

Foundation Course on Business Forecasting

Talk @XIM University

Dr. Tanujit Chakraborty

Sorbonne University, Abu Dhabi, UAE

Sorbonne Centre for Artificial Intelligence, Paris, France

Research Areas: Time Series Forecasting, Machine Learning, Econometrics, Health Data Science

June 13, 2024

One-Day Course Roadmap

- History and Motivation
- Introduction to Time Series
- Business Forecasting and Applications
- Naïve Model
- Exponential Smoothing Method
- ARIMA
- Neural Networks for Time Series
- Python Implementation (next session)
- References

What people think I forecast?

When I go to any university, and I tell people that my job is time series forecasting and machine learning, usually one of two things happens:

- ...like, weather forecasting?

- Lots of domain knowledge and specialized models exist
- We leave it to meteorologists

- ...so, can you predict the stock market and we all get rich?

- I'll tell you how, and we're all going to be rich!
- Try it on your own risk!

What I forecast?

- Epidemic time series (e.g., dengue, malaria, hepatitis, etc.)
- Sales forecasting in the supply chain, retail at pharmacy companies
- Forecasting in climate
 - Air quality
 - El Nino
 - Seismic events
- Key macroeconomic variables (inflation, unemployment, exchange rate, etc.)
- ...

Can these be forecasted?

1. daily electricity demand in 3 days' time
2. Google stock price tomorrow
3. Google stock price in 6 months' time
4. maximum temperature tomorrow
5. total sales of drugs in pharmacies next month

Something is easy to forecast if:

1. we have a good understanding of the factors that contribute to it
2. there is a lot of data available
3. the future is somewhat similar to the past
 - ID assumption: samples are identically distributed
4. the forecasts cannot affect the thing we are trying to forecast.
 - self-fulfilling prophecies (election polls)
 - controlled systems
 - Big bull effect in stock markets / bitcoin prices

Past of Forecasting



- ❖ In ancient Babylon, forecasters would foretell the future based on the distribution of maggots in a rotten sheep's liver.



- ❖ Beginning in 800 BC, a priestess known as the [Oracle of Delphi](#) would answer questions about the future at the Temple of Apollo on Greece's Mount Parnassus.

Forecasters are to blame!

- ❖ Forecasters had a tougher time under the emperor Constantius, who issued a decree in AD357 forbidding anyone “to consult a soothsayer, a mathematician, or a forecaster -- May curiosity to foretell the future be silenced forever.”



- ❖ News report on 16 August 2006: A Russian woman is suing weather forecasters for wrecking her holiday. A court in Uljanovsk heard that Alyona Gabitova had been promised 28 degrees and sunshine when she planned a camping trip to a local nature reserve, newspaper Nowyje Iswestija said.
- ❖ But it did nothing but pour with rain the whole time, leaving her with a cold. Gabitova has asked the court to order the weather service to pay the cost of her travel.

Reputations can be made and lost

Some Misconceptions (Low Expectations): Our forecasts will always be inaccurate, so we should concentrate our efforts elsewhere.

“I think there is a world market for maybe five computers.

(Chairman of IBM, 1943)

“There is no reason anyone would want a computer in their home.”

(President, DEC, 1977)

“There’s no chance that the iPhone is going to get any significant market share. No chance.”

(Steve Ballmer, CEO Microsoft, April 2007)

“We’re going to be opening relatively soon ... The virus ... will go away in April.”

(Donald Trump, February 2020)

"Prediction is very difficult, especially if it's about the future!" - Niels Bohr

Reputations can be made and lost

Some Misconceptions (High Expectations): If only we had the latest forecasting technology, then all our problems could be solved.

Don't be fooled: Covid won't be cured by a panacea

Philip Ball

**The
Guardian**



'Cure-alls' such as vitamin D and ivermectin seem appealing. But the truth is, specific diseases demand specific medicines



ELSEVIER

International Journal of Forecasting

Volume 38, Issue 2, April–June 2022, Pages 423–438



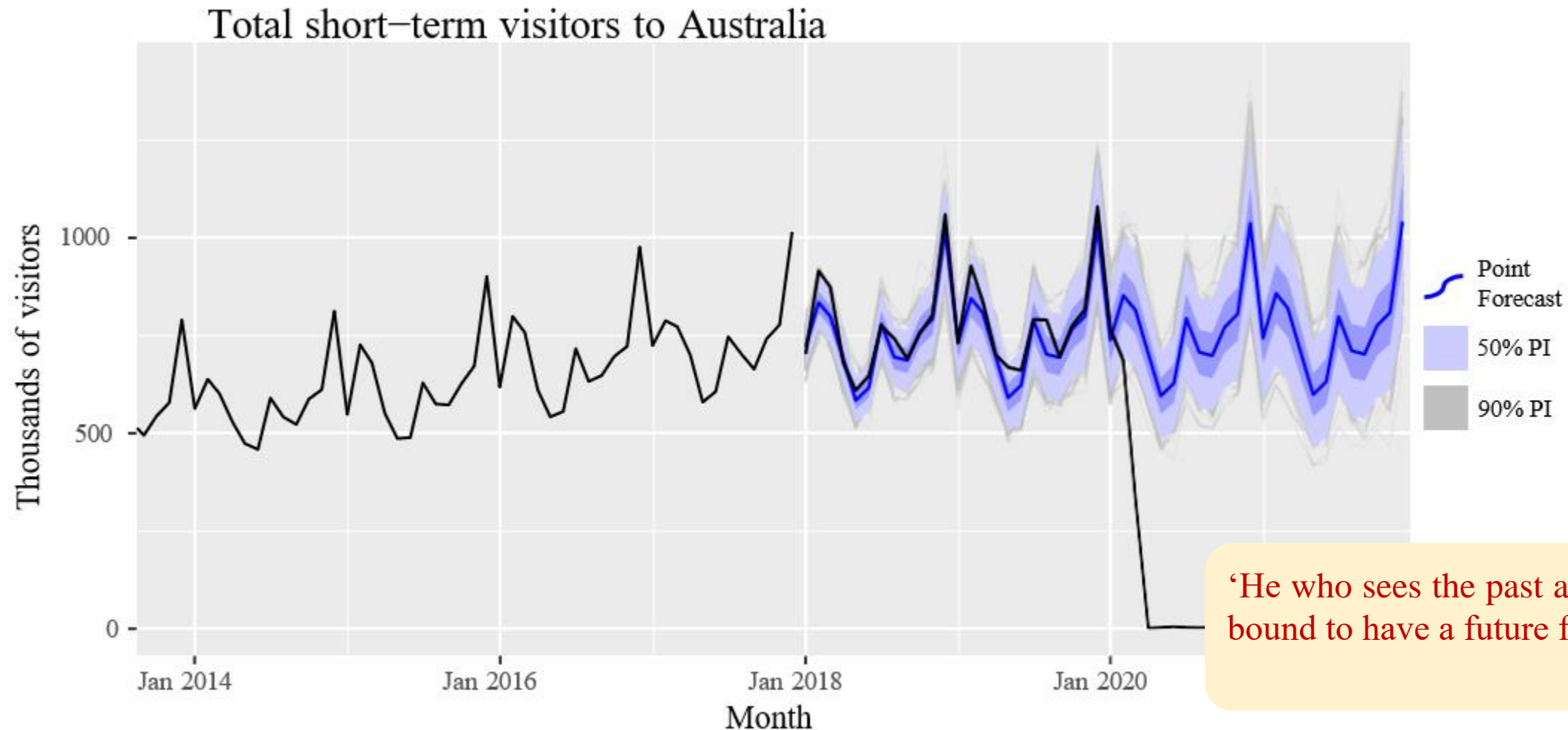
Forecasting for COVID-19 has failed

John P.A. Ioannidis^a  , Sally Cripps^b, Martin A. Tanner^c

- Poor data input
- Wrong modeling assumptions
- Lack of incorporation of epidemiological features
- Poor past evidence on effects of available interventions

- Lack of transparency
- Consideration of only one or a few dimensions of the problem at hand
- Lack of expertise in crucial disciplines
- Groupthink and bandwagon effects

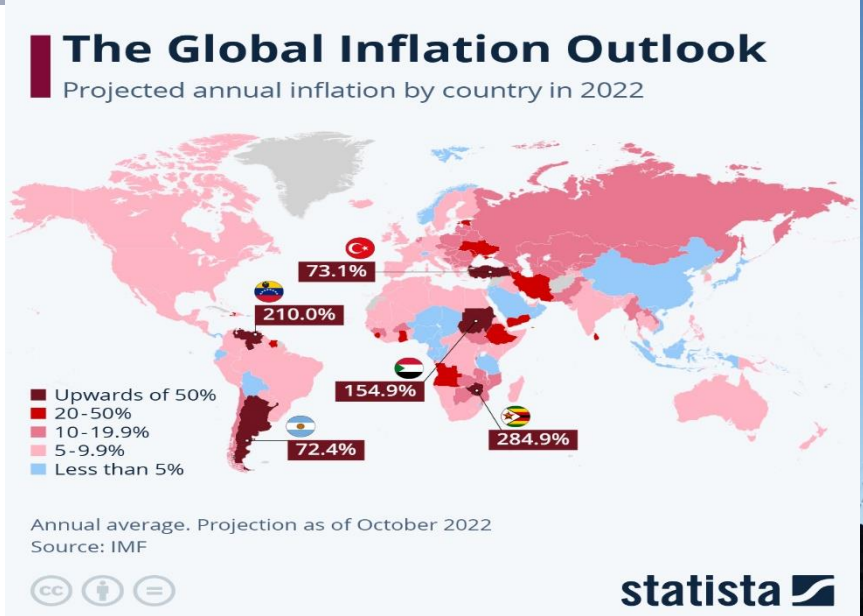
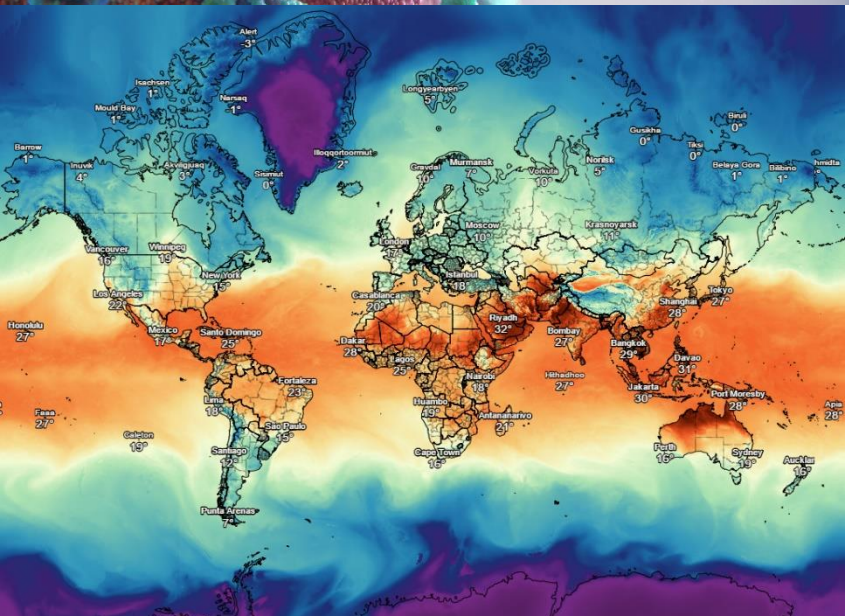
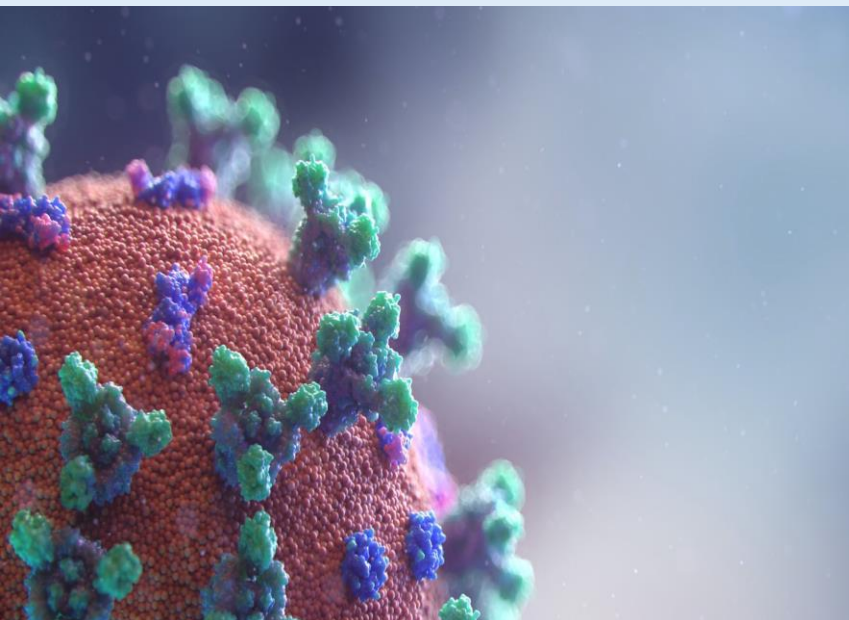
Uncertainty and Forecasting



‘He who sees the past as surprise-free is bound to have a future full of surprise.’

- Amos Tversky

What can we forecast?



What is Time Series?

Introduction

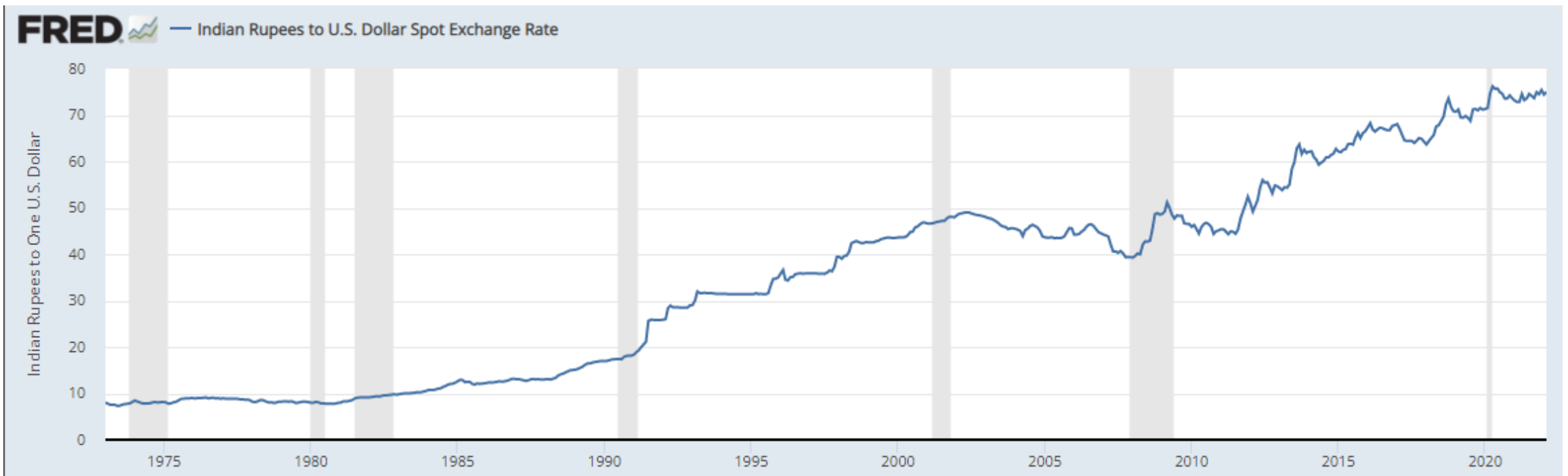
- **Time series** is a set of observations, each one being recorded at a specific time. (e.g., Annual GDP of a country, Sales figure, etc.)
- **Discrete time series** is one in which the set of time points at which observations are made is a discrete set. (e.g., All above including irregularly spaced data)
- **Continuous time series** are obtained when observations are made continuously over some time intervals. (e.g., ECG graph)
- **Forecasting** is estimating how the sequence of observations will continue in to the future. (e.g., Forecasting of major economic variables like GDP, Unemployment, Inflation, Exchange rates, Production and Consumption)
- **Forecasting** is very difficult, since it's about the future! (e.g., forecasts of daily cases of COVID-19)

Use of Time Series Data

- To develop forecast model
 - What will be the rate of inflation in next year?
- To estimate dynamic causal effects
 - If the rate of interest increases, what will be the effect on the rates of inflation and unemployment in 3 months? in 12 months?
 - What is the effect over time on electronics good consumption due to a hike in the excise duty?
- Time dependent analysis
 - Rates of inflation and unemployment in the country can be observed over a time period.

Time Series Data: Exchange Rate data

- **Time-series data:** The data collected **on the same observational unit** at multiple time periods
- Source: **FRED ECONOMICS DATA** (Shaded areas indicate US recessions)
- Units: **Indian Rupees to One U.S. Dollar, Not Seasonally Adjusted**
- Frequency: **Monthly (Averages of daily figures)**



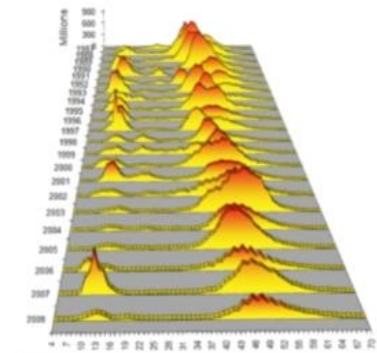
Shape of Time Series Data

- We usually think that the data is one-dimensional.
- It only consists of the time and the data associated with it (Temp.).

	New York City
1990-01-01	1
1990-01-02	5
1990-01-03	9
1990-01-04	13

- But it can be Multidimensional.

	New York City	London	Tokyo	Paris
1990-01-01	1	2	3	4
1990-01-02	5	6	7	8
1990-01-03	9	10	11	12
1990-01-04	13	14	15	16



Time Series Components

- **Trend (T_t)** : pattern exists when there is a long-term increase or decrease in the data.
- **Seasonal (S_t)** : pattern exists when a series is influenced by seasonal factors (e.g., the quarter of the year, the month, or day of the week).
- **Cyclic (C_t)** : pattern exists when data exhibit rises and falls that are not of fixed period (duration usually of at least 2 years).
- **Decomposition** : $Y_t = f(T_t; S_t; C_t; I_t)$, where Y_t is data at period t and I_t is irregular component at period t .
- **Additive decomposition** : $Y_t = T_t + S_t + C_t + I_t$
- **Multiplicative decomposition** : $Y_t = T_t * S_t * C_t * I_t$
- **A stationary series is** roughly horizontal, constant variance and no patterns predictable in the long-term.

Stationary Time Series

Stationary Series:

- A series free from trend and seasonal patterns
- A series exhibits only random fluctuations around mean
- A stationary time series exhibits similar statistical behavior in time and this is often characterized by a constant probability distribution in time

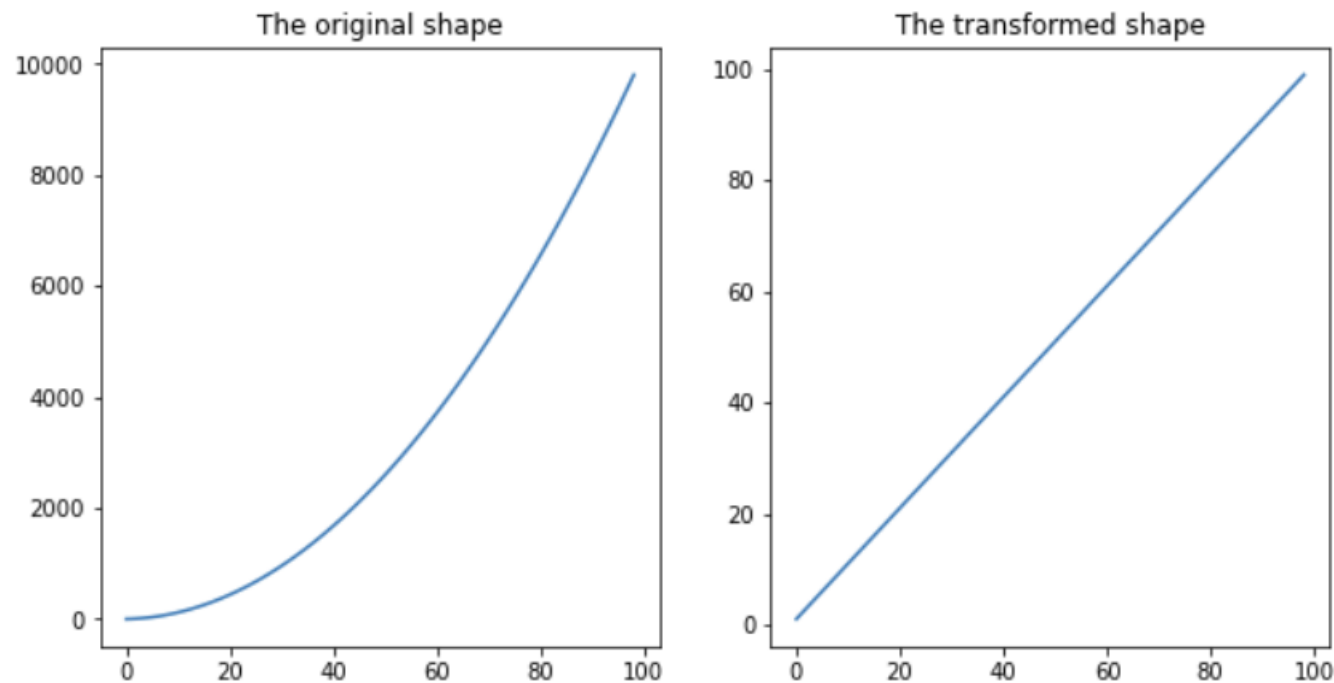
Unit root test Augmented Dickey Fuller Test (ADF) :

- Checks whether any specific patterns exists in the series H_0 : data is non-stationary
- H_1 : data is stationary
- A small p-value suggest data is stationary

Transformation in Time Series

Power Transformation

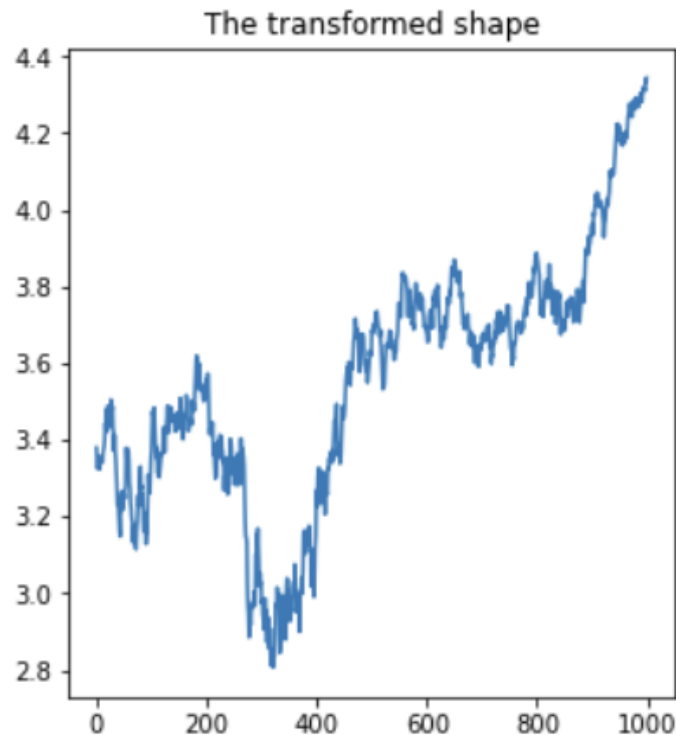
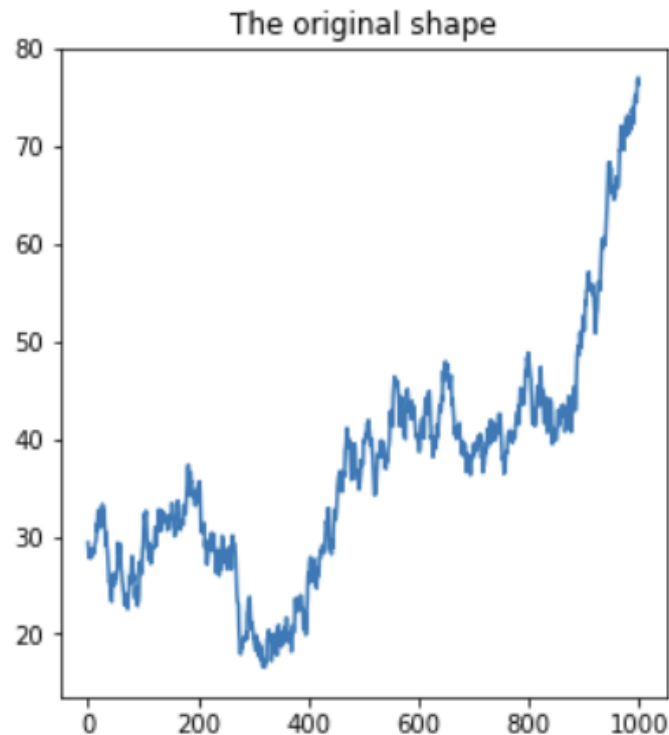
- This means, we will take the power of **some** value for each values on the time-series. Here, we will **need** to understand the nature of the data first before applying any transformation.



- We have done this because **some model returns better results when the "form is linear"**. Thus, that time we want to transform them in to some linear form. And for that you need to understand the "nature" of the data.

Log Transformation

- This is the fundamental transformation of all.
- This can be used to "squash" the data in the smaller range.
- This can be the **default** transformation too!
- It has the frequent application in *finance*.



Box-cox Transformation

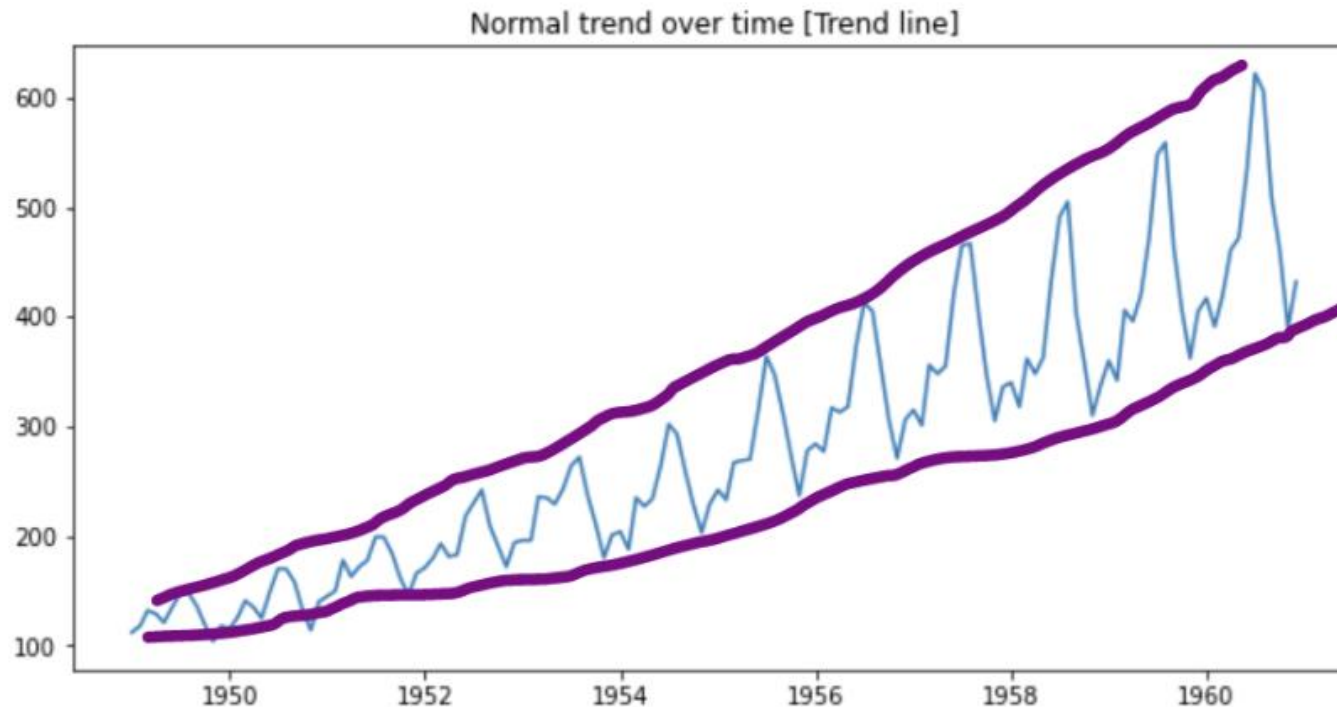
- This is the **generalization** of the ‘Power’ and ‘Log’ transformation.
- It is used to achieve ‘**normality**’ in the *non-normal* variables.
- Unifies the power and log transforms.
- λ is chosen automatically by the *boxcox()* function in *Scipy*
- “Estimating Box-Cox power transformation parameter via goodness of fit tests”.

$$y'(t) = \frac{y(t)^\lambda - 1}{\lambda} \quad \text{if } \lambda \neq 0$$

$$y'(t) = \log y(t) \quad \text{if } \lambda = 0$$

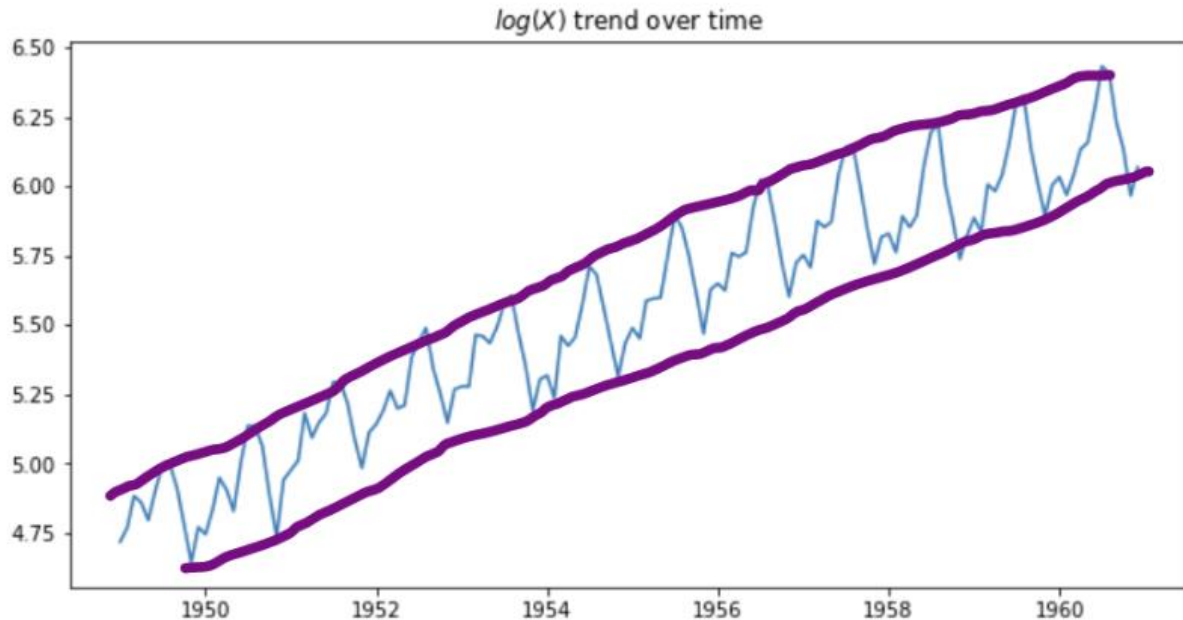
$$\text{Since, } \lim_{\lambda \rightarrow 0} \frac{\lambda^\lambda - 1}{\lambda} = \ln x$$

Air Passengers data



- Here, we can see that the trend is **increasing** and **too** the magnitude.
- There is the **seasonality**: On each year of some specific months the amount rises and then dips
- The amount is increasing i.e., the seasonality is same but the magnitude is increasing over time: *which the model has to consider.*

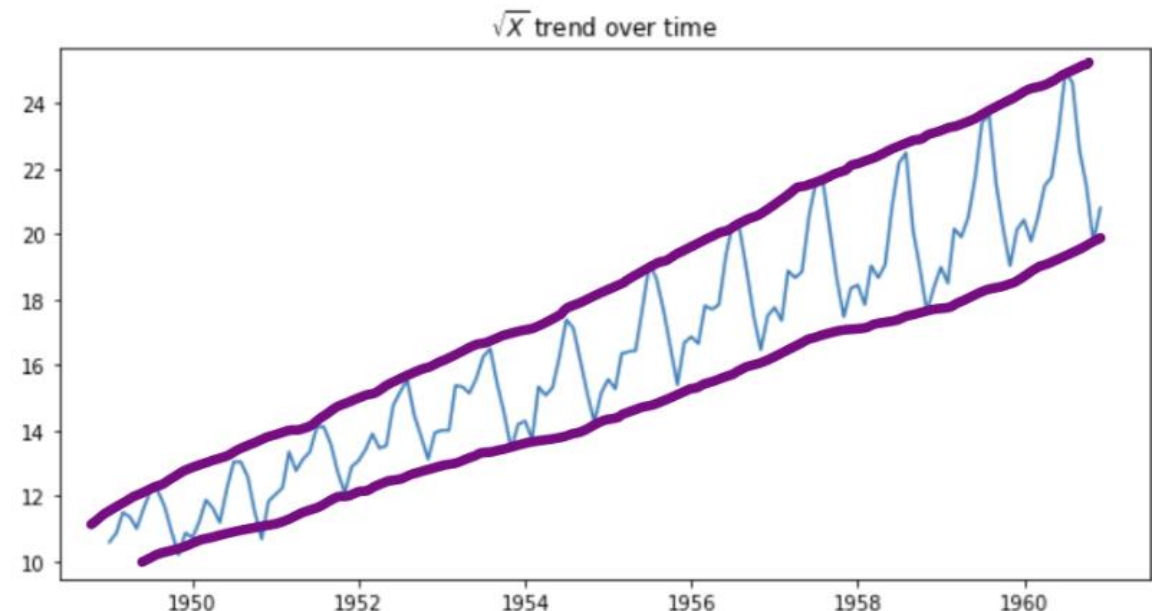
Transformation on Air Passengers data



It is more **even on the both sides?** like, the magnitude is not changing much as it did in the normal data.

Other Transformation: **Fourier, Wavelet**

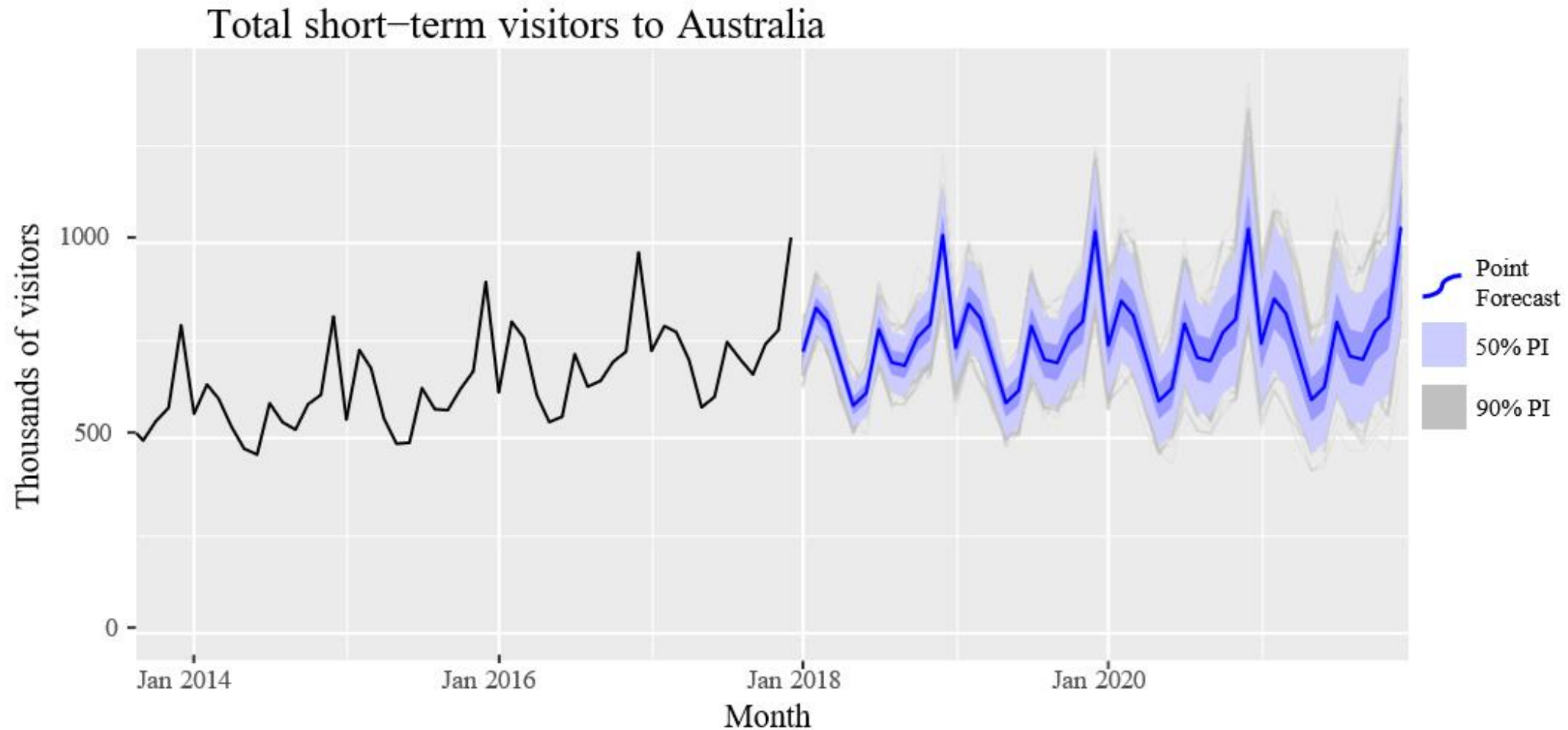
It still has some increase in the magnitude but **not much as the original data.**



Forecasting

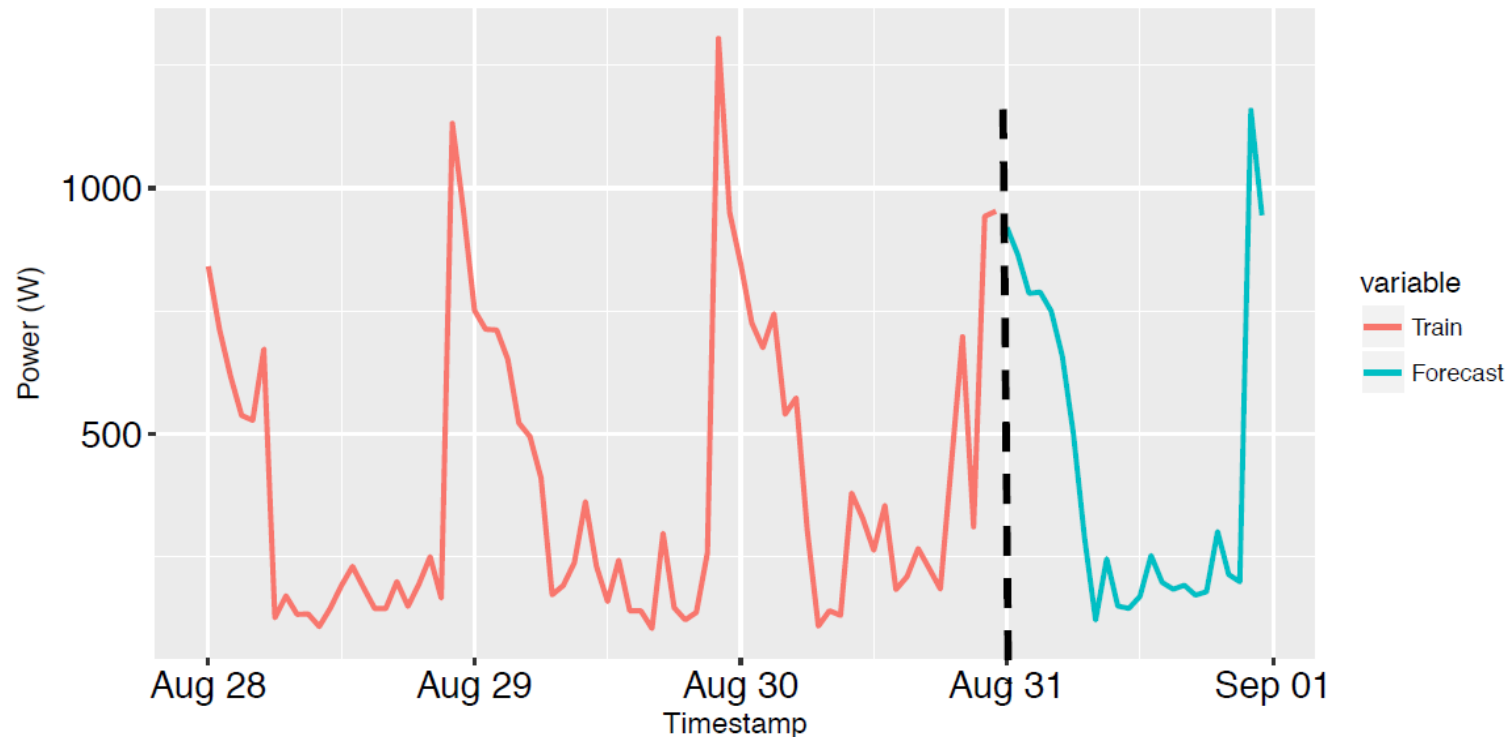
What is Forecasting?

Forecasting is estimating how the sequence of observations will continue into the future.



Forecasting: Assumptions

- **Time series Forecasting:** Data collected at regular intervals of time (e.g., Electricity Forecasting).
- **Assumptions:** (a) Historical information is available;
(b) Past patterns will continue in the future.



Types of Time Series Tasks

1. One-Step Forecasting

2. Multi-Step Forecasting

- Incremental Multi-step
- Multi-output Multi step

One-Step Forecast

Here we will forecast for only one time - Very next frame.

Types of Tasks

- 1-step forecast



For time series, I'll use x and y interchangeably, unlike supervised learning where x = input and y = target



Multi-Step Forecast

The **most of the time** we would be needing such forecast.

— The term used is: "**Forecast horizon**" = Multiple steps in the future.

Now, there are **2** ways to produce such "**Multi step**" forecasts.

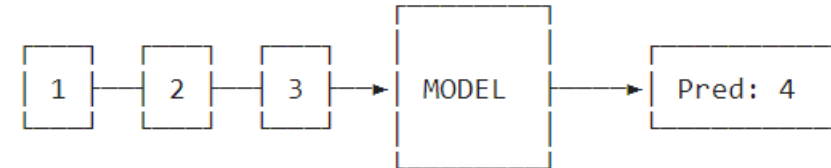
1. **Incremental method** (which can be done with any 1 step predictor)
2. **Multi-output forecast** (limited to certain models)

Remember: These are just the methods not the types of models.

Incremental Multi-Step Forecast

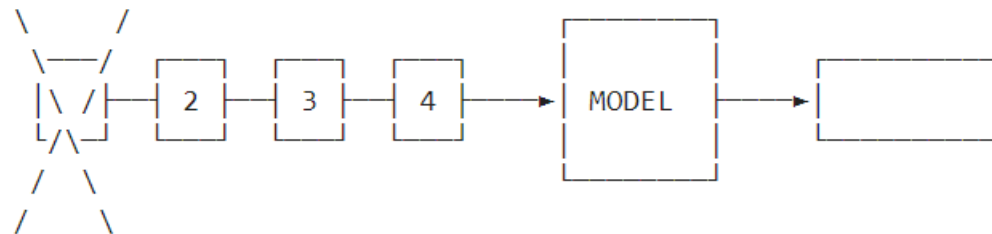
The model can be trained on the fixed n days. So, suppose

Here, our $p = 3$ and $h = 2$.



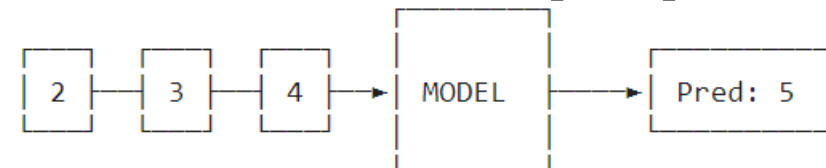
- p = How many days do you want to base your forecast on?
- h = How many days do you want to forecast?

Now, in the first iteration we have forecast 4th day. But still one day is yet to be forecast so we will take day 4 as the input.



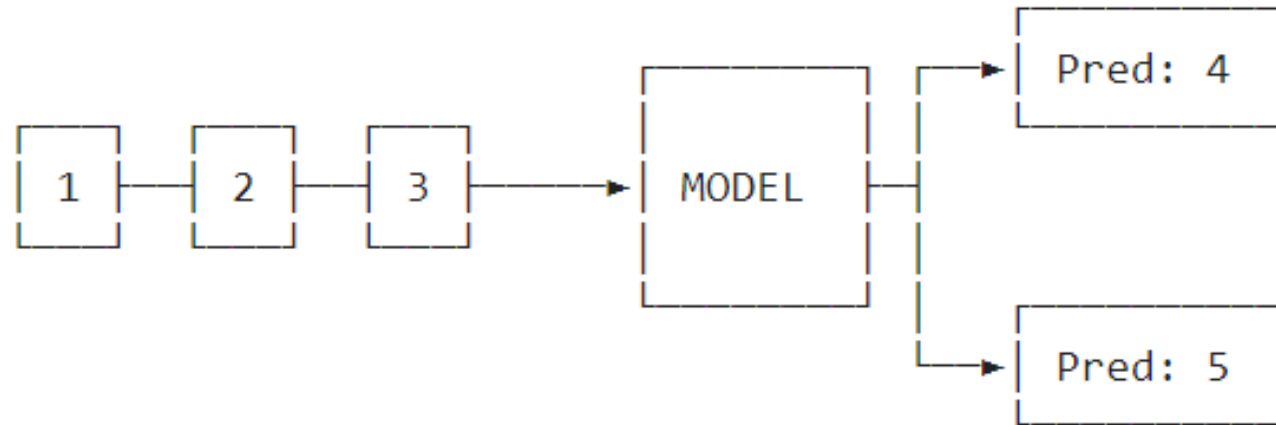
So, instead of increasing the number of days, we will **eliminate the first ones** to keep the p consistent.

This is how **incremental multi-step** works in a nutshell.



Multi Output Forecast

Here, we feed the time-series in and get the **h days out and at once**.



Multi-step Forecasting

Short -term	Involves predicting events only a few time periods (days, weeks, months) into future
Medium - term	Involves predicting events to one or two years into future
Long - term	Involves predicting events to many years into future

Type	Uses
Short –term & Medium-term	Budgeting and selecting new research for development projects
Long - term	Strategic planning

Business Forecasting Process

1. **Problem definition:** Involves understanding how the forecast will be used, expectations of the customer or user of the forecast, how often forecast need to be revised, what level of forecast accuracy is required to make good business decision.
2. **Data collection:** Consists of obtaining the relevant history for the variable(s) to be forecasted, including historical information on potential predictor variables.
3. **Preliminary data analysis:** Needed for selection of the appropriate forecasting model, to identify the patterns such as trends, seasonal and other cyclic components. Numerical summaries such as sample mean, standard deviation, percentiles, stationarity, nonlinearity, and auto correlations need to be computed and evaluated. Unusual observations or potential outliers need to be identified and flagged for possible further study. If predictor variables are involved, scatter plots of each pair of variables should be examined.

Business Forecasting Process

4. **Model selection and fitting:** Choosing one or more forecasting models and fitting the model to the data (estimating the unknown parameters of the model).
4. **Model validation:** Evaluation of the forecasting model to determine how it is likely to perform in the intended application. The magnitude of forecast error need to be examined not only on historical data but also on fresh or new data.
5. **Forecast model deployment:** Consists of using the model to forecast the future values of the variable of interest for the customer.
6. **Monitoring forecasting model performance:** An ongoing activity after the model has been deployed to ensure that the model still performing satisfactorily. Monitoring of forecast errors is an essential part of good forecasting system design.

Auto-regression Analysis

Auto Regression Analysis

- Regression analysis for time-ordered data is known as **Auto-Regression Analysis**
- **Time series data** are data collected on the same observational unit at multiple time period.



Example: Indian rate of price inflation

Modeling with Time Series Data

- Correlation over time
 - Serial correlation, also called autocorrelation
 - Calculating standard error
- To estimate dynamic causal effects
 - Under which dynamic effects can be estimated?
 - How to estimate?
- Forecasting model
 - Forecasting model build on regression model



Can we predict the trend (USD vs. INR) at a time, say 2025?

Some Notations and Concepts

- Y_t = Value of Y in a period t
- Data set $[Y_1, Y_2, \dots, Y_{T-1}, Y_T]$: T observations on the time series random variable Y

There are four ways to have the time series data for AutoRegression analysis

- Lag: The first lag of Y_t is Y_{t-1} , its j -th lag is Y_{t-j}
- Difference: The first difference of a series, Y_t is its change between period t and $t - 1$,
that is, $y_t = Y_t - Y_{t-1}$
- Log difference: $y_t = \log(Y_t) - \log(Y_{t-1})$
- Percentage: $y_t = \frac{Y_{t-1}}{Y_t} \times 100$

Related Concepts and Notations

Assumptions

1. **Uniform:** We consider only consecutive, evenly spaced observations

For example, say monthly data in 2010-2021 for each year, and without any missing month(s); no other data, for example, on daily basis for a year is admissible.

2. **Stationarity:** A time series Y_t is stationary if its probability distribution does not change over time, that is, if the joint distribution of $(Y_{i+1}, Y_{i+2}, Y_{i+3}, \dots, Y_{i+T})$ does not depend on i .

Stationary property implies that history is relevant. In other words, stationary requires the future to be like the past (in a probabilistic sense).

Auto-regression analysis assumes that Y_t is both uniform and stationary.

Auto-correlation coefficient

The correlation of a series with its own lagged values is called autocorrelation (also called serial correlation)

Formula: j^{th} Autocorrelation

The j^{th} autocorrelation, denoted by ρ_j is defined as

$$\rho_j = \frac{\text{Cov}(Y_t, Y_{t-j})}{\sigma_{Y_t} \sigma_{Y_{t-j}}}$$

where, $\text{COV}(Y_t, Y_{t-j})$ is the j^{th} **auto-covariance**

Covariance

Formula: $Cov(Y_t, Y_{t-j})$

The covariance between the variables Y_t and Y_{t-j}

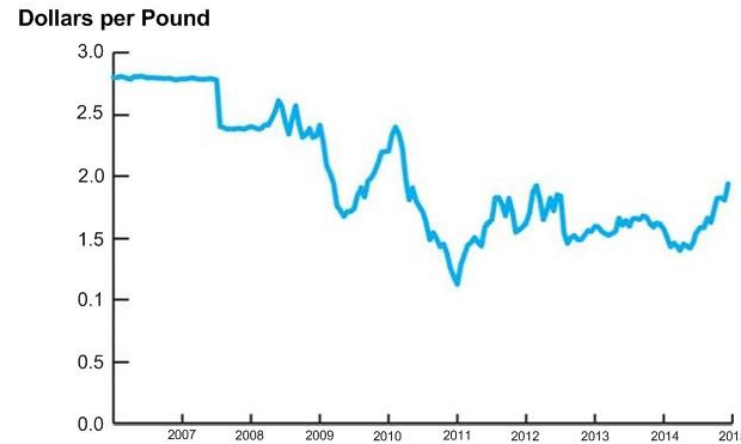
$$Cov(Y_t, Y_{t-j}) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

n is the number of observations

Y_{t-j}	\dots	Y_t
x_1		y_1
x_2		y_2
\dots		\dots
x_j		y_j
\cdot		\cdot
\cdot		\cdot
\cdot		\cdot
x_n		y_n

Example: Autocorrelation

Example



- For the given data, say $\rho_1 = 0.84$ between two given consecutive years
 - **This implies that the Dollars per Pound is highly serially correlated**
- Similarly, we can determine ρ_2 , ρ_3, \dots etc.

Popular Forecasting Techniques

Random Walk

Introduction

- In time series analysis, a random walk is a stochastic process where future values are determined by previous values plus a random shock.
- Understanding the random walk is crucial for modeling and forecasting in various fields.

What is it?

- It is about *those* stocks which follow a **random walk** i.e. its price is not predictable.
- "*Sometimes the best-fitting model is in fact a random walk*"
- It is either to go up or down randomly having 50-50 chance on each direction
- It is *impossible* predict the next as there is 50-50 chances

Characteristics of a Random Walk

- **Constant Drift:**

- A random walk typically exhibits a constant drift or trend over time.
- The drift reflects the average rate of change in the series.

- **Unpredictable Movement:**

- Future values of a random walk are unpredictable.
- Each step depends solely on the current position and a random shock, making forecasting challenging.

- **No Auto-correlation:**

- A random walk typically has no autocorrelation.
- The correlation between observations at different lags is close to zero.

Gaussian Random Walk

- It is where the up and down values come from the Gaussian distribution
- So, here the prices don't go 1 unit up or down but go any value from the Gaussian distribution, thus called: Gaussian Random Walk

$$new = old + e$$

- Since there is unpredictability, we can only know that the error $e \in N(0, \sigma^2)$
- **Something interesting that we can do with log**
- The general formulae for new price (in random walk):

$$new = old + \mu + e$$

- μ = Drift - This would control the trend of the time series.
- e = The noise $\in N(0, \sigma^2)$

Naïve Forecasting Methods

Average Method

- Here, the forecasts of all future values are equal to the average (or "mean") of the historical data. If we let the historical data be denoted by y_1, \dots, y_T , then we can write the forecasts as

$$\hat{y}_{T+h|T} = \bar{y} = (y_1 + \dots + y_T)/T.$$

- The notation $\hat{y}_{T+h|T}$ is a short-hand for the estimate of y_{T+h} based on the data y_1, \dots, y_T .

Naive Method

- For naïve forecasts, we simply set all forecasts to be the value of the last observation.

That is,

$$\hat{y}_{T+h|T} = y_T.$$

- This method works remarkably well for many economic and financial time series.

Seasonal Naive Method

- A similar method is useful for highly seasonal data. In this case, we set each forecast to be equal to the last observed value from the same season (e.g., the same month of the previous year). Formally, the forecast for time $T + h$ is written as

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)},$$

where m = the seasonal period, and k is the integer part of $(h - 1)/m$ (i.e., the number of complete years in the forecast period prior to time $T + h$). This looks more complicated than it really is.

- For example, with monthly data, the forecast for all future February values is equal to the last observed February value. With quarterly data, the forecast of all future Q2 values is equal to the last observed Q2 value (where Q2 means the second quarter). Similar rules apply for other months and quarters, and for other seasonal periods.

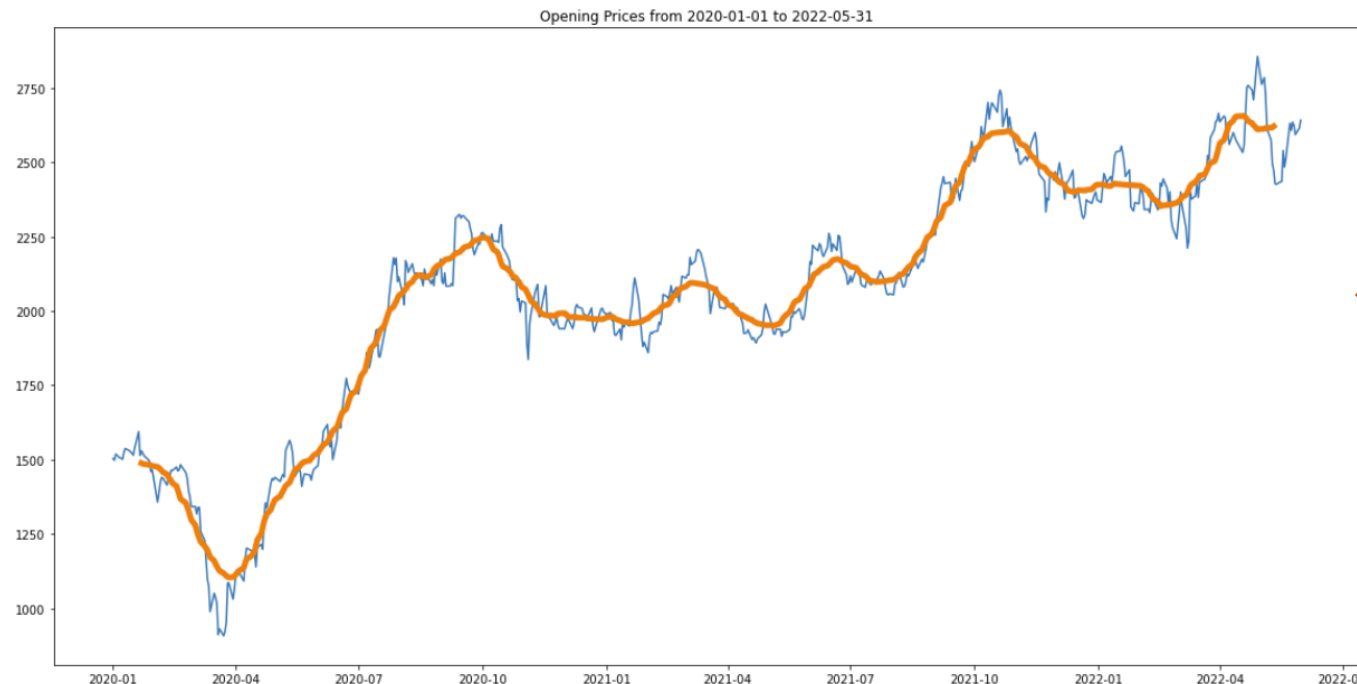
Exponential Smoothing

Finally at Non-Naive Forecasting!

- **Simple Exponential Smoothing** (SES)
 - Non Trending, Non Seasonal (Simplest)
- **Holt Model**
 - Trending, Non Seasonal (Average)
- **Holt Winters**
 - Trending, Seasonal (Amazing)

The Simple Moving Averages

- Simple is simple. No argument.
- It is used to show the *zaggy* data in a smooth manner
- Helps to give the overall trend of the time-series
- Used to show the intuition of the data without showing much information



Too Simple right?

Can you think of
something better!

Exponentially Weighted Moving Average

- This is the Exponentially Weighted Moving Average where the weight is given exponentially lower as the data point gets older.
- The moving average is designed as such that older observations are given lower weights. The weights fall exponentially as the data point gets older – hence the name exponentially weighted.

$$\text{EWMA} = \alpha x_t + (1 - \alpha)\bar{x}_{t-1}$$

Why Exponential?

Current Situation:

$$\text{EWMA} = \alpha x_t + (1 - \alpha) \bar{x}_{t-1}$$

Which includes:

$$\text{EWMA} = \alpha x_t + (1 - \alpha) [\alpha x_{t-1} + (1 - \alpha) \bar{x}_{t-2}]$$

Which Becomes:

$$\text{EWMA} = \alpha x_t + (1 - \alpha) \alpha x_{t-1} + (1 - \alpha)^2 \bar{x}_{t-2}$$

Again:

$$\text{EWMA} = \alpha x_t + (1 - \alpha) \alpha x_{t-1} + (1 - \alpha)^2 [\alpha x_{t-2} + (1 - \alpha) \bar{x}_{t-3}]$$

Which becomes:

$$\text{EWMA} = \alpha x_t + (1 - \alpha) \alpha x_{t-1} + (1 - \alpha)^2 \alpha x_{t-2} + (1 - \alpha)^3 \bar{x}_{t-3}$$

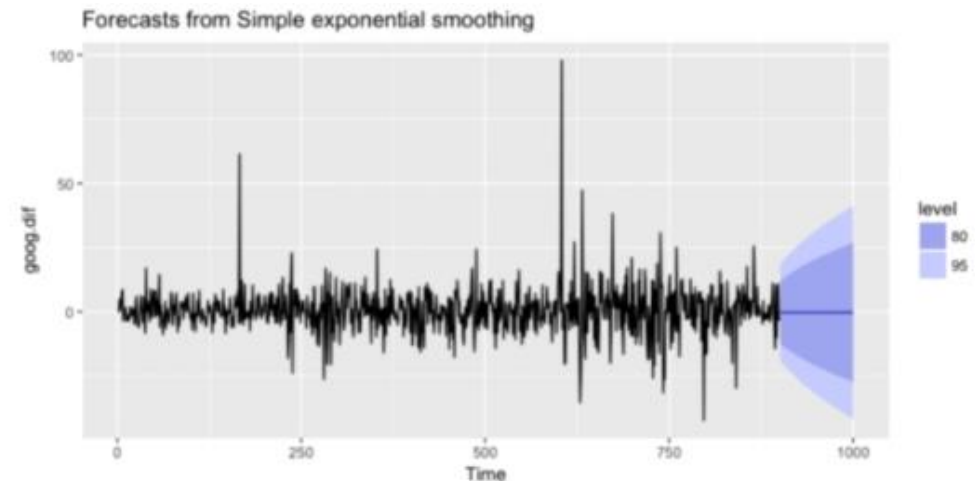
We can see that:

$$(1 - \alpha) \rightarrow (1 - \alpha)^2 \rightarrow (1 - \alpha)^3 \rightarrow \dots \rightarrow (1 - \alpha)^{t-1}$$

SES Model

- No Trend, No Seasonality model
- Assumes that there are some **fluctuations** in the data & those fluctuations are around some constant value.
- Thus, the model tries to learn what the average value is - by using the EWMA method.
- To forecast (I repeat), to forecast the value **it assumes that the same EWMA value will be propagated to the future**, because that was what the value was back in time. (around which the values were fluctuating)
- That constant value is called: The Level in this ETS terminology

$$\text{level}(t + h) = \text{EWMA}(\text{time series})$$



Holt Linear Trend Model

- There is Trend, No Seasonality
- See, the trend has to be linear - either positive or negative (assuming not like the cosine ~ curve! (:))
- It uses **2** EWMA's!! *For the level For the trend*
- In the forecast, it is just the linear combination Level and Trend - and that is our linear equation!

$$y = \beta_0 + \beta_1 x$$

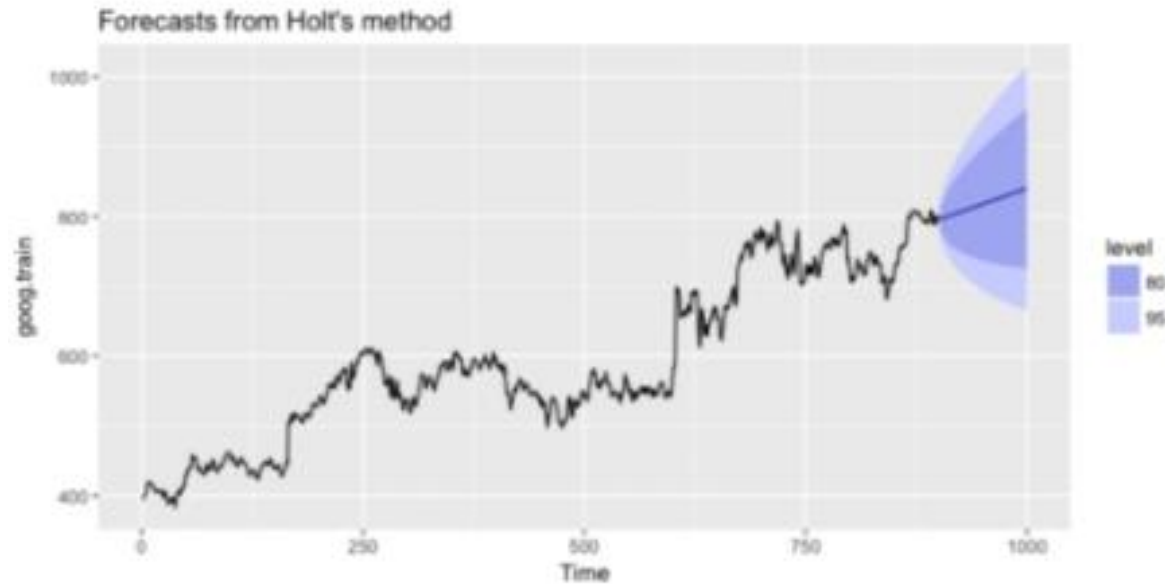
$$\text{forecast} = \text{level} + \text{trend} \times h$$

- Again, h : is the number of steps in the future
- So the 2 EWMA's are:

$$\text{level}(t + h) = \text{EWMA}(\text{level of time-series})$$

$$\text{trend}(t + h) = \text{EWMA}(\text{trend of time-series})$$

Holt Linear Trend Model



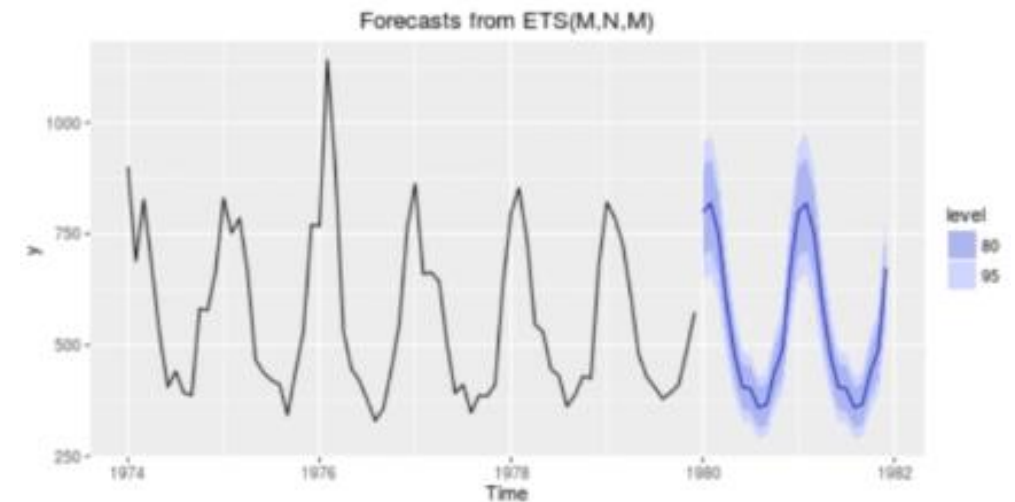
- As can be seen in the image, the line goes in the upward direction.
- Which is the result of the intercept = level and 5 slope = trend. (Better than the SES model, just the horizontal line)

Holt Winter Model

- There is seasonality, There is trend
- It means, the constant value will stay constant for each year's each season.
- If the selling of book in April 2021 is +5 then it will be +5 in the April 2022 (might increase or decrease based on the trend and level) but as per to say... it will be +5 .

$$\begin{aligned}
 \text{level}(t + h) &= EWMA(\text{level of time-series}) \\
 \text{trend}(t + h) &= EWMA(\text{rend of time-series}) \\
 \text{seasonal}(t + h) &= EWMA(\text{seasonal of time-series})
 \end{aligned}$$

- And, there can be different ways to forecast
 1. Either: Forecast = Trend + Level + Season
 2. Either: Forecast = Trend x Level + Season
 3. Either: Forecast = Trend + Level x Season
 4. Either: Forecast = Trend x Level x Season



Auto-Regression Model

Auto-Regression Model

Definition

An autoregressive model (also called AR model) is used to model a future behavior for a time-ordered data, using data from past behaviors.

- Essentially, it is a linear regression analysis of a dependent variable using one or more variables(s) in a given time-series data.

$$Y_t = f(Y_{t-1}, Y_{t-2}, \dots, Y_{t-p})$$

Auto-Regression Model for Forecasting

Definition

- A natural starting point for forecasting model is to use past values of Y , that is, Y_{t-1}, Y_{t-2}, \dots to predict Y_t .
- An auto-regression is a regression model in which Y_t is regressed against its own lagged values.
- The number of lags used as regressors is called the **order of auto-regression**.
 - In first order auto-regression (denoted as $AR(1)$), Y_t is regressed against Y_{t-1}
 - In p^{th} order auto-regression (denoted as $AR(p)$), Y_t is regressed against, $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$

p^{th} order Auto-regression Model

Formula: p^{th} Order Auto-regression Model

In general, the p^{th} order auto-regression model is defined as

$$Y_t = \beta_0 + \sum_{i=1}^p \beta_i Y_{t-i} + \varepsilon_t$$

where, $\beta_0, \beta_1, \dots, \beta_p$ is called auto-regression coefficients and ε_t is the noise term or residue and in practice it is assumed to Gaussian white noise

For example, $AR(1)$ is $Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t$

The task in AR analysis is to derive the "best" values for $\beta_i, i = 0, 1, \dots, p$ given a time series $[Y_1, Y_2, \dots, Y_{T-1}, Y_T]$

Computing AR Coefficients

Computing AR(p) model

- A number of techniques known for computing the AR coefficients
- The most common method is called Least Squares Method (LSM)
- The LSM is based upon the **Yule-Walker equations**

$$\begin{bmatrix}
 1 & \rho_1 & \rho_2 & \rho_3 & \rho_4 & \cdots & \cdots & \rho_{p-2} & \rho_{p-1} \\
 \rho_1 & 1 & \rho_1 & \rho_2 & \rho_3 & \cdots & \cdots & \rho_{p-3} & \rho_{p-2} \\
 \rho_2 & \rho_1 & 1 & \rho_1 & \rho_2 & \cdots & \cdots & \rho_{p-4} & \rho_{p-3} \\
 \rho_3 & \rho_2 & \rho_1 & 1 & \rho_1 & \cdots & \cdots & \rho_{p-5} & \rho_{p-4} \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \cdots & \vdots & \vdots \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \cdots & \vdots & \vdots \\
 \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \rho_{p-4} & \rho_{p-5} & \cdots & \cdots & \rho_1 & 1
 \end{bmatrix}
 \begin{bmatrix}
 \beta_1 \\
 \beta_2 \\
 \beta_3 \\
 \vdots \\
 \vdots \\
 \beta_{p-1} \\
 \beta_p
 \end{bmatrix}
 =
 \begin{bmatrix}
 \rho_1 \\
 \rho_2 \\
 \rho_3 \\
 \vdots \\
 \vdots \\
 \rho_{p-1} \\
 \rho_p
 \end{bmatrix}$$

ARIMA Model

Autoregressive Integrated Moving Average

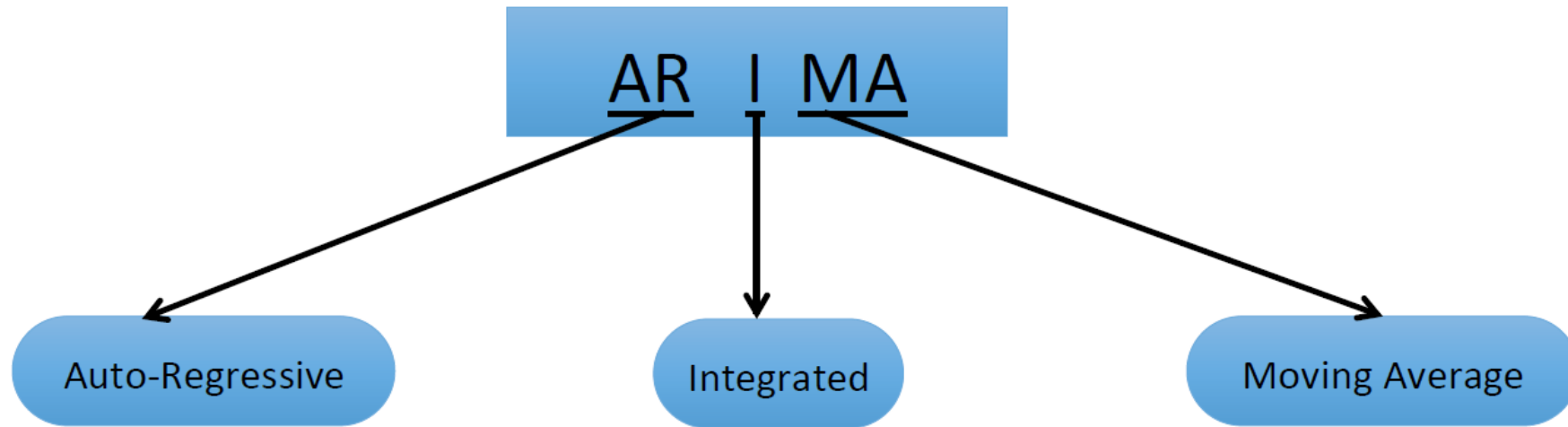
- The ARIMA model, introduced by [Box and Jenkins \(1976\)](#), is a linear regression model indulged in tracking linear tendencies in stationary time series data.
- **AR:** autoregressive (lagged observations as inputs) **I:** integrated (differencing to make series stationary) **MA:** moving average (lagged errors as inputs).
- The model is expressed as ARIMA (p, d, q) where p, d and q are integer parameter values that decide the structure of the model.
- More precisely, p and q are the order of the AR model and the MA model respectively, and parameter d is the level of differencing applied to the data.
- The mathematical expression of the ARIMA model is as follows:

$$y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

where y_t is the actual value, ε_t is the random error at time t , ϕ_i and θ_j are the coefficients of the model.

- It is assumed that ε_{t-1} ($\varepsilon_{t-1} = y_{t-1} - \hat{y}_{t-1}$) has zero mean with constant variance, and satisfies the i.i.d. condition.
- Three basic Steps: Model identification, Parameter Estimation, and Diagnostic Checking.

ARIMA model



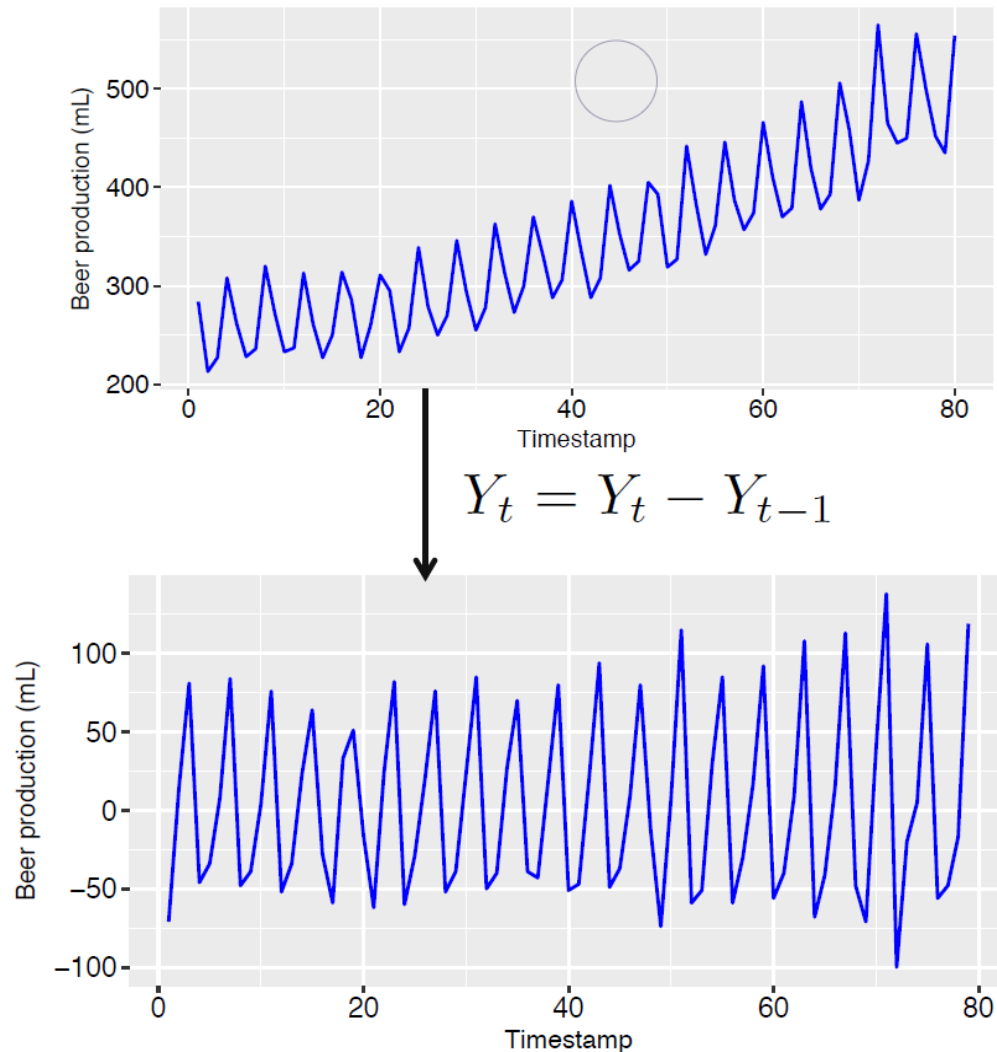
$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \epsilon_t \quad [\text{Order } p]$$

$$Y_t = Y_t - Y_{t-1} \quad [\text{Order } d]$$

$$Y_t = \beta_0 + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \cdots + \phi_q \epsilon_{t-q} \quad [\text{Order } q]$$

ARIMA is defined by a tuple (p, d, q)

Differencing in ARIMA Model



Differencing order (d):
Number of times differencing is done

ACF / PACF Plots

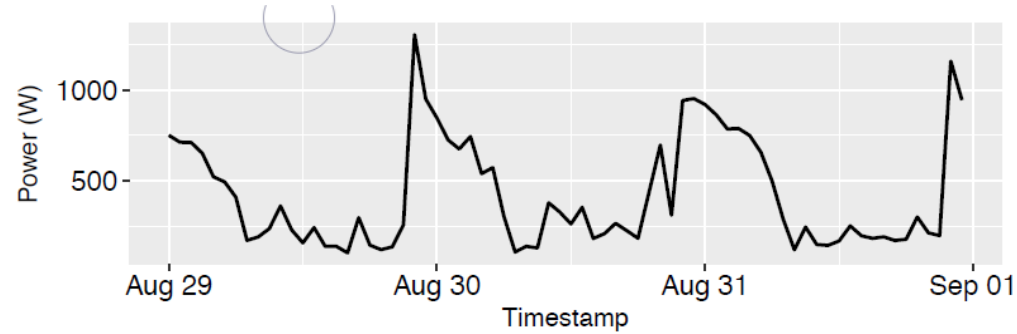
1. Auto-Correlation Function (ACF) Plot:

- Correlation coefficients of time-series at different lags
- Defines q order of MA model

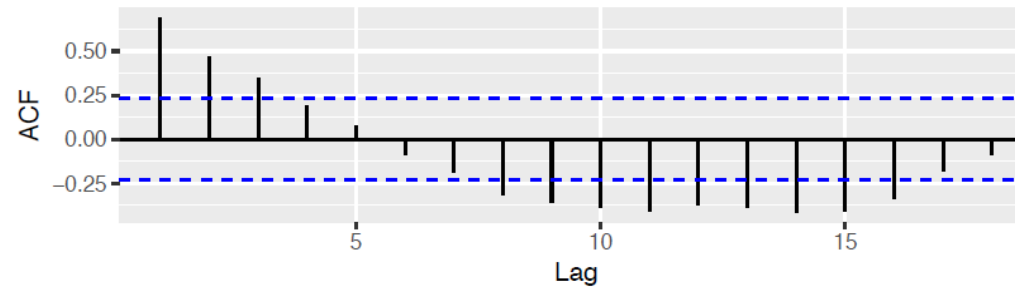
2. Partial Auto-correlation Function (PACF) Plot:

- Partial correlation coefficients of time series at different lags
- Defines p order of AR model

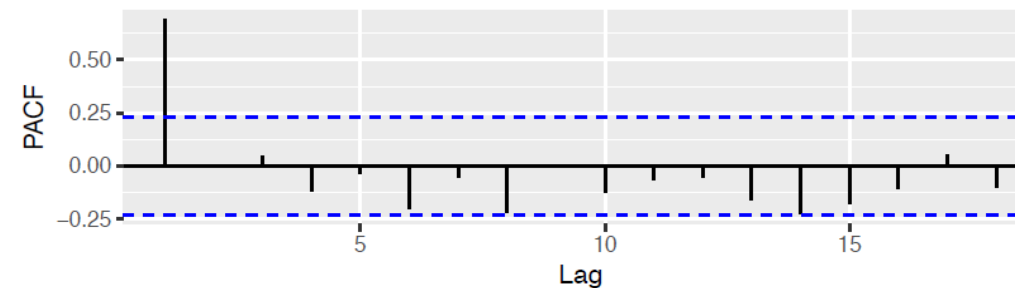
ACF / PACF Plots : Example



Data



ACF Plot



PACF Plot

Multivariate Forecasting

- VAR models (vector autoregressive models) are used for multivariate time series. The structure is that each variable is a linear function of past lags of itself and past lags of the other variables.
- As an example suppose that we measure three different time series variables, denoted by $x_{(t,1)}$, $x_{(t,2)}$, and $x_{(t,3)}$.
- The vector autoregressive model of order 1, denoted as VAR(1), is as follows:

$$x_{t,1} = \alpha_1 + \phi_{11}x_{t-1,1} + \phi_{12}x_{t-1,2} + \phi_{13}x_{t-1,3} + w_{t,1}$$

$$x_{t,2} = \alpha_2 + \phi_{21}x_{t-1,1} + \phi_{22}x_{t-1,2} + \phi_{23}x_{t-1,3} + w_{t,2}$$

$$x_{t,3} = \alpha_3 + \phi_{31}x_{t-1,1} + \phi_{32}x_{t-1,2} + \phi_{33}x_{t-1,3} + w_{t,3}$$

- Each variable is a linear function of the lag 1 values for all variables in the set.

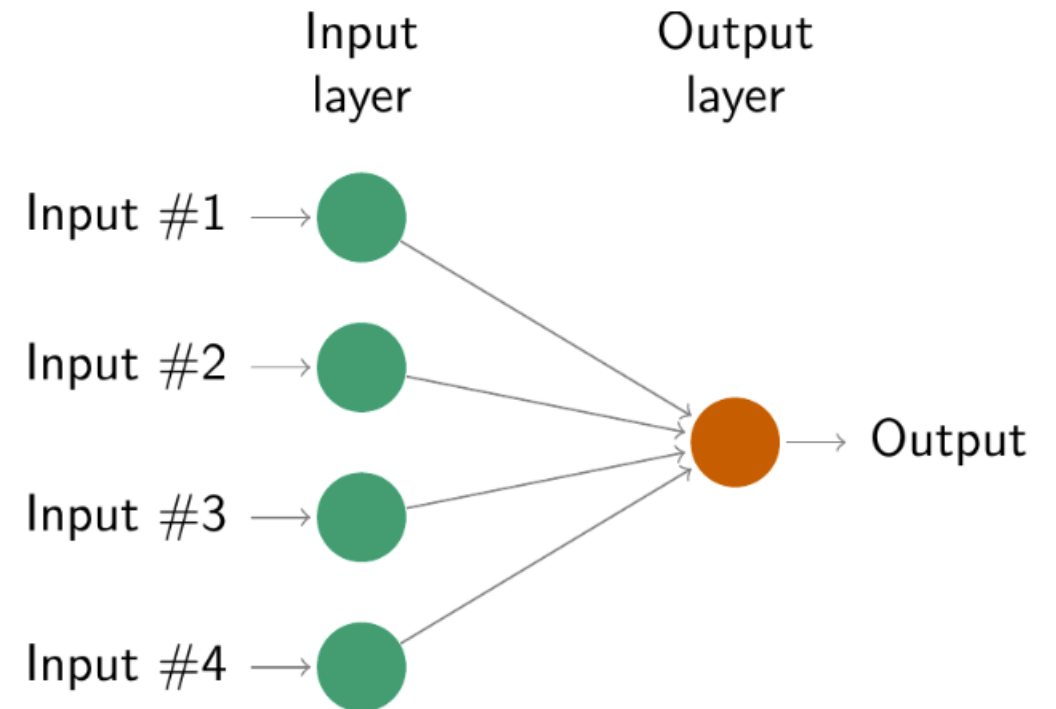
- In a VAR(2) model, the lag 2 values for all variables are added to the right sides of the equations, In the case of three x-variables (or time series) there would be six predictors on the right side of each equation, three lag 1 terms and three lag 2 terms.
- In general, for a VAR(p) model, the first p lags of each variable in the system would be used as regression predictors for each variable.
- VAR models are a specific case of more general VARMA models. VARMA models for multivariate time series include the VAR structure above along with moving average terms for each variable. More generally yet, these are special cases of ARMAX models that allow for the addition of other predictors that are outside the multivariate set of principal interest.

It arose from macroeconomic data where large changes in the data permanently affect the level of the series. $\mathbf{x}_t = \phi \mathbf{x}_{t-1} + \mathbf{w}_t$

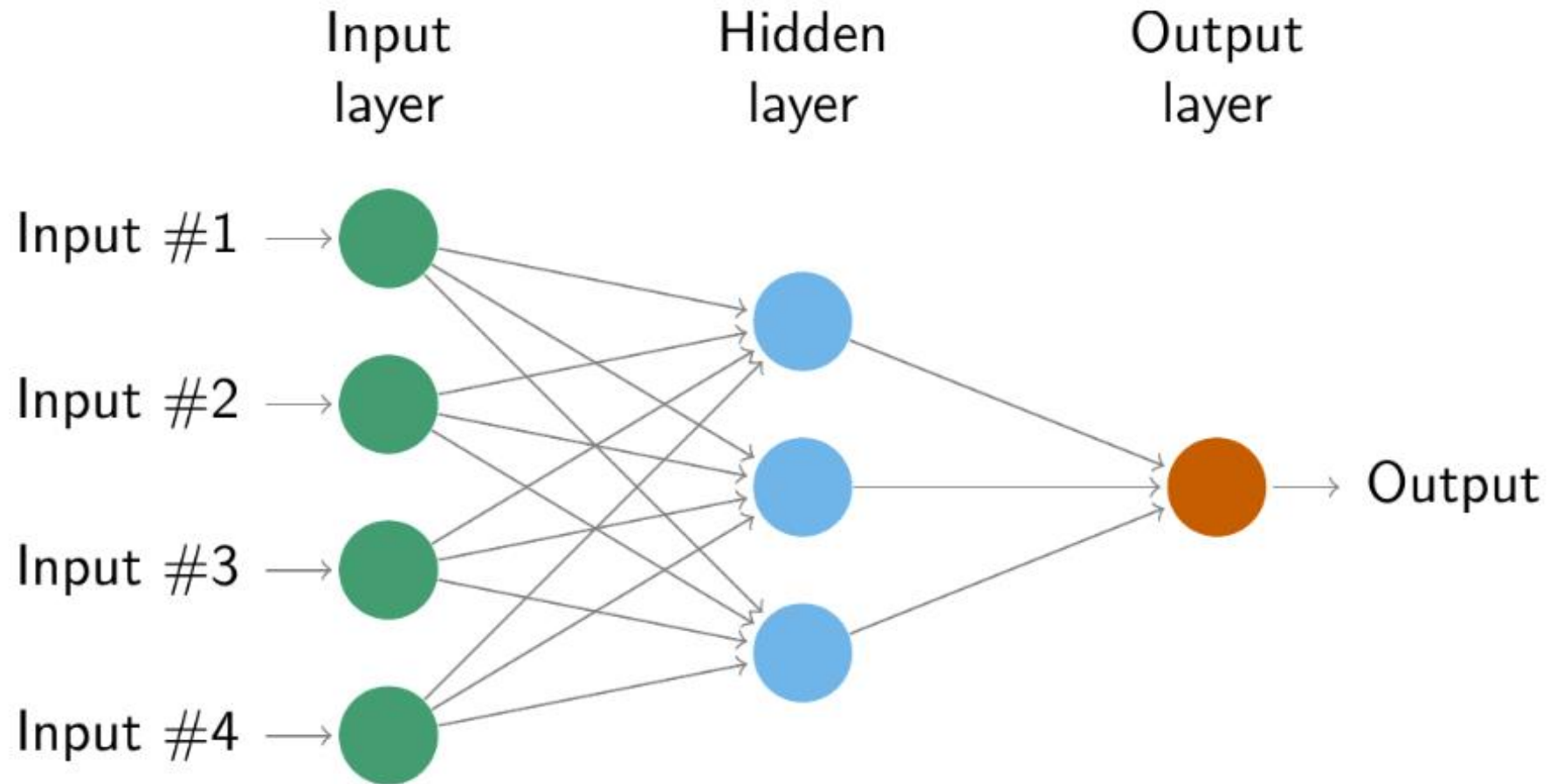
Autoregressive Neural Network

Neural Network

- Artificial neural networks are forecasting methods that are based on simple mathematical models of the brain.
- They allow complex nonlinear relationships between the response variable and its predictors.
- A neural network can be thought of as a network of "neurons" which are organized in layers.
- The predictors (or inputs) form the bottom layer, and the forecasts (or outputs) form the top layer. There may also be intermediate layers containing "hidden neurons".



Feed-forward Neural Network



Feed-forward Neural Network

- Feed-forward neural network comprise of a series of hidden layers between the input and output layer.
- The outputs of the nodes in one layer are inputs to the next layer. The inputs to each node are combined using a weighted linear combination. The result is then modified by a nonlinear function before being output.

$$z_j = b_j + \sum_{i=1}^4 w_{i,j} x_i.$$

- In the hidden layer, this is then modified using a nonlinear function such as a sigmoid,

$$s(z) = \frac{1}{1 + e^{-z}},$$

to give the input for the next layer. This tends to reduce the effect of extreme input values, thus making the network somewhat robust to outliers.

ARNN for Forecasting

- With time series data, lagged values of the time series can be used as inputs to a neural network, typically named as Autoregressive neural network (ARNN)
- $\text{ARNN}(p, k)$ is a feed-forward network with one hidden layer, with p lagged inputs and k nodes in the hidden layer.
- $\text{ARNN}(p, 0)$ model is equivalent to an $\text{ARIMA}(p, 0, 0)$ model, but without the restrictions on the parameters to ensure stationarity.
- With seasonal data, it is useful to also add the last observed values from the same season as inputs.
- For time series, the default is the optimal number of lags (according to the AIC) for a linear $\text{AR}(p)$ model. If k is not specified, it is set to $k = (p + 1)/2$ (rounded to the nearest integer).
- When it comes to forecasting, the network is applied iteratively.

Forecast Evaluation

Forecast Evaluation

Performance metrics such as mean absolute error (MAE), root mean square error (RMSE), and mean absolute percent error (MAPE) are used to evaluate the performances of different forecasting models for the unemployment rate data sets:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2};$$

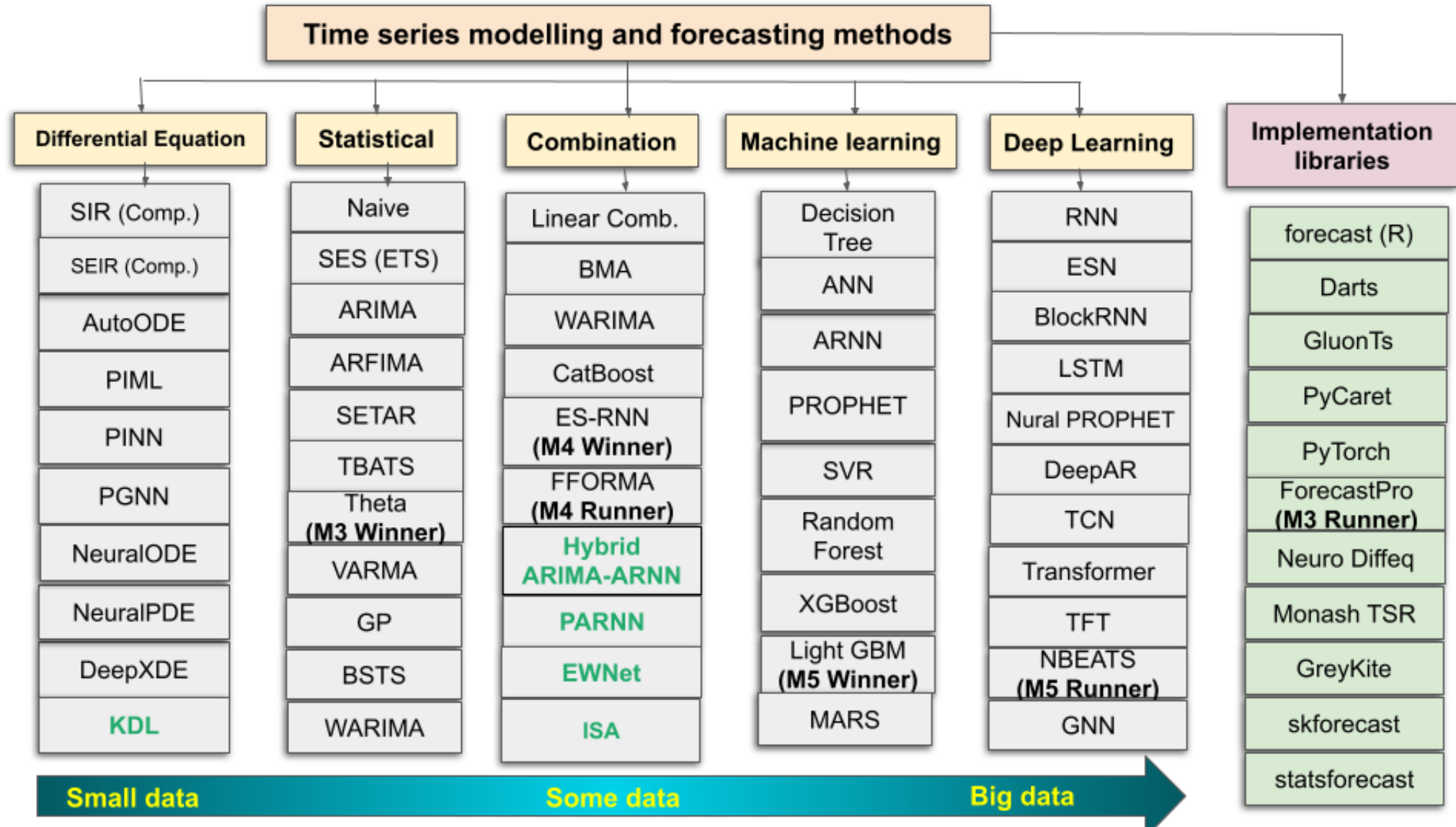
$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|;$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|,$$

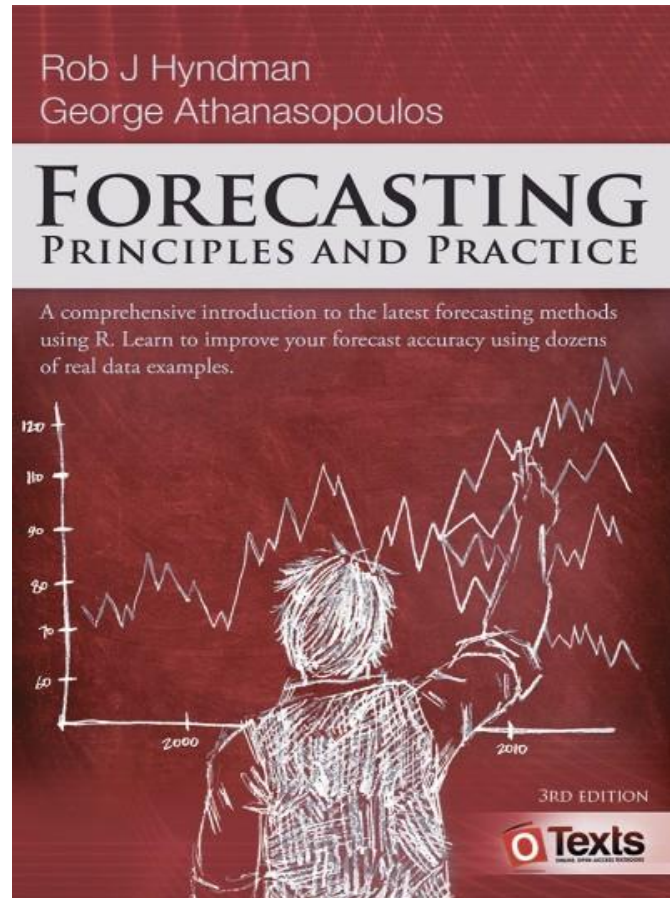
Where y_i is the actual output, \hat{y}_i is the predicted output, and n denotes the number of data points.

By definition, the lower the value of these performance metrics, the better is the performance of the concerned forecasting model.

Quick Survey on Forecasting Tools



Reference Book



Scan for Slides

Read Online: <https://otexts.com/fpp3/>