Basics Concepts in Time Series

Data Science Course

Dr. Ariel Mantzura

2024-05-16

Topics to be Covered in this Lecture

- Introduction
- General definition of time series
- Components of a time series
- Decomposition of time series to components.
- Simple methods for forecasting

Introduction

- Time series analysis is a statistical technique used to analyze data points collected at regular intervals over time.
- It enables us to understand the patterns, trends, and relationships within the data, and to make predictions about future values based on historical observations.
- Time series analysis encompasses a range of statistical techniques and models designed to analyze and interpret time series data, including autoregressive models, moving average models, exponential smoothing methods and more.

Introduction - Characteristics of time series

- **Time Dimension**: Time series data is characterized by the sequential ordering of observations, with each observation associated with a specific time point or interval.
- **Temporal Dependence**: Time series data often exhibits temporal dependence, where the value of a data point depends on its previous values.
- This autocorrelation makes time series analysis distinct from other types of data analysis.

Introduction - Examples of Time Series Analysis

- **Financial Markets**: Time series analysis is widely used in finance to predict stock prices, currency exchange rates, and commodity prices. By analyzing historical price movements and trading volumes, investors can make informed decisions about buying, selling, or holding assets.
- **Economic Forecasting**: Economists use time series analysis to forecast key economic indicators such as GDP growth, inflation rates, and unemployment rates. Understanding the cyclical patterns and long-term trends in economic data helps policymakers formulate effective monetary and fiscal policies.

Introduction - examples

- **Demand Forecasting**: Businesses use time series analysis to forecast demand for products and services. By analyzing historical sales data, seasonality, and external factors such as advertising campaigns and economic conditions, companies can optimize inventory management and production planning.
- **Meteorology and Climate Science**: Time series analysis is essential in meteorology and climate science for forecasting weather patterns, temperature trends, and natural disasters such as hurricanes and droughts. By analyzing historical weather data and atmospheric conditions, meteorologists can provide accurate weather forecasts and early warnings.

Introduction - examples

- **Healthcare and Epidemiology**: Time series analysis is used in healthcare to monitor patient vital signs, disease outbreaks, and the effectiveness of medical treatments. By analyzing patient data over time, doctors can detect abnormalities, track disease progression, and make personalized treatment recommendations.
- **Engineering and Manufacturing**: Time series analysis is applied in engineering and manufacturing to monitor equipment performance, detect anomalies, and optimize production processes. By analyzing sensor data and machine metrics, engineers can identify potential failures, minimize downtime, and improve product quality.

Introduction

- Time series analysis is a powerful tool for understanding and predicting patterns in sequential data.
- By leveraging historical observations and statistical models, analysts can gain valuable insights into various phenomena across different domains and make data-driven decisions to drive innovation and efficiency.

General Definition of Time Series

- A time series is a sequence of data points, typically measured at successive time intervals.
- Mathematically, a time series can be represented as a collection of observations indexed by time, denoted as $\{y_t\}$ where t represents the time index.
- Each observation y_t corresponds to a specific time point or interval t
- In mathematical notation:

$$\{y_t\} = \{y_0, y_1, y_2, \ldots, y_t, \ldots, \ldots\}$$

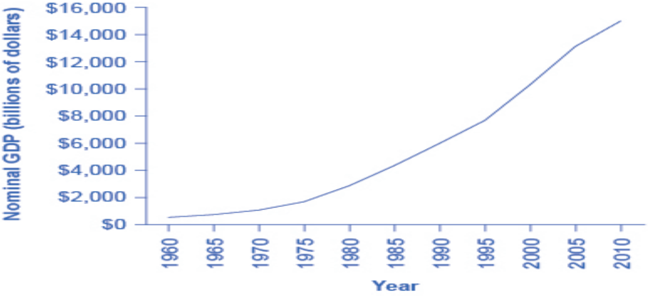
  represents the value of the time series at time t.
- Time series data can be collected at regular intervals (e.g., daily, monthly, quarterly).

Components of a time series

- The 4 main components of a time series are:
  - Trend
  - Cycle
  - Seasonality
  - Irregularity

Components of a time series - Trend

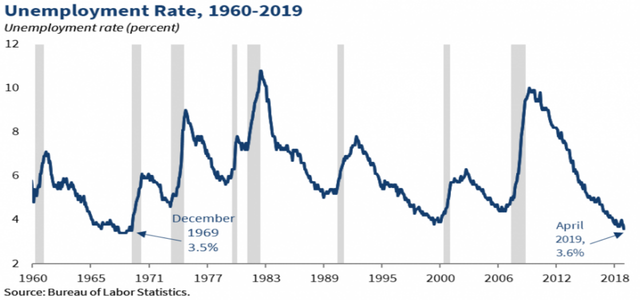
- The trend is defined as the 'long term' movement in a time series, and is a reflection of the underlying level.
- It is the result of influences such as population growth, price inflation and general economic changes. The following graph presents a series in which there is an obvious upward trend over time:

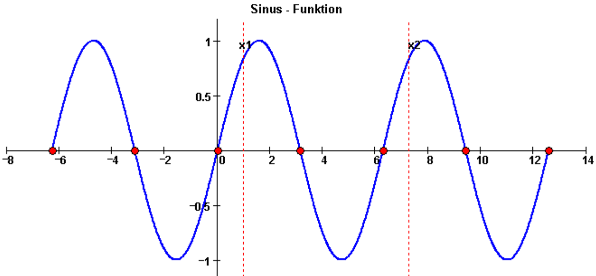


Components of a time series - Cycle

- Many time series oscillate around a trend or the mean if there is no trend.
- In most cases, it takes multiple periods for a series to complete a cycle.
- The cycles in most time series are aperiodic - i.e. the timing and and duration of a series being below or above the trend are irregular.
- Finally, in practice, trend is indistinguishable from a cycle.
- Thus, usually we talk about a recent (local) trend or a trend-cycle.

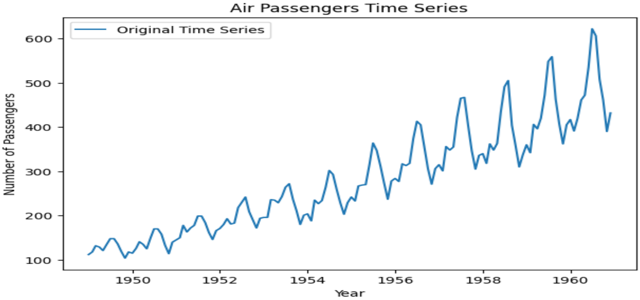
Components of a time series - Cycle





Components of a time series - Seasonality

- A seasonal effect is a systematic and calendar related effect.
- Some examples include the higher electricity consumption in summer due to hot weather or increase in retail series before Pesach or Rosh Hashanah.
- Other seasonal effects include trading day effects and moving holiday.



Components of a time series - Seasonality

- The seasonal component consists of effects that are reasonably stable with respect to timing, direction and magnitude.
- It arises from systematic, calendar related influences such as:
  - Natural Conditions - For instance, weather fluctuations that are representative of the season
  - Business and Administrative procedures - For instance, students looking for a job in summer.
  - Social and Cultural behaviour - Pesach or Rosh Hashanah.

Components of a time series - Irregularity

- The irregular component, also known as the residual or noise, represents the random fluctuations or unpredictable variations in a time series that cannot be explained by the trend, seasonality, or cycle components.
- It reflects the residual variation in the data after accounting for the systematic patterns captured by the other components.
- In essence, the irregular component represents the random "noise" or unmodeled factors in the time series that are not part of any discernible pattern or trend.

Classical Decompostions of time series

- **Additive Decomposition**: In an additive decomposition, the time series is expressed as the sum of its trend, cycle, seasonal, and irregular components.

$$y_t = T_t + C_t + S_t + I_t$$

- T_t is the trend component at time t.
- C_t is the cycle component at time t.
- S_t is the seasonal component at time t.
- I_t is the Irregular component at time t.

Classical Decompostions of time series

- **Multiplicative Decomposition**: In a multiplicative decomposition, the time series is expressed as the product of its trend, cycle, seasonal, and irregular components.

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

- T_t is the trend component at time t.
- C_t is the cycle component at time t.
- S_t is the seasonal component at time t.
- I_t is the Irregular component at time t.

Moving Average - Centered

- A centered moving average is symmetrical around the observation of interest.
- For example: a moving average of length 13 around observation t will average upon the following observations:

$$t-6, t-5, t-4, ..., t, t+1, t+2, ...t+6$$

- The moving average at point t is:

$$\frac{1}{13}\sum_{i=t-6}^{t+6} x_i$$

- In general:

$$\frac{1}{l}\sum_{i=t-\frac{l-1}{2}}^{t+\frac{l-1}{2}} x_i$$

- Where l is odd and the length of the average.

Moving Average - One Sided

- A One sided average is asymmetrical around the observation of interest.
- For example: a moving average of length 13 for observation t will average upon the following observations:

$$t-13, t-5, t-4, ..., t$$

- The moving average at point t is:

$$\frac{1}{13}\sum_{i=t-13}^{t} x_i$$

- In general:

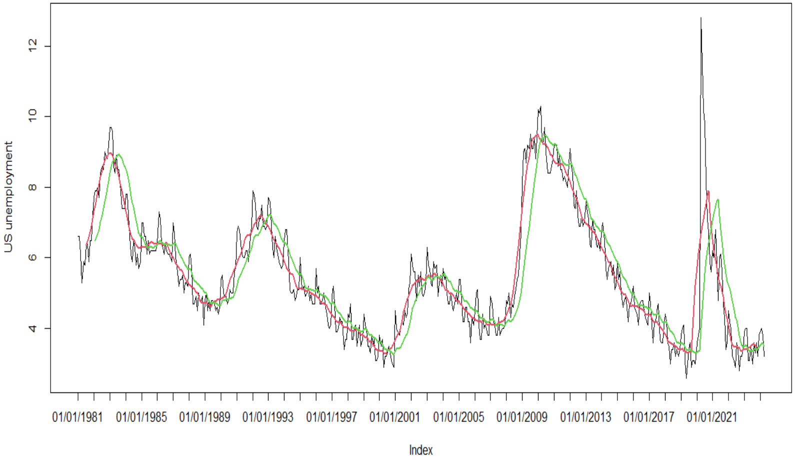
$$\frac{1}{l}\sum_{i=t-l}^{t} x_i$$

- Where l is odd and the length of the average.

Exercise

- Create a series (noise) with 1000 observation from a normal distribution with mean=0 and sd=0.1.
- Create another series names sig with 1000 observations 0.5*sin((1:1000)/50)
- Write a function called mav wich calculates a centered and one sided moving average.
- The arguments of the functions are x(data),l=length of moving average and logical center={0,1}
- Run the function on the data and plot results of both methods.

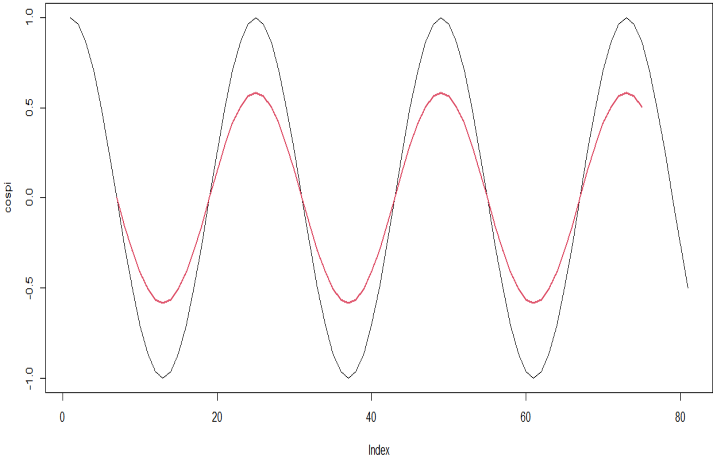
Moving Average - phase shift



Moving Average - The endpoint problem

- Notice that there are no values for either the first 6 months or the last 6 months.
- This is because there is insufficient data at the ends of the series to apply a symmetric filter.
- For this reason, symmetric filters cannot be used at either end of a series.
- This is known as the endpoint problem.

Moving Average - Amplitude Reduction



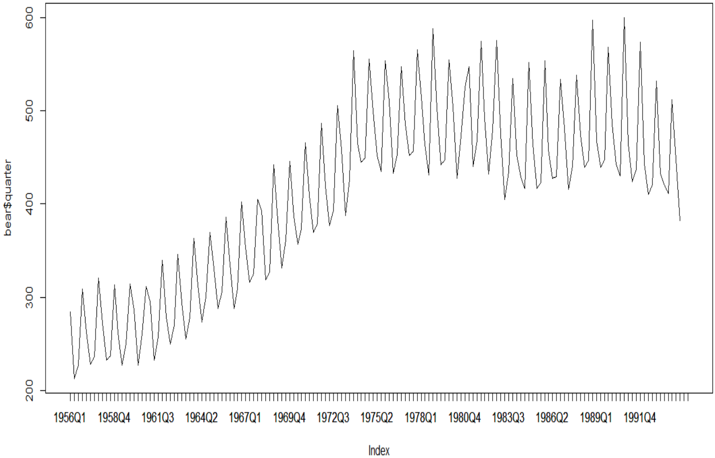
Additive Decomposition

- In some time series, the amplitude of both the seasonal and irregular variations do not change as the level of the trend rises or falls.
- In such cases, an additive model is appropriate.
- In the additive model, the observed time series y_t is considered to be the sum of three independent components: the seasonal s_t , the trend T_t and the irregular I_t .
- That is

$$y_t = Trend_t + Seasonal_t + Irregular_t = T_t + S_t + I_t$$

.

Additive Decomposition



Multiplicative Decomposition

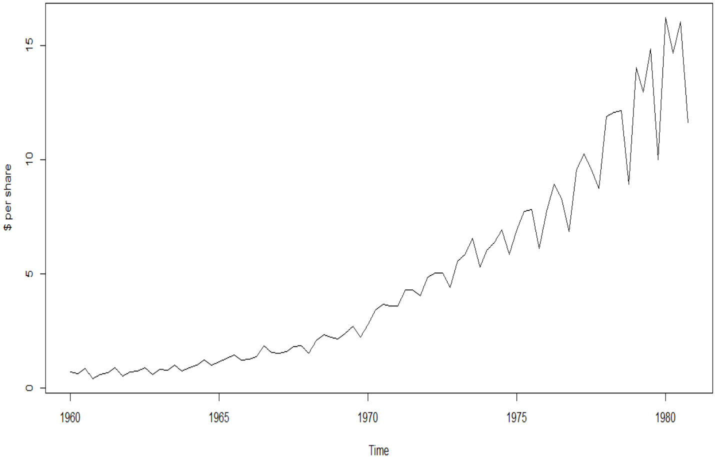
- In many time series, the amplitude of both the seasonal and irregular variations increase as the level of the trend rises.
- In this situation, a multiplicative model is usually appropriate.
- In the multiplicative model, the original time series is expressed as the product of trend, seasonal and irregular components.

$$Observed_t = Trend_t \cdot Seasonal_t \cdot Irregular_t = T_t \cdot S_t \cdot I_t$$

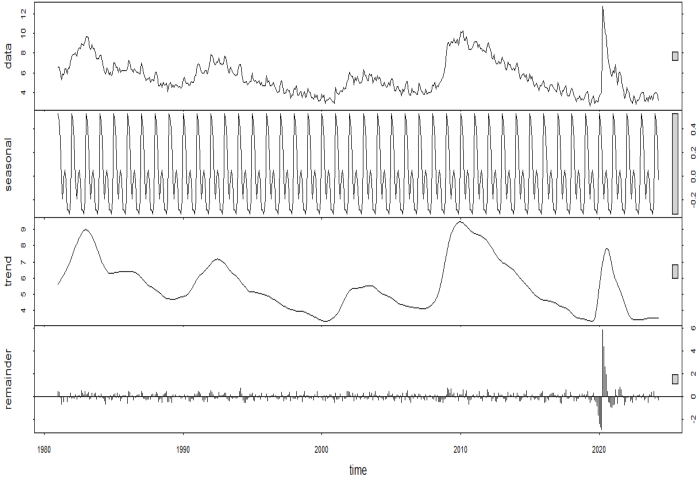
.

- Under this model, the trend has the same units as the original series, but the seasonal and irregular components are unitless factors, distributed around 1.

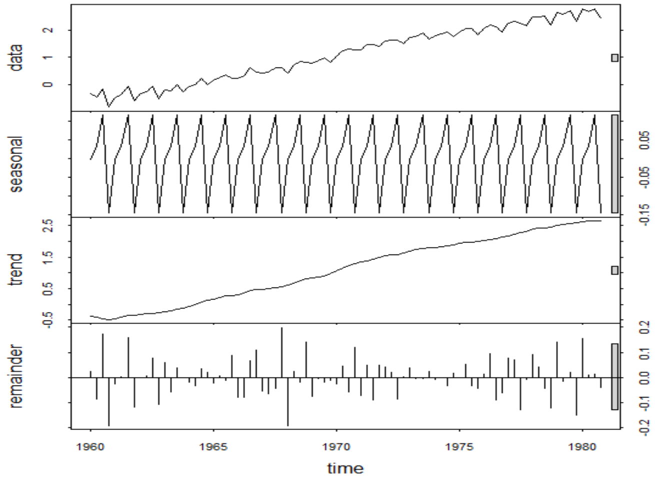
Multiplicative Decomposition



Additive decomposition



Multiplicative decomposition



Simple Autoregressive models - white noise

- The building block for AR models is the white noise process, which will be denoted as ϵ_t.
- In the least general case

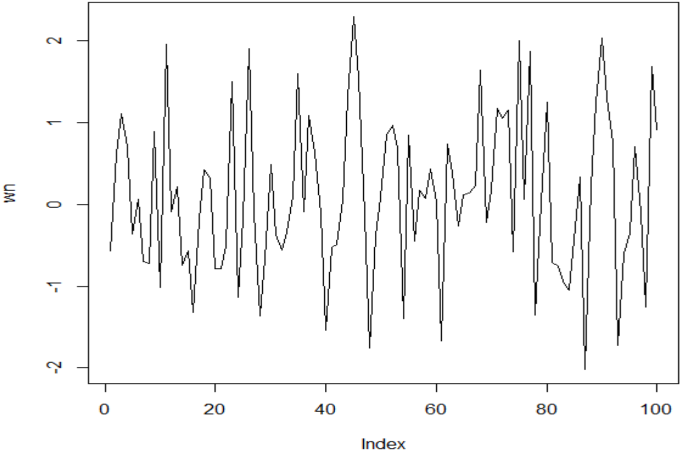
$$\epsilon_t \sim i.i.d \quad N(0, \sigma_\epsilon^2)$$

,

- Notice three implications of this assumption:

1. $E(\epsilon_t) = E(\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, ...) = 0$
2. $E(\epsilon_t \epsilon_{t-j}) = Cov(\epsilon_t \epsilon_{t-j}) = 0$
3. $Var(\epsilon_t) = \sigma_\epsilon^2$

Simple Autoregressive models - white noise



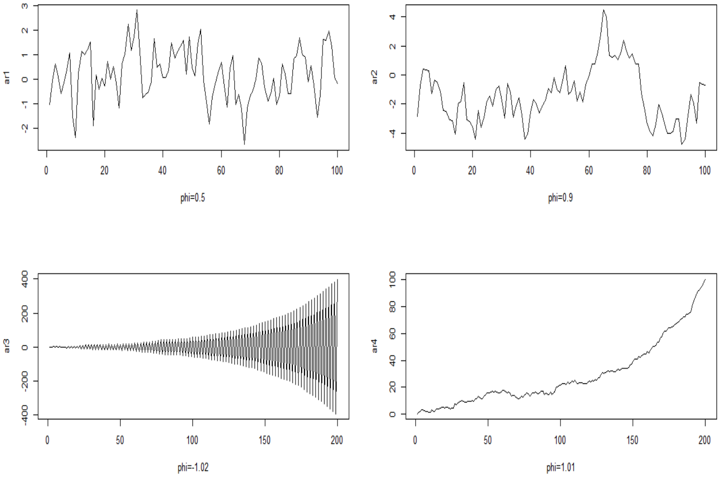
Simple Auto regressive models - AR(1) Model

- The most basic Auto regressive model is the AR(1) model which has the following form:

$$Y_t = c + \phi Y_{t-1} + \epsilon_t$$

- As you can see, you can simulate a sequence $\{y_t\}$ given a sequence of realizations of the white noise process $\{\epsilon_t\}$, and a starting value for y_0.
- The Dynamics of this model highly depends on the parameter ϕ.

Simple Auto regressive models - AR(1) Model



mean variance and Autocovariance.

- The Expectation of the t_{th} observation of a time series is:

$$\mu_t = E(Y_t) = \int_{-\infty}^{\infty} y f_{Y_t}(y) dy$$

- The Variance of the t_{th} observation of a time series is

$$\gamma_{0t} = Var(Y_t) = E(Y_t^2) - E(Y_t)^2.$$

- The Covariance of Y_t and Y_{t-j} is:

$$\gamma_{jt} = Cov(Y_t, Y_{t-j}) = E(Y_t - \mu_t)(Y_{t-j} - \mu_{t-j})$$

- This term is also called Auto Covariance.

Auto Correlation

- The AutoCorrelation is defined as:

$$\rho_{jt} = \frac{Cov(Y_t Y_{t-j})}{\sqrt{Var(Y_t)}\sqrt{Var(Y_{t-j})}}$$

Stationarity

- We say that the model (or process) describing Y_t is **weakly stationary** or **covariance stationary** if:
  - $\mu = E(Y_t)$ does not depend on t.
  - $\gamma_0 = Var(Y_T)$ does not depend on t.
  - $\gamma_j = Cov(Y_t, Y_{t-j})$ does not depend on t but can depend on j which is the distance between he observations.

Stationarity

- If an $AR(1)$ process i.e.

$$Y_t = c + \phi_t Y_{t-1} + \epsilon_t$$

  is stationary we can calculate its mean variance and covariance as follows:

- $$E(Y_t) = E(\phi_t Y_{t-1} + \epsilon_t)$$

  so that

$$\mu = \phi\mu + c \Rightarrow \mu = \frac{c}{1-\phi}$$

  for $|\phi| < 1$.
- When $|\phi| \geq 1$ than it can be shown that the process is not stationary.

Stationarity

- We get that

$$Y_t = \mu(1-\phi) + \phi_t Y_{t-1} + \epsilon_t$$

- We can write this a bit differently:

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + \epsilon_t$$

- The variance is calculated as follows:

$$E(Y_t - \mu)^2 = \phi^2 E(Y_{t-1} - \mu)^2 + 2\phi E(Y_{t-1} - \mu)\epsilon_t + E(\epsilon_t^2)$$

- Beacause of stationarity this can be written:

$$\gamma_0 = \phi^2 \gamma_0 + \sigma_\epsilon^2$$

- So that

$$\gamma_0 = \frac{\sigma_\epsilon^2}{1-\phi}$$

Stationarity

- We can further multiply

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + \epsilon_t$$

by $Y_{t-j} - \mu$ on both sides, take expectations and get the autocovariance:

$$E(Y_t - \mu)(Y_{t-j}) = \phi E(Y_{t-1} - \mu)(Y_{t-j} - \mu) + E(\epsilon)(Y_{t-j} - \mu)$$

and get

$$\gamma_j = \phi\gamma_{j-1}$$

- If we continue this way we get

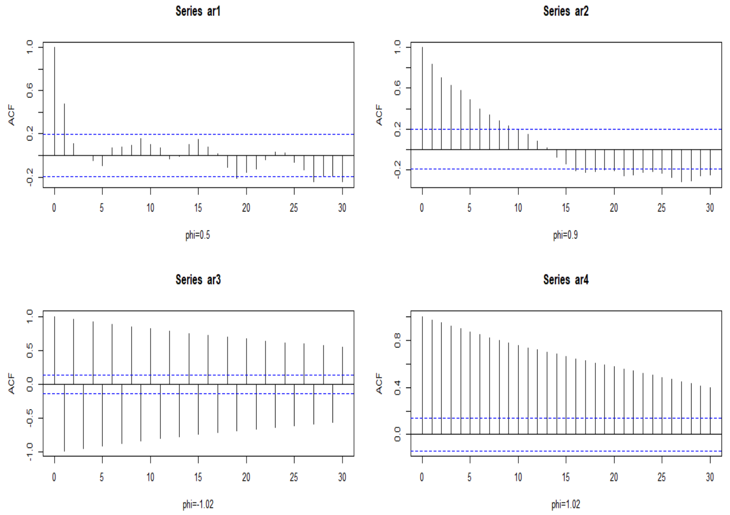
$$\gamma_j = \phi^j \gamma_0$$

- To get the Autocorrelation we simply divide by γ_0 and get:

$$\rho_j = \phi^j$$

.

Autocorrelation function and plot



Predicting a time series

- Once we are convinced that some data follows an AR(1) process we can estimate the ϕ paramater using a simple OLS to get $\hat{\phi}$
- We then can predict the next observation as follows:

$$\hat{y}_{t+1|t} = \hat{\phi} y_t$$

- We can continue this process and predict h steps forward:

$$\hat{y}_{t+h|t} = \hat{\phi}^h y_t$$

Special case - The Random Walk

- The Random walk is a special case of AR(1)
- Is this process stationary
- Let us examine this according to the definition.
- $$E(Y_t) = E(Y_{t-1} + \epsilon) \Rightarrow E(Y_t) = E(Y_{t-1}) ... = E(Y_0)$$
- $$Var(Y_t) = Var(Y_{t-1} + \epsilon_t) \Rightarrow Var(Y_{t-1}) + \sigma^2 ... = Var(Y_0) + t\sigma^2$$
- Therefor the process is not covariance stationary.