Time Series Analysis

This syllabus is designed for individuals aiming to develop expertise in time series analysis, relevant for both aspiring and working data analysts and data scientists. It progressively builds your knowledge from foundational concepts to advanced techniques, enabling you to tackle real-world time series problems and create impactful data analysis projects.

By the end of this course, you will be able to:

- Understand the core concepts of time series data and its characteristics.
- Pre-process and clean time series data effectively.
- Visualize time series data to identify trends, seasonality, and patterns.
- Apply statistical methods for time series analysis, including trend analysis, seasonality decomposition, and autocorrelation.
- Build and evaluate time series forecasting models using techniques like ARIMA, SARIMA, and Prophet.
- Implement advanced time series analysis methods with a focus on practical applications.

1. Time Series data

Time series data is a collection of observations or measurements collected or recorded over a sequence of time intervals.

These data points are usually collected at consistent intervals, like daily, hourly, or monthly.

Date	Open	High	Low	Close	Adj Close	Volume
2024-01-02	187.149994	188.440002	183.889999	185.639999	185.152283	82488700
2024-01-03	184.220001	185.880005	183.429993	184.250000	183.765945	58414500
2024-01-04	182.149994	183.089996	180.880005	181.910004	181.432098	71983600
2024-01-05	181.990005	182.759995	180.169998	181.179993	180.703995	62303300
2024-01-08	182.089996	185.600006	181.500000	185.559998	185.072495	59144500
2024-01-09	183.919998	185.149994	182.729996	185.139999	184.653595	42841800
2024-01-10	184.350006	186.399994	183.919998	186.190002	185.700836	46792900
2024-01-11	186.539993	187.050003	183.619995	185.589996	185.102417	49128400
2024-01-12	186.059998	186.740005	185.190002	185.919998	185.431549	40444700
2024-01-16	182.160004	184.259995	180.929993	183.630005	183.147568	65603000

Here are some real-life examples:

- **Stock prices:** The price of a stock fluctuates throughout the day. Recording the price every minute creates a time series.
- **Sales data:** A retail store might track its daily sales figures, forming a time series.
- **Social media activity:** The number of likes, shares, or comments on a social media post can be tracked over time, creating a time series.
- **Weather data:** Temperature, humidity, and rainfall can all be recorded at regular intervals, forming separate time series.

1.1 Key characteristics:

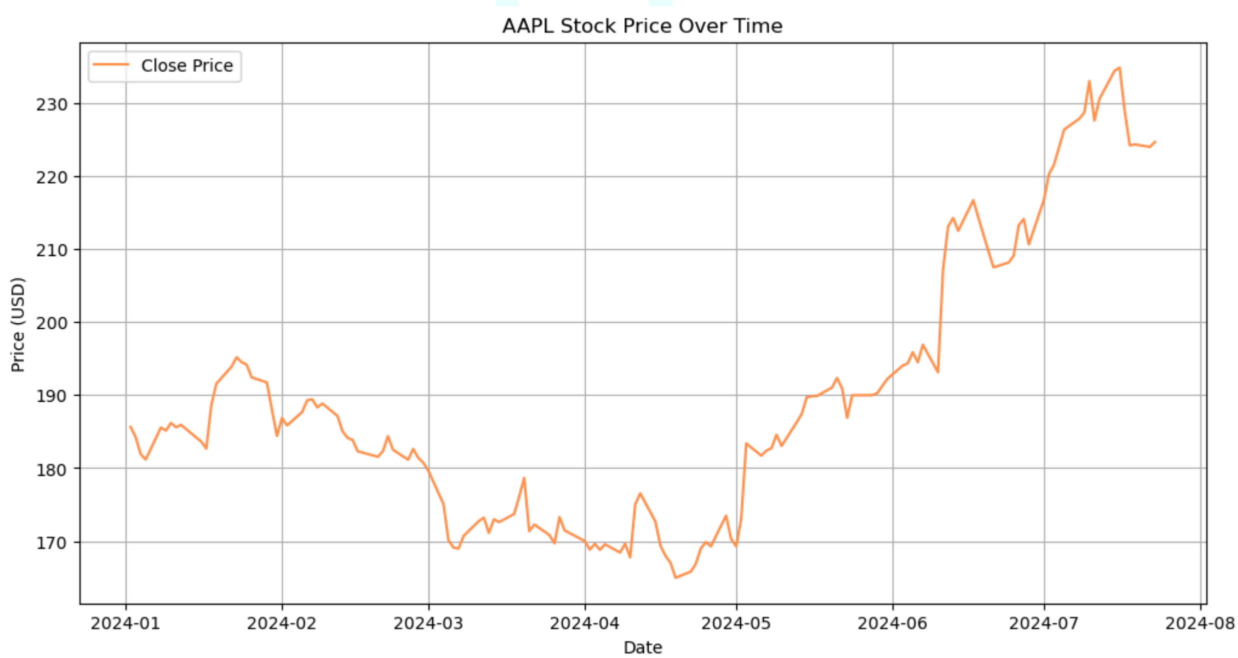
- **Chronological Order:** Time series data is ordered based on time, with each data point corresponding to a specific time period, such as hours, days, months, or years.

- **Sequential Dependence:** The value of each observation in a time series is often dependent on the values of previous observations. This sequential dependence can manifest as trends, seasonality, or cyclic patterns.

- **Temporal Components:** Time series data often contains various temporal components such as trend, seasonality, cyclical patterns, and irregular fluctuations (noise or randomness).

- **Constant Frequency:** Time series data typically has a constant frequency, meaning that observations are recorded at regular intervals. However, in practice, missing values or irregular intervals may occur.

- **Dynamic Nature:** Time series data can be dynamic, meaning that patterns and relationships may change over time due to external factors or underlying dynamics.



For Apple Inc. (AAPL), we obtained historical stock price data from January 1, 2023, to January 1, 2024, and we plotted the closing prices during that period of time.

2. Time Series Analysis

Time series analysis is a statistical technique used to extract meaningful insights and patterns from time series data. It involves analyzing the past behavior of a time series to forecast future values or understand the underlying structure of the data.

Goals: The ultimate aim is twofold:

- **Understanding the past:** This involves uncovering the underlying structure of the data, like trends and seasonal variations.
- **Forecasting the future:** By leveraging the patterns we discover, we can predict future values of the time series. This is crucial for applications like business planning, resource allocation, and risk management.

3. Time Series Decomposition

It involves breaking down a time series data into its constituent parts, revealing the underlying patterns that contribute to the overall behavior.

Time series decomposition separates these elements:

- **Trend:** This captures the long-term direction of the data, whether it's increasing, decreasing, or staying flat.

  Example: An upward trend in the stock prices of a tech company over several years due to continuous innovation and market expansion.

- **Seasonality:** Seasonality refers to regular and predictable patterns or cycles in a time series that repeat at fixed intervals, such as daily, monthly, or yearly.

  Example: Increased sales during holiday seasons like Christmas.

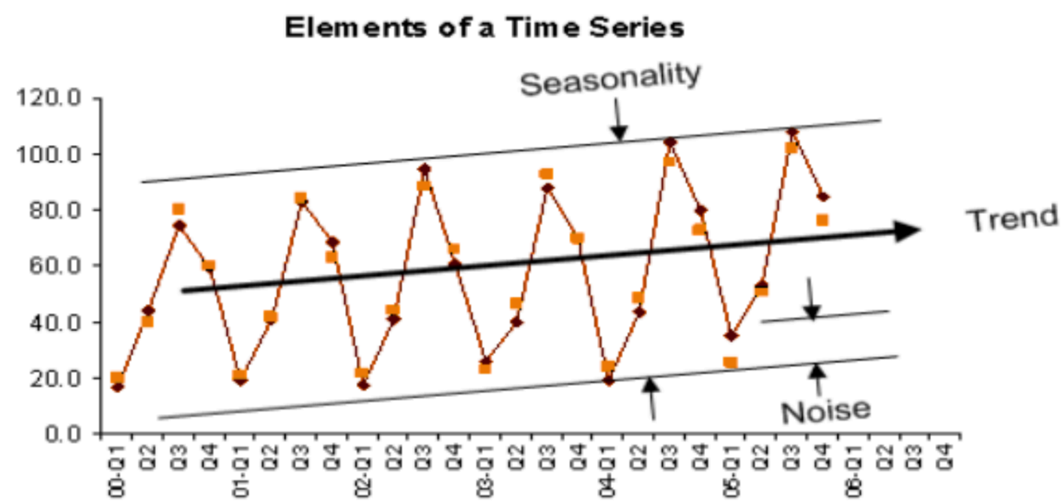
- **Cyclicity:** Cyclic patterns are long-term fluctuations in a time series that are not as regular or predictable as seasonal patterns. They occur over periods longer than a year and are often associated with economic or business cycles.

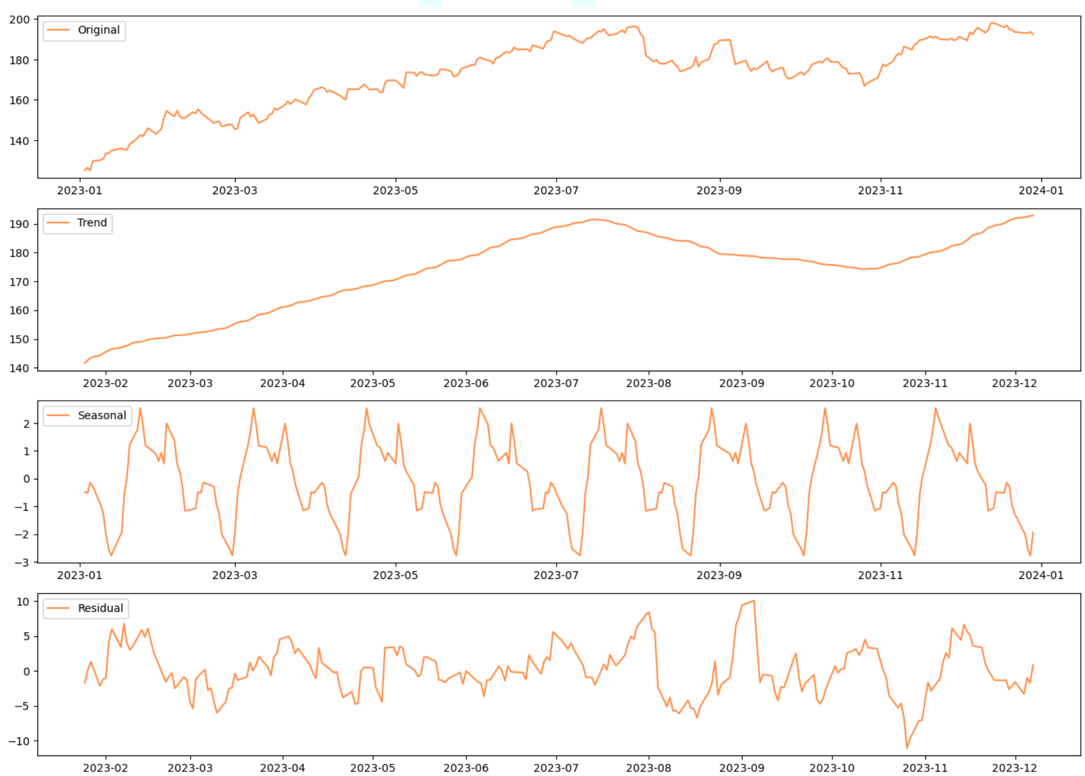
  Examples:

  - **Economic Cycles:** Periods of expansion and recession in an economy.
  - **Business Cycles:** Fluctuations in sales and profits of a company over multiple years due to market conditions.

- **Residuals (or Noise):** This component represents the random fluctuations in the data that cannot be explained by the other components.

  Example: Sudden drops or spikes in stock prices due to unexpected news or events.





There are two main types of decomposition methods:

Additive model: In this model, the trend, seasonality, and residual components are assumed to be added together to form the original time series. This applies when the seasonal variations remain constant regardless of the overall trend.

$$Y_t = T_t + S_t + R_t$$

Multiplicative model: Here, the seasonal component is multiplied by the trend-cycle component to obtain the original series. This is suitable for situations where the seasonal fluctuations are proportional to the level of the trend.

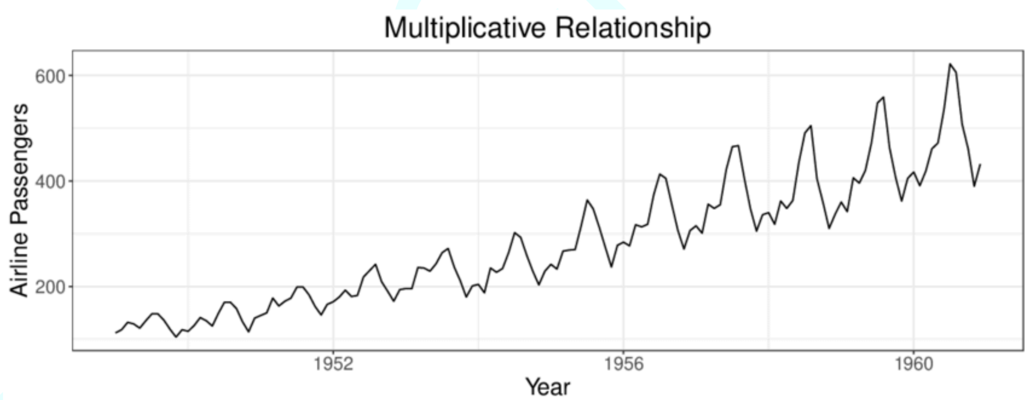
$$Y_t = T_t * S_t * R_t$$

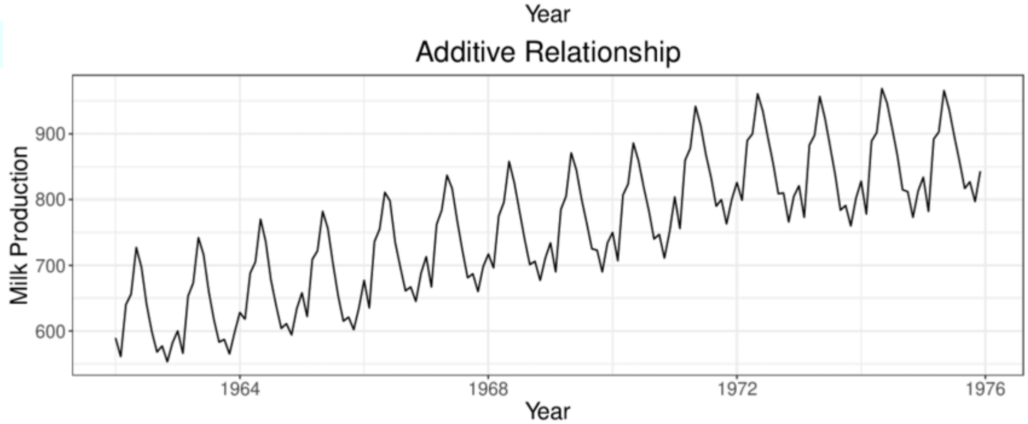
Here:

S represents the Seasonal variation

T encodes Trend plus Cycle

R describes the Residual or the Error component



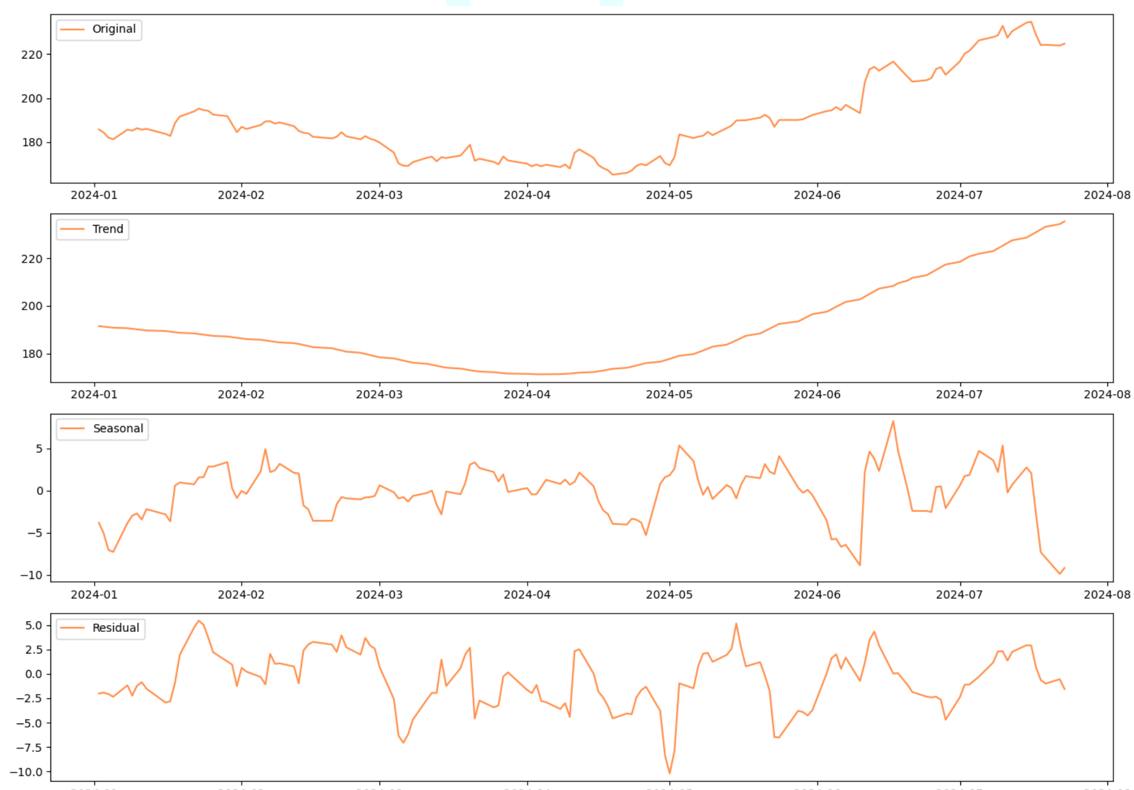


4. STL Decomposition (Seasonal and Trend Decomposition using LOESS)

It is a robust and flexible method for decomposing a time series into its seasonal, trend and residual components. It uses a statistical technique called LOESS (locally estimated scatterplot smoothing).

4.1 How does it differ from classical decomposition?

1. Classical methods typically assume fixed seasonal patterns. STL can handle a wider variety of seasonal patterns, including those that change over time.
2. Classical methods offer limited control over the decomposition process. STL provides users with more flexibility to adjust smoothing parameters for the trend and seasonal components.
3. Classical methods can be more susceptible to outliers in the data, potentially distorting the decomposed components. STL exhibits better robustness in handling outliers.
4. STL can only handle additive decomposition. In a multiplicative relationship, classical decomposition with a multiplicative model might be necessary.



4.2 Choosing the right decomposition method: The best decomposition method depends on your specific data and the relationships between the components. If you suspect a multiplicative relationship between seasonality and trend, classical decomposition with a multiplicative model might be a better choice. However, for most cases, STL's flexibility and robustness make it a very effective tool.

5. Stationarity

Time series is considered stationarity if its statistical properties such as mean (no trend), variance, and autocorrelation are constant over time. The properties of one section of a data are much like the properties of the other sections of the data.

5.1 Why do we need stationarity?

- It is important because many statistical methods and models (like forecasting models), assume that the time series data are stationary.

- It helps prediction become easier because the statistical properties in future will be the same as they are now.

Assume the mean of each distribution is the same. The fact can be seen in the following way: Let's imagine we have ten identical machines making a certain product, like a lightbulb. Assume that every machine is operated for an hour. It is now simple to see that the total (average) output produced by ten machines is equal to the total (average) output produced by one machine operating for ten hours.

5.2 Types of Stationarity

1. **Strict Stationarity**
   A time series Y_t is strictly stationary if the joint distribution of any collection of values $Y_{t1}, Y_{t2}, ...Y_{tn}$ is the same as the joint distribution of $Y_{t1+k}, Y_{t2+k}, ...Y_{tn+k}$ for all t1, t2, …… tn and all k. It means joint probability distribution remains unchanged when shifted along any time period.

2. **Weak Stationarity**
   A time series is weakly stationary if its mean, variance, and autocorrelation structure are time-invariant (constant over time).

5.3 Testing for Weak Stationarity

1. **Augmented Dickey Fuller (ADF) Test:**
   This is a widely used test for stationarity that checks for the presence of a unit root in the data. A unit root essentially means the data has a non-stationary trend, and differencing might be necessary.

   The ADF test works by fitting a specific model to the data and then evaluating the null hypothesis (H0) that there is a unit root (non-stationary) against the alternative hypothesis (Ha) that the series is stationary.

   Null Hypothesis (H0): The time series has a unit root (non-stationary).
   Alternative Hypothesis (H1): The time series is stationary.

   Decision Criteria:
   - If the ADF statistic is less than the critical value at a given significance level, we reject the null hypothesis.
   - A low p-value (typically less than 0.05) indicates strong evidence against the null hypothesis, suggesting the time series is stationary.

   ```
   ADF Test Results:
   ADF Statistic: 0.097
   p-value: 0.966
   Critical Values:
      1%: -3.478
      5%: -2.883
      10%: -2.578
   ```

   - The ADF statistic is positive and very close to zero.
   - The p-value is extremely high (0.9658), which is much greater than any typical significance level.
   - The ADF statistic is significantly greater than all critical values.

   Therefore, we fail to reject the null hypothesis.

2. **Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test:**
   This test takes a different approach by focusing on the null hypothesis (H0) that the series is stationary, with the alternative hypothesis (Ha) that it's non-stationary.

The KPSS test statistic measures the variance of the cumulative sum of the residuals from fitting a constant mean model to the data. Higher values of the test statistic indicate stronger evidence against stationarity.

Null Hypothesis (H0): The time series is stationary.
Alternative Hypothesis (H1): The time series is non-stationary.

Decision Criteria:

- If the KPSS statistic is greater than the critical value at a given significance level, we reject the null hypothesis.
- A low p-value (typically less than 0.05) indicates strong evidence against the null hypothesis, suggesting the time series is non-stationary.

```
KPSS Test Results:
KPSS Statistic: 0.5
p-value: 0.01
Critical Values:
   10%: 0.119
   5%: 0.146
   2.5%: 0.176
   1%: 0.216
```

- The KPSS statistic is significantly larger than all critical values.
- The p-value is very low (0.01), which is less than any typical significance level.

Therefore, we reject the null hypothesis. This means there is strong evidence to suggest that the time series is **non-stationary**.

5.4 Testing for Strict Stationarity

After testing for weak stationarity, which is a necessary condition for strict stationarity, we can compare the empirical distributions of different segments of the time series using statistical tests like the Kolmogorov-Smirnov test.

The Kolmogorov-Smirnov (K-S) test is a non-parametric test used to compare the cumulative distribution functions of two samples. It can be used to assess the distribution of a part of a series with another part of the same series.

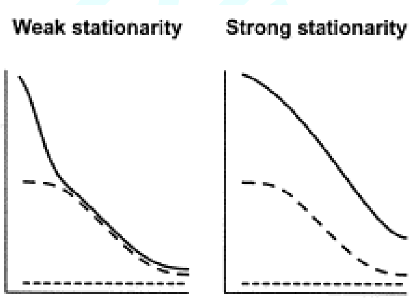
A high p-value suggests that there is no significant difference between the empirical distribution function.

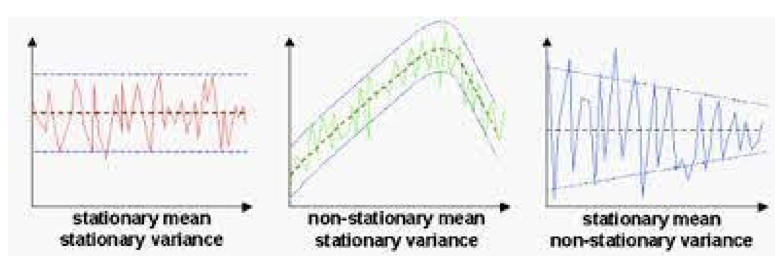
Strict Stationary Series - K-S Test Statistic: 0.068, p-value: 0.6109
Weak Stationary Series - K-S Test Statistic: 0.184, p-value: 0.0004

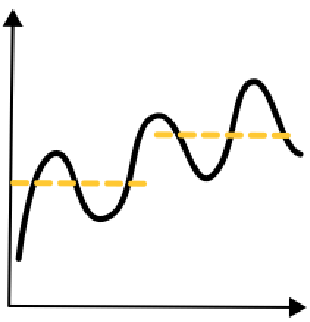
5.5 Which one to choose?

Most common time series models (e.g., ARIMA) assume weak stationarity. If the goal is forecasting, weak stationarity is usually sufficient. If modeling the entire distribution of the series, strict stationarity is ideal.

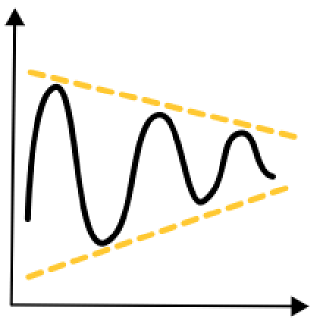
For instance, in the finance sector, where analysts mainly pay attention to return variance, autocorrelation, and mean, weak stationarity might be adequate for modeling stock returns. However, strict stationarity, which takes into consideration the joint distribution of price movements over many time periods, can be more suited when modeling the full stock price process.



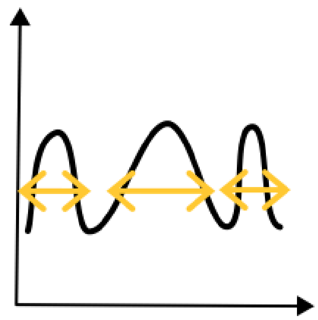




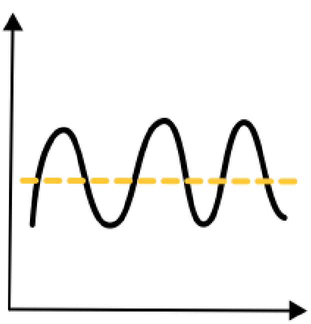
Mean dependent on time



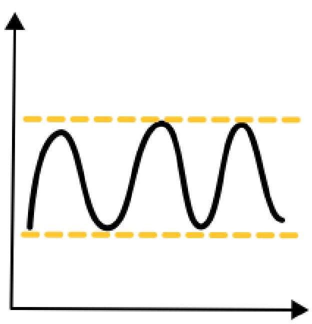
Variance dependent on time



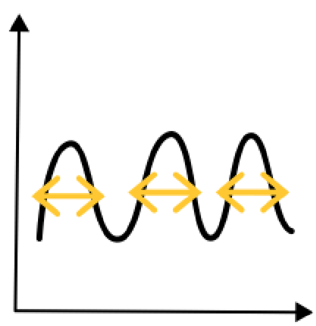
Covariance dependent on time



Mean independent on time



Variance independent on time



Covariance independent on time

5.6 Making a time series, Stationary

1. **Differencing**
   Differencing is a method to remove trends and seasonality from a time series. It involves subtracting the previous observation from the current observation.

   First-order differencing: $Y'_t = Y_t - Y_{t-1}$

   Second-order differencing (when the data will not appear to be stationary after first order differencing): $Y''_t = Y'_t - Y'_{t-1}$

2. **Transformation**
   Transformation stabilizes the variance of a time series. Common transformations include:

   Logarithmic transformation: Taking the logarithm of the data can be helpful for series with large variations or exponential trends.

Power transformation: Raising the data to a specific power can sometimes stabilize the variance of the series.
Box-Cox transformation: This is a more general transformation that can encompass both logarithmic and power transformations.

3. **De-trending**
   It involves removing a trend component from the time series, which can be done using regression or differencing.

   There are various detrending methods, including:
   Linear detrending: This method fits a straight line to the data and subtracts it from the original series, essentially removing the linear trend.
   Moving average detrending: This method calculates the moving average of the series and subtracts it from the original data to remove the trend component.

4. **Seasonal Adjustment**
   It involves removing the seasonal component from a time series. This can be done using techniques like STL decomposition.

5.7 Choosing the right method:

The best method for making your time series stationary depends on the nature of the non-stationarity present in your data. Here are some general guidelines:

- **Trends:** If the data has a trend (upward or downward), differencing or detrending techniques are suitable options.
- **Seasonality:** For seasonal patterns, seasonal differencing or decomposition methods are more appropriate.
- **Variance:** If the time series shows exponential growth or multiplicative seasonality. Transformations like logarithms can help stabilize the variance if it's an issue.
- **Over-differencing:** Be cautious of over-differencing your data. While it might remove trends, it can also introduce spurious seasonality and make the data less interpretable.
- **Domain knowledge:** Consider your understanding of the data generating process when choosing a transformation method. For instance, a logarithmic transformation might not be suitable if negative values are not meaningful in your context.

6. White Noise and Random Walk:

6.1 White noise

A time series with no pattern, trend, or seasonality is known as white noise. It is made up of independently dispersed random values that have the same mean, variance, and probability distribution as a whole. Additionally uncorrelated is white noise, which denotes that there is no linear relationship between the numbers at various time intervals.

- Constant or Zero Mean:
- Constant Variance: $Var(w_t) = \sigma^2$
- No Autocorrelation: $Cov(w_t, w_{t-k}) = 0 \, for \, k \neq 0$

6.2 Characteristics: Unpredictable: No predictable patterns, Completely random. If you find the time series is a white noise, there is no need to go for the next step as it can not be used for prediction or forecasting.

Imagine flipping a fair coin repeatedly. Each flip, heads or tails, is completely independent of the one before it.

6.3 Random walk

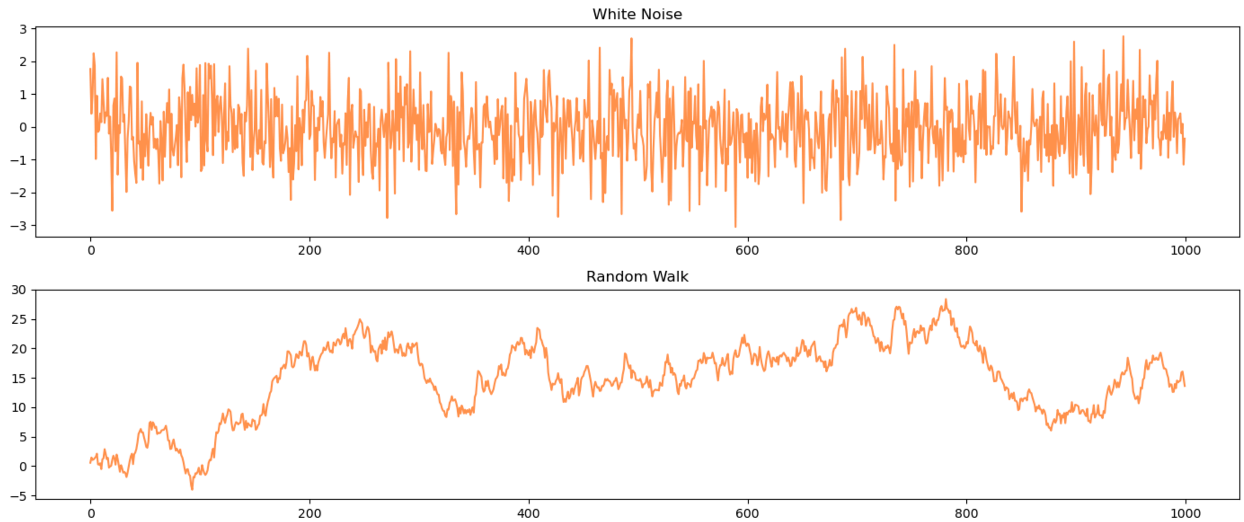
A time series with a cumulative pattern but no trend or seasonality is called a random walk. It is non stationary time series where the value at time t is the value a time t-1 plus an error term (white noise)

$$y_t = y_{t-1} + e_t$$

6.4 Characteristics:

- Non Stationary: Mean and variance change over time.
- No predictable patterns: It lacks predictable patterns, trends or seasonality
- Dependence on latest observation: Each value is heavily dependent on the previous value plus white noise.
- Its first difference is stationary

The behavior of financial assets, including stock prices, exchange rates, and interest rates, is frequently modeled using a random walk.



6.5 Identifying White Noise and Random Walk

1. By seeing its graph and statistical characteristics.
2. Look at its plot and autocorrelation function (ACF). The plot of a white noise series will show no apparent pattern or direction, while the plot of a random walk series will show a drifting or wandering pattern. The ACF of a white noise series will show no significant spikes, except at lag zero, while the ACF of a random walk series will show significant spikes that decay slowly over time.
3. You can use the Ljung-Box test to check whether the series has significant autocorrelations, or the Dickey-Fuller test to check whether the series is stationary or not. A white noise series will have no significant autocorrelations and will be stationary, while a random walk series will have significant autocorrelations and will be non-stationary.

```
ADF Test for White Noise:
ADF Statistic: -32.462559337689974
p-value: 0.0
Critical Value (1%): -3.437
Critical Value (5%): -2.864
Critical Value (10%): -2.568

KPSS Test for White Noise:
KPSS Statistic: 0.23806270020371642
p-value: 0.1
Critical Value (10%): 0.347
Critical Value (5%): 0.463
Critical Value (2.5%): 0.574
Critical Value (1%): 0.739

ADF Test for Random Walk:
ADF Statistic: -2.79787305269428
p-value: 0.05857823023908064
Critical Value (1%): -3.437
Critical Value (5%): -2.864
Critical Value (10%): -2.568

KPSS Test for Random Walk:
KPSS Statistic: 1.2935550026767026
p-value: 0.01
Critical Value (10%): 0.347
Critical Value (5%): 0.463
Critical Value (2.5%): 0.574
Critical Value (1%): 0.739

Ljung-Box Test for White Noise:
      lb_stat  lb_pvalue
10  14.025574   0.171828

Ljung-Box Test for Random Walk:
        lb_stat  lb_pvalue
10  8828.660312        0.0
```

7. Time Series Forecasting Models

7.1 Autoregressive (AR) Model

An autoregressive (AR) model uses information from past values to predict future values. The AR model specifies that the output variable depends linearly on its own previous values.

The order of the AR model, denoted by p, represents the number of lagged observations included in the model.

y_t is called an AR or order p, AR(p), if it satisfies

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots\ldots\ldots \varphi_p y_{t-p} + \varepsilon_t$$

Where ε_t is white noise and c is constant

7.2 Moving Average (MA) Model

Moving average represents the value of a time series as a linear function of past error terms (or residuals).

The order of the MA model, denoted by q, represents the number of lagged forecast errors included in the model.

Rather than using past values of the forecast variable in a regression, a moving average model uses past forecast errors in a regression-like model.

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots\ldots\ldots \theta_p \varepsilon_{t-q}$$

Where ε_t is white noise and ε_{t-i} are the lagged values of the error terms

7.3 Autoregressive Moving Average (ARMA) Model

It combines the Autoregressive (AR) and Moving Average (MA) models to analyze and forecast time series data. It is a powerful tool for capturing both the temporal dependencies in the data and the error terms.

ARMA (p,q)

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots\ldots\ldots \varphi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots\ldots\ldots \theta_p \varepsilon_{t-q}$$

7.4 Autoregressive Integrated Moving Average (ARIMA) Model

It combines autoregressive (AR), integrated (I), and moving average (MA) components to handle non-stationary data. The I part involves differencing the data to make it stationary.

ARIMA(p,d,q)

$$\Delta y_t = \varphi_1 \Delta y_{t-1} + \varphi_2 \Delta y_{t-2} + \ldots\ldots\ldots \varphi_p \Delta y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots\ldots\ldots \theta_p \varepsilon_{t-q}$$

7.5 Seasonal Autoregressive Integrated Moving Average (SARIMA) Model

The SARIMA model extends the ARIMA model to support the modeling of seasonal data. It includes additional seasonal terms to capture seasonal patterns.

A SARIMA model is denoted as SARIMA (p,d,q) (P,D,Q) s

Where P, D and Q are orders for the seasonal part, s is the period.

7.6 Vector AutoRegressive (VAR) Model

The VAR model is a generalization of the univariate autoregressive (AR) model to multiple parallel time series. It captures the linear interdependencies among multiple time series.

7.7 Vector Moving Average (VMA) Model

The Vector Moving Average (VMA) model is a multivariate extension of the Moving Average (MA) model used in univariate time series analysis. The VMA model captures the linear dependencies among multiple time series and their lagged error terms.

7.8 Vector AutoRegressive Moving Average (VARMA) Model

The VARMA model combines the Vector Autoregressive (VAR) and Vector Moving Average (VMA) models. It captures the relationships among multiple time series and their past values and past errors.

7.9 Vector AutoRegressive Integrated Moving Average (VARIMA) Model

The VARIMA model extends the VARMA model to include integration (differencing), similar to how ARIMA extends ARMA. It is used when the time series data are non-stationary and need differencing to achieve stationarity.

8. Smoothing Methods

Smoothing techniques are used to reduce noise and highlight significant patterns, such as cycles and trends. By using these techniques, time series data can be improved for forecasting and subsequent analysis by being more stable and readable.

8.1 Importance of Smoothing

- **Noise Reduction:** Smoothing reduces the short-term fluctuations and noise in the data, making the underlying patterns clearer.
- **Trend Identification:** It helps in identifying long-term trends and seasonal patterns in the data.

- **Forecasting:** Smoothed data is often used as a basis for forecasting future values in the time series.

8.2 Moving Average (MA)

The Moving Average method calculates the average of the time series data (Y_{t-i}) within a specific number of periods, called the window size (N).

$$MA_t = \frac{1}{N}\sum_{i=0}^{N-1} Y_{t-i}$$

1. **Simple Moving Average (SMA):** Averages over a fixed window size.
2. **Weighted Moving Average (WMA):** Weights the data points, giving more importance to recent observations.
3. **Exponential Moving Average (EMA):** Applies exponentially decreasing weights to older observations.

8.3 Exponential Smoothing (ES)

It assigns exponentially decreasing weights to past observations, giving more importance to recent data points.

$$S_t = \alpha X_t + (1 - \alpha) S_{t-1}$$

where , S_t = smoothed value at time t

X_t = observed value at time t

S_{t-1} = previous smoothed statistic

α = smoothing parameter between 0 and 1

Selecting alpha: A high value of α gives more weight to the current observation, while a low value of α gives more weight to the previous forecast.

1. **Single Exponential Smoothing (SES):** Suitable for data with no trend or seasonality

2. **Double Exponential Smoothing or Holt's linear (DES):** Accounts for linear trends in the data
3. **Triple Exponential Smoothing or Holt-Winters (TES):** Handles both trend and seasonality

Difference between EMA and ES:

EMA	**ES**
α is calculated based on the chosen window size N. $\alpha = 2/(N + 1)$	α is user-defined and flexible for different methods.
Primarily focuses on smoothing data for trend analysis without addressing seasonality.	Can handle trends and seasonality.

When to use which:

- Moving Average:
  - Best for Data with less variability, no clear trend, or seasonality.
  - Great for short-term smoothing.
  - Does not react well to sudden shifts or trends.
  - When you need a simple, short-term smoothing without much concern for trends or seasonal patterns.
  - or straightforward tasks where immediate reactions to new data are less critical (e.g., noise reduction).

- Exponential Smoothing:
  - Faster adaptation to new trends compared to moving averages.
  - Can model both trend and seasonality effectively (DES and TES).
  - For forecasting or scenarios where you need a model that can adapt to recent changes without lagging too much.

9. Granger causality test

The Granger causality test is a statistical hypothesis test used to determine if one time series can predict another time series. It is based on the idea that if one time series X Granger-causes another time series Y, then past values of X should help in forecasting future values of Y better than past values of Y alone.

This test is mostly used when you want to forecast multivariate data using models VAR, VMA, VARMA, VARIMA.

Example: Suppose you want to determine if today's Apple's stock price can help forecast tomorrow's Tesla's stock price.

10. ACF and PACF

10.1 ACF

Autocorrelation is the correlation of a time series with a lagged version of itself. A time lag refers to the amount of time by which the data is shifted. It measures the degree to which a current value in the series is related to past values. This concept is fundamental in time series analysis as it helps to understand the internal structure and dependency in the data.

10.2 PACF

The Partial Autocorrelation Function (PACF) is a tool used in time series analysis to measure the correlation between observations of a time series separated by various time lags, but unlike the Autocorrelation Function (ACF), it controls for the values of the time series at all shorter lags. The PACF helps in identifying the direct relationship between observations at specific lags, eliminating the influence of the intervening lags.

10.3 Key Difference from ACF:

ACF captures the direct and indirect correlations of the series with its lags.

PACF captures only the direct correlations of the series with its lags.

The error bands on the above ACF plot are represented by blue bars, and anything falling inside of them is not statistically significant. Correlation readings outside of this range are therefore most likely a result of a correlation and not a statistical anomaly. By default, the confidence interval is set to 95%.

Because the signal is always perfectly correlated with itself, it makes logical that ACF is

Model	ACF	PACF
AR	Tails off	Cuts off after lag p
MA	Cuts off after lag p	Tails off
ARMA	Tails off	Tails off

- Tails off means geometric decay or sine functions like graphs.

AR order (p): How many lags is required to predict values

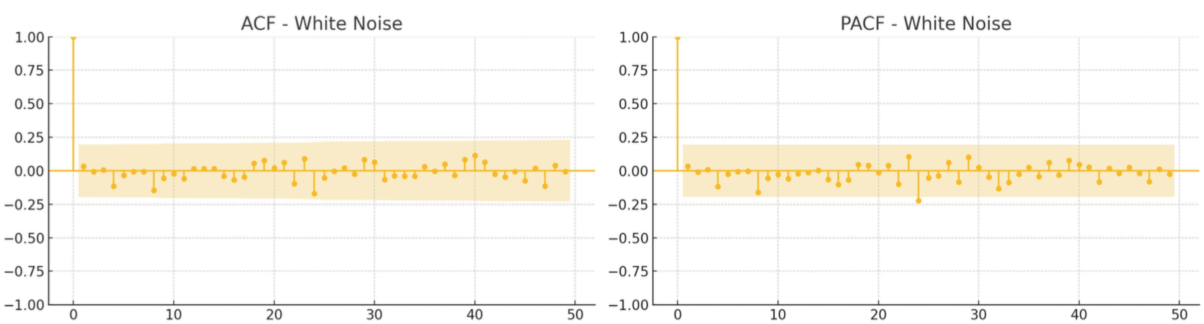
MA order(q): How many previous errors should be considered to predict the values

Differencing order (d): Number of times the data needs to be differenced to make it stationary

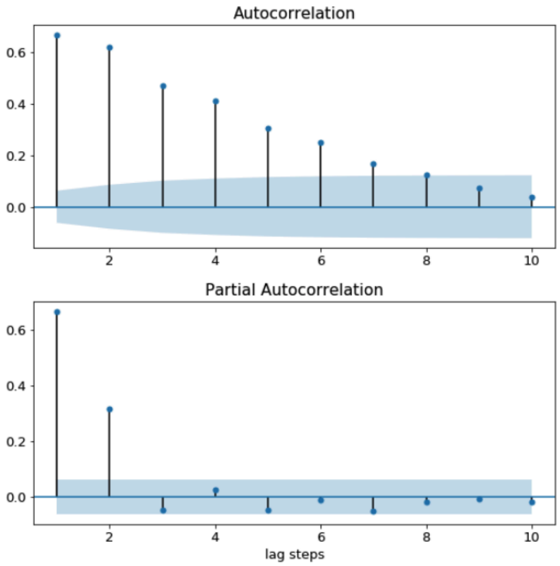
Seasonal Period (m): Number of periods in one seasonal cycle

10.4 Identifying Models from ACF and PACF

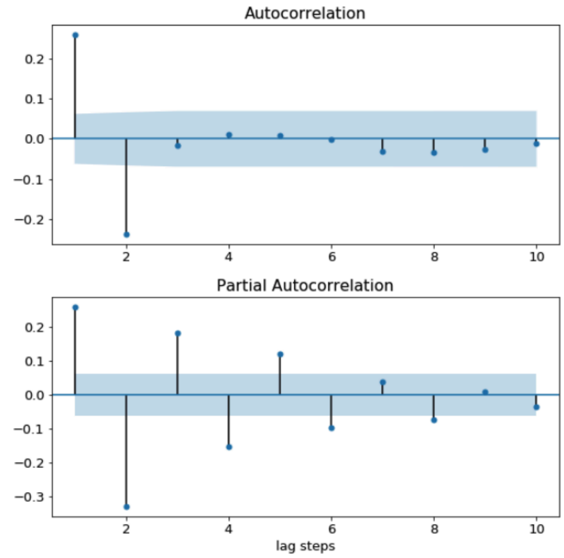
sx



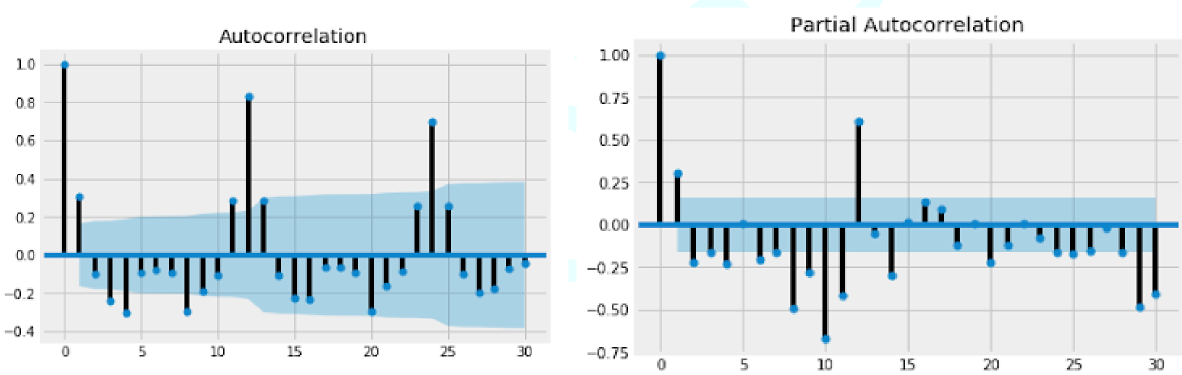
White Noise: For the White Noise series, both the ACF and PACF show no significant lags, indicating no correlation at any lag.



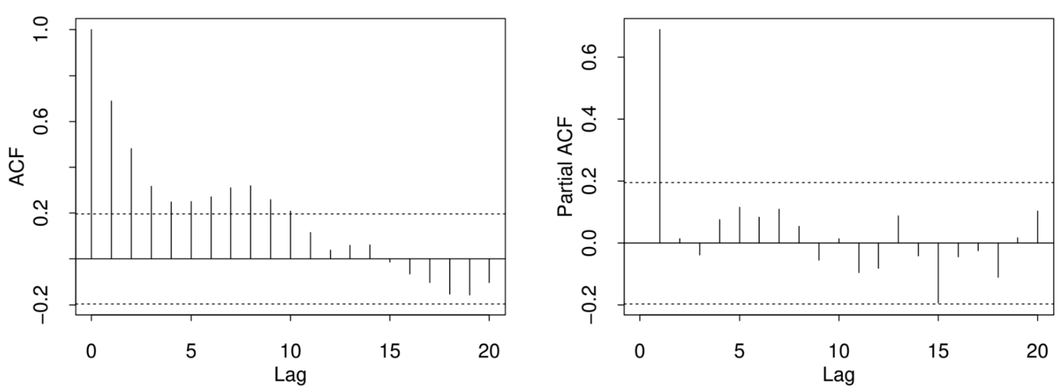
AR(2)



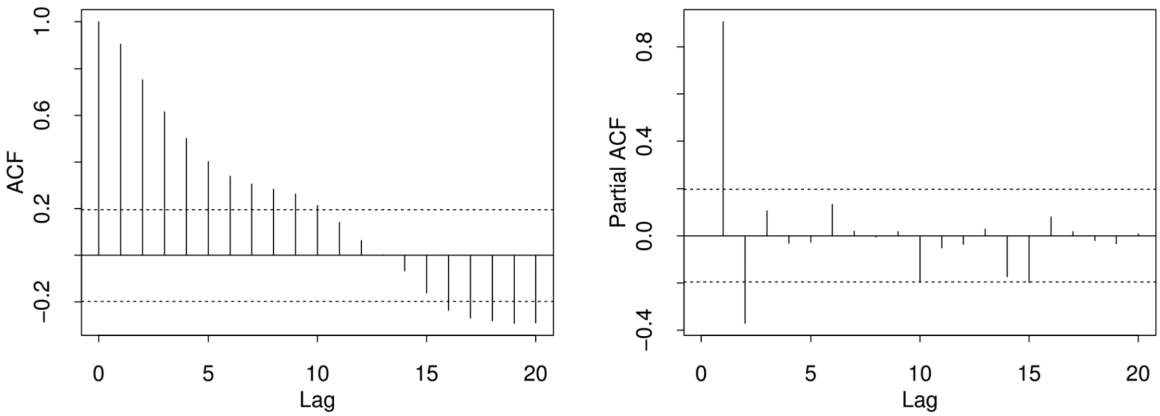
MA(2)



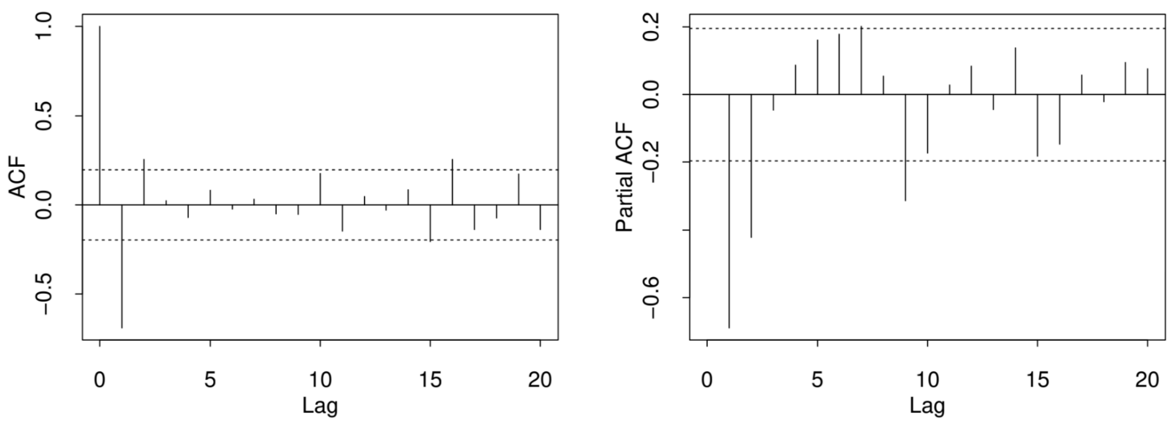
SARIMA (1,1,1,12)



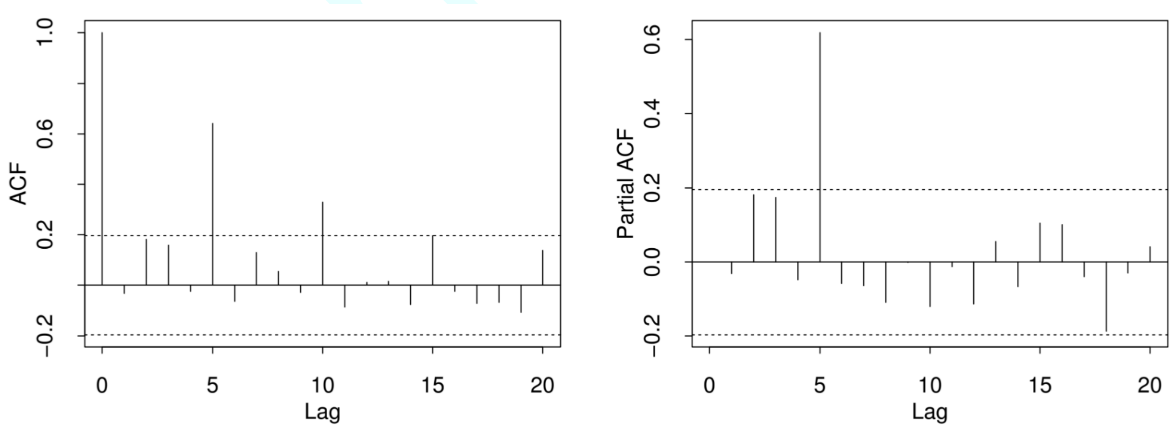
AR(1)



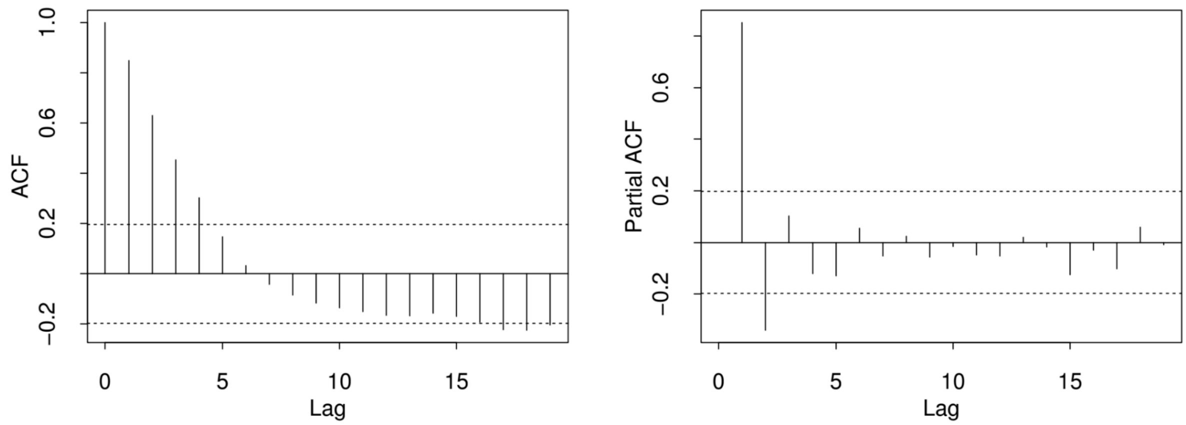
AR(2)



AR(2) / MA(2) - THIS ONE

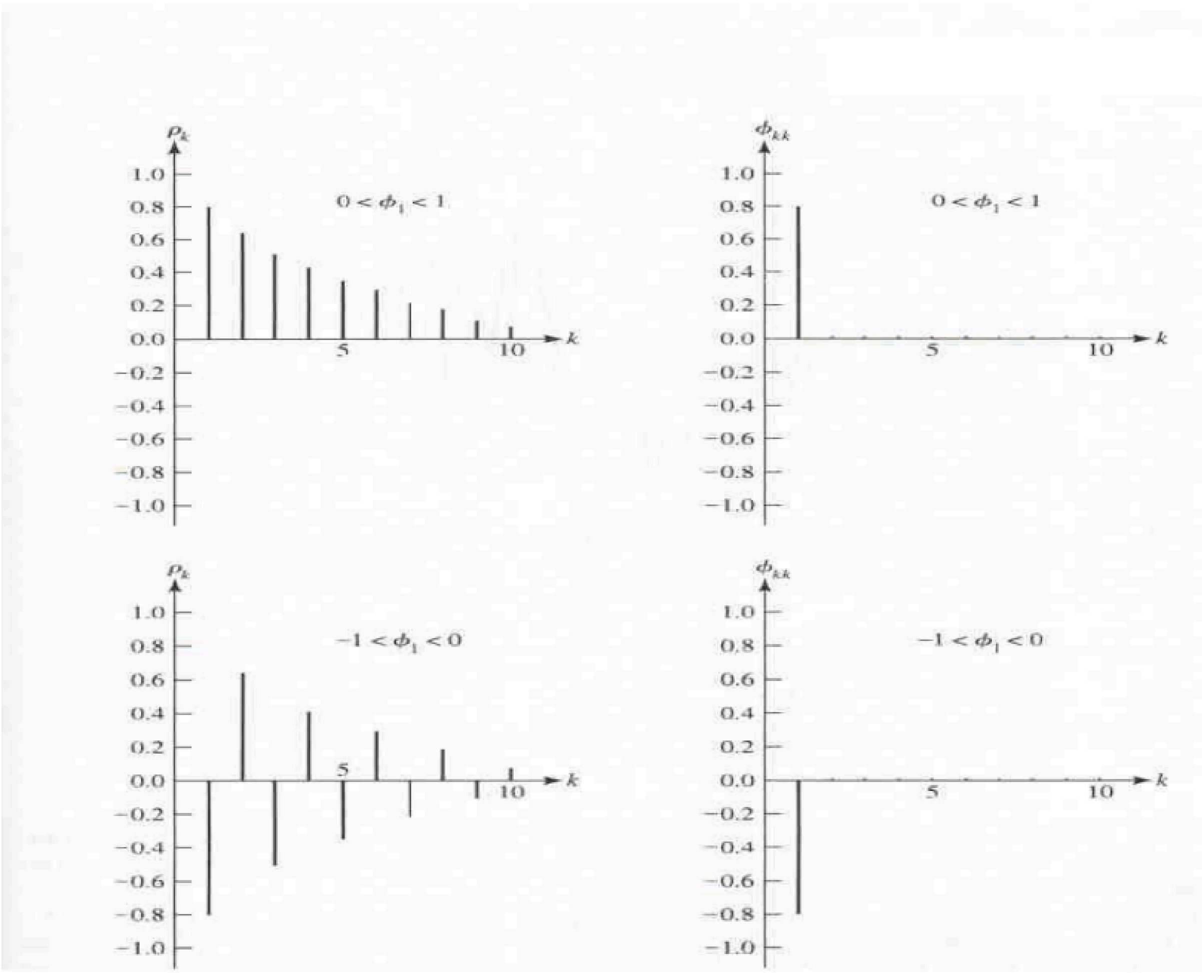


Seasonal AR(1)



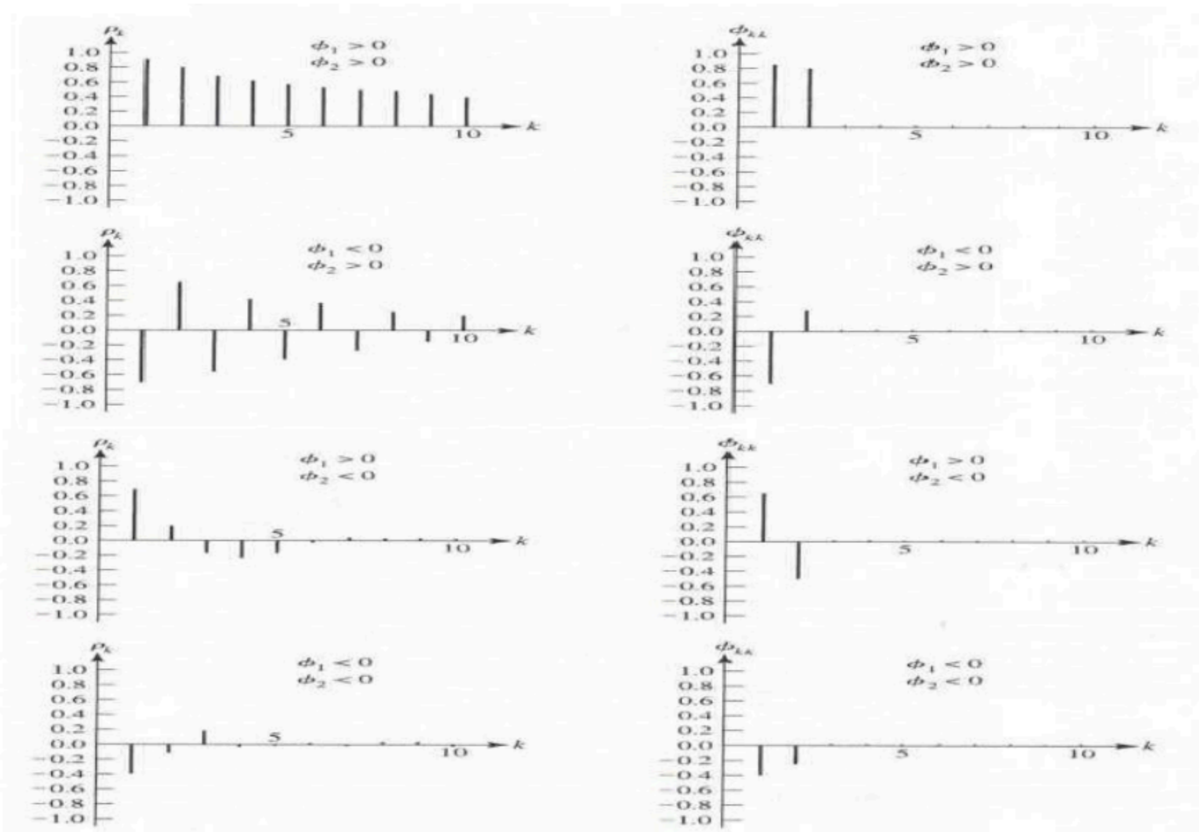
ARMA(2,4)

AR(1) PROCESS



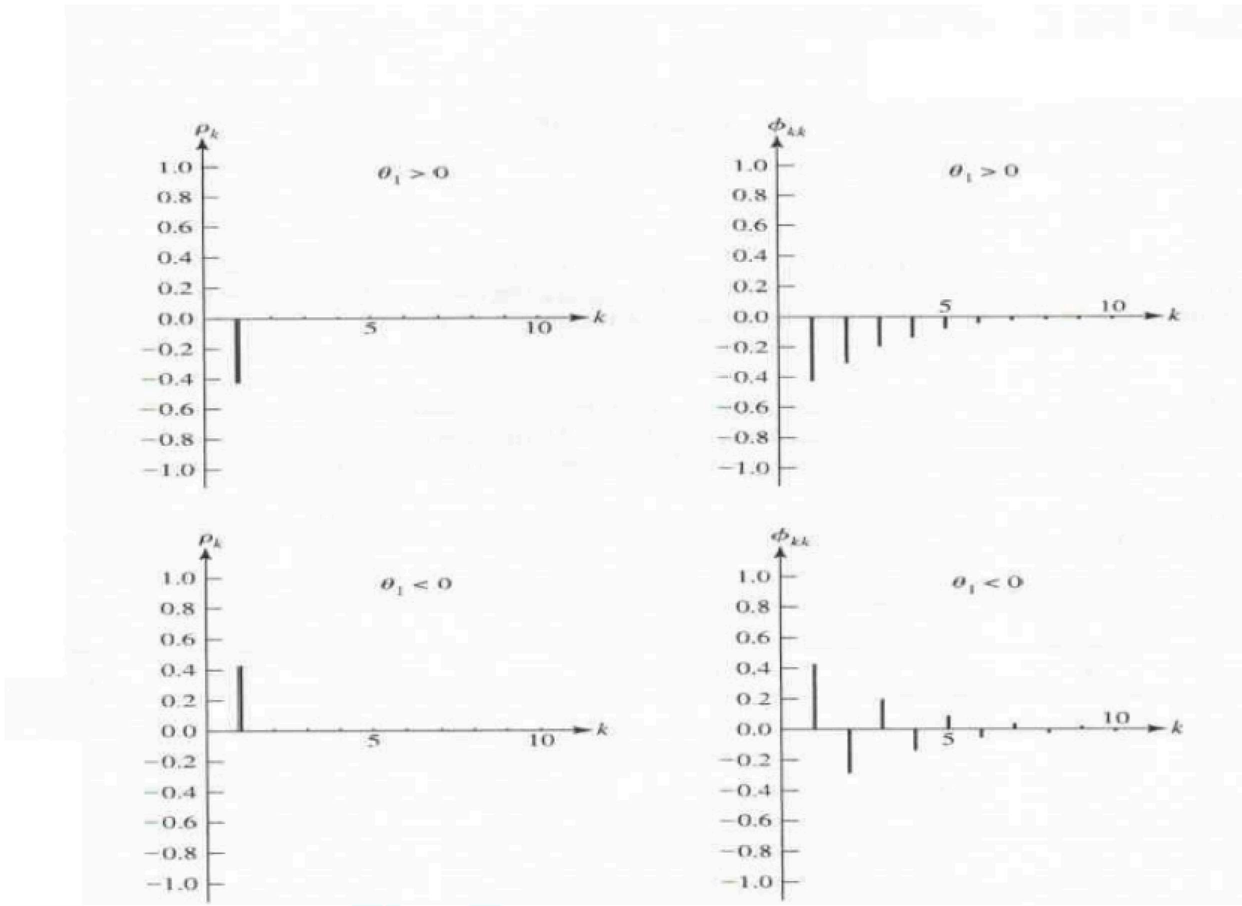
ρ_k
$0 < \phi_1 < 1$
ϕ_{kk}
$0 < \phi_1 < 1$
ρ_k
$-1 < \phi_1 < 0$
ϕ_{kk}
$-1 < \phi_1 < 0$
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
5
10
k

AR(2) PROCESS



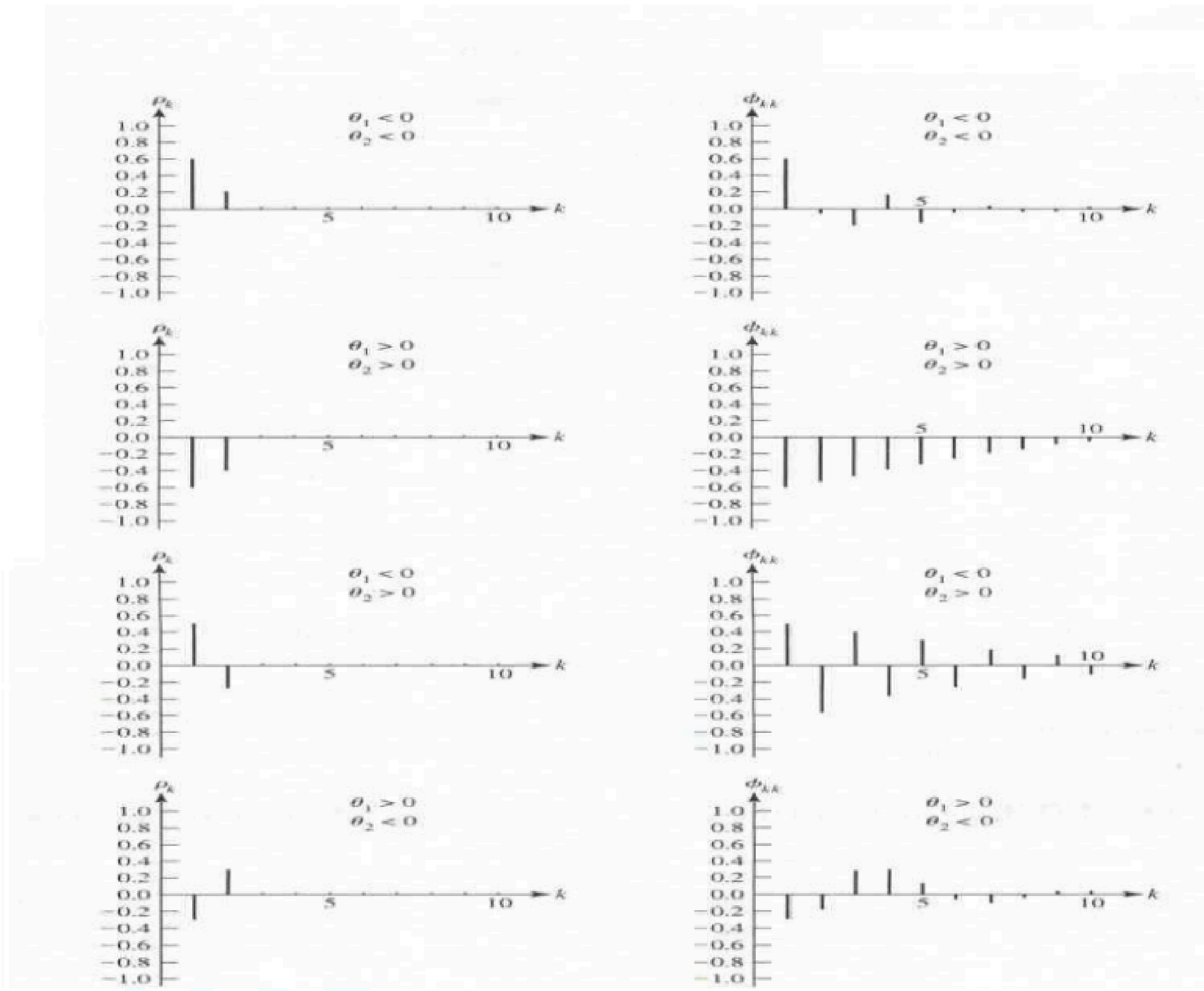
ρ_k
$\phi_1 > 0$
$\phi_2 > 0$
ϕ_{kk}
$\phi_1 > 0$
$\phi_2 > 0$
$\phi_1 < 0$
$\phi_2 > 0$
$\phi_1 < 0$
$\phi_2 > 0$
$\phi_1 > 0$
$\phi_2 < 0$
$\phi_1 > 0$
$\phi_2 < 0$
$\phi_1 < 0$
$\phi_2 < 0$
$\phi_1 < 0$
$\phi_2 < 0$
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
5
10
k

MA(1) PROCESS

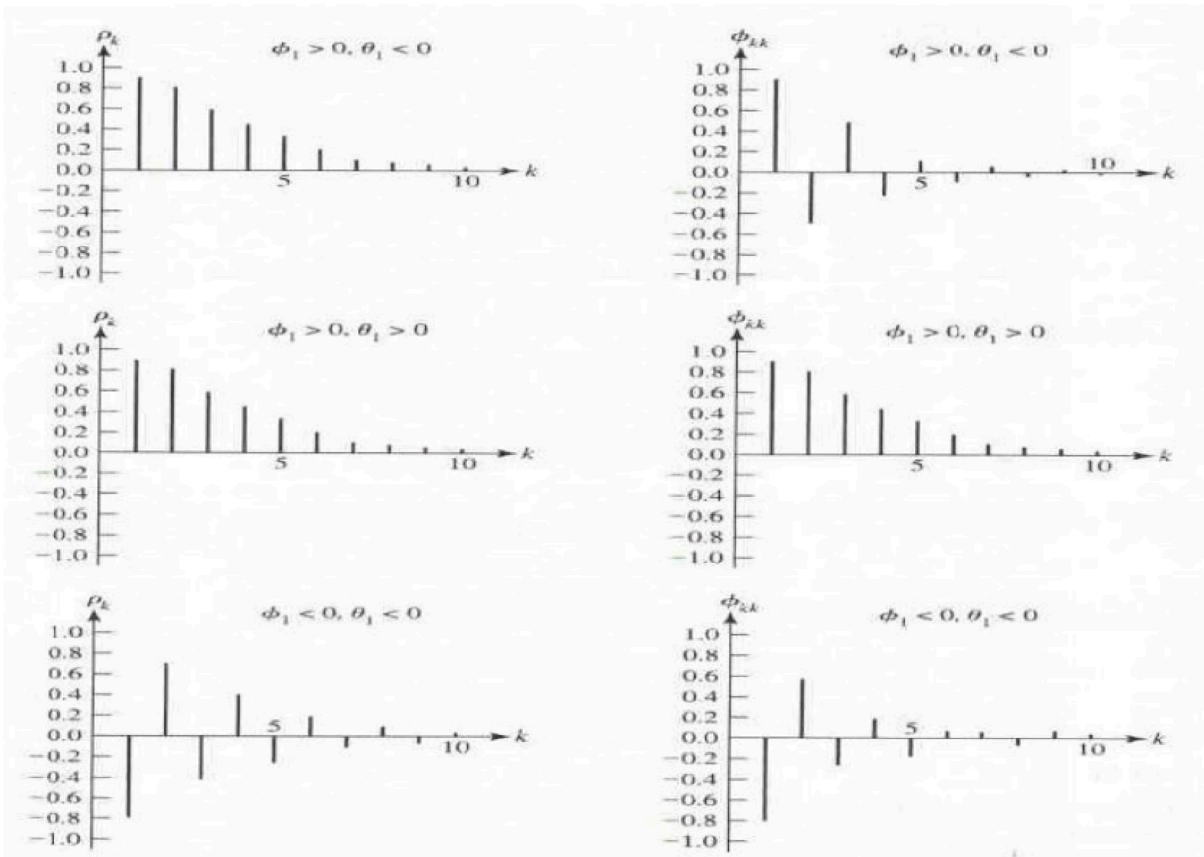


ρ_k
$\theta_1 > 0$
ϕ_{kk}
$\theta_1 > 0$
ρ_k
$\theta_1 < 0$
ϕ_{kk}
$\theta_1 < 0$
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
5
10
k

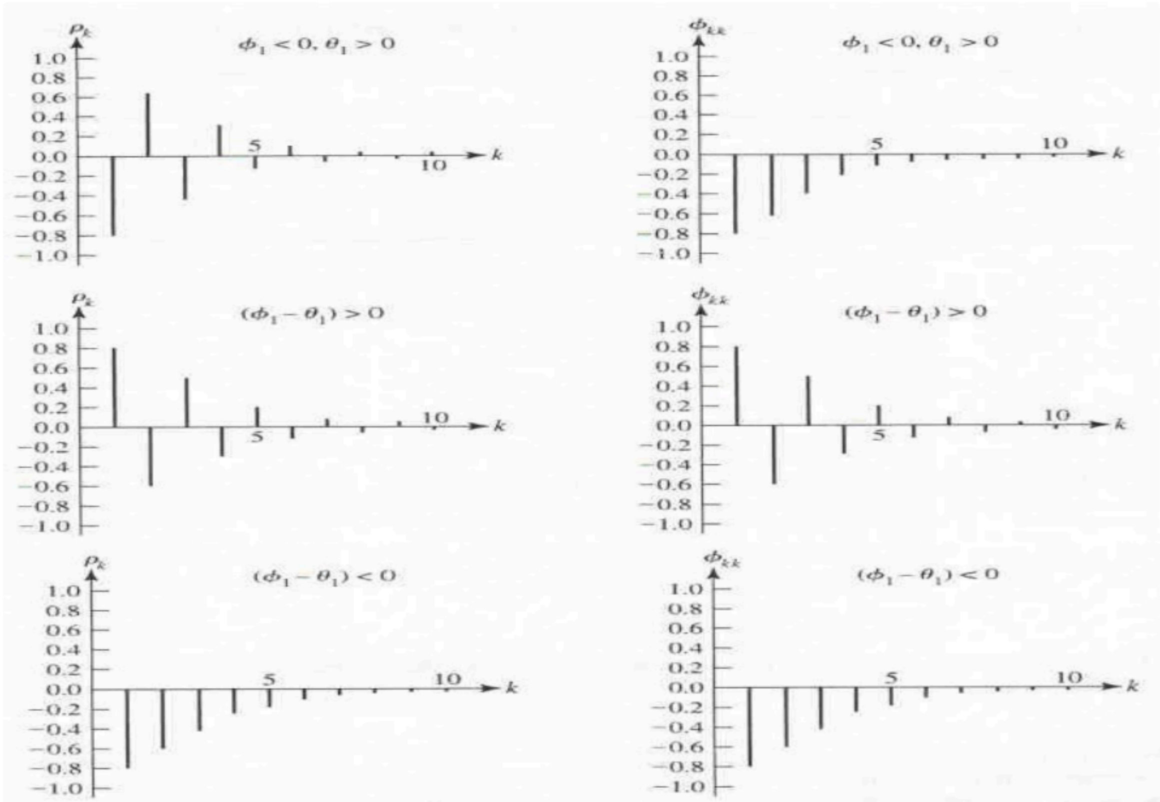
MA(2) PROCESS

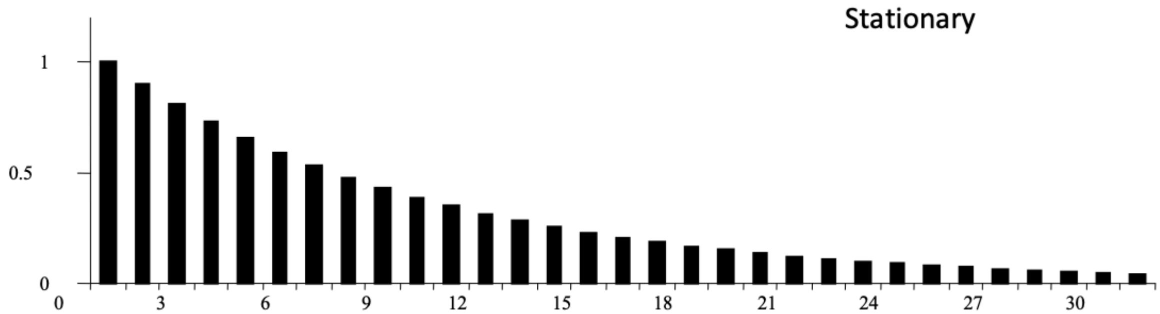

ρ_k
$\theta_1 < 0$
$\theta_2 < 0$
ϕ_{kk}
$\theta_1 > 0$
$\theta_2 > 0$
$\theta_1 < 0$
$\theta_2 > 0$
$\theta_1 > 0$
$\theta_2 < 0$
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
5
10
k

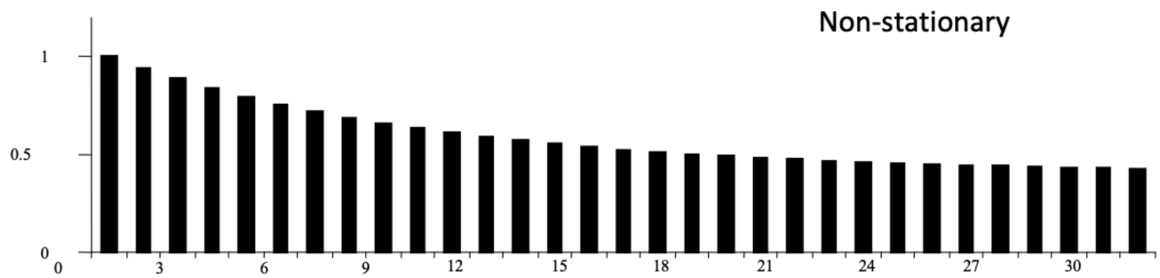
ARMA(1,1) PROCESS



ARMA(1,1) PROCESS (contd.)







- The order of the model is decided when the model errors resemble white noise.

11. Model Evaluation Metrics

11.1 Mean Absolute Error (MAE): Measures the average magnitude of errors in a set of predictions, without considering their direction.

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |Y_t - \widehat{Y}_t|$$

It treats all errors equally, making it less sensitive to outliers compared to RMSE and MSE.

Forecasting Retail Sales: To assess the average deviation between predicted and actual sales without overemphasizing larger errors.

Example: Evaluating a demand forecasting model where consistent small errors are acceptable.

11.2 Mean Squared Error (MSE): Measures the average of the squares of the errors, emphasizing larger errors more than smaller ones.

MSE = $\frac{1}{n}\sum_{t=1}^{n}(Y_t - \widehat{Y}_t)^2$

Useful for detecting large errors and is sensitive to outliers.

Energy Consumption Forecasting: To minimize the impact of large deviations in predicted vs. actual energy usage, where large errors could indicate significant inefficiencies.

11.3 Root Mean Squared Error (RMSE): The square root of the MSE, bringing it back to the original scale of the data.

RMSE = $\sqrt{MSE}$

Useful when large errors are particularly undesirable.

Weather Prediction: To evaluate the accuracy of temperature forecasts, where larger errors (e.g., predicting 30°C instead of 20°C) are more significant.

Example: Assessing a model predicting patient vital signs, where large deviations could indicate serious health issues.

11.4 Mean Absolute Percentage Error (MAPE): Measures the accuracy as a percentage of the actual values.

MAPE = $\frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \widehat{Y}_t}{Y_t}\right| \times 100$

Example: Comparing the accuracy of sales forecasts across different product categories with varying sales volumes.

11.5 Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC):

Used primarily for model selection in statistical modeling especially time series analysis, balancing goodness of fit with model complexity. It is best for selecting the parameters for the statistical models like AR, ARIMA, SARIMA etc. Other metrics like MAE, MSE, RMSE, and MAPE are used to evaluate model performance.

Decision criteria: The model with the lowest AIC or BIC is preferred.

AIC: Often preferred when the primary concern is finding the best predictive model regardless of the number of parameters.

BIC: Preferred when the goal is to find the true model among a set of candidates, with a stronger preference for simpler models.

Suppose you have a time series and you fit several ARIMA models with different orders (p, d, q):

Model	AIC	BIC
ARIMA(1,1,1)	120.5	125.3
ARIMA(2,1,1)	118.7	126.8
ARIMA(1,1,2)	119.8	125.6

Model Comparison:

AIC: ARIMA(2,1,1) has the lowest AIC (118.7).

BIC: ARIMA(1,1,1) has the lowest BIC (125.3).

Model Selection:

If prioritizing prediction accuracy and willing to handle more complexity, choose ARIMA(2,1,1).

If prioritizing simplicity and avoiding overfitting, choose ARIMA(1,1,1).

12. Data Preprocessing

Data preprocessing is a crucial step in time series analysis as it ensures the data is clean, consistent, and suitable for modeling. This step includes handling missing values, dealing with outliers, and transforming the data to meet the assumptions of various time series models.

12.1 Handling Missing Values

1. **Imputation:** Replace missing values with the mean, median, or mode of the column or forward or backward fill.
2. **Interpolation:** Replace missing values based on the surrounding observations using linear, polynomial and spline interpolation.
3. **Predictive Modeling:** Using statistical or machine learning models to predict missing values based on the values of other variables.

12.2 Making data stationary

Removing trend and seasonality. (Methods are given under the Stationarity Topic)

12.3 Handling Outliers

1. **Transformation:** Log or Box-Cox Transformations
2. **Imputation:** Replace with mean, median or value based on surrounding. Capping (limit values to a certain range) or Winsorization (certain percentile values)
3. **Smoothing or rolling average**

12.4 Resampling

1. **Downsampling**: Aggregating data to a lower frequency (e.g., from daily to monthly).
2. **Upsampling**: Increasing the frequency of data (e.g., from monthly to daily), often requiring interpolation to fill in gaps.

Resampling Frequencies: D (daily), W (weekly), M (monthly), Q (quarterly), Y (yearly)