DS C31 (NLP/DL) Track
Live online Session
UpGrad, IIIT-B

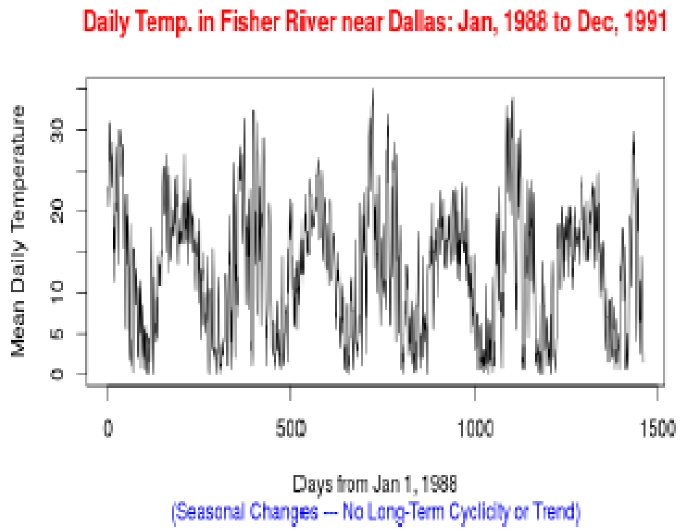
19 December 2021, Sunday, 7pm-8:30pm

Bangalore

Time Series Analysis / Forecasting

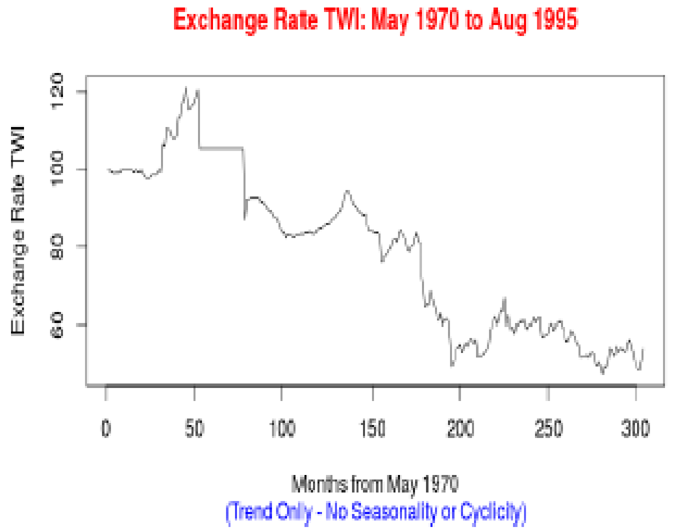
Dr. V. Ramasubramanian
Professor
International Institute of Information Technology - Bangalore (IIIT-B)
Bangalore, India

- Daily Stock market figures
- Sales / Revenue / Profitability figures of companies by year
- Demographic / Development data (population, birthrate, infant mortality figures, literacy, per-capita income, school enrollment figures) by year
- Blood Pressure and other body vitals by time units (in hours, minutes, ...) appropriate to the severity of the patient
- Ecology / Geology data — progression of the intensity of the tremors of an earthquake with time, pollution figures by year, annual average temperature, thickness of the glacial ice-sheet etc. by year indicating global warming, annual rainfall figures, ...)
- Epidemiology data — spread (number of people affected) of an epidemic with time
- Speech data (signal parameters evolving over time) — these need to be modeled for speech recognition and of course to build applications such as the popular Siri on iPhones

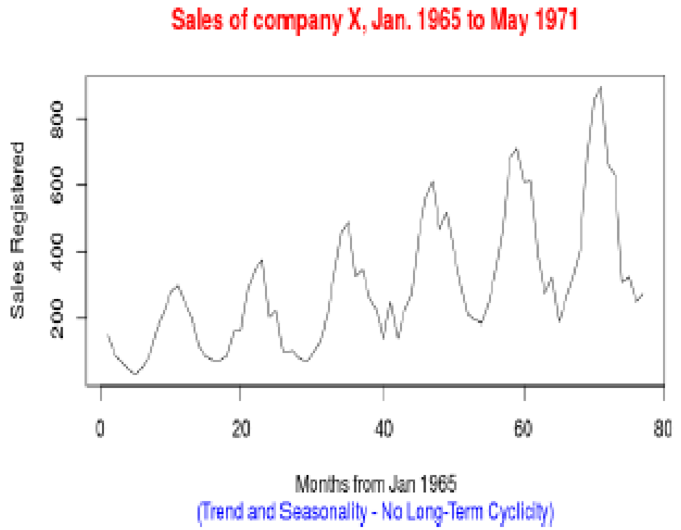


(a) Seasonal Changes in Temperature — No Long-Term Cyclicity or Trend

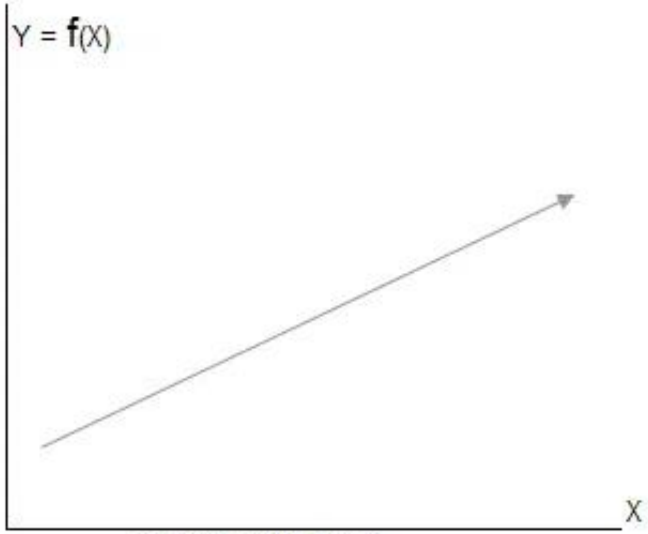
(b) Large Cyclicity (between 1300 and 1800) and Prominent Trend - No Seasonality



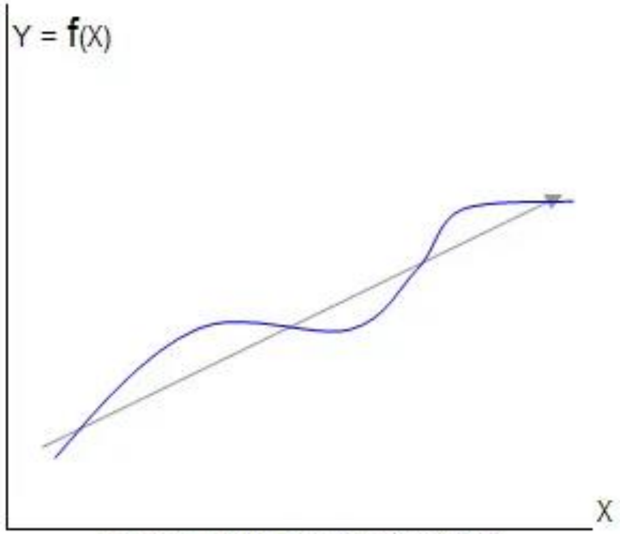
(c) Trend Only - No Seasonality or Long-Term Cyclicity



(d) Trend and Seasonality - No Long-Term Cyclicity

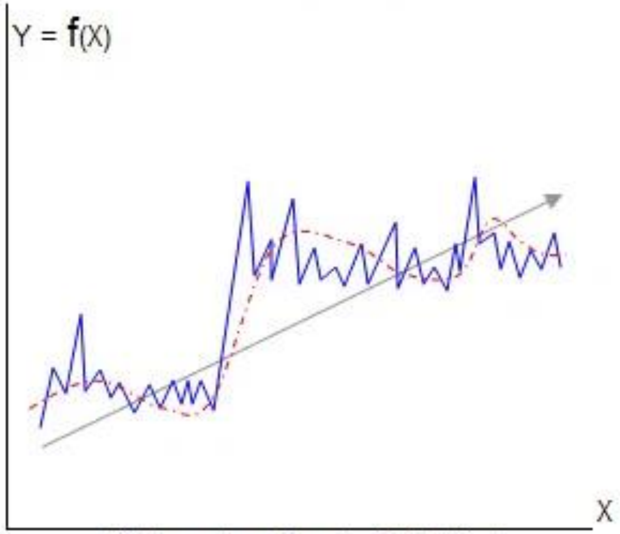


(1) Long-term Trend



(2) Long-term Trend with Cyclical Variations/Movements

(3) Long-term Trend with Cyclical and Seasonal Variations/Movements



(4) Long-term Trend with Cyclical, Seasonal, and Random Variations/Movements

$$(X_0, X_1, \ldots X_t, \ldots X_T),$$

stationarize the data

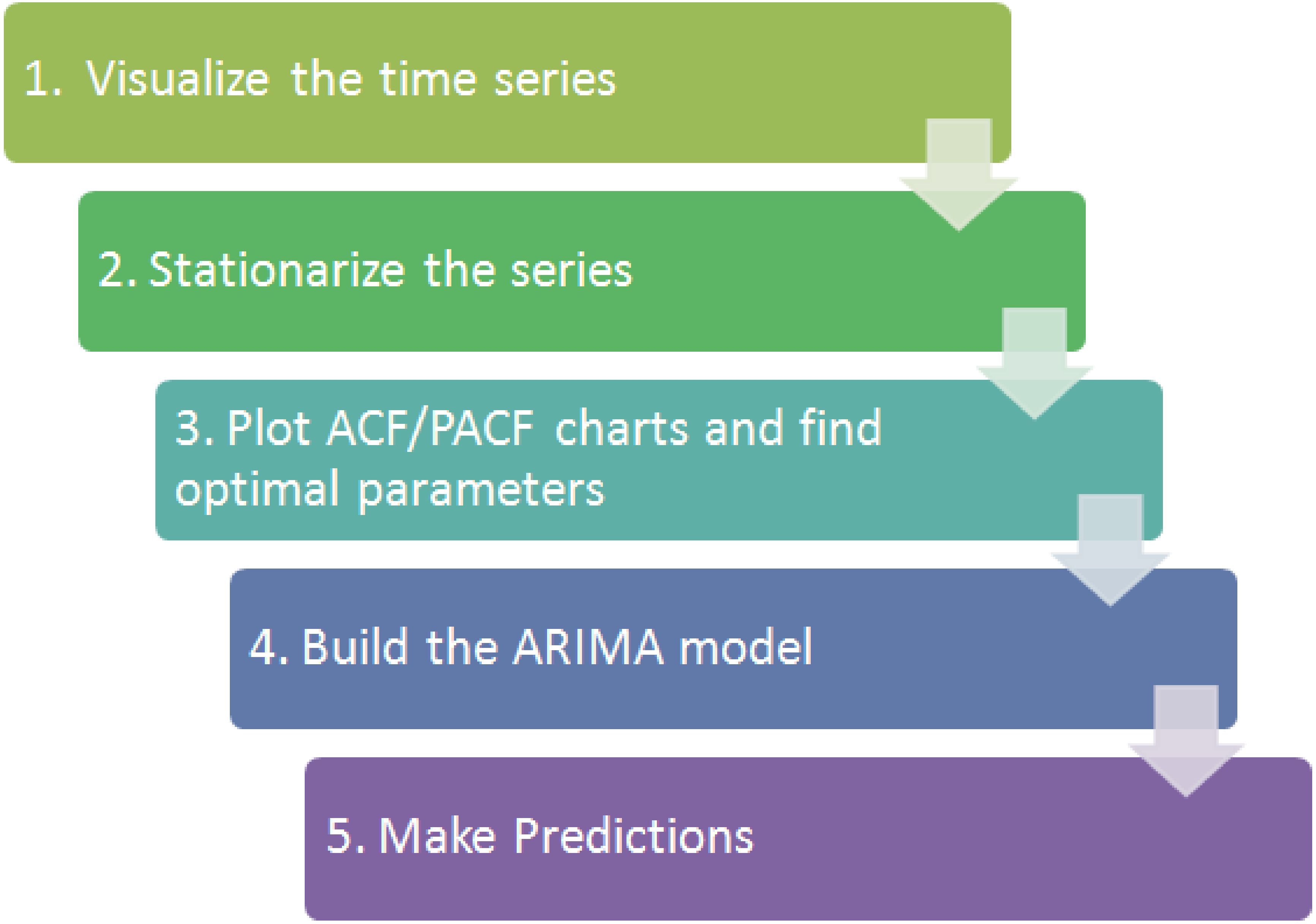
The modified stationary time series obtained after removing the part of the data that characterizes the trend, seasonality and cyclicity in the data is often called the *residual* series, or simply the residue. The residual series is therefore of the form:

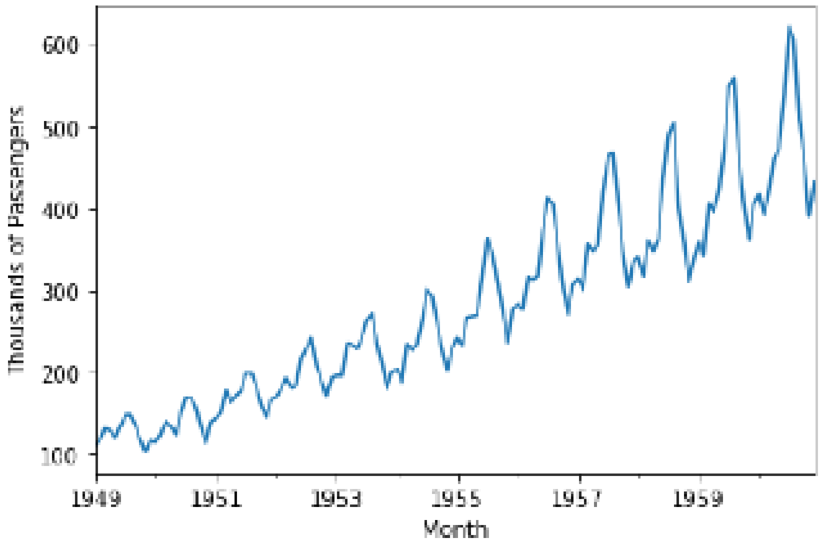
$$(Y_0, Y_1, \ldots, Y_T) \quad \text{where } Y_t = S(X_t) \quad M(t) \approx Y_t$$

Test if the residual time series is stationary. If it is not, then iterate on the earlier step to tweak the seasonal, trend and cyclicity models (the transformation S) to arrive at a stationary residual series. Model the residual series as a stationary time series using the methods we will discuss later in this module. Assume the stationary model is M where $M(t) \approx Y_t$ for every time stamp t.

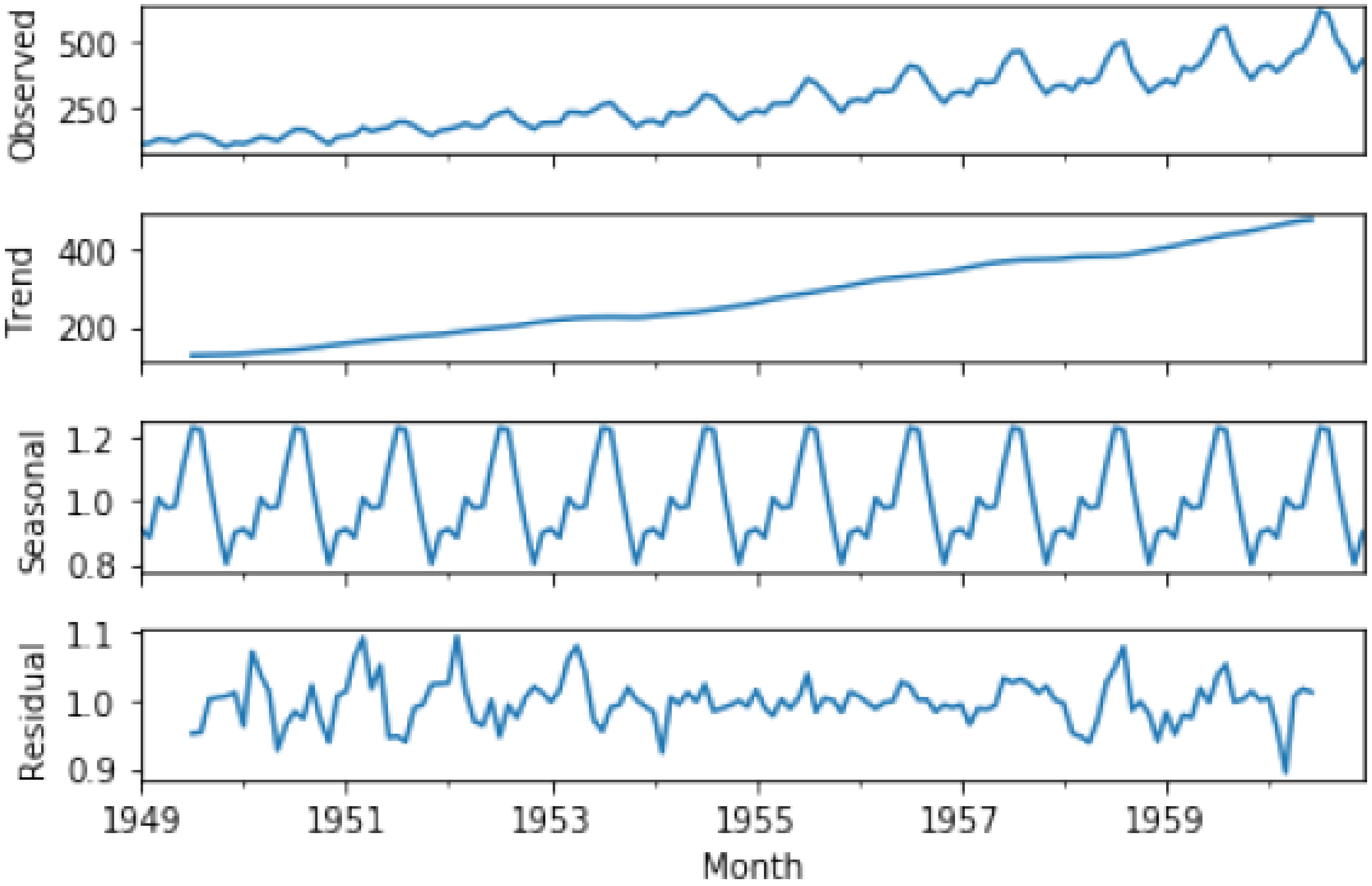
Construct the final model by applying the inverse transformation $S^{-1}(.)$ on the stationary model.

$$\epsilon_t = X_t - S^{-1}(M(t)), 0 \leq t \leq T$$

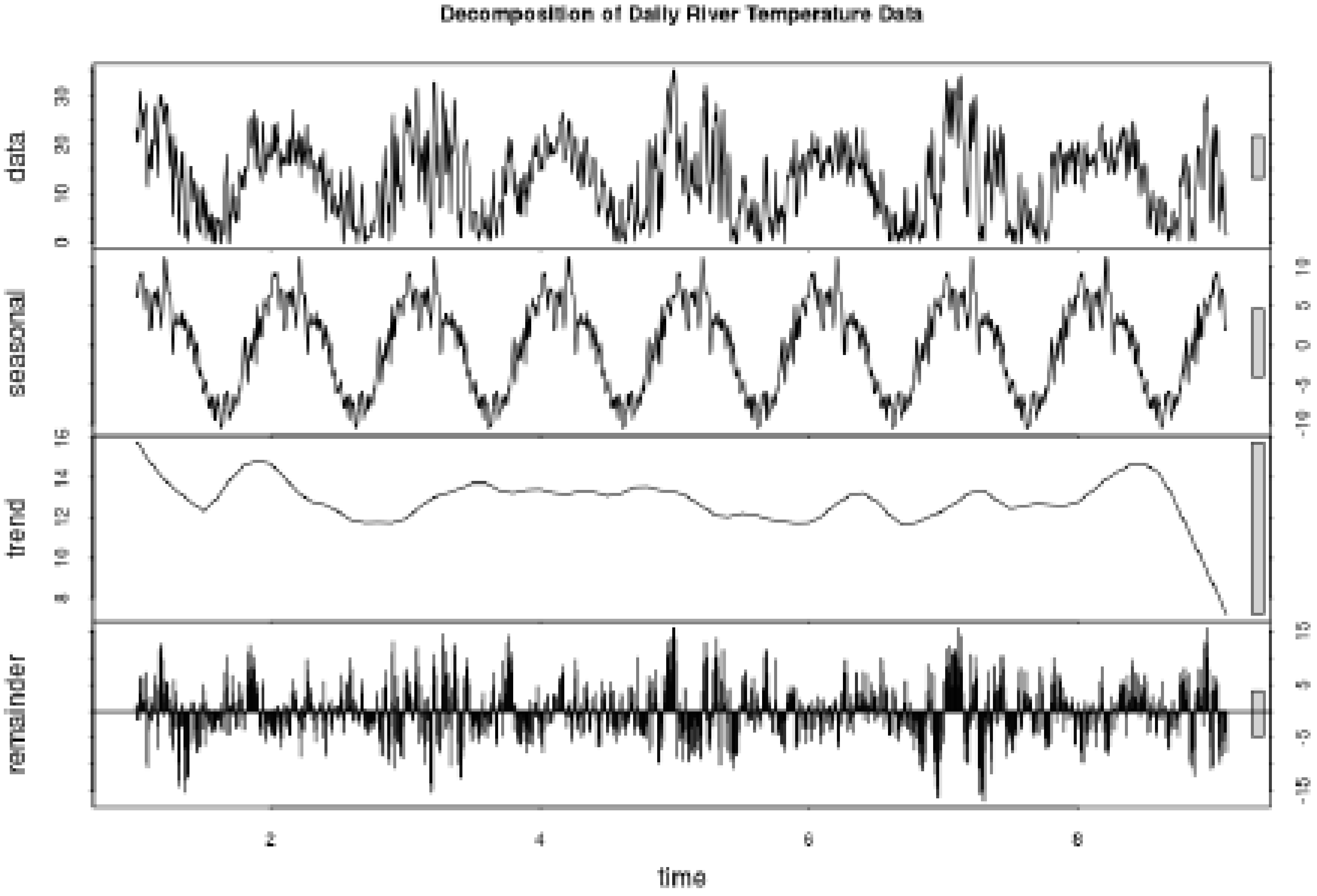

1. Visualize the time series
2. Stationarize the series
3. Plot ACF/PACF charts and find optimal parameters
4. Build the ARIMA model
5. Make Predictions



Thousands of Passengers
100
200
300
400
500
600
1949
1951
1953
1955
1957
1959
Month



Observed
250
500
Trend
200
400
Seasonal
0.8
1.0
1.2
Residual
0.9
1.0
1.1
1949
1951
1953
1955
1957
1959
Month



Decomposition of Daily River Temperature Data
data
seasonal
trend
remainder
time
0
10
20
30
10
5
0
-5
-10
16
14
12
10
8
10
5
-5
-15
2
4
6
8

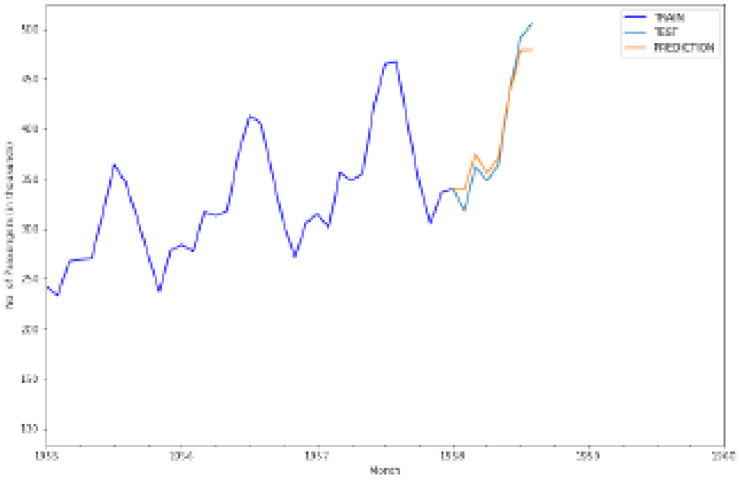
$$y_t = f(t; \beta) + \varepsilon_t$$

β is the vector of unknown parameters and ε_t are the uncorrelated errors.

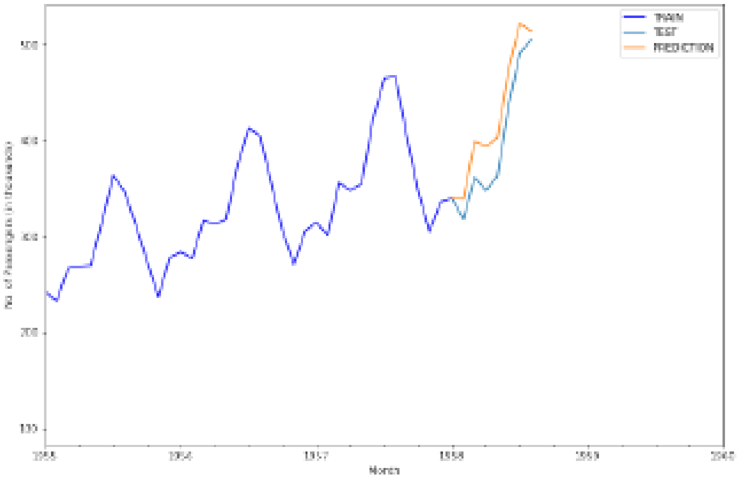
- Forecasting - predicting the future from the past
- Given an observed value Y, predict Y_{t+1} using $Y_1 \ldots Y_t$
- In other words, learn f such that

$$Y_{t+1} = f(Y_1, \ldots, Y_t)$$

- One Step Forecasting
  - Given data upto time t, predict value only for the next one step *i.e.* at $t+1$
- Multi Step Forecasting
  - Given data upto time t, predict values for two or more steps *i.e.* at $t+1, t+2, t+3, \ldots$



One Step Prediction for 8 steps



Multi Step Prediction for 8 steps

Note how close prediction is to true value in case of one step prediction

Standard evaluation metrics for time series forecasting are;

- Mean Absolute Error (MAE)
- Mean Absolute Percentage Error (MAPE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Normalized Root Mean Squared Error (NRMSE)

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$$

$$MAPE = \frac{100\%}{n}\sum_{j=1}^{n}\left|\frac{y_j - \hat{y}_j}{y_j}\right|$$

$$MSE = \frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2$$

Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

Normalized Root Mean Squared Error (NRMSE)

$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}}{Z}$$

where Z is the normalization factor

- NRMSE allows for comparison between models across different datasets
- Common normalization factors:
  - Mean: Preferred when same preprocessing and predicted feature
  - Range: sensitive to sample size
  - Standard Deviation: suitable across datasets as well as predicted features

Stationary Time Series Analysis

Stationarity

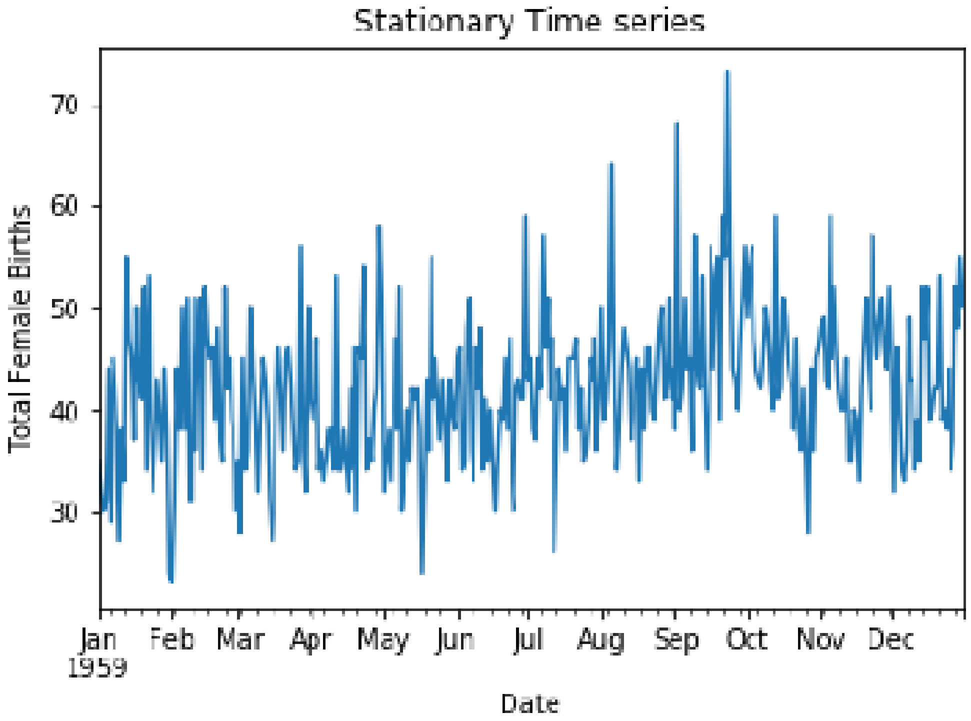
A (strongly) time-invariant series is one where the joint distribution of the set of values ($X_1, . . . ,X_n$) is identical to the joint distribution of the values ($X_{t+1}, . . . ,X_{t+n}$) for any t, n

weak stationarity

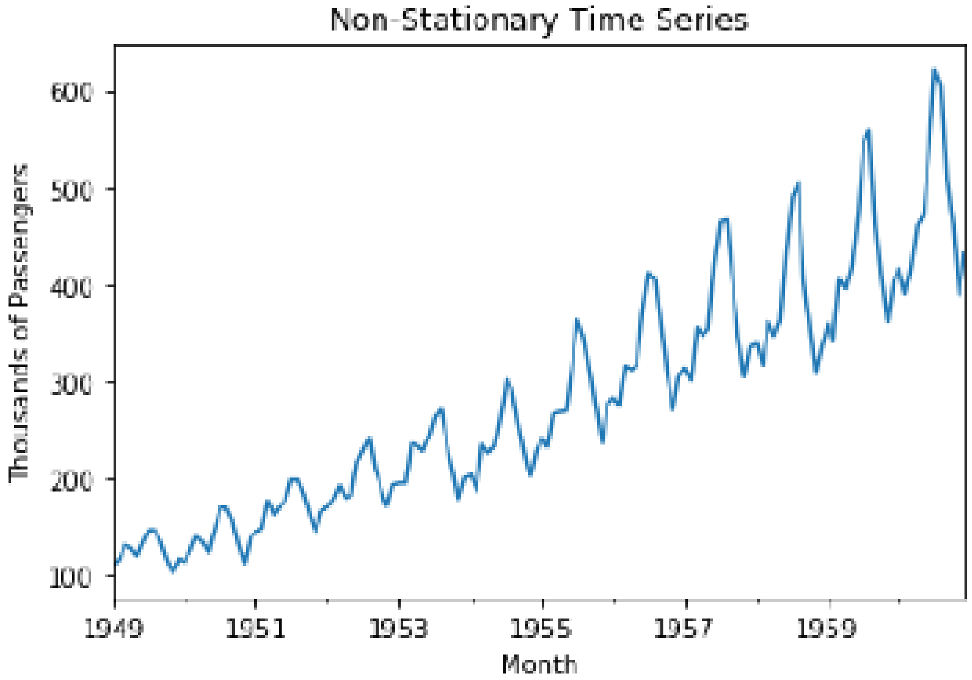
A time-series is stationary in a limited sense — that the relationship between any pair of values (some time-steps, including zero, away) in the time series is independent of the timestamp of the first event.

- Strict stationarity
  - $P(Y_t) = P(Y_{t+k})$ and $P(Y_t, Y_{t+k})$ is independent of t
  - Mean and variance time invariant
- Weak Stationarity
  - In this case, mean constant, variance constant
  - $Cov(Y_1, Y_{1+k}) = Cov(Y_2, Y_{2+k}) = Cov(Y_3, Y_{3+k}) = \gamma$
  - *i.e.* Covariance only depends on lag value k

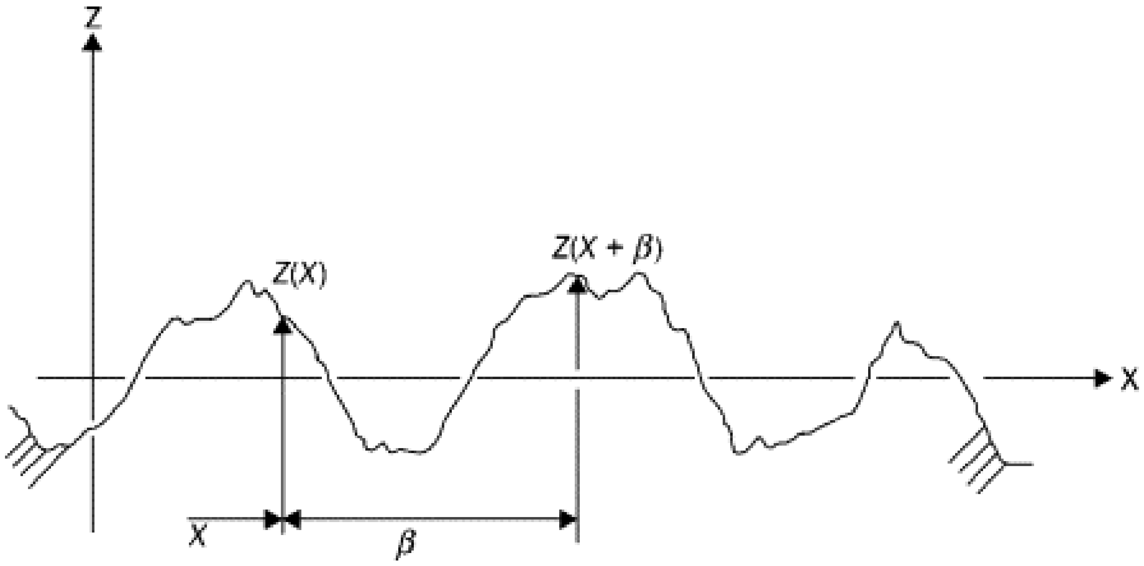
- A time series is **stationary** if it does not exhibit any trend or seasonality



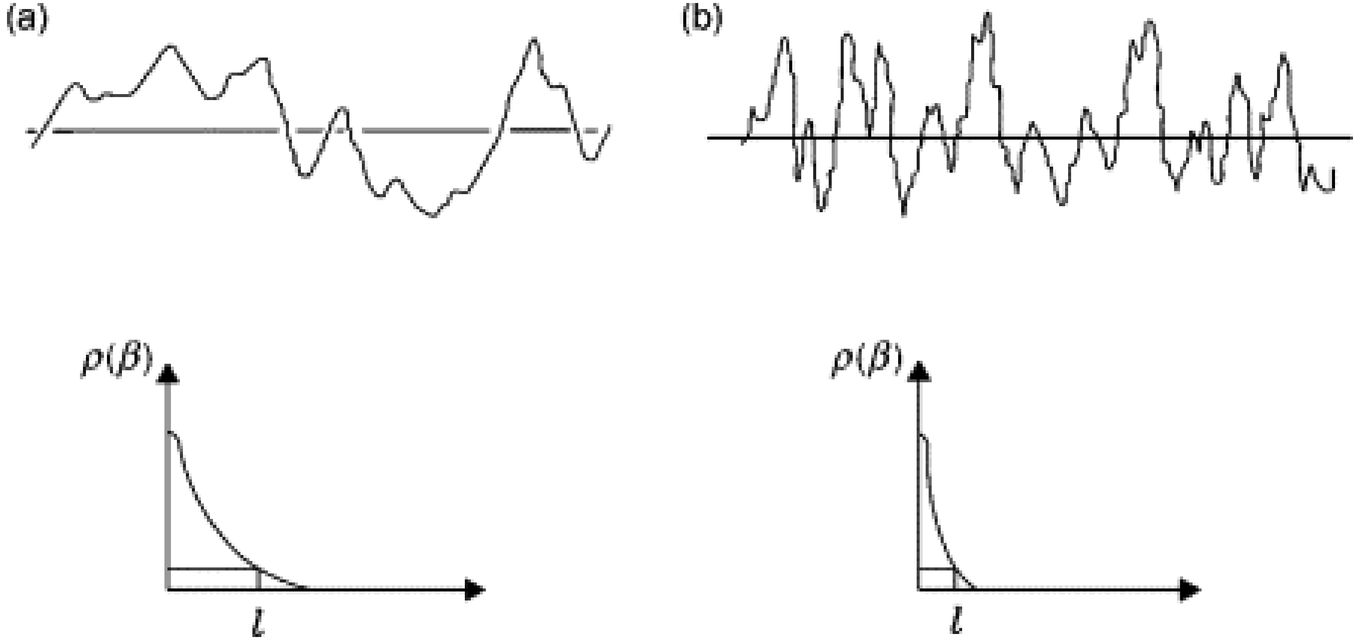
Stationary Time Series

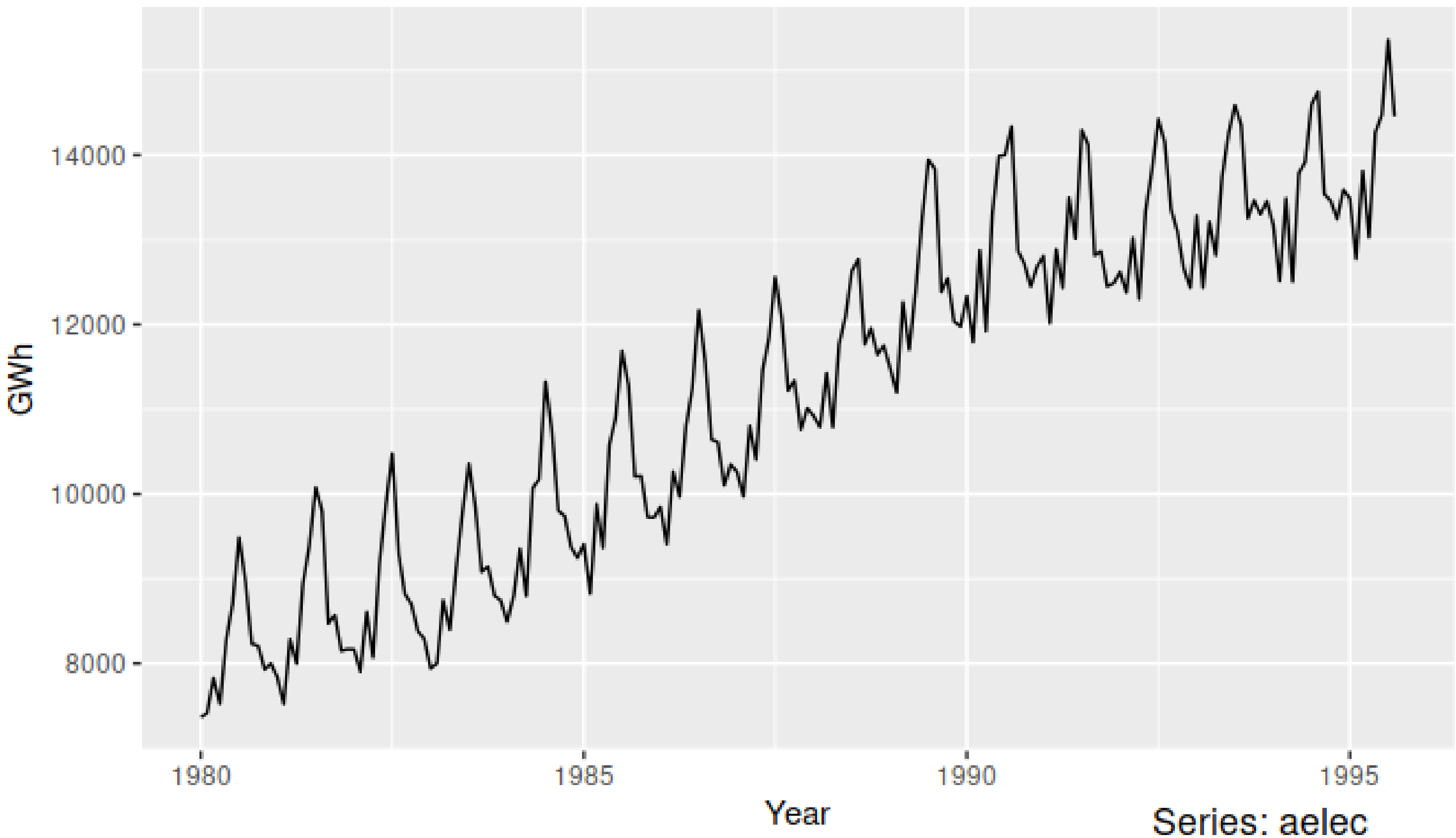


Non-Stationary Time Series



$$\rho(\beta) = \frac{1}{\sigma^2}\left\{\lim_{L\to\infty}\frac{1}{L}\int_0^L Z(x).\,Z(x+\beta)\,dx\right\}$$



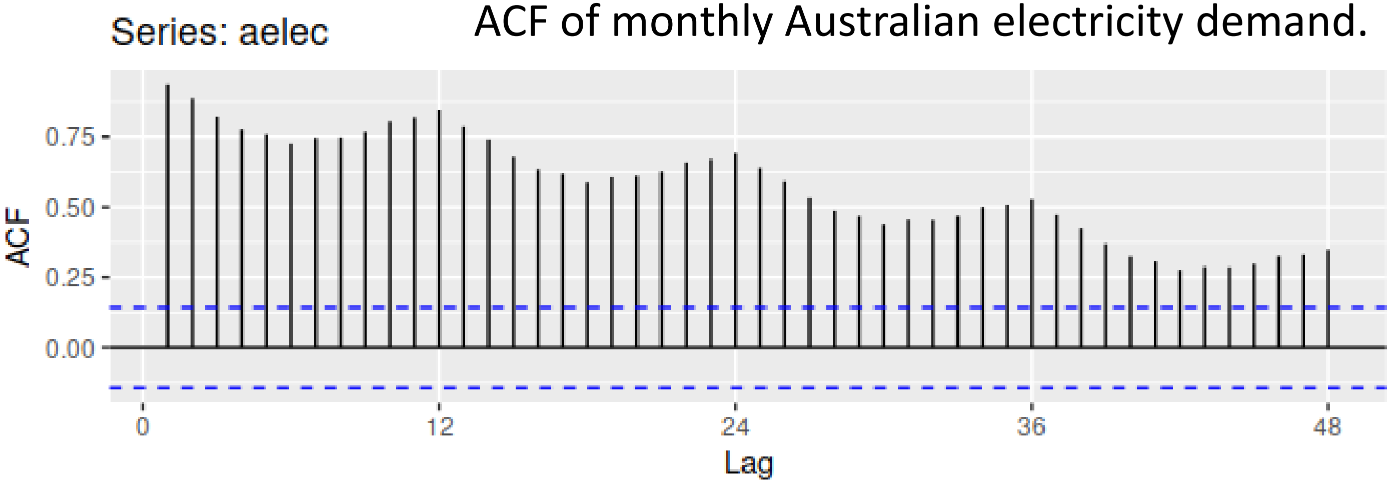


Monthly Australian electricity demand from 1980–1995.

When data are both trended and seasonal, we see a combination of these effects.

The monthly Australian electricity demand series shows both trend and seasonality. Its ACF is shown below

ACF of monthly Australian electricity demand.



The slow decrease in the ACF as the lags increase is due to the trend, while the “scalloped” shape is due the seasonality.

Covariance Function for the time series is defined as

$$\sigma_X(i,j) = E\left[\left(X_i - \mu_X(i)\right)\left(X_j - \mu_X(j)\right)\right]$$

The time series is said to be (weakly) stationary if $E[X_t]$ and the covariance function $\sigma(t+h, t)$ for any fixed *lag h* is independent of the time *t*

Autocorrelation function (ACF) $\rho(h)$

$$\rho(h) = \eta(h)/\eta(0)$$

$$\eta(h) \equiv \sigma(t+h, t).$$

Sample Mean — this approximates the mean of X_t at any t

$$\overline{X} = \frac{1}{T+1}\sum_{t=0}^{T} X_t$$

Sample Autocovariance Function — Expectation is replaced by an average over the length of the series; mean at t replaced by $\overline{X}$

$$\overline{\eta}(h) = \frac{1}{T+1}\sum_{t=0}^{T-h}(X_{t+h}-\overline{X})(X_t-\overline{X})$$

Sample ACF

$$\overline{\rho}(h) = \frac{\overline{\eta}(h)}{\overline{\eta}(0)}$$

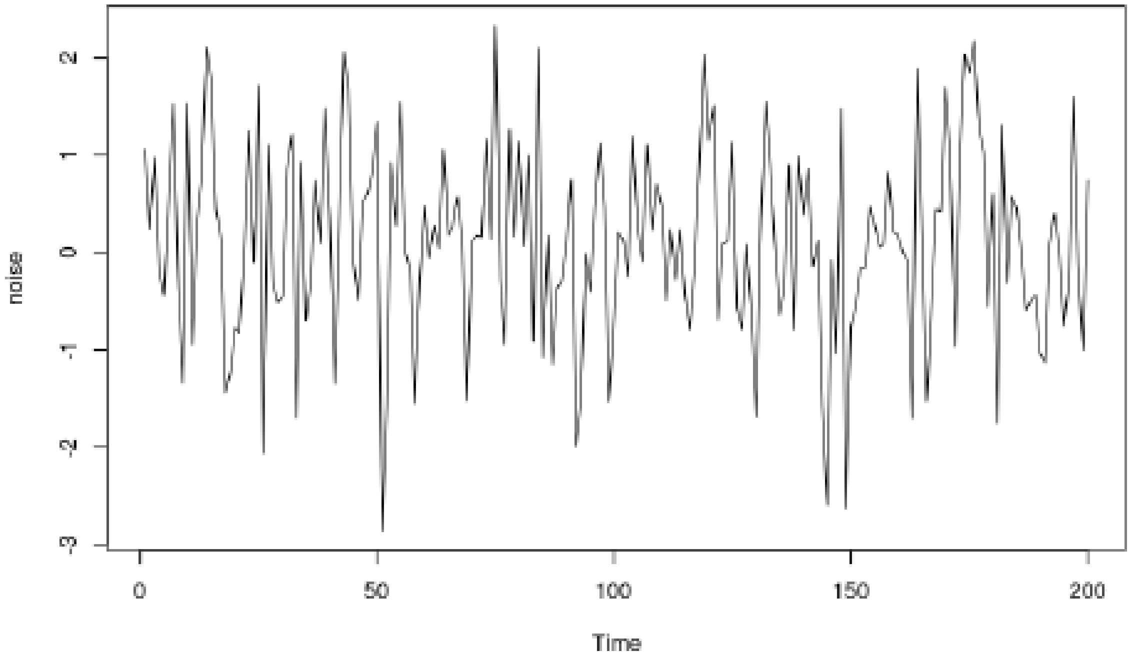
Partial Autocorrelation Function (PACF).

$$\pi_X(h) = \frac{Cov(X_{t+h}, X_t \mid X_{t+1}, \ldots, X_{t+h-1})}{\sqrt{Var(X_{t+h} \mid X_{t+1}, \ldots, X_{t+h-1}).Var(X_t \mid X_{t+1}, \ldots, X_{t+h-1})}}$$

The **partial autocorrelation** at ***lag h*** is the correlation that results after removing the effect of any correlations due to the terms at shorter lags.

White Noise

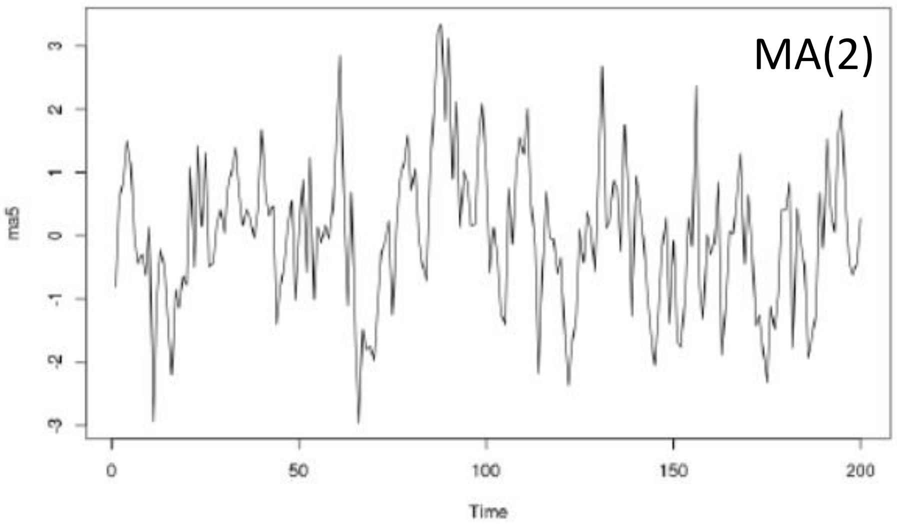
$$\rho(h) = \eta(h) = \begin{cases} 1 & \text{if } h = 0 \\ 0 & \text{Otherwise} \end{cases}$$



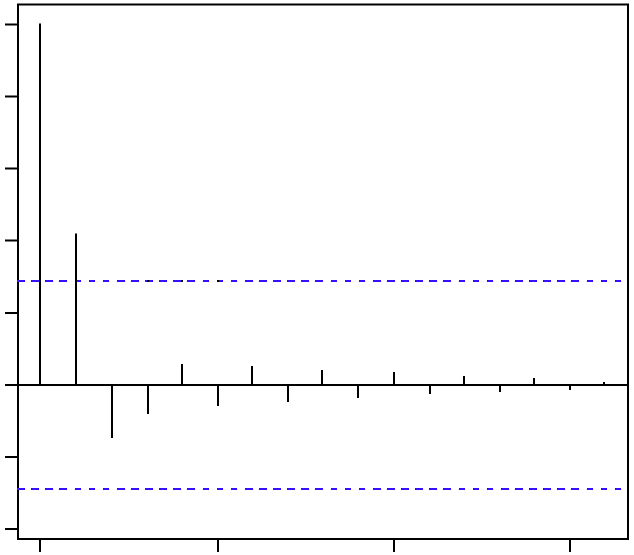
Moving Average (MA) Time Series

Moving Average time series $\{X_t\}$ of *Order* h (denoted $MA(h)$) is

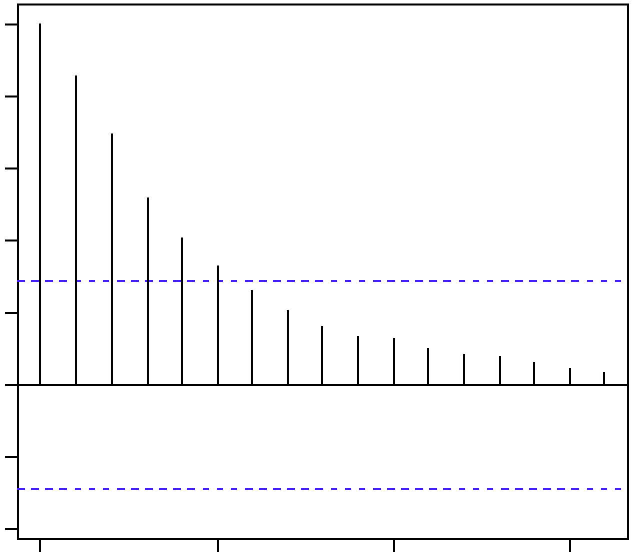
$$X_t = \mu + Y_t + \sum_{i=1}^{h} \alpha_i Y_{t-i} \qquad \{Y_t\} = \mathcal{W}(0, \sigma^2)$$



ACF

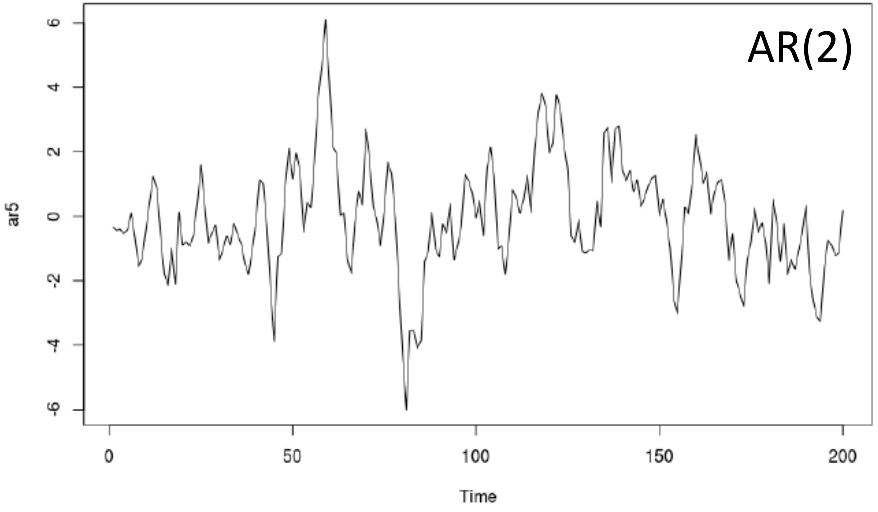


PACF



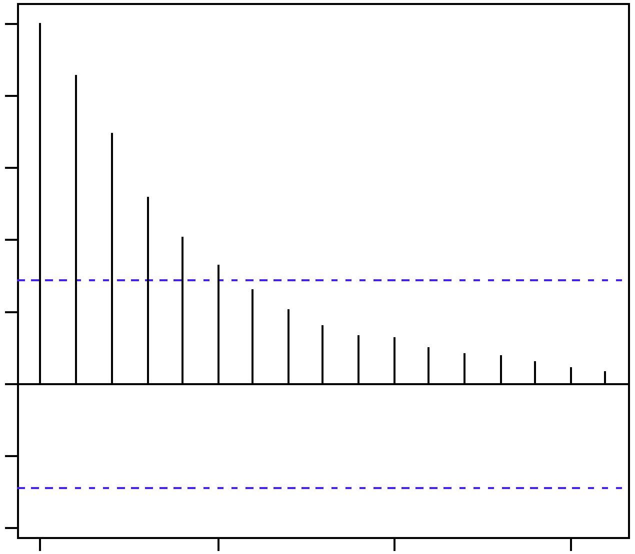
Auto Regressive (AR) Time Series

$AR(h)$ of $order\,h$ as a series $\{X_t\}_0^T$ where

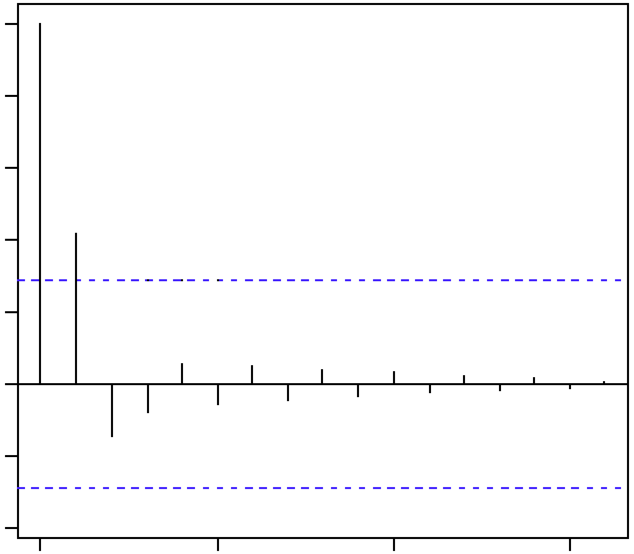


$$X_t = \mu + \sum_{i=1}^{h} \alpha_i X_{t-i} + Z_t, \quad Z_t = \mathcal{W}(0, \sigma^2)$$

ACF



PACF



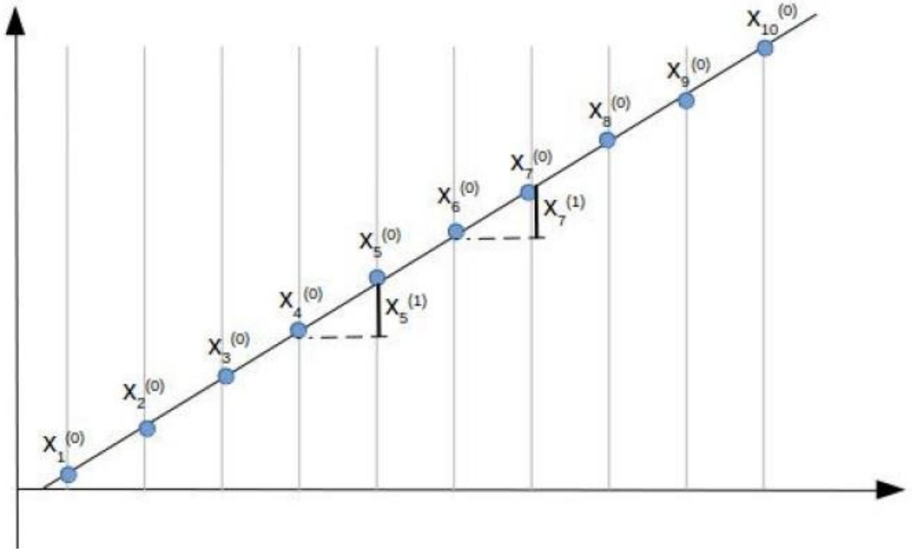
$ARMA(p,q)$ Models

$$X_t = \mu + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{i=1}^{q} \beta_i Z_{t-i} + Z_t$$

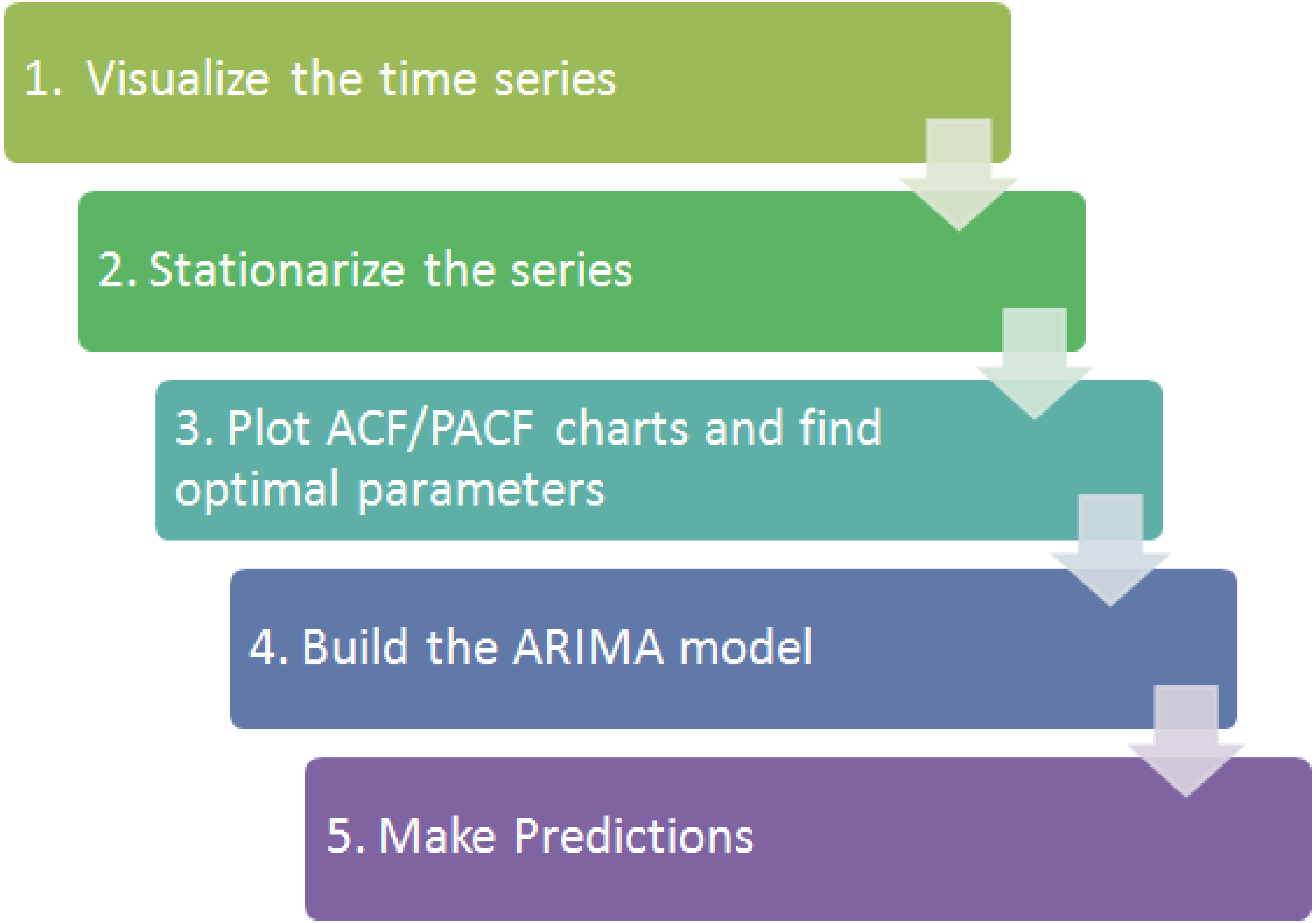
Process	**ACF** ($\rho_X(h)$)	**PACF** ($\pi_X(h)$)
$AR(p)$	Infinite waning exponential or sinusoidal tail	$\pi_X(h) = 0$ for $h > p$
$MA(q)$	$\rho_X(h) = 0$ for $h > q$	Infinite waning exponential or sinusoidal tail
$ARMA(p,q)$	Like $AR(p)$ for $h > q$	Like $MA(q)$ for $h > p$

ARIMA Models for Non-Stationary Series Analysis

- Stands for Auto Regressive Integrated Moving Average
- In ARIMA, the AR and MA are same as ARMA
- However, I indicates the amount of difference done
- If differencing done once, it is called I(1)
- Thus an ARIMA(p,d,q) model is a combination of AR(p) and MA(q) with I(d)



- A time series which is non-stationary can be converted to a stationary time series by differencing
- $Y_t' = Y_t - Y_{t-1}$
- If still not stationary, do second order differencing
- $Y_t'' = Y_t' - Y_{t-1}' = Y_t - 2Y_{t-1} + Y_{t-2}$


1. Visualize the time series
2. Stationarize the series
3. Plot ACF/PACF charts and find optimal parameters
4. Build the ARIMA model
5. Make Predictions

Thank you