Time Series and Forecasting

Muskaan Jain (MA23C022)

Department of Mathematics

Indian Institute of Technology, Madras

What is Time Series ?

- A Time Series is a sequential set of data points, usually measured over successive times.
- It is mathematically defined as a set of vectors $x(t)$,t=0,1,2, . . .

Components of Time Series

- **Trend:** It is a long term movement in a time series.
- **Seasonal:** Seasonal variations in a time series are fluctuations within a year during the season.
- **Cyclic:** The cyclical variation in a time series describes the medium-term changes in the series, caused by circumstances, which repeat in cycles. The duration of a cycle extends over longer period of time, usually two or more years.
- **Irregular:** Random variations in a time series are caused by unpredictable influences, which are not regular and also do not repeat in a particular pattern.

ACF and PACF

- Autocorrelation refers to the way the observations in a time series are related to each other and is measured by a simple correlation between current observation (x_t) and the observation p periods from the current one (x_{t-p}).

$$r_p = \frac{\sum_{t=p+1}^{N}(x_t - \bar{x})(x_{t-p} - \bar{x})}{\sum_{t=1}^{N}(x_t - \bar{x})^2}$$

- Partial Autocorrelations are used to measure the degree of association between x_t and x_{t-p} when the effects at other time lags are removed.

ACF and PACF

- **ACF (Auto-Correlation Function)**
  - The correlation between the observation at the current time spot and the observations at previous time spots.
- **PACF (Partial Auto-Correlation Function)**
  - The correlation between observations at two time spots given that we consider both observations are correlated to observations at other time spots. For example, today's stock price can be correlated to the day before yesterday, and yesterday can also be correlated to the day before yesterday. Then, PACF of yesterday is the "real" correlation between today and yesterday after taking out the influence of the day before yesterday.

Partial Autocorrelation Function (PACF)

PACF Estimation:
The PACF is estimated by controlling the effects of other lags, generally using linear regression. The PACF at a given lag is the coefficient of that lag obtained from the linear regression. The regression includes all the lags between the current time period and the given lag as independent variables. For instance, PACF at lag 3 can be estimated as:

$$Y_t = B_0 + B_1 Y_{t-1} + B_2 Y_{t-2} + B_3 Y_{t-3}$$

Here, $B_3 =$ **PACF at lag 3**

White Noise

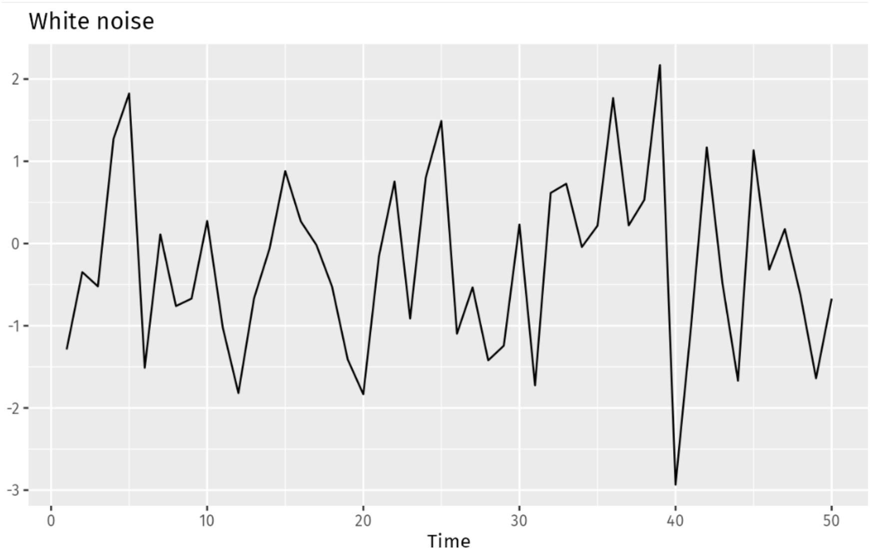
- $E(y_t) = \mu$

  $\text{var}(y_t) = \sigma^2$

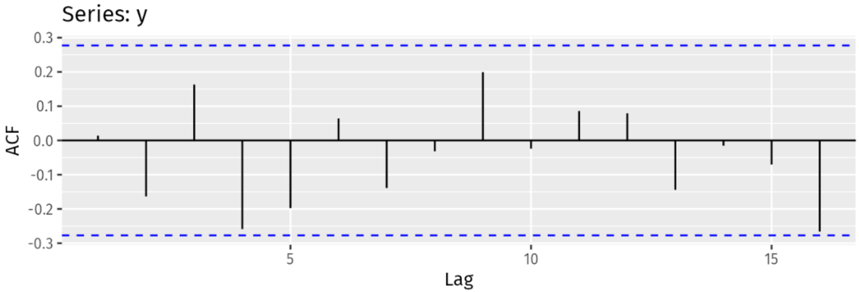
  $$\gamma_{t-r} = \begin{cases} \sigma^2 & \text{if } t = r \\ 0 & \text{otherwise} \end{cases}$$

- The white noise has constant mean and variance, and zero auto-covariance except at lag 0.
- The scatter plot of such series indicates no pattern and hence forecasting would be of no use.
- If $\mu = 0$ and other conditions hold, then it is known as zero mean white noise.

White Noise



White noise
2
1
0
-1
-2
-3
0
10
20
30
40
50
Time



Series: y
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
ACF
5
10
15
Lag

Strict Stationarity

- The foundation of time series analysis is stationarity. A time series $\{r_t\}$ is said to be **strictly stationary** if the joint distribution of $(r_{t_1}, \dots, r_{t_k})$ is identical to that of $(r_{t_1+t}, \dots, r_{t_k+t})$ for all t, where k is an arbitrary positive integer and $(t_1, \dots, t_k)$ is a collection of k positive integers. In other words, strict stationarity requires that the joint distribution of $(r_{t_1}, \dots, r_{t_k})$ is invariant under time shift.
- This is a very strong condition, requiring all moments (mean, variance, higher-order moments) and distributions to be constant over time.

Weak Stationarity

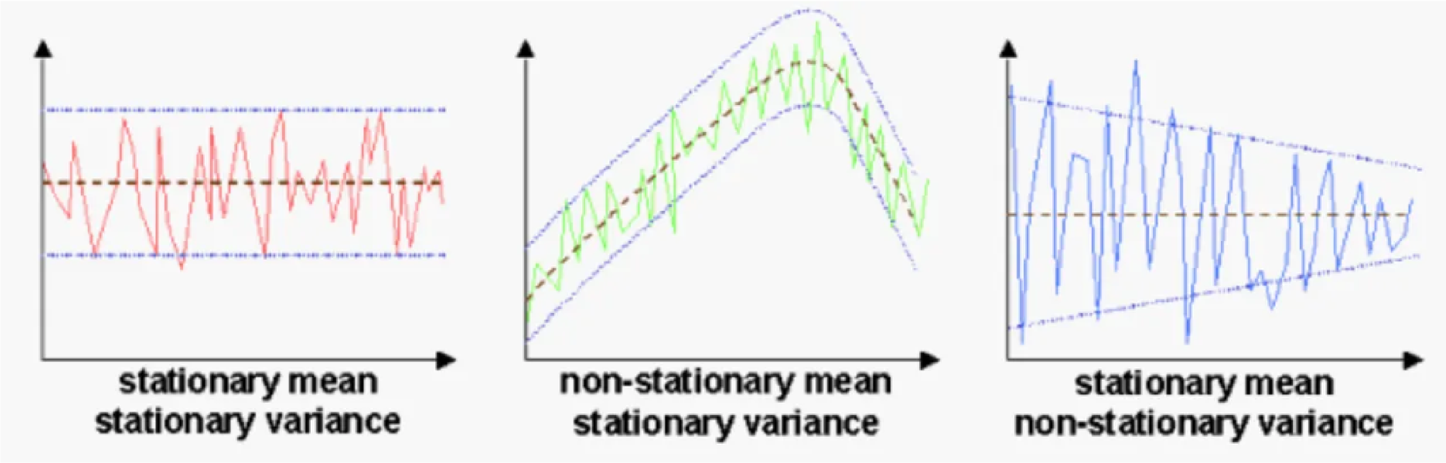
A time series $\{x_t\}$ is weakly stationary if both the mean of x_t and the covariance between x_t and $x_{t-\ell}$ are time-invariant, where ℓ is an arbitrary integer.
More specifically, $\{x_t\}$ is weakly stationary if:

(a) $E(x_t) = \mu$, which is a constant, and

(b) $\mathrm{Cov}(x_t, x_{t-\ell}) = \gamma_\ell$, which only depends on ℓ.

The weak stationarity implies that the time plot of the data would show that the T values fluctuate with constant variation around a fixed level.

Stationarity



Relation between Weak Stationary and Strict Stationary time series

Strict stationarity requires that the entire probability distribution of the series (all moments, including mean, variance, skewness, kurtosis, etc.) remains the same over time. Weak stationarity, on the other hand, only requires that the first two moments (mean and variance) and autocovariance remain constant over time, without demanding that the full distribution is invariant.
A weakly stationary time series may exhibit changes in higher-order moments or its underlying distribution over time, which means it can fail to meet the strict requirements of strict stationarity. Therefore, while every **strictly stationary** time series is **weakly stationary**.

Stationarity and Gaussian Time series

For a Gaussian time series, weak stationarity implies strict stationarity. This is because if a time series is Gaussian and weakly stationary (i.e., it has a constant mean and variance, and its autocovariance depends only on the lag), the higher-order moments are automatically determined by the first two moments. As a result, the full distribution remains constant over time.

A **Gaussian time series** is fully described by its mean and covariance. In this case , weak stationarity is sufficient to guarantee strict stationarity. Therefore, every weakly stationary Gaussian time series is also strictly stationary.

Modelling

The selection of a proper model is extremely important as it reflects the underlying structure of the series and this fitted model in turn is used for future forecasting.
A time series model is said to be linear or non-linear depending on whether the current value of the series is a linear or non-linear function of past observations.

Moving Average (MA)

The following represents MA(q) model:

- Variable as the linear combination of the error terms
  $y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q}$

  MA(q) model uses past errors as the explanatory variables.

- Expectation $E(y_t) = \mu$

- $\text{Var}(y_t) =$ Constant, also constant autocovariance
  $y_t = \mu + \sum_{i=1}^{q} \theta_i u_{t-i} + u_t$

- y_t depends on the white noise error terms

Autoregressive Model (AR)

An autoregressive model is one where the current value of a variable, y, depends upon only the values that the variable took in previous periods plus an error term.

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

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + u_t$$

Autoregressive Moving Average Model (ARMA)

There are situations where the time-series may be represented as a mix of both **AR** and **MA** models referred as ARMA(p,q).
The general form of such a time-series model, which depends on p of its own past values and q past values of white noise disturbances, takes the form:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_p y_{t-p} + u_t + \Phi_1 u_{t-1} + \Phi_2 u_{t-2} + \ldots + \Phi_q u_{t-q}$$

Deriving mean property

Use Stationarity to derive AR or MA model's properties.
For example, we have the following AR(1) model:

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t \tag{1}$$

Assuming that the series is weakly stationary, we have

$$E(r_t) = \mu, \quad \mathrm{Var}(r_t) = \gamma_0, \quad \text{and} \quad \mathrm{Cov}(r_t, r_{t-j}) = \gamma_j,$$

where μ and γ_0 are constant and γ_j is a function of j, not t. Taking the expectation of Eq. (1) and because $E(a_t) = 0$, we obtain

$$E(r_t) = \phi_0 + \phi_1 E(r_{t-1}).$$

Deriving mean property

Under the stationarity condition, $E(r_t) = E(r_{t-1}) = \mu$ and hence

$$\mu = \phi_0 + \phi_1 \mu \quad \text{or} \quad E(r_t) = \mu = \frac{\phi_0}{1 - \phi_1}.$$

This result has two implications for r_t. First, the mean of r_t exists if $\phi_1 \neq 1$. Second, the mean of r_t is zero if and only if $\phi_0 = 0$. Thus, for a stationary AR(1) process, the constant term ϕ_0 is related to the mean of r_t and $\phi_0 = 0$ implies that $E(r_t) = 0$.
Next, using $\phi_0 = (1 - \phi_1)\mu$, the AR(1) model can be rewritten as

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + a_t. \quad (2)$$

By repeated substitutions, the prior equation implies that

$$\begin{aligned} r_t - \mu &= a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \cdots \\ &= \sum_{i=0}^{\infty} \phi_1^i a_{t-i}. \quad (3) \end{aligned}$$

Deriving Variance Property

We have

$$\mathrm{Cov}(r_{t-1}, a_t) = 0.$$

This can be seen from the fact that r_{t-1} occurred before time t and a_t does not depend on any past information. Taking the square, then the expectation of Eq. 2, we obtain

$$\mathrm{Var}(r_t) = \phi_1^2 \mathrm{Var}(r_{t-1}) + \sigma_a^2,$$

where σ_a^2 is the variance of a_t and we make use of the fact that the covariance between r_{t-1} and a_t is zero.

Deriving Variance Property

Under the stationarity assumption, $\text{Var}(r_t) = \text{Var}(r_{t-1})$, so that

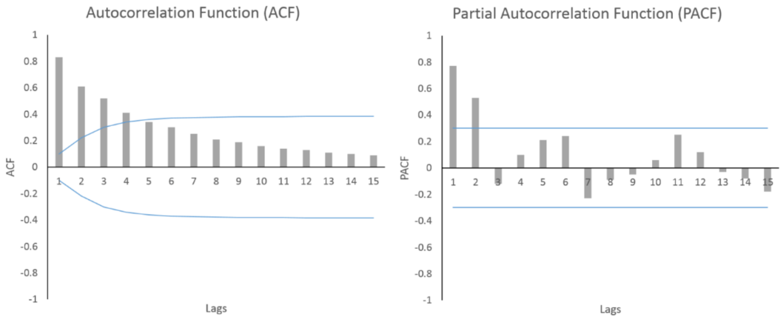
$$\text{Var}(r_t) = \frac{\sigma_a^2}{1 - \phi_1^2}$$

provided that $\phi_1^2 < 1$. The requirement of $\phi_1^2 < 1$ results from the fact that the variance of a random variable is bounded and non-negative. Consequently, the weak stationarity condition implies that $1 < \phi < 1$.

Rules of Using ACF and PACF

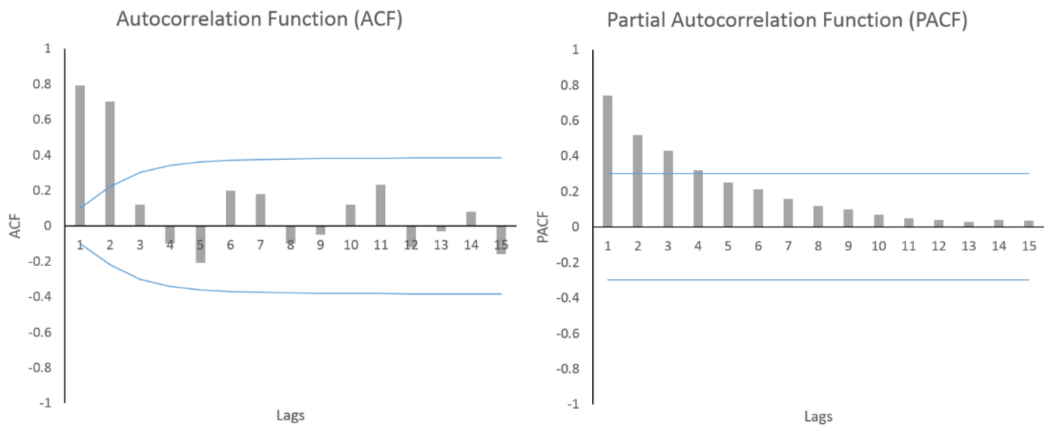
- Use **PACF** to determine the terms used in the AR model. Only significant terms will be chosen. The number of terms determines the order of the model.
  - For example, if the PACF of yesterday's stock price is significant. All other days' PACFs are not significant. Then, yesterday's stock price will be used to predict today's stock price. The AR model is called the first-order autoregression model.
- Use **ACF** to determine the terms used in the MA model.
- Choose a model by using PACF and ACF charts together.

Interpreting ACF and PACF Plots



A gradual geometrically declining ACF and a PACF that is significant for only a few lags indicate an AR process. In the figures, we can see that ACF is geometrically declining with lags. The PACF has 2 significant lags followed by a drop in PACF values and they become insignificant. With 2 significant PACF lags and gradually falling ACF, we can say that the series is an AR(2) process.

Interpreting ACF and PACF Plots



An MA (Moving Average) process is characterized by a PACF that decreases gradually in a geometric pattern, while the ACF shows significant values only at a few initial lags. This pattern is the inverse of what is observed in an AR process.

In this case, the PACF decays geometrically, and the ACF has significant values at the first two lags before becoming insignificant. This behavior suggests an MA(2) process.

Stationarity and Differencing

- A series which is **non-stationary** can be made stationary after **differencing**.
- A series which is stationary after being differentiated once is said to be **integrated of order 1** and is denoted by **I(1)**.
- In general, a series which is stationary after being differentiated d times is said to be **integrated of order** d, denoted **I(d)**.
- Therefore, a series which is stationary without differencing is said to be **I(0)**.

ARIMA

1. Distinction between **ARMA** and **ARIMA** is the *integration* component which brings us back to the subject of stationarity.
2. In reality, most economic variables are non-stationary hence they have to go through a transformation process called **differencing** before they become stationary.
3. The transforming process is also called **integration**.
4. So **ARIMA** informs the researcher or reader that the series in question has gone through an integration process before being used for any analysis.
5. Hence, the moment a nonstationary variable is differenced before becoming stationary, such is known as an *integrated* variable.

Understanding ARIMA (p, d, q)

ARIMA (p, d, q) tells us the number of lags of the dependent variable (p), how many times the variable is differenced to become stationary (d), and the number of lags of the error term (q).

Such that:

1. **ARIMA(1, 1, 2)** indicates the model has: one lag of the dependent variable (**1**), the variable being used is of first-difference stationary (**1**) and two lags of the error term (**2**).
2. **ARIMA(1, 0, 1)** indicates the model has: one lag of the dependent variable (**1**), the variable being used is of level stationary (**0**) and one lag of the error term (**1**).
3. **ARIMA(1, 0, 1)** = **ARMA (1, 1)** if the series is stationary in level.

SARIMA Model

Some financial time series such as quarterly earning per share of a company exhibits certain cyclical or periodic behavior. Such a time series is called a seasonal time series.
The **SARIMA** (Seasonal AutoRegressive Integrated Moving Average) model is an extension of the ARIMA model that supports seasonality in the time series.

SARIMA Notation:

$$\text{SARIMA}(p, d, q)(P, D, Q, s)$$

This notation consists of two parts:
Non-seasonal part: (p, d, q)

- **p**: The number of *autoregressive (AR)* terms. This captures how past values influence the current value.
- **d**: The number of *differencing* operations needed to make the series stationary.
- **q**: The number of *moving average (MA)* terms. This captures how past forecast errors influence the current value.

Seasonal part: (P, D, Q, s)

- **P**: The number of *seasonal autoregressive* (SAR) terms. This captures how past seasonal values (from the same season in previous cycles) influence the current value.
- **D**: The number of *seasonal differencing* operations needed to remove seasonal trends.
- **Q**: The number of *seasonal moving average* (SMA) terms. This captures how past forecast errors from the same season influence the current value.
- **s**: The length of the *seasonal cycle* (e.g., 12 for monthly data with an annual seasonality).

Seasonal Differencing

Seasonal differencing is a technique used to remove the seasonal patterns in a time series. It helps make the data stationary by eliminating regular, repeating patterns (such as quarterly or annual cycles) that can obscure the underlying trend and noise in the data.

Seasonal Differencing

In seasonal differencing, each data point in the series is replaced by the difference between that data point and the data point from the same period in the previous cycle (season). This removes the influence of repeating seasonal patterns.

The general formula for seasonal differencing with period s is:

$$y_t = x_t - x_{t-s}$$

Where:

- x_t is the original value at time t.
- x_{t-s} is the value at the same seasonal period in the previous cycle (e.g., $t - 12$ for monthly data with annual seasonality).
- y_t is the seasonally differenced value at time t.