Python\AWS \ML Training:

Machine Learning
Day 9

SummitWorks™
GLOBAL SOLUTION ARCHITECTS

Authorized & published by Summitworks Technologies Inc

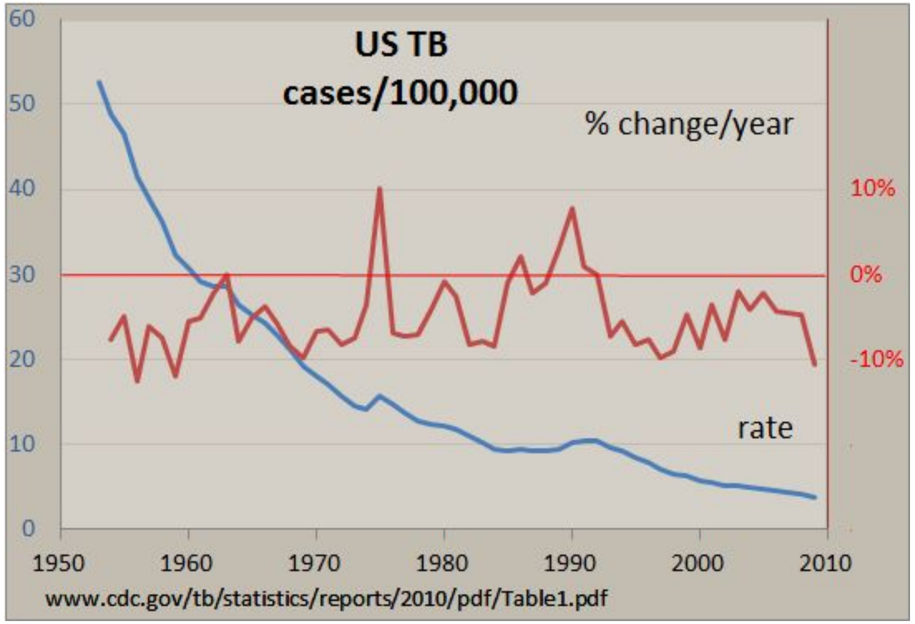
Time Series Analysis

Time Series Analysis

- Time series analysis
- Importance of time series analysis
- Component of time series analysis
- Trend
- Seasonality
- Cyclical Patterns
- Irregular
- White noise
- AR model
- MA model
- ARMA model
- Stationarity
- Approach to remove non-stationarity
- Detrending
- Differencing
- Auto correlation function ACF
- Partial Auto Correlation function PACF

Time Series Analysis

- A time series can be defined as a set of data dependent on time
- Time acts as an independent variable to estimate dependent variables
- A time series is a set of observation taken at specified times (usually at equal intervals)
- A time series defined by the values Y1,Y2.. of a variable Y at time t1,t2.. is given by formula: Y - F(t)



Importance of Time Series Analysis

- Using in forecasting problem such as business forecasting
- It studies the past behavior of the phenomenon.
- It is used in planning for the future operations
- It is used in evaluating the current accomplishments

Time Series Terminology

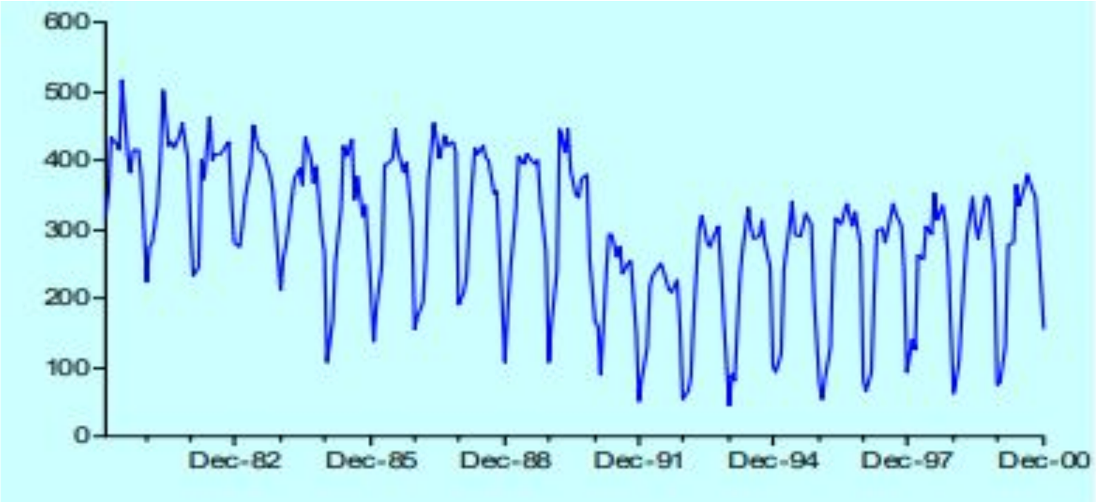
- **A time series** is a sequence of numerical data points in successive order
- **Lag**: How far a part two value of series are
- **Differencing**: Taking different between two values of series
- **Prediction symbol ^** in time series. For example, **ŷ** represents predicted value y

Components of Time Series Analysis

- Components of time series analysis
  - **Trend**
    - A gradual shift or movement to relatively higher or lower values over a long period of time
    - If a general pattern is upward, we call it uptrend
    - If a general pattern is down, we call it a downtrend
    - If there were no trend, we call it horizontal trend or stationary trend

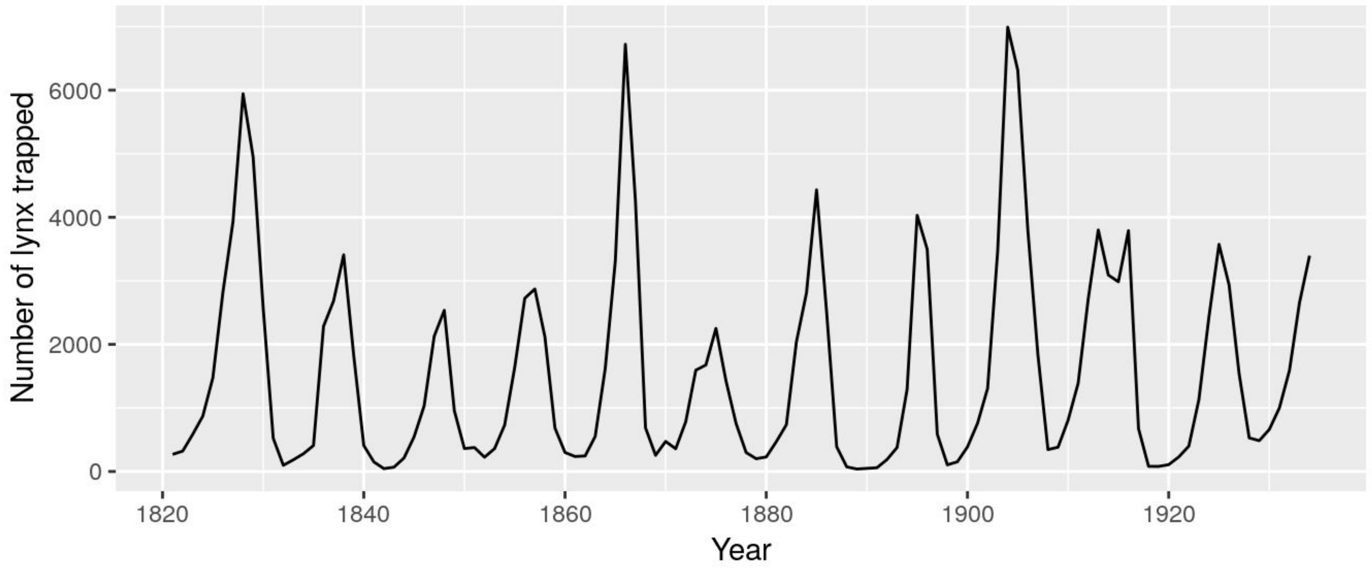
Components of Time Series Analysis

- Components of time series analysis
  - **Seasonality** (regular of peaks and drops - can predict using time)
    - It is shown as upward of downward swings
    - It shows repeating pattern within fixed time period
    - It is usually observed within one year
    - In other words, Regular wavelike fluctuations of constant length, repeating themselves within a period of no longer than a year. Random noise or error in a time series.



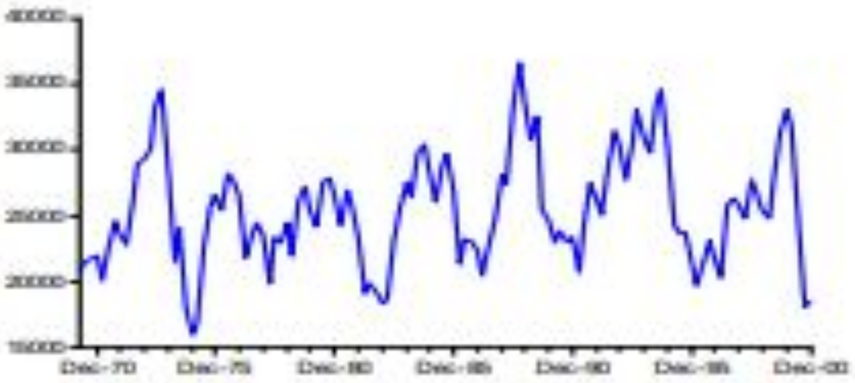
Components of Time Series Analysis

- Components of time series analysis (cont)
    - **Cyclical** pattern (has peak and dop at unpredictable time - happened for other reasons rather than time)
        - The data don't have fixed point
        - it is much harder to predict
        - Wavelike movements, quasi-regular fluctuations around the long-term trend, lasting longer than a year. Example: Business cycles
        - In this example, These show clear aperiodic population cycles of approximately 10 years. The cycles are not of fixed length – some last 8 or 9 years and others last longer than 10 years.



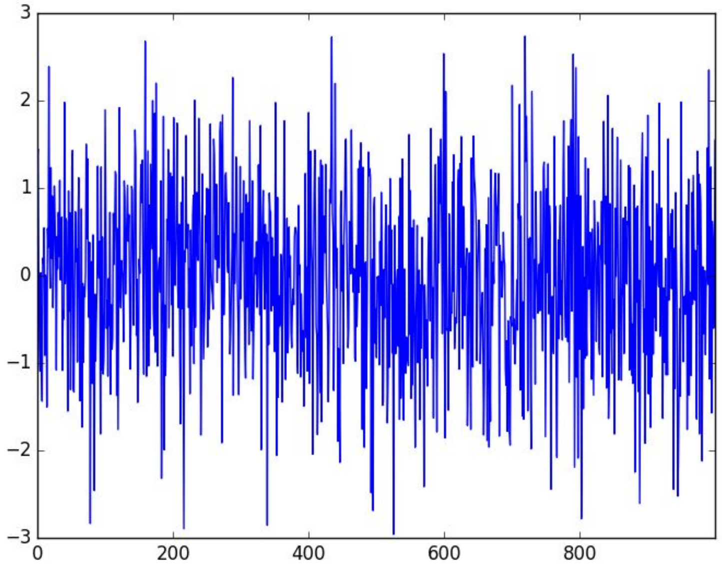
Components of Time Series Analysis

- Components of time series analysis (cont)
  - **Irregular** (peak and drop with no pattern of variation)
    - It is erratic, unsystematic, and residual fluctuations
    - It is observed in short duration and it shows nonrepeating
    - It causes by a random variation or unforseen events
    - Presence of white noise
    - In other world, the movement left after explaining the trend, seasonal and cyclical movements



White Noise

- It describes the assumption that each element in a series is a random draw from our population.
- A white noise time series is unpredictable
- Autoregressive and Moving Average models correct for violations of this white noise assumption
- A time series is white noise if it meets following criteria:
  - Mean = 0
  - Standard deviation is constant with time
  - Correlation between lags is 0
- White noise is important
  - Since a model can be understand
    - Y_t = Signal + noise (error)



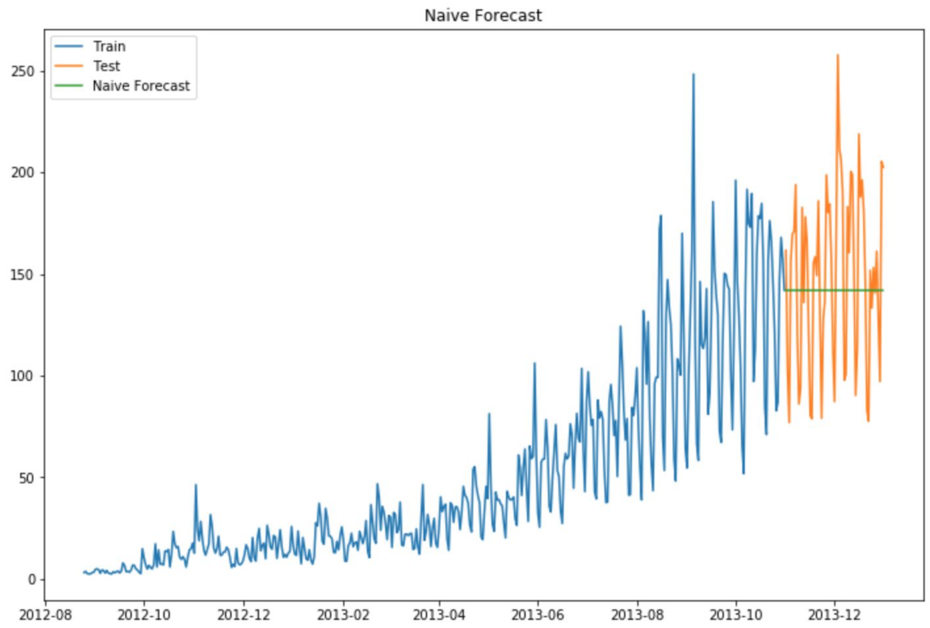
Simple Naive

- We call Y_t is the value of *y* at current time
- We call Y_{t-1} is the value of *y* at last period time
- We call Y_{t+1} is the value of *y* at next period time
- The simple naive model says the what happened last time period (last year, last month, yesterday) will happen this time. Get the value of last day to predict next day

$$\hat{Y}_{t+1} = Y_t$$

Simple Naive

- We can observe that Naive method isn't suited for datasets with high variability. It is best suited for stable datasets.
- If we test the accuracy using Root Mean Square Error, we can see the result is not good

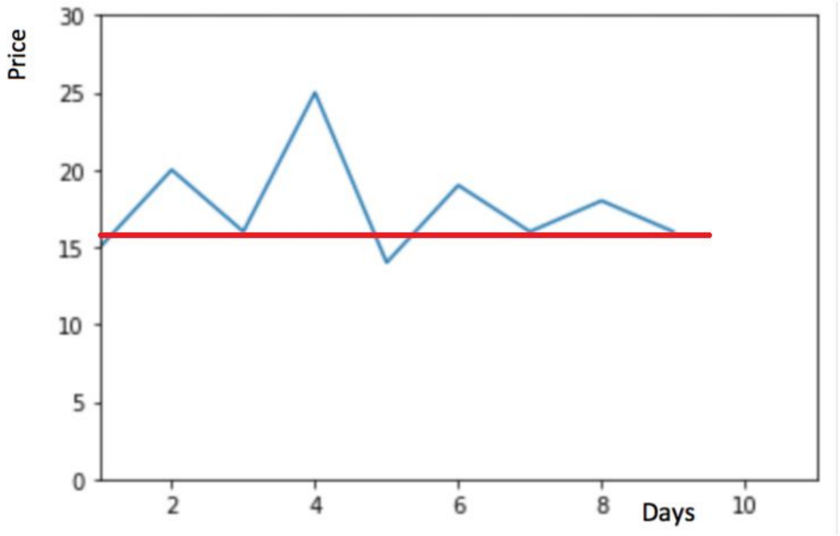


Simple Average

- Simple Average technique is forecasting technique which forecasts the expected value equal to the average of all previously observed points is called

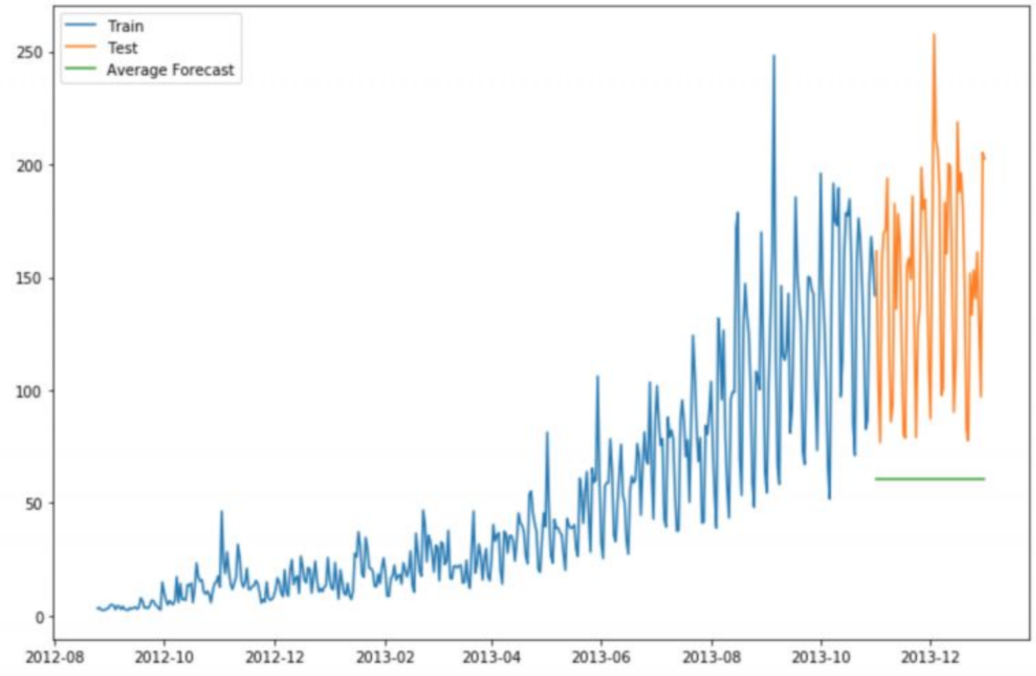
$$\hat{y}_{x+1} = \frac{1}{x} \sum_{i=1}^{x} y_i$$

- It is good in this case:



Simple Average

- However, in this situation, it is not good



MA Model

- The moving-average model specifies that the output variable depends linearly on the current and various past values of a stochastic (imperfectly predictable) term.
- It accounts for the possibility of relationship between a variable and the residuals from previous periods
- MA(q) is a moving average model with q lags:

$$y_t = \mu + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

- Where
  - θq is the coefficient for the lagged error term in time t-q
  - μ is a constant
  - yt is the dependent variable at time t
  - ϵt is the error at time t

MA Model

- An Moving Average (MA) model algorithm example on dataset:
- Step 1: Compute the average (with period or lag of 3)
    - Every prediction is a average of 3 time before it
    - (t-1,t-2 and t-3)

Month	Demand	Forecast
Jan	120	
Feb	103	
Mar	105	
Apr	84	109
May	114	97
Jun	90	101
Jul	100	96
Aug	113	101
Sep	99	101
Oct	108	104
Nov	109	107
Dec	88	105
Jan	91	102
Feb	96	96
Mar	113	92
Apr	84	100
May	98	98
Jun	87	98
Jul	91	90
Aug	119	92
Sep	99	99
Oct	106	103
Nov	89	108
Dec	107	98

Prediciton

AR Model

- An autoregressive (AR) model is a representation of a type of random process;
- The autoregressive model specifies that the output variable depends linearly on its own previous values and on a stochastic term.
- Value of a variable in one period is related to its value in previous periods
- AR(p) is an autoregressive model with p lags:

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

Where:

- μ is a constant
- γp is the coefficient for the lagged variable
- y_t is the dependent variable at time t
- y_{t-1} is the independent variable at previous time period
- ϵt is the error at time t

AR Model

- An autoregressive (AR) model algorithm example on dataset:
- Step 1: Create the lagged time of current time (t-1,t-2,t-3)
- Step 2: Run multiple regression:
  - t is target
  - t-1,t-2 and t-3 is a features

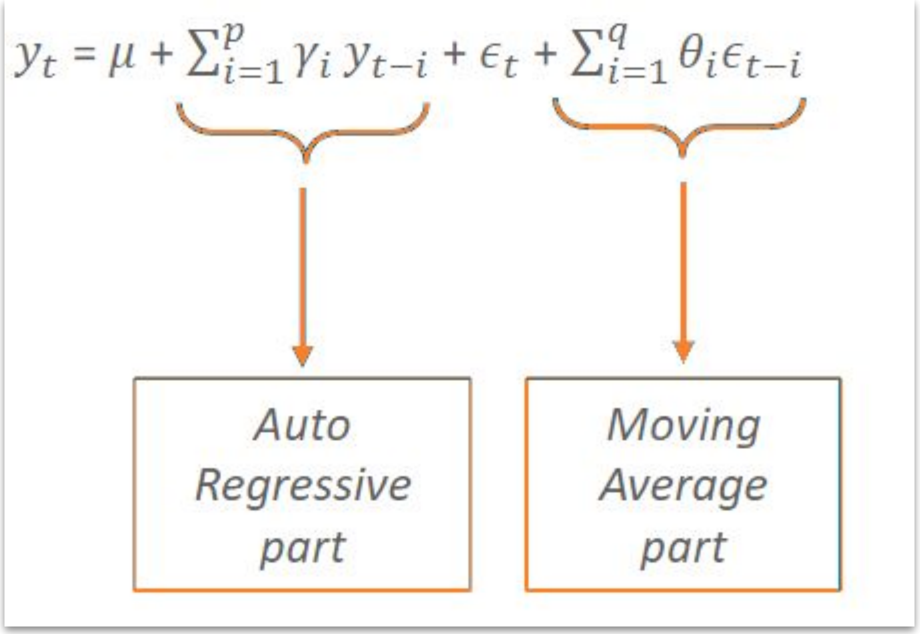
	Coefficient
Intercept	0.777058
t-1	0.959978
t-2	0.008595
t-3	0.02791

- Make prediction for next time stamp based on coefficient of t-1: **y(t) = 0.9599 * y(t-1)**

Date	t	t-1	t-2	t-3
11/25/2016	120.38	120.84	121.47	121.77
11/28/2016	120.41	120.38	120.84	121.47
11/29/2016	120.87	120.41	120.38	120.84
11/30/2016	118.42	120.87	120.41	120.38
12/1/2016	115.1	118.42	120.87	120.41
12/2/2016	115.4	115.1	118.42	120.87
12/5/2016	117.43	115.4	115.1	118.42
12/6/2016	117.31	117.43	115.4	115.1
12/7/2016	117.95	117.31	117.43	115.4
12/8/2016	118.91	117.95	117.31	117.43
12/9/2016	119.68	118.91	117.95	117.31
12/12/2016	117.77	119.68	118.91	117.95
12/13/2016	120.31	117.77	119.68	118.91
12/14/2016	120.21	120.31	117.77	119.68
12/15/2016	120.57	120.21	120.31	117.77
12/16/2016	119.87	120.57	120.21	120.31
12/19/2016	119.24	119.87	120.57	120.21
12/20/2016	119.09	119.24	119.87	120.57

ARMA Model

- Using AR and MA models sometimes is not enough since they are lack of accuracy. As a result, ARMA model was created as a combination of AR and MA models
- It combine both p autoregressive terms and q moving average terms.
  - First part, it is the exactly the same with Auto-Regression model
  - Second part, it is from Moving Average model (How much the predicted is off from actual result)



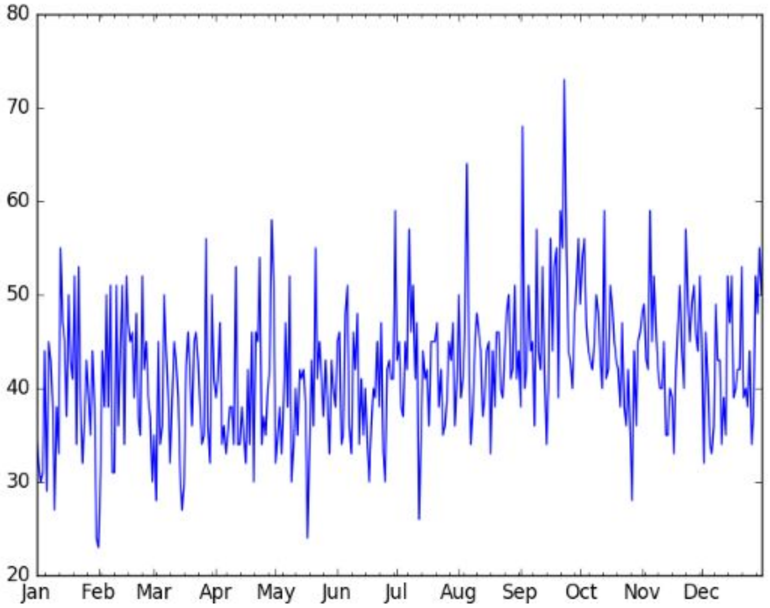
ARMA Model

- Above equation is a perfect formula that represents relationship between y_t and y_{t-1}. There will is no access ϵ_t when we make a real prediction. So, the real prediction will not include this ϵ_t
- Prediction formula is at following:

$$\hat{y}_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

Stationarity

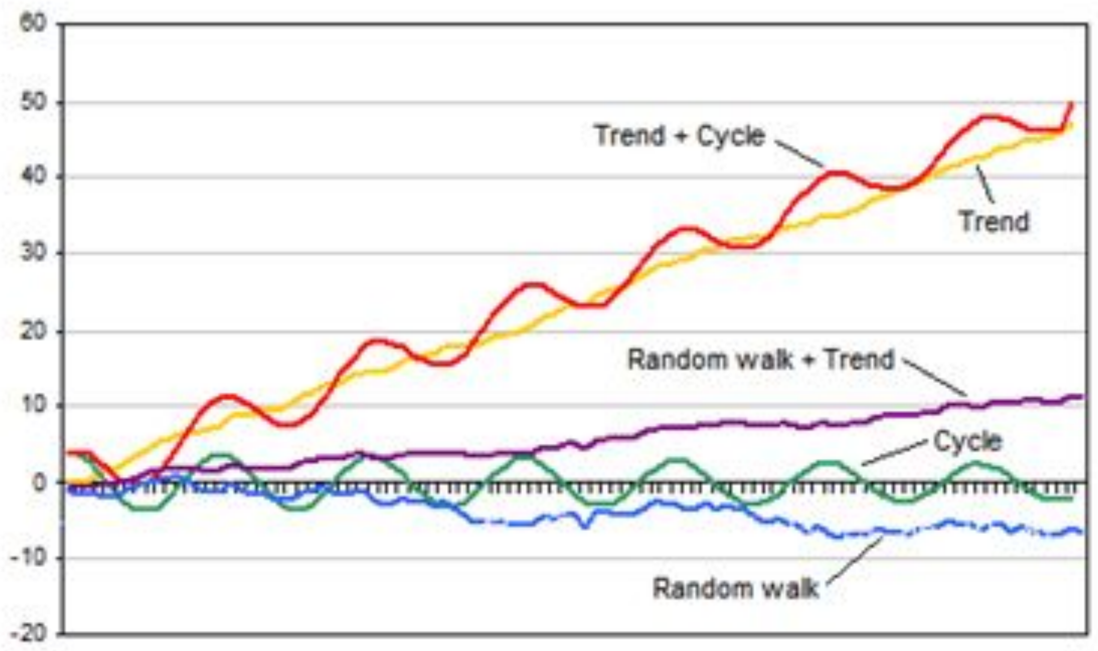
- ARMA model is valid only if the variables are non-stationary
- A stationary time series is one whose statistical properties such as mean, variance, autocorrelation, etc. are all constant over time.
- In other words, the stationarity is a process with mean and variance do not change over time and do not have a trends (there is no seasonality)
- With this definition, It is clear that a white noise process is stationary (but the correlation of white noise is 0, so it is unpredictable).



Approaches to remove Non-Stationarity

- There are two approaches to remove non-stationarity
  - Detrending
  - Differencing

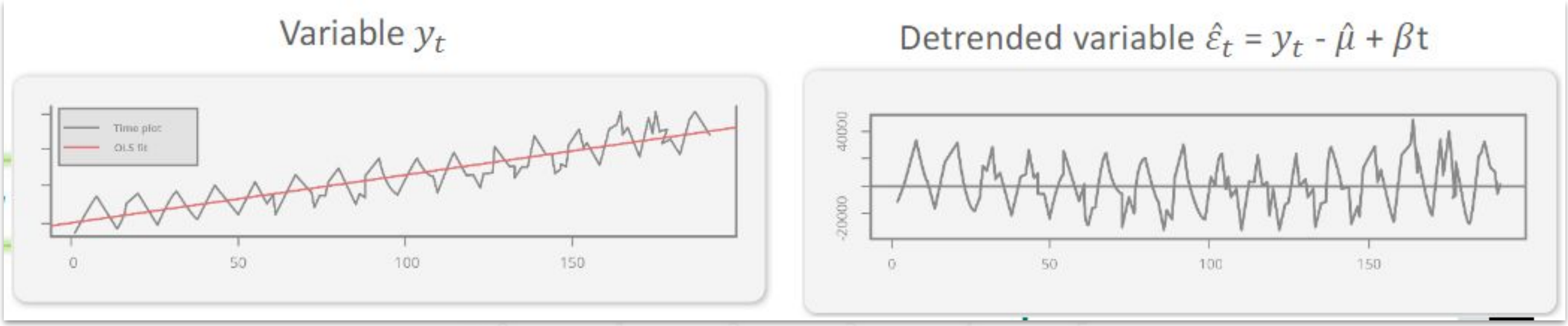
Table 1 Non-stationary behavior



Detrending

- A variable can be detrended by regressing the variable on a time trend and obtaining the residuals (detrended variable is variable y_t subtract the regression)

$$y_t = \mu + \beta t + \epsilon_t$$



Differencing

- **Differencing** is a very standard way to remove trend and/or seasonality of a series.
- Use the concept of differenced variable:

$$\Delta y_t = y_t - y_{t-1}$$

- Lag-1 differencing (first order):
  - Taking the difference between two neighboring values
  - It is useful for removing trend, since the series of difference or jump values will no longer contain a trend
- Lag-M differences:
  - Taking the difference between values in the same season in the previous circle
  - It is useful for removing seasonality, where M is a number of seasons in our data
- Double differences (second order):
  - Perform differencing operation twice (find differences on the differenced series)
  - It is useful for removing both trend and seasonality

ARIMA model

- **ARIMA(p,d,q)** denotes an ARMA model with:
    - **p** autoregressive lags,
    - **q** moving average lags
    - and difference in the order of **d**

Argument Dicky-Fuller Test (ADFT)

- You can estimate the above model for stationarity by testing the significance of the γ coefficient:
  - If the null hypothesis is not rejected, γ*=0, then y_t is not stationary
  - Difference the variable and repeat the test to see if the differenced variable is stationary
  - If the null hypothesis is rejected, γ*>0, then y_t is stationary

Assume an AR(1) model.

The model is non-stationary or a unit root is present if $|\rho| = 1$

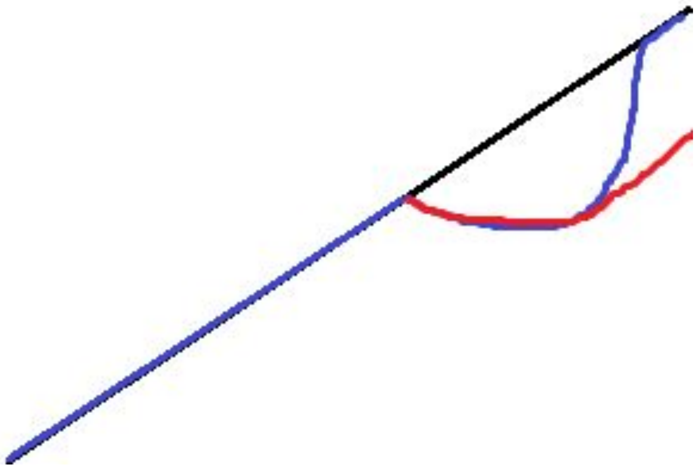
$$y_t = \rho y_{t-1} + \epsilon_t$$

$$y_t - y_{t-1} = \rho y_{t-1} - y_{t-1} + \epsilon_t$$

$$\Delta y_t = (\rho - 1) y_{t-1} + \epsilon_t = \gamma y_{t-1} + \epsilon_t$$

What is Unit root?

- A unit root (also called a unit root process or a difference stationary process) is a stochastic trend in a time series, sometimes called a “random walk with drift”;
    - If a time series has a unit root, it shows a systematic pattern that is unpredictable.
    - A possible unit root in the figure:
        - The red line shows the drop in output and path of recovery if the time series has a unit root. The red line remains permanently below the trend
        - Blue shows the recovery if there is no unit root and the series is trend-stationary. The blue line returns to meet and follow the dashed trend line

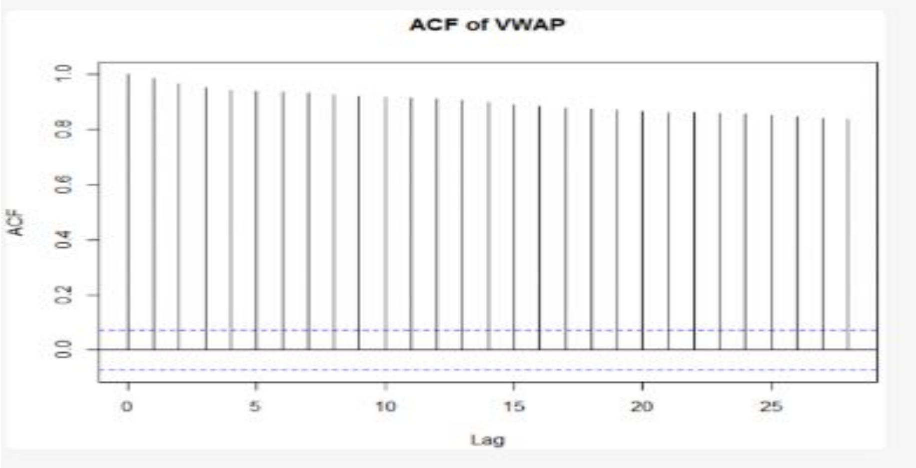


Auto-Correlation Function - ACF

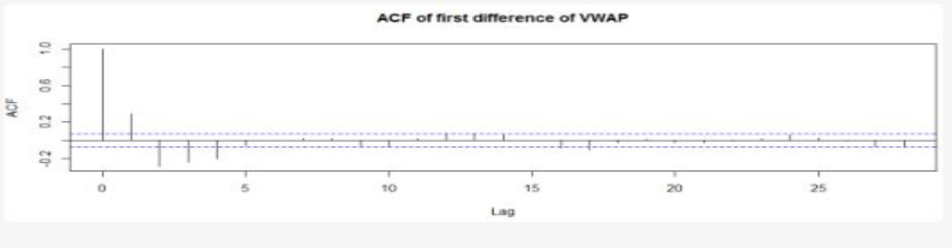
- ACF is the proportion of the covariance of yt and yt-k to the variance of a dependent variable yt

$$\mathrm{ACF(k)} = \rho_k = \frac{\mathrm{Cov}(y_t, y_{t-k})}{\mathrm{Var}(y_t)}$$

- Gives the gross correlation between yt and yt-k



ACF of non-stationary series



ACF of stationary series

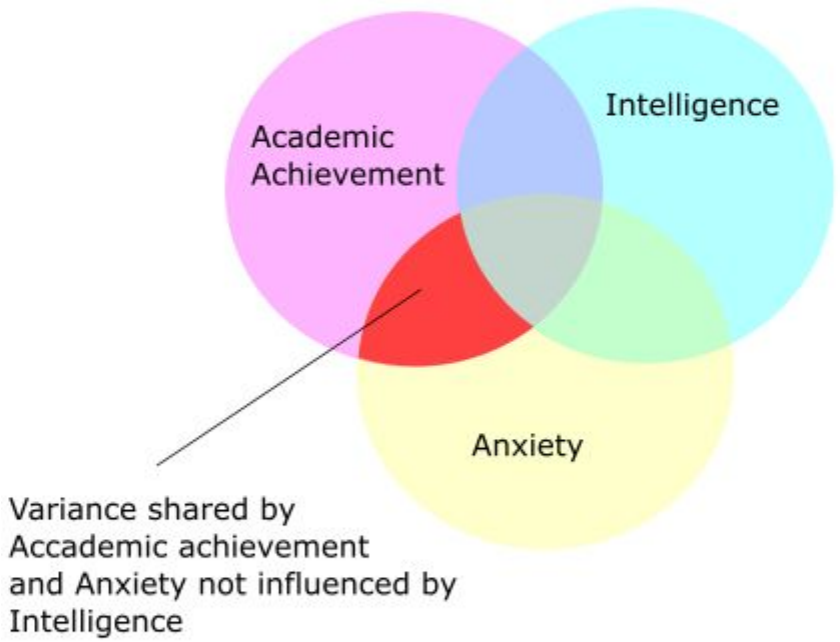
Partial Correlation

- Partial correlation is the relationship between two variables while controlling for a third variable. The purpose is to find the unique variance between two variables while eliminating the variance from a third variables.
- In other words, Partial correlation measures the degree of association between two random variables, with the effect of a set of controlling random variables removed.
- Call A,B and C are variable. The partial correlation between A and B controlled for C

$$r_p = r_{AB.C} = \frac{r_{AB} - r_{AC}r_{BC}}{\sqrt{(1-r_{AC}^2)(1-r_{BC}^2)}}$$

Partial Correlation

- The partial correlation between the two variables, academic achievement and anxiety controlled for intelligence
- While computing this correlation, the effect of intelligence on both the variables, academic achievement and anxiety, was removed



- The correlation between anxiety (B) and academic achievement (A) is – 0.369.
- The correlation between intelligence (C) and academic achievement (A) is 0.918.
- The correlation between anxiety (B) and intelligence (C) is – 0.245

$$r_{AB.C} = \frac{r_{AB} - r_{AC} r_{BC}}{\sqrt{(1 - r_{AC}^2)(1 - r_{BC}^2)}} =$$

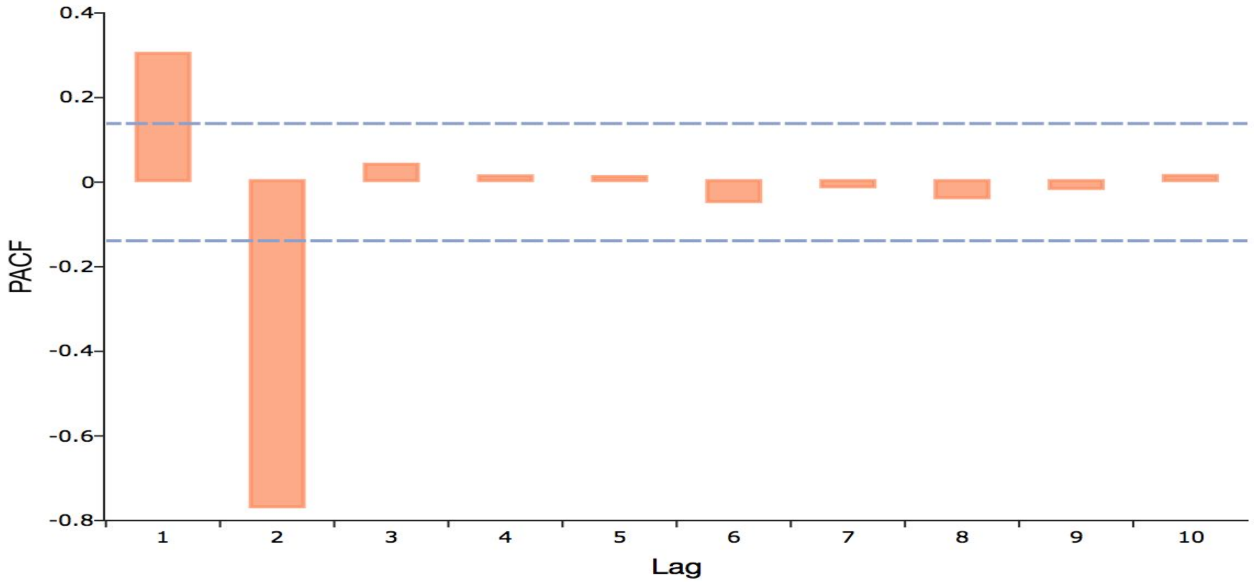
$$= \frac{-0.369 - (0.918 \times -0.245)}{\sqrt{(1 - 0.918^2)(1 - (-0.245^2))}} = \frac{-.1441}{0.499} = -0.375$$

Partial Auto-Correlation Function - PACF

- Simple correlation between y_t and y_{t-k} minus the part explained by the intervening lags

$$\rho_k^* = \mathrm{Corr}|\, y_t - E^*(y_t \,|\, y_{t-1},\ldots,y_{t-k+1}),\, y_{t-k})|$$

- Where $E^*(y_t | y_{t-1},\ldots,y_{t-k+1})$ is the minimum mean squared predictor of y_t by $y_{t-1},\ldots,y_{t-k+1}$
- For an AR(1) model, PACF(k) = γ for first lag



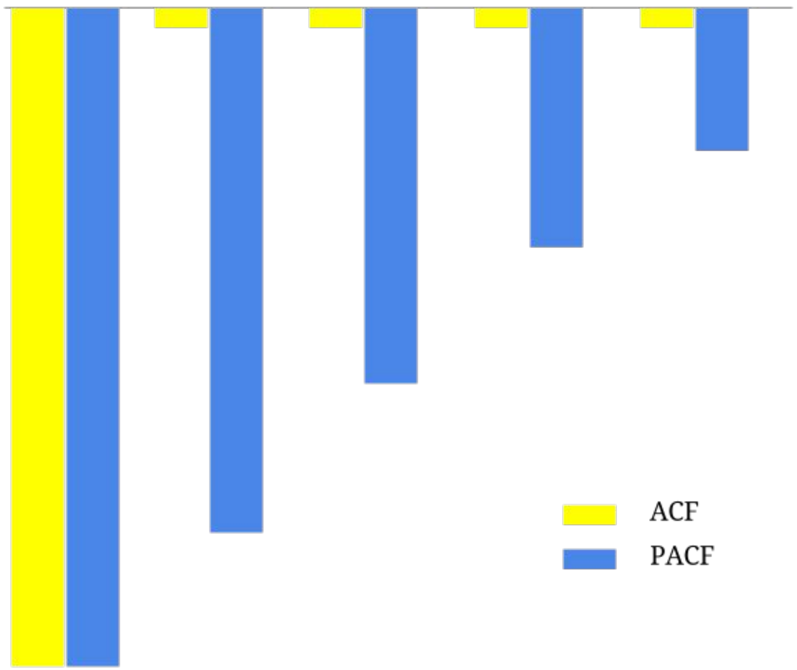
Detecting AR & MA using ACF and PACF plots

- ACF and PACF can be used to detecting which models can be used

	ACF	PACF
AR	Geometric	Significant till p lags
MA	Significant till p lags	Geometric
ARMA	Geometric	Geometric

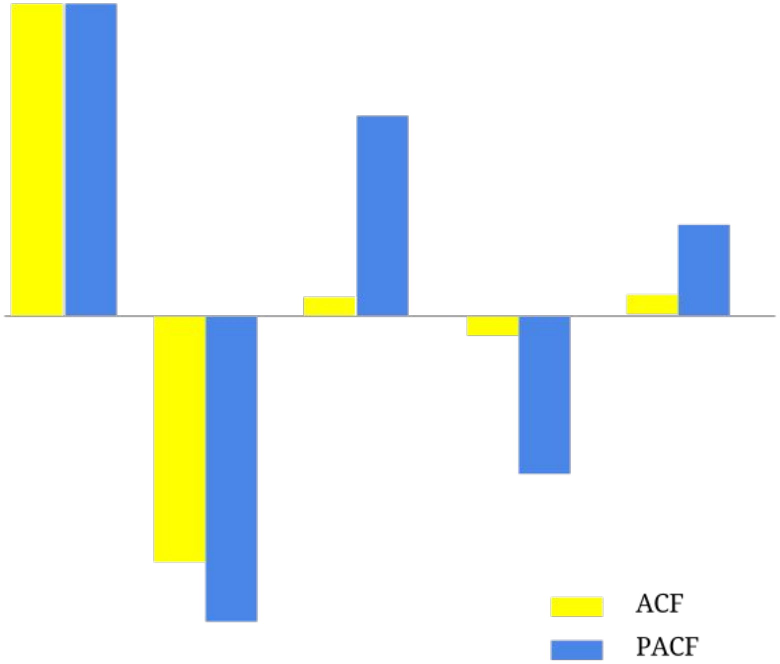
Detecting AR & MA using ACF and PACF plots

- On this plot the ACF is significant only once, while the PACF is geometric.
  - Hence it is an MA(1) process.



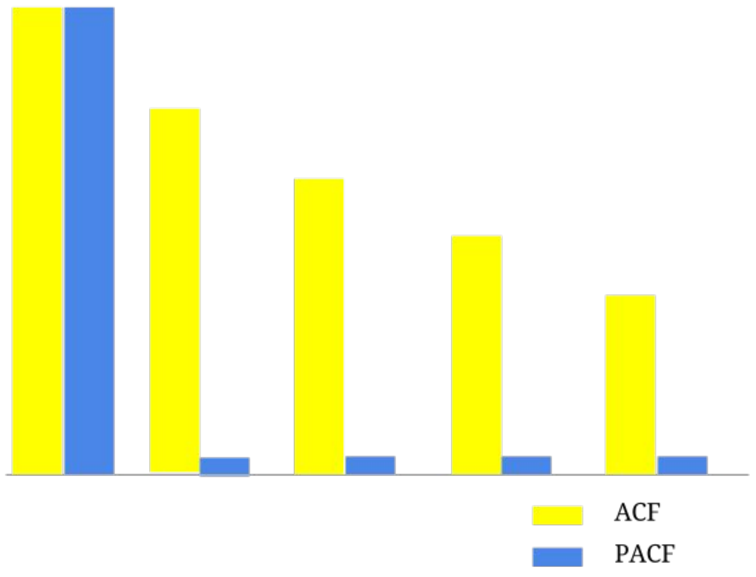
Detecting AR & MA using ACF and PACF plots

- In this example the ACF is significant in the first and second lags, while the PACF follows a geometric decay.
  - Hence it is an MA(2) process.



Detecting AR & MA using ACF and PACF plots

- ACF decays geometrically, and the PACF shows only one significant lag
  - Hence it is an AR(1) process.



Demo

A case study and solution using time series in Python

Q & A