exponential

- Objective is to estimate the overall time series as a combination of long term trend and seasonality
  - Additive Model: $X_t = F_t + S_t + \varepsilon_t$
  - Multiplicative Model: $X_t = F_t \times S_t \times \varepsilon_t$
  - Multiplicative model can be converted to additive model by taking log. Other models are also possible
- 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 .eg. Differencing

Time Series ➡ Decomposition view



Time Series ⇨ Decomposition view

- Assess the trend component by smoothing or curve fitting (regression with time)
- Assess/ account for seasonality i.e. de-seasonalize 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

Time Series ➡ Definition of Stochastic Process

- A stochastic process is a collection of family of random variables **$\{X_t, t \in T\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$,**

where Ω - Sample space

$\mathcal{F}$ - all collection of subsets A of Ω and

P – probability measure

T – Index set

Index set:

- The index set **T** is a collection of all time functions that can result from random experiment, usually the index **T** denote time. X_t are independent and identically distributed (iid) random variables with mean μ_t and variance σ_t^2. Each realization of this process gives an ensemble or a data set.

In stochastic time series,

$$Y_t,\ t \in Z = \{\pm 1, \pm 2, \ldots\}$$

is a family of random variables, Y_t denoting the value of the characteristic of interest at time t.

Thus, $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$ is seen as a realized value of the random vector $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)'$ with joint probability density function $f_{\mathbf{Y}}(y)$.

Time Series ➡ Definition of Stochastic Process

The joint distribution function of a finite random variables

$$\{Y_{t_1}, \ldots Y_{t_n}\}, t_1 < t_2 < \ldots < t_n$$

from the collection $\{Y_t, t \in T\}$ is

$$F_{Y_{t_1}, \ldots Y_{t_n}}(y_1, \ldots, y_n) = P[Y_{t_1} \le y_{t_1}, \ldots, P[Y_{t_n} \le y_{t_n}],$$

$$(y_{t_1}, \ldots, y_{t_n}) \in R^n$$

Time Series ➡ Stationary Stochastic Process in Time series

A special kind of stochastic process is based on the assumption that the process is in a particular state of equilibrium. This type of assumption is called stationarity.

A stochastic process is strictly stationary if its properties are unaffected by a change of origin.

Thus, a time series stochastic process is said to be strictly stationary if the joint probability density function of m observations $Y_{t_1}, \ldots Y_{t_m}$ made at time $t_1, \ldots, t_m$ is same as that of

m observations $Y_{t_1+h}, \ldots Y_{t_m+h}$ made at time points $t_1+h, \ldots, t_m+h$ for any h. That is,

$$F_{Y_{t_1}, \ldots Y_{t_m}}(y_{t_1}, \ldots, y_{t_m}) = F_{Y_{t_1+h}, \ldots Y_{t_m+h}}(y_{t_1+h}, \ldots, y_{t_m+h})$$

That is,

$F_{Y_{t_1}, \ldots Y_{t_m}}(y_{t_1}, \ldots, y_{t_m}) = F_{Y_{t_1+h}, \ldots Y_{t_m+h}}(y_{t_1+k}, \ldots, y_{t_m+h})$

For all possible non-empty finite distinct sets

$(t_1, \ldots, t_m)$ and $(t_{1+h}, \ldots, t_{m+h})$ in the index set T and all $y_{t_1}, \ldots, y_{t_m}$ in the range of random variables Y_t.

innovate achieve lead

Time Series ⇨ Stochastic Process

Joint Distribution of the Y's dependent only on their **relative positions** (not affected by time shift)

$(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_n})$ has the same distribution as $(Y_{t_{1+h}}, Y_{t_{2+h}}, \ldots, Y_{t_{n+h}})$. For example (Y_8, Y_{11}) has same distribution as (Y_{20}, Y_{23}).

- **For reliable prediction to be made, the time series should be stationary (No systematic change such as seasonality, trend i.e. only random fluctuations.**

When m = 1, strict stationary implies that the pdf $f(y_t)$ is the same for all t, is say f(y). The stochastic process f(y) has a constant mean

$$\mu = E(Y_t) = \int_{-\infty}^{\infty} y f(y) dy$$

and a constant variance

$$\sigma^2 = E(Y_t - \mu)^2 = \int_{-\infty}^{\infty} (y - \mu)^2 \mathrm{f(y)dy}$$

provided the mean and variance exists.

Time Series ⇨ Stationary Stochastic Process in Time series

The mean (μ) and variance (σ^2) are estimated as

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\hat{\sigma}_y^2 = \frac{1}{n}\sum_{i=1}^{n} (y_i - \bar{y})^2$$

Time Series ➡ Stationary Stochastic Process in Time series

Thus, a stationary process remains a equilibrium about a common mean value.

However, in industry, business (ex. Stock price) and economics many time series are better represented as non-stationary and in particular, having no natural mean.

Time Series ➡ Stationary Stochastic Process in Time series

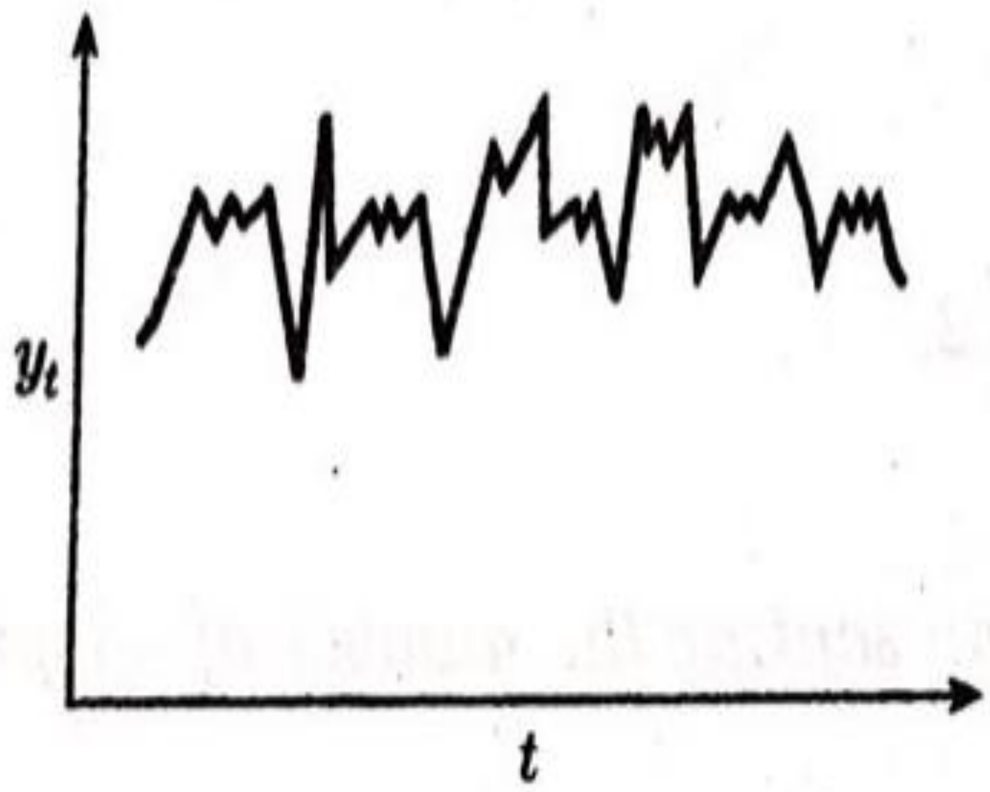


Fig 15.1 A stationary time series.

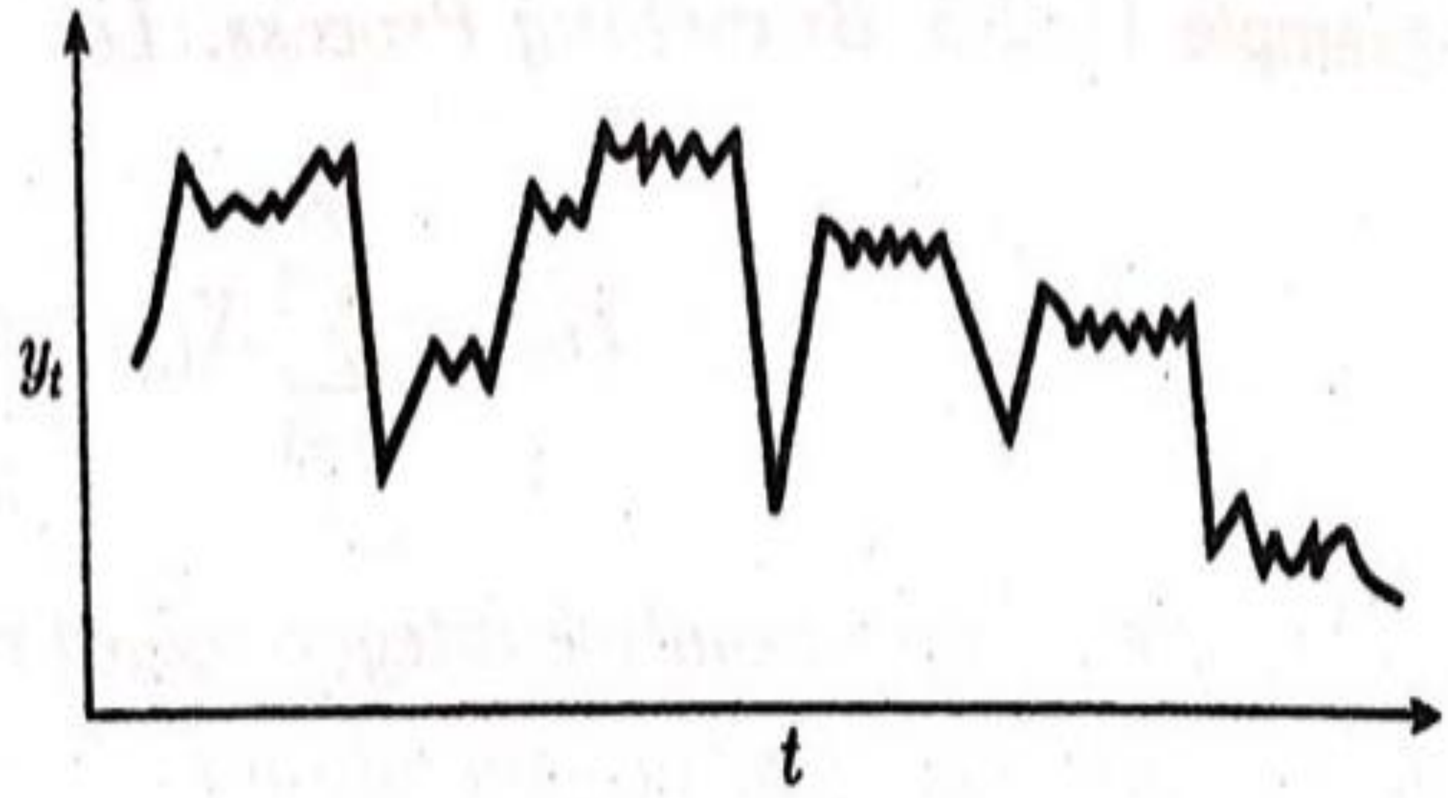


Fig 15.2 A non-stationary time series.

Time Series ➡ Stationary Stochastic Process in Time series

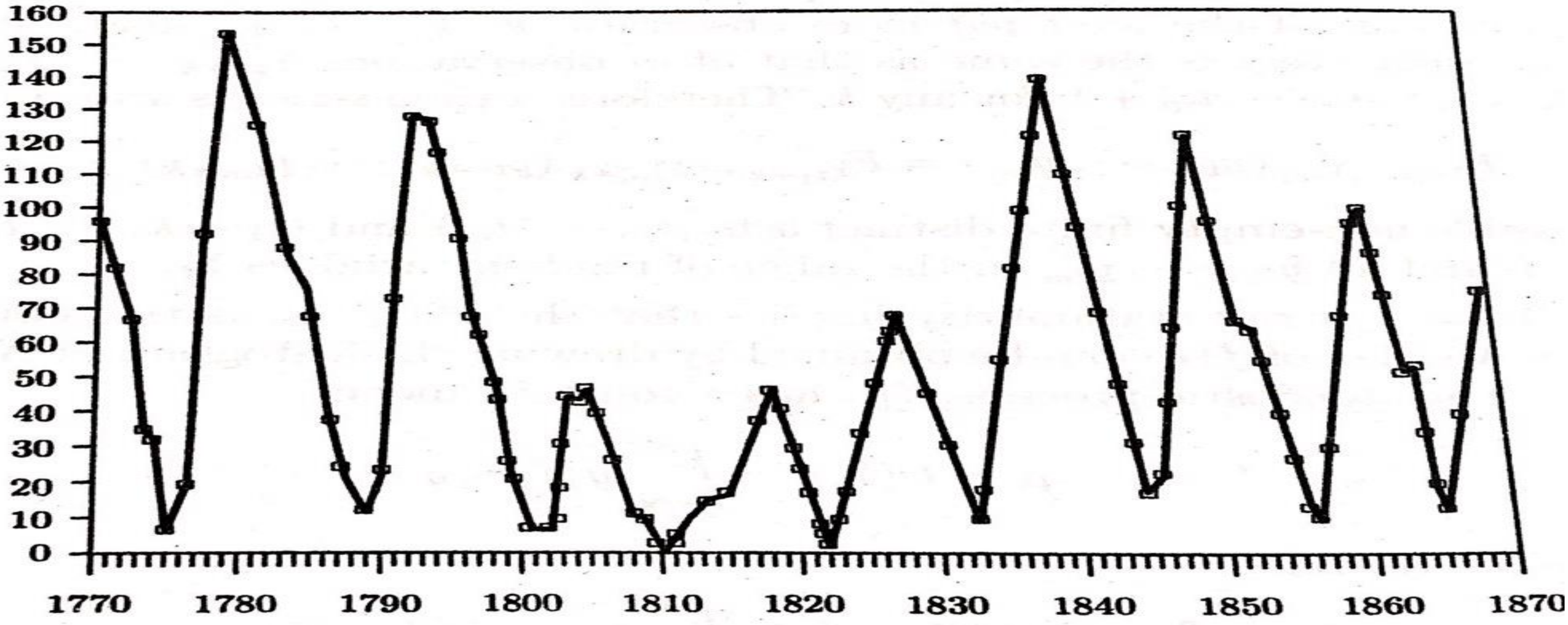


Fig 15.3 The Wolfer sunspot numbers, 1770–1869.
(Source: Box and Jenkins, 1976)

Time Series ➡ Stationary Stochastic Process in Time series

However, many non-stationary time series can be so modified that the reduced to time series obeys the original series.

The two main component which cause lack of stationarity are trend and seasonality. In fitting the stationary time series model, we therefore, assume that the trend and seasonality have been eliminated from the original series.

- Stationarity implies that the joint distribution of pair of observations y_{t_1}, y_{t_2} viz.,

$$f(y_{t_1}, y_{t_2}) = f(y_{t_1+h}, y_{t_2+h})$$

For any integer h. That is, the joint distribution of any pair of observations on time points which differ by a constant quantity is the same for such pairs.

- Thus

$$f(y_{t_1}, y_{t+h}) = f(y_{t_1}, y_{t_1+h}) = f(y_{t_2}, y_{t_2+h})$$

have the same distribution.

The form of $f(y_{t_1}, y_{t+h})$ can be inferred by plotting values of (y_{t_1}, y_{t+h}); $t = 1, 2, \ldots.,$ i.e., values of y_t separated by lag h.

Time Series ➡ Autocovariance and Autocorrelation

Figures a 15.4, 15.5 illustrate scatter diagrams for lag k ($k = 1, 2$).

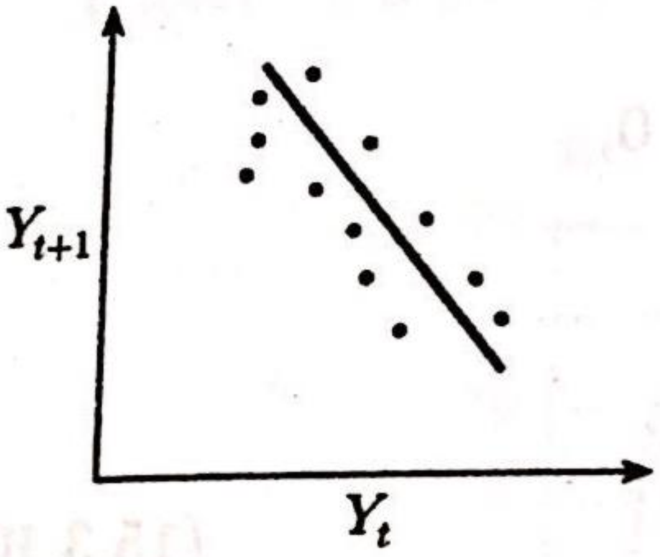


Fig 15.4(a)

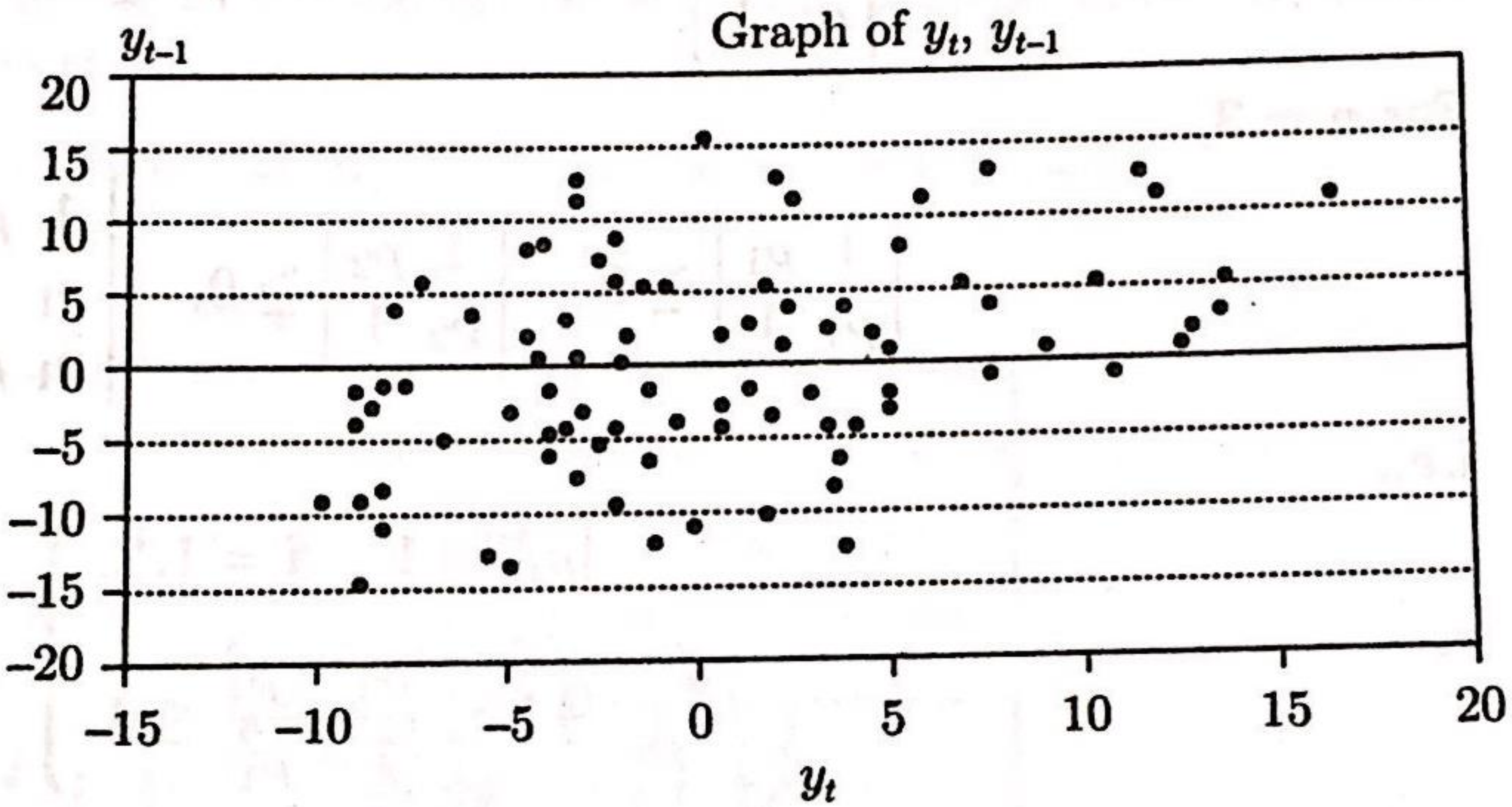


Fig 15.4(b) $Y_t = 0.6\varepsilon_{t-1} + \varepsilon_t$

Time Series ⇨ Autocovariance and Autocorrelation

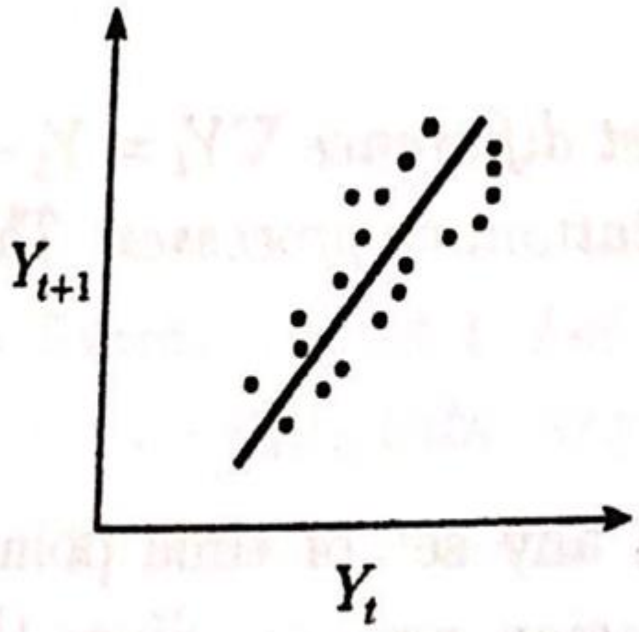


Fig 15.5(a)

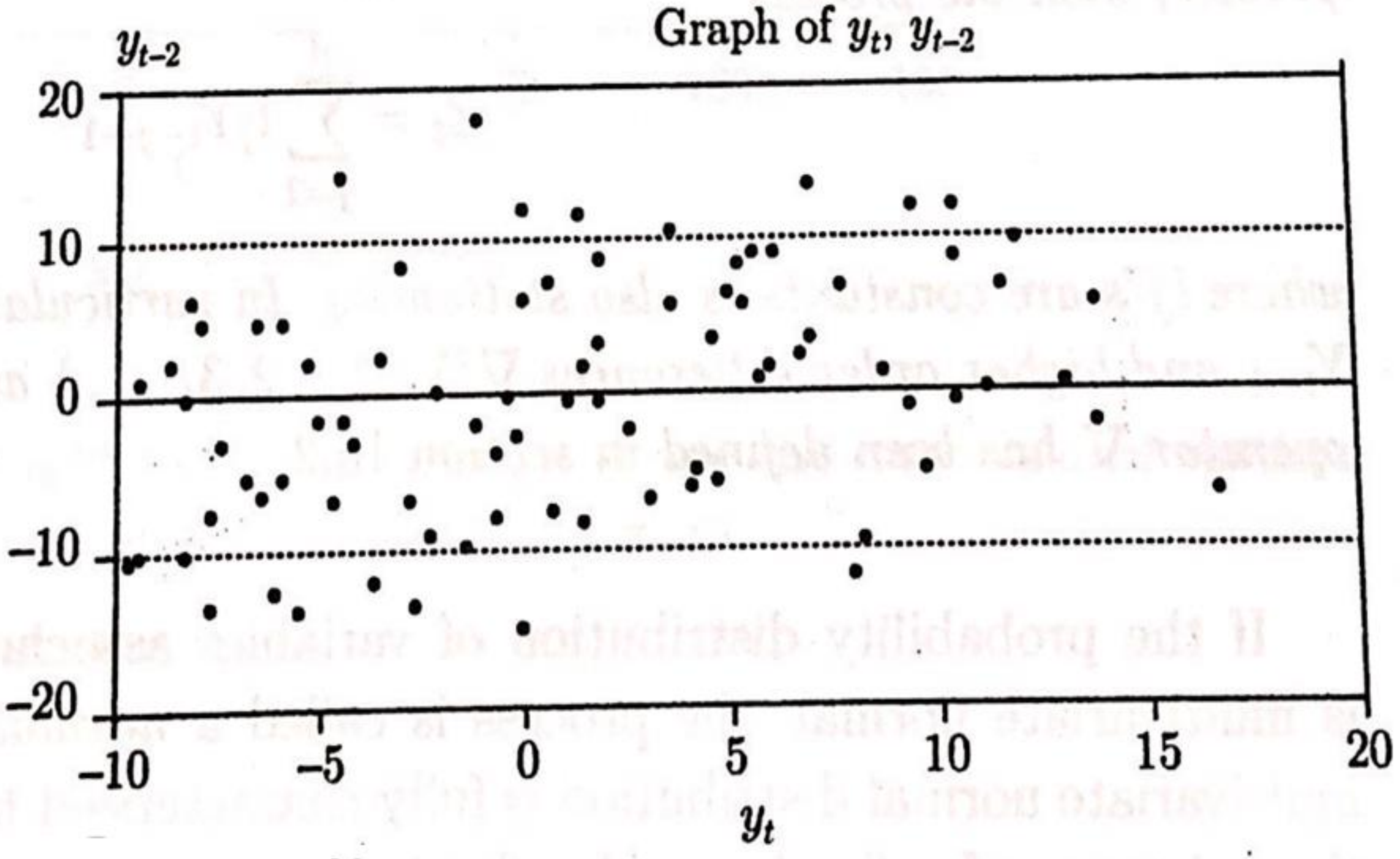


Fig 15.5(b) Y_t=$0.7\varepsilon_t+\varepsilon_{t-1}$

Time Series ⇨ Autocovariance and Autocorrelation

- Let $\{Y_t\}$ be a Time Series with E $\{Y^2\} < \infty$. The mean function of $\{Y_t\}$ is $\mu_y(t) = E(Y_t)$.
- The covariance function of $\{Y_t, Y_{t+h}\}$ is
- $\gamma_h = \mathrm{Cov}\,(Y_t, Y_{t+h}) = E[(Y_t-\mu)(Y_{t+h}-\mu)]$
- In general

$$\gamma_y(r, s) = \mathrm{Cov}\,(Y_r, Y_s) = E[\{Y_r-\mu_y(r)\}\{Y_s-\mu_y(s)\}]$$

for all integers r and s

- Clearly,

$$\gamma_0 = \text{Cov}\,(Y_t, Y_{t+0}) = E[(Y_t - \mu)(Y_{t+0} - \mu)] = \sigma^2_y$$

$$|\gamma_h| \leq \gamma_0, \forall\, h = 1, 2, \ldots$$

- because, $|Cov(Y_t, Y_{t+h})| \leq \sqrt{V(Y_t)V(Y_{t+h})}$

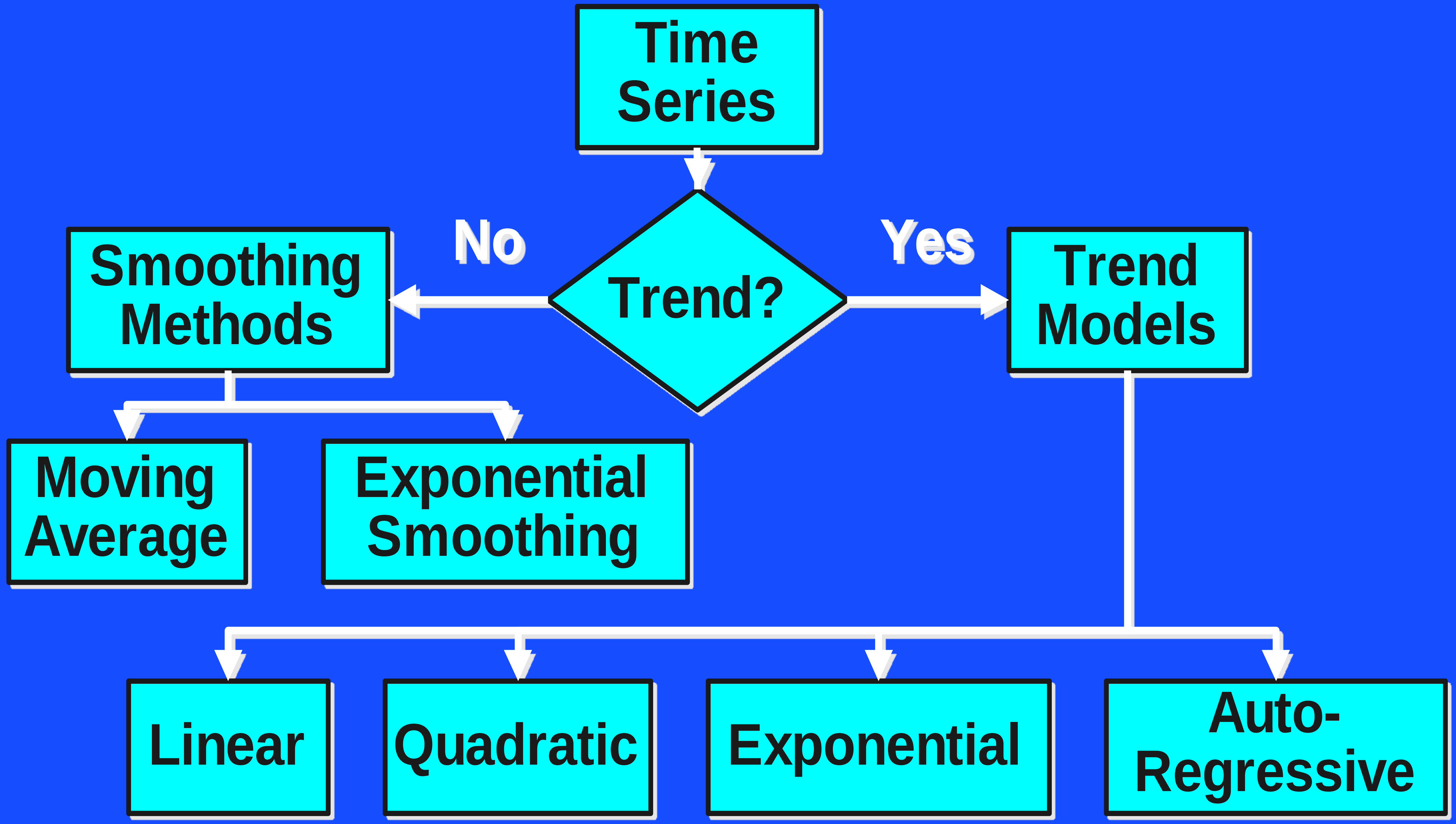
- **Let {Y_t}** be a stationery Time Series. The Auto Co-Variance Function (ACVF) of **{Y_t}** at lag h is

$$\gamma_y(t, t+h) = \mathrm{Cov}\,(Y_t, Y_{t+h}).$$

The Autocorrelation Function (ACF) of **{Y_t}** at lag h is

$$\rho_h = \frac{\gamma_h}{\gamma_0} = \frac{\mathrm{Cov}(Y_t, Y_{t+h})}{\sigma^2}$$

$\rho_{0=1}$


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

Time Series ⇨ Autocorrelation

Just as correlation measures the extent of a linear relationship between two variables, autocorrelation measures the linear relationship between *lagged values* of a time series.

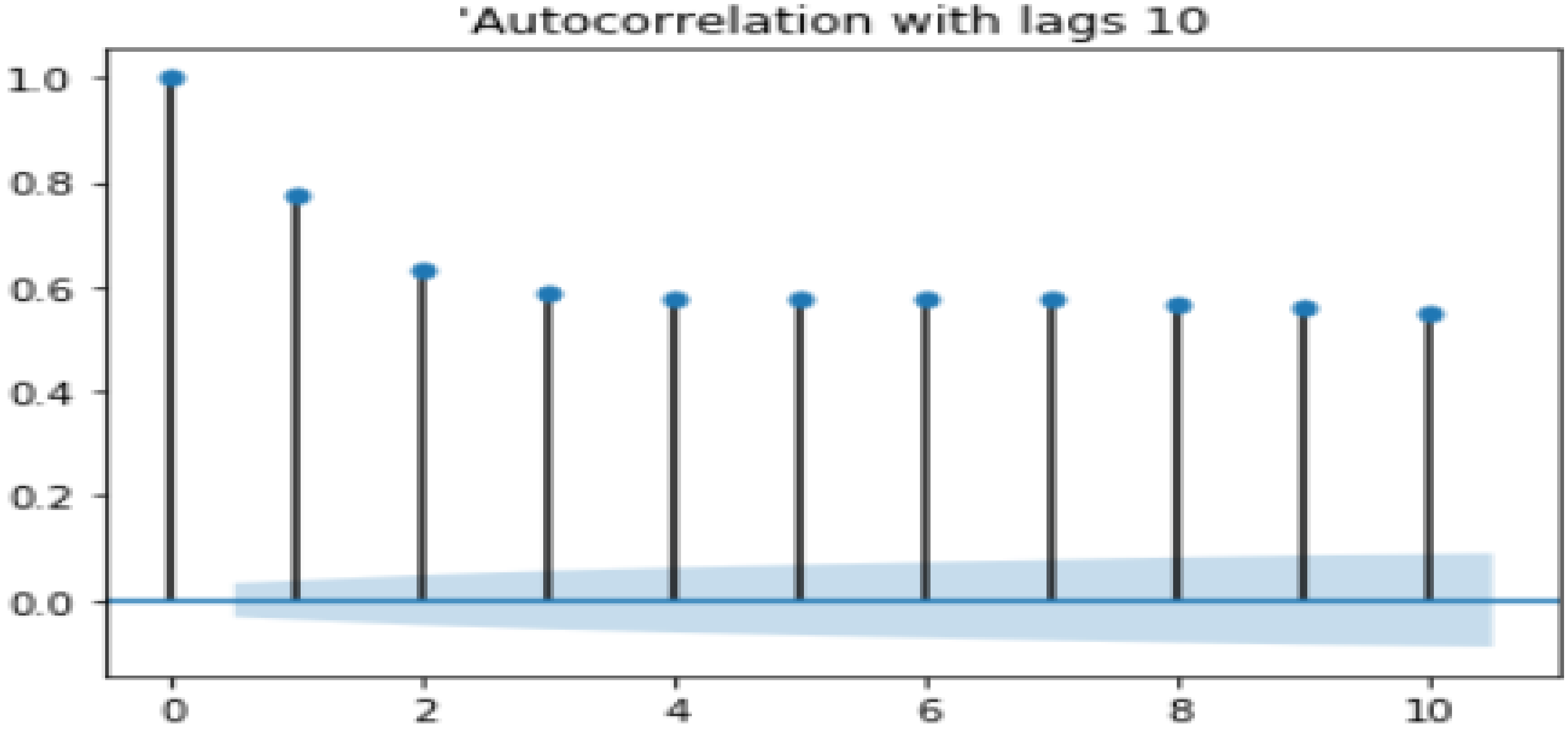
There are several autocorrelation coefficients, corresponding to each panel in the lag plot. For example, r_1 measures the relationship between y_t and y_{t-1}, r_2 measures the relationship between y_t and y_{t-2}, and so on

The value of r_k can be written as

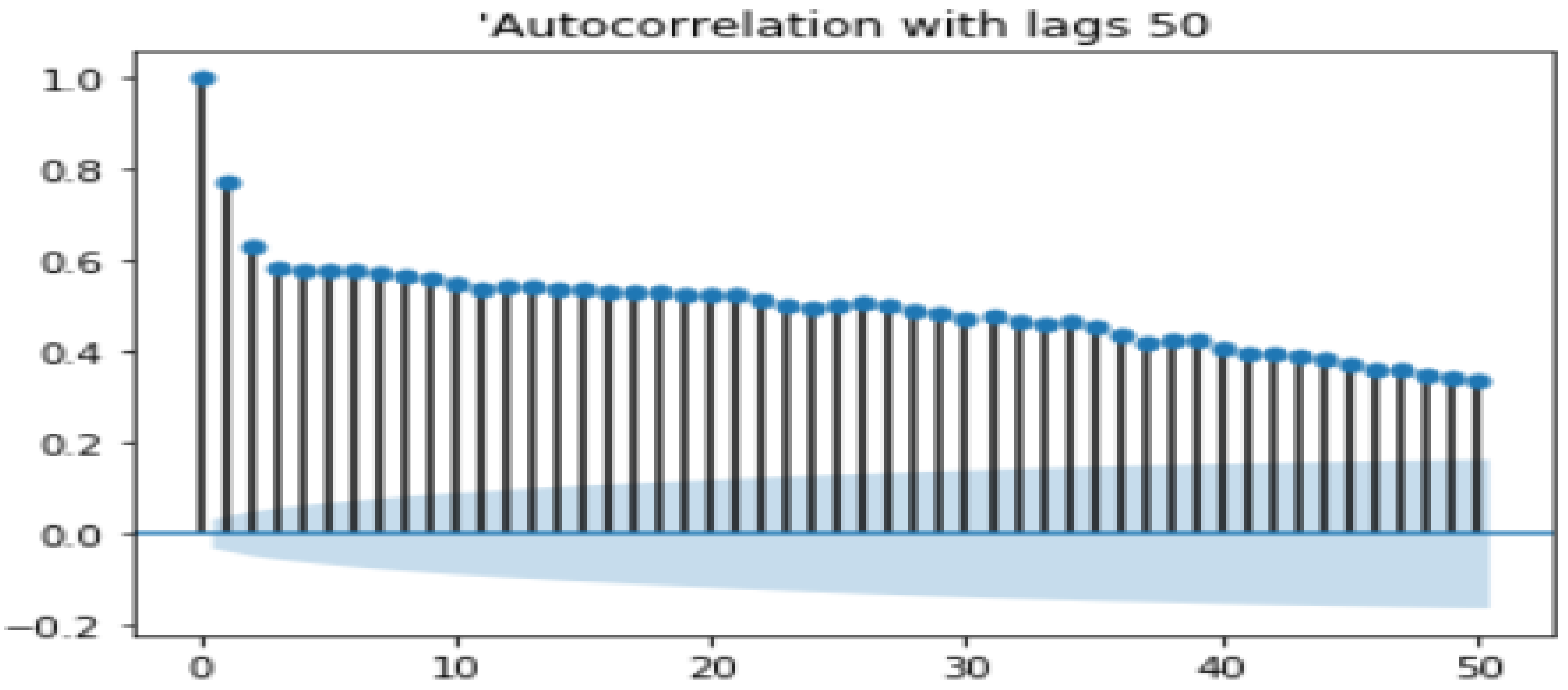
$$r_k = \frac{\sum_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T} (y_t - \bar{y})^2}$$

where T is the length of time series.

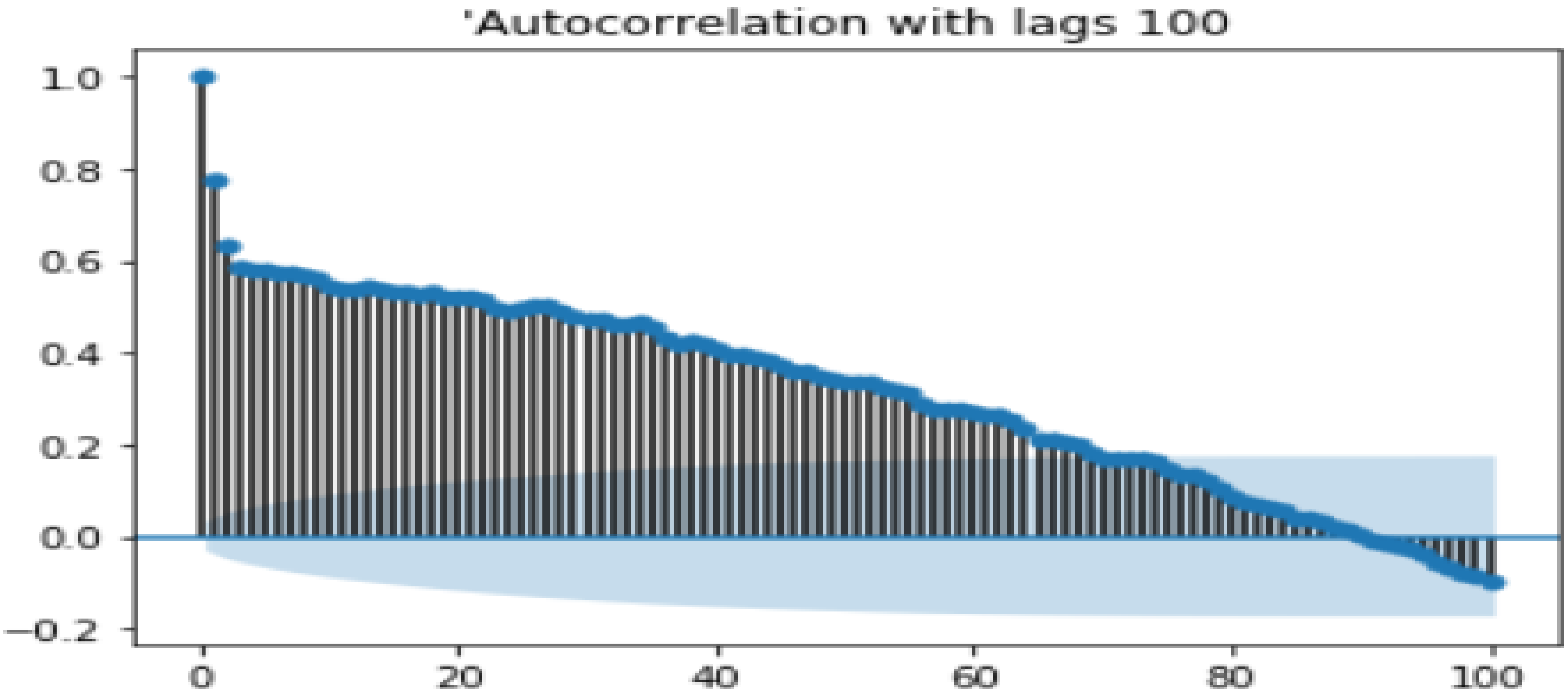
Time Series ➔ Autocorrelation



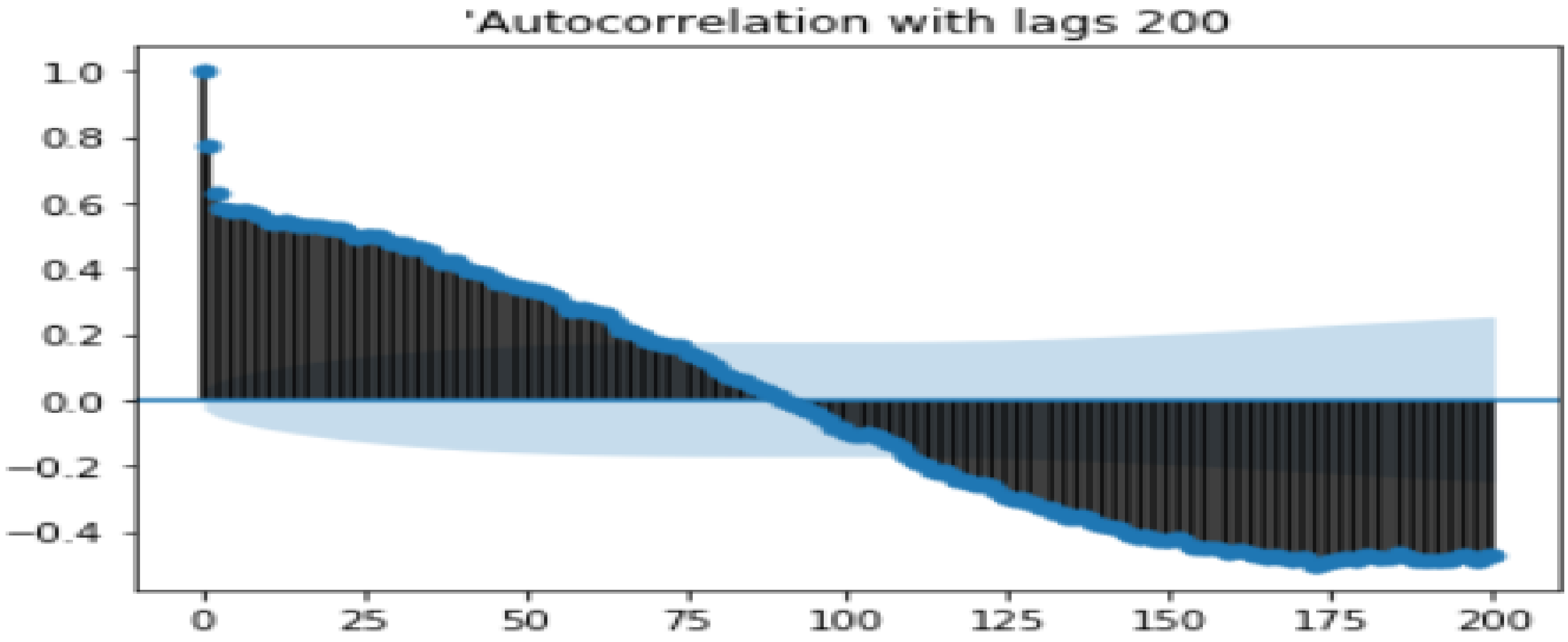
Time Series → Autocorrelation



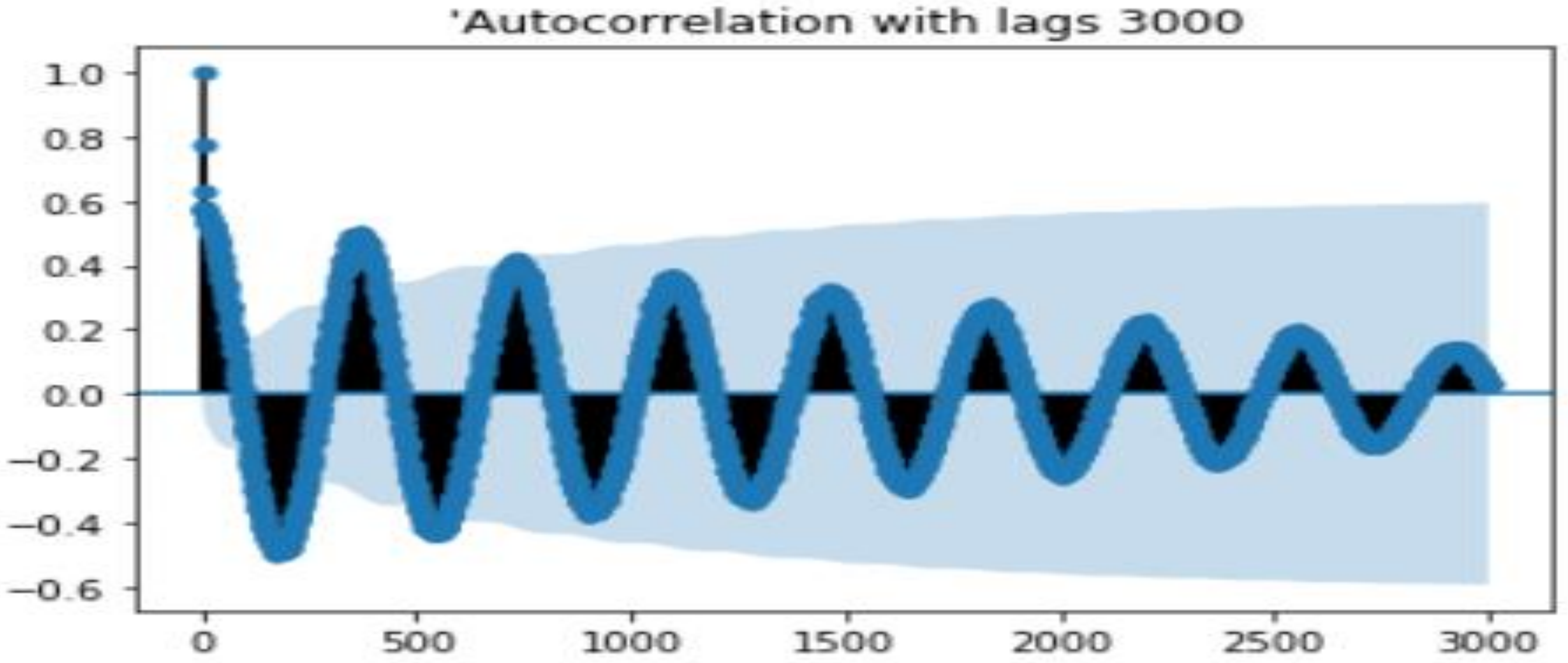
Time Series → Autocorrelation



Time Series ➡ Autocorrelation



Time Series → Autocorrelation



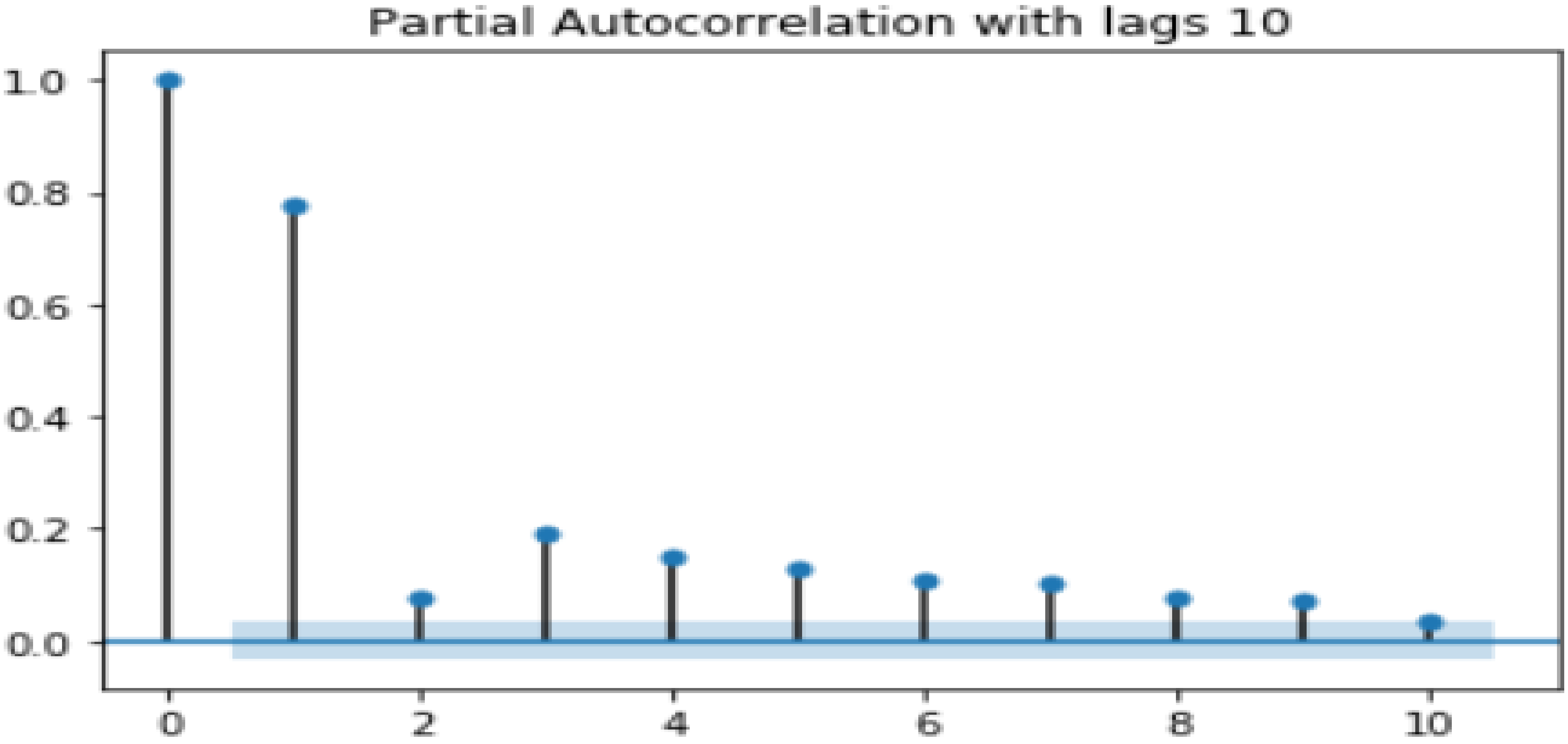
Time Series → Partial – Autocorrelation Function (PACF)

- The partial correlation between two variables is a conditional correlation taking into account their dependence on all other remaining variables
  - Eg. A third order (lag) partial autocorrelation is

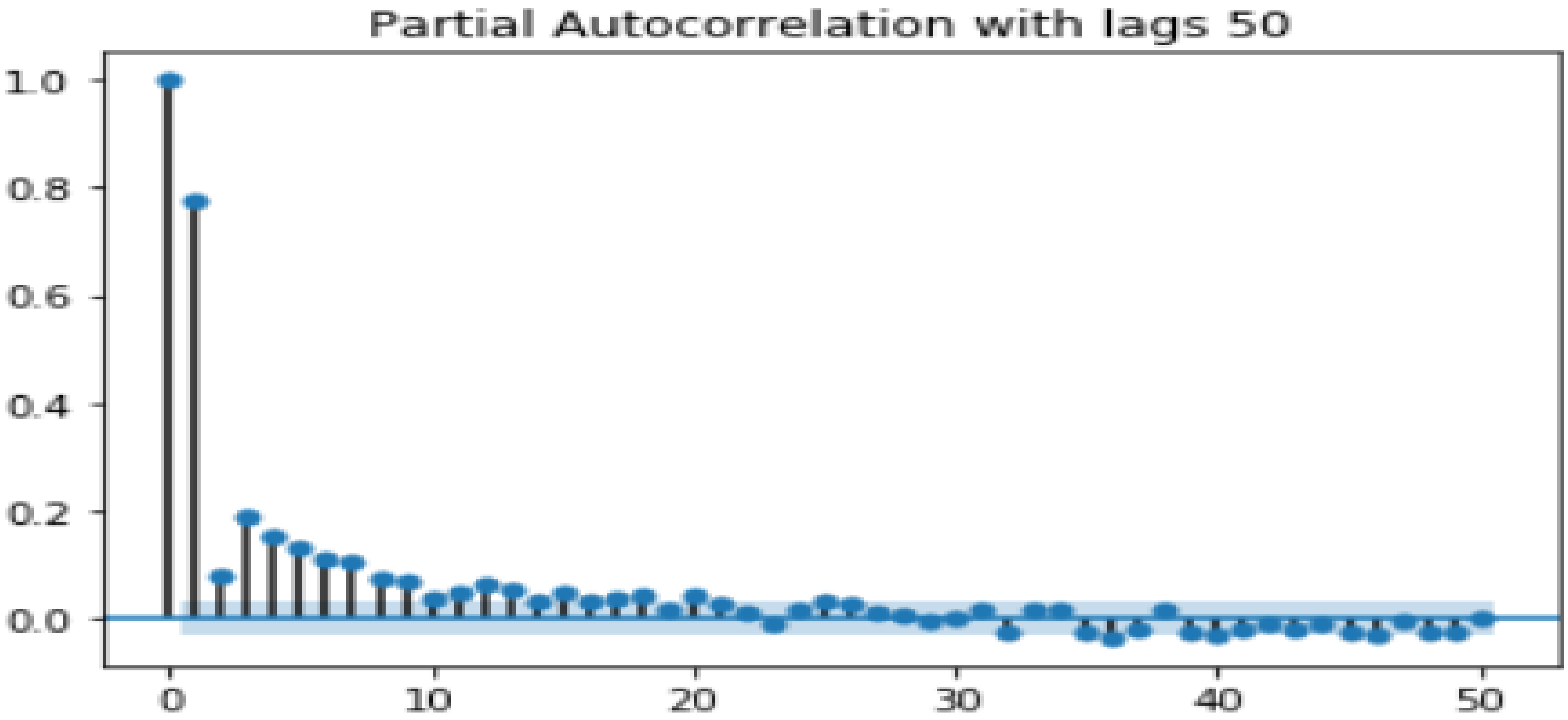
$$\frac{Cov(X_t, X_{t-3} | X_{t-1}, X_{t-2})}{\sqrt{Var(X_t | X_{t-1}, X_{t-2}) Var(X_{t-3} | X_{t-1}, X_{t-2})}}$$

- The first order PACF & ACF are same

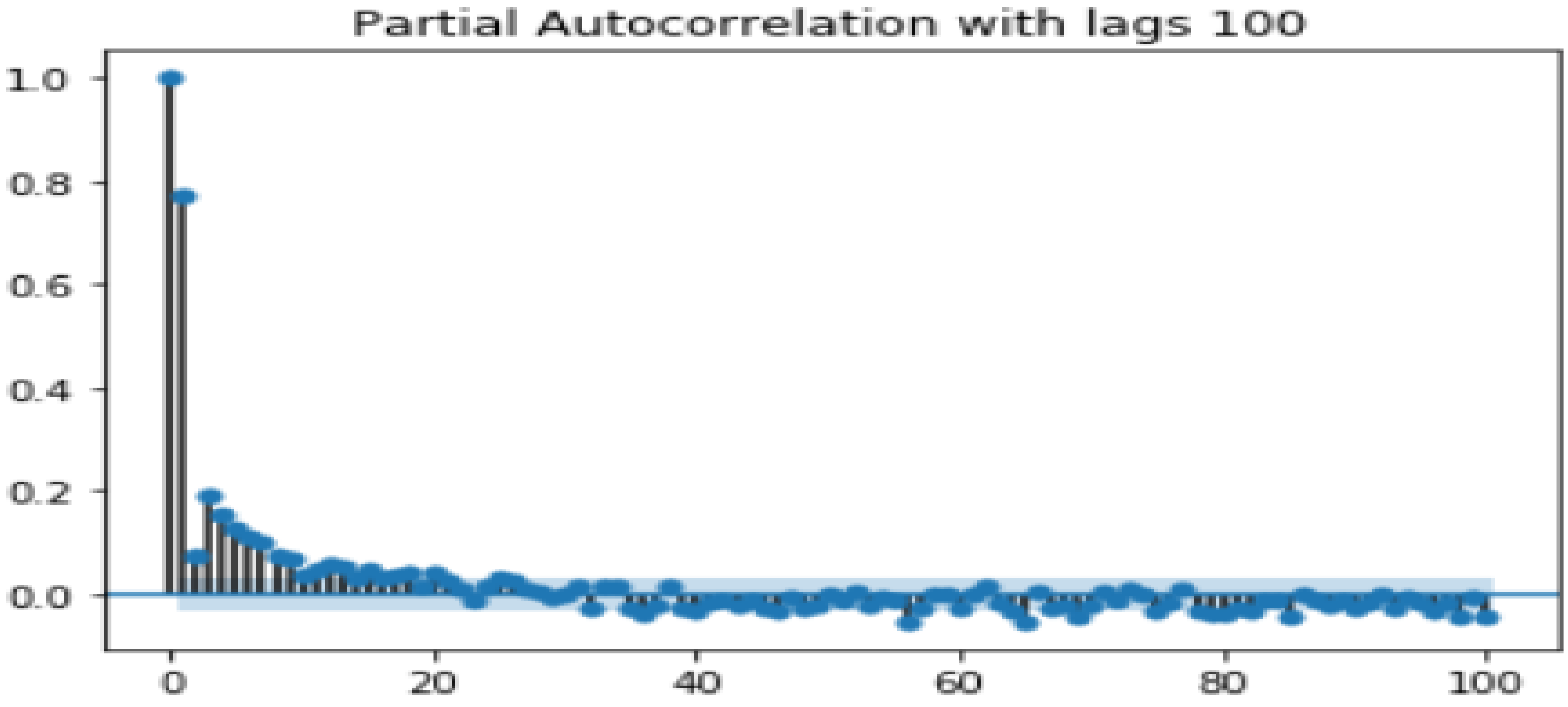
Time Series → Partial autocorrelation



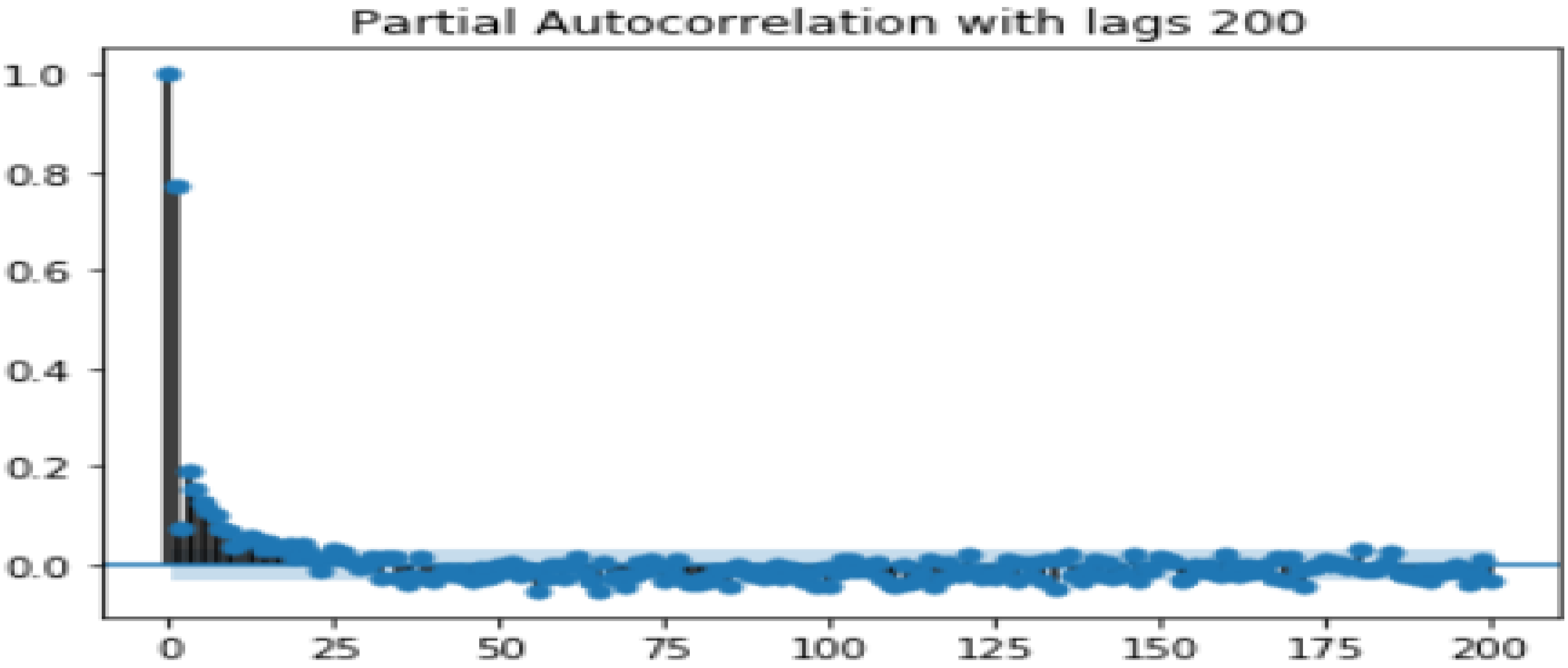
Time Series ➡ Partial autocorrelation



Time Series → Partial autocorrelation



Time Series → Partial autocorrelation



Time Series ⇨ Measurement of forecast error

- To try to determine which one of these forecasting methods gives the “best” forecast, we evaluate the “success” of each method using a ***performance measure***
- Each performance measure begins by evaluating the ***forecast errors*** Δ_t for each time period t given by

Time Series ➡ Measurement of forecast error

What was forecasted
to happen at time t using the
forecast technique

Forecast
Error
at time t

$$\Delta_t = y_t - F_t$$

What actually
happened
at time t

Time Series ⇨ Performance Measures

Mean Square Error (MSE)

$$MSE = \frac{\sum(\Delta_t)^2}{n}$$

Mean Absolute Percent Error (MAPE)

$$MAPE = \frac{\sum \frac{|\Delta_t|}{y_t}}{n} 100$$

Mean Absolute Deviation (MAD)

$$MAD = \frac{\sum|\Delta_t|}{n}$$

Largest Absolute Deviation (LAD)

$$LAD = Max\,|\Delta_t|$$

Time Series ➔ Which performance should be used?

- Choice of the modeler
- **Advantages**
  - MSE gives greater weight to larger deviations (which could result from outliers)
  - MAD gives less weight to larger deviations
  - MAPE gives less overall weight to a large deviation if the time series value is large
  - LAD tells us if all deviations fall below some threshold value
- Although we illustrate all 4 techniques, in general focus will be on MSE and MAD.

innovate achieve lead

Time Series ➡ Smoothing Techniques – Moving Averages

- The Moving Average basically filters out rapid fluctuations .i.e. high frequency noise. Thus it acts as a low-pass filter.
- We can choose the index of the MA filter by trial and error. Comparing the Forecast Accuracy for index selection.
- For our eg., we find that MSE for the 2-Qtr filter is better

Quarter/Year	Utilization (%)	2-Qtr MA	3-Qtr MA	4-Qtr MA
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			
	MSE			

Time Series ➔ Smoothing Techniques – Exponential

- Done the analysis on the previous problem for 3 different values of α
- For α=0.7, it is better
- The exponential smoothing does not have a significant initial delay like the Moving Av. where the initial k values are not smoothed
- Formula can be rearranged as

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

Quarter/ Year	Utilization (%)	F_t (α=0.2)	F_t (α=0.4)	F_t (α=0.7)
1/2003	82.50			
2/2003	81.30	82.50	82.50	82.50
3/2003	81.30	82.26	82.02	81.66
4/2003	79.00	82.07	81.73	81.41
1/2004	76.60	81.45	80.64	79.72
2/2004	78.00	80.48	79.02	77.54
3/2004	78.40	79.99	78.61	77.86
4/2004	78.00	79.67	78.53	78.24
1/2005	78.80	79.34	78.32	78.07
2/2005	78.70	79.23	78.51	78.58
3/2005	78.40	79.12	78.59	78.66
4/2005	80.00	78.98	78.51	78.48
1/2006	80.70	79.18	79.11	79.54
2/2006	80.70	79.49	79.74	80.35
3/2006	80.80	79.73	80.13	80.60
	MSE	**3.85**	**2.40**	**1.58**

Time Series → Autoregressive (AR) models

- An **autoregressive** (**AR**) **model** is a representation of some type of random process; used to describe certain time-varying processes. The autoregressive model specifies that the output variable depends linearly on its own previous values and on a stochastic (random) error term; thus the model is in the form of a stochastic difference equation.

- The Autoregressive model of order p denoted by AR (p) is given by

$$y_t = c_t + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + \varepsilon_t$$

where $\phi_1, \phi_2, \phi_3, \ldots \phi_p$ are the parameter of the model; c is constant and ε_t is the **white noise**.

Time Series ➔ White Noise

Time series that show no autocorrelation are called **white noise.** That is, if the variables are independent and identically distributed with a mean of zero. This means that all variables have the same variance (σ^2) and each value has a zero correlation with all other values in the series.

A white noise called **white** when it has the same intensity at every frequency. Its name is derived by analogy to light, which is **called** "**white**" when it contains all visible frequencies **noise.**

The term **noise**, in this context, came from signal processing where it was used to refer to unwanted electrical or electromagnetic energy that degrades the quality of signals and data

Time Series ⇨ White Noise

For white noise series, we expect each autocorrelation to be close to zero. Of course, they will not be exactly equal to zero as there is some random variation. For a white noise series, we expect 95% of the spikes in the ACF to lie within $\pm 2/\sqrt{T}$ where T is the length of the time series.

It is common to plot these bounds on a graph of the ACF (the blue dashed lines above). If one or more large spikes are outside these bounds, or if substantially more than 5% of spikes are outside these bounds, then the series is probably not white noise

- **Moving-average model**: The **moving-average model** specifies that the output variable depends linearly on the **current and various past** values of a stochastic term.

- Moving average model of order q (MA(q)):

$$y_t = w_t + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \cdots + \theta_q y_{t-q} + \varepsilon_t$$

Time Series ➔ Moving Average (MA) Model

- **Moving-average model** of order q (MA(q)):

$$X_t = w_t + \sum_{i=1}^{q} \theta_i w_{t-q} + \varepsilon_t$$

where: $\theta_1, \theta_2, \ldots, \theta_q$ are constants with $\theta_q \neq 0$; and w_t is **Gaussian white noise** $w_t\,(0, \sigma^2_w)$.

Note: **Gaussian noise**, named after Carl Friedrich **Gauss**, is **statistical noise** having a probability density function (PDF) equal to that of the normal distribution, which is also known as the **Gaussian** distribution. In other words, the values that the **noise** can take on are **Gaussian** distributed.

- **Autoregressive–moving-average model**. In the statistical analysis of time series Auto-Regressive Moving Average (ARMA) models provide a parsimonious description of a (weakly) stationary stochastic process in terms of two polynomials, one for the **autoregression** (AR) and the second for the **moving average** (MA).

Time Series ⇨ Autoregressive Moving Average (ARMA) Model

- The AR and MA models dynamics can be combined into what is called an *autoregressive moving-average (ARMA) model*.

- The ARMA (1, 1) is

$$y_t = \varphi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

$$\varepsilon_t \approx WN(0, \sigma^2)$$

Time Series ⇨ Autoregressive Moving Average (ARMA) Model

- Higher order ARMA processes involve additional lags of X and epsilon.
- The ARMA (1, 1) is

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p}$$

$$+ \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}, \ or$$

$$\Phi(L) y_t = \Theta(L) \varepsilon_t$$

Time Series ➡ Autoregressive Integrated Moving Average (ARIMA) Model

- ARIMA is an acronym that stands for Auto-Regressive Integrated Moving Average. Specifically,
  - AR ➡ Auto-regression: A model that uses the dependent relationship an observation and some number of **lagged observations**.
  - I ➡ Integrated: The use of **differencing** of raw observations in order to make the time series stationary.
  - MA ➡ Moving Average: A model that uses the dependency between an observation and **residual error** from a moving average model applied to lagged observations.

- Each of these are explicitly specified in the model as a parameter.
- Note that AR and MA are two widely used **linear models** that work on stationary time series and I is a **pre-processing to stationarize** time series if needed.

Time Series → **Autoregressive Integrated Moving Average (ARIMA) Model**

- **Rationale** - The first task is to provide a reason why we're interested in a particular model, as quants. Why are we introducing the time series model? What effects can it capture? What do we gain (or lose) by adding in extra complexity?
- **Definition** - We need to provide the full mathematical definition (and associated notation) of the time series model in order to minimise any ambiguity

Time Series ⇨ Choice of p, d and q

- **Second Order Properties** - We will discuss (and in some cases derive) the second order properties of the time series model, which includes its mean, its variance and its autocorrelation function
- **Correlogram** - We will use the second order properties to plot a correlogram of a realisation of the time series model in order to visualise its behaviour.

Time Series ➔ Choice of p, d and q

- **Simulation** - We will simulate realisations of the time series model and then fit the model to these simulations to ensure we have accurate implementations and understand the fitting process.

Time Series ⇨ Choice of p, d and q

- **Real Financial Data** - We will fit the time series model to real financial data and consider the correlogram of the residuals in order to see how the model accounts for serial correlation in the original series.
- **Prediction** - Create n - step ahead forecasts of the time series model for particular realisations in order to ultimately produce trading signals

Time Series ⇨ Choice of p, d and q

- Look at autocorrelation graph of data (will help if MA model is appropriate).
- Look at partial autocorrelation graph of data (will help if MR model is appropriate).
- Look at extended autocorrelation chart of data (will help if a combination of MA and AR models is needed).

Time Series → Choice of p, d and q

- Try Akaike’s information criterion (AIC) on a set of models and investigate the models with the lowest AIC values
- Try the Schwartz Bayesian information criterion (BIC) and investigate the models with the lowest BIC values
- All the above criterion to choose p, d and q are available as package in R.

Reference

- Peter J Brockwell and Richard A Davis. Introduction to Time Series and Forecasting, 2/e, Springer
- Douglas C. Montgomery, and Cheryl L. Jennings, Murat Kulahcin. *Introduction to* Time Series Analysis and Forecasting, 2/e, Wiley

Thank you