Time Series Analysis

1)Correlation of Two Time Series:

The correlation coefficient is a measure of how muchtwo series vary together.

1. Correlation of one means that the two series have a perfect linear relationship with no deviation.
2. High correlation means that the two series strongly vary together.
3. Low correlation means that the two series vary together, but there is weak association.
4. A negative correlation means they vary in opposite directions, but still with a linear relationship

Two trending series may show a strong correlation even if they are completely unrelated. This is referred to as "spurious correlation". That's why when you look at the correlation of say, two stocks, you should look at the correlation of their returns and not their levels.

Compute percent change using pct_change()

returns = stocks_and_bonds.pct_change()

Compute correlation using corr()

correlation = returns['SP500'].corr(returns['US10Y'])

1.1)Relationship Between R-Squared and Correlation:

- [corr(x, y)] 2 = R 2 (or R-squared)
- sign(corr) = sign(regression slope)
- In last example:
  - R-Squared = 0.753
  - Slope is positive
  - correlation = + √0.753 = 0.868

2)Autocorrelation:

Autocorrelation is the correlation of a single time series with lagged copy of itself. It is also called single correlation

- Correlation of a time series with a lagged copy of itself

Series	Lagged Series
5	
10	5
15	10
20	15
25	20
⋮	⋮

The AC would be the correlation of the series with the same series lagged by the one day.

When return have negative autocorrelation, we say it is ”mean reverting”. If a series has positive correlation, we say its a “trend-following”

2.1)Autocorrelation Function:

ACF shows not only at one lag autocorrelation, but entire autocorrelation function for different lags.

Any significant non-zero autocorrelations implies that the series can be forcast from the past.

The ACF can also be usefull for selecting a parsimonious model for fitting the data.

plot_acf is the statsmodels functions for plotting the autocorrelation function.

- Import module:

```
from statsmodels.graphics.tsaplots import plot_acf
```

- Plot the ACF:

```
plot_acf(x, lags= 20, alpha=0.05)
```

The alpha argument sets the width of the confidence intervel. A Confidence Interval is a range of values we are fairly sure our true value lies in.

2.2)Confidence Interval of ACF

- Argument alpha sets the width of con dence interval
- Example: alpha=0.05
    - 5% chance that if true autocorrelation is zero, it will fall outside blue band
- Confidence bands are wider if:
    - Alpha lower
    - Fewer observations
- Under some simplifying assumptions, 95% con dence bands are $\pm 2/\sqrt{N}$
- If you want no bands on plot, set alpha=1

3)White Noice:

A white noise time series is simply a sequence of uncorrelated random variables that are identically distributed. Stock returns are often modeled as white noise. Unfortunately, for white noise, we cannot forecast future observations based on the past - autocorrelations at all lags are zero.

White Noise is a series with:

- Constant mean
- Constant variance
- Zero autocorrelations at all lags

Special Case: if data has normal distribution, then Gaussian White Noise

4)Random Walk:

In random walk, today's price is equal to yesterday's price plus some noice.

- Today's Price = Yesterday's Price + Noise : $P_t = P_{t-1} + \varepsilon_t$
  - Change in price is white noise : $P_t - P_{t-1} = \varepsilon_t$
- Can't forecast a random walk
- Random walk with drift: $P_t = \mu + P_{t-1} + \varepsilon_t$
  - Change in price is white noise with non-zero mean: $P_t - P_{t-1} = \mu + \varepsilon_t$
- Incidentally, if prices are in logs, the difference in log price is one way to measure return.
- To test whether the time series is random walk, you can regress current values on lagged values. If the slop coefficient (beta) is not significantly different from one then we cannot reject the null hypothesis that the series is random walk. However is slop is less than one we can reject null hypothesis.

4.1)Statistical Test for Random Walk

- Random walk with drift $P_t = \mu + P_{t-1} + \varepsilon_t$
- Regression test for random walk $P_t = \alpha + \beta P_{t-1} + \varepsilon_t$
- Test: $H_0: \beta = 1$ (random walk) $H_1: \beta < 1$ (not random walk)

```
# Run Augmented Dickey-Fuller Test on SPX data
results = adfuller(df['SPX'])
```

```
# Print p-value
print(results[1])
```

If the p-value is less than 5%, we can reject the null hypothesis that the series is a random walk with 95% confidence.

5)Stationarity:

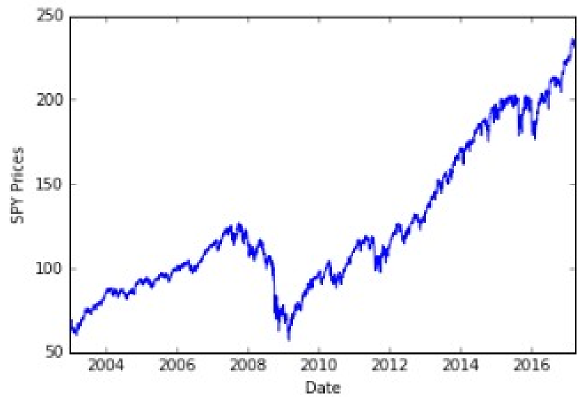
It means join distribution of the observations do not depend on time.

- Strong stationarity: entire distribution of data is time-invariant
- Weak stationarity: mean, variance and autocorrelation are time- invariant (i.e., for autocorrelation, corr (X_t , $X_{t-\tau}$) is only a function of τ)
- It a series is not stationary it is difficult to model.
- Seasonal series are also not stationary.
- Why Do We Care?
  - If parameters vary with time, too many parameters to estimate
  - Can only estimate a parsimonious model with a few parameters

Many TS can be made stationary through simple tranformation

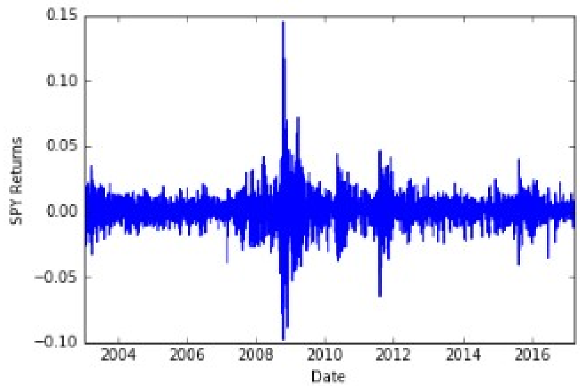
- Random Walk

```
plot.plot(SPY)
```



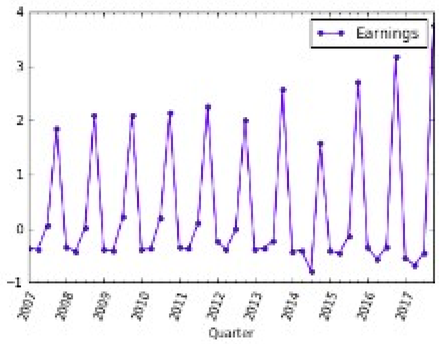
- First difference

```
plot.plot(SPY.diff())
```



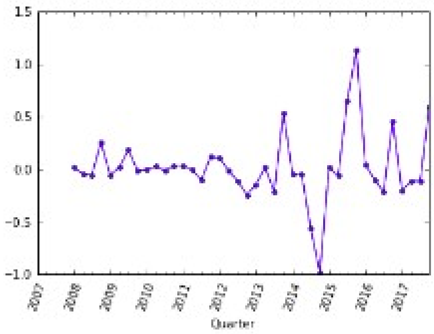
- Seasonality

```
plot.plot(HRB)
```



- Seasonal difference

```
plot.plot(HRB.diff(4))
```



6)Auto Regression Model: $R_t = \mu + \varphi * R_{t-1} + \varepsilon_t$

Since only one lagged value on right hand side, this is called:

- AR model of order 1, or
- AR(1) model

If phi is 1 then the process is random walk and if If phi is 1 then the process is white noice.

For sationary, -1<phi (φ)<1

- AR(1): $R_t = \mu + \varphi_1 R_{t-1} + \varepsilon_t$
- AR(2): $R_t = \mu + \varphi_1 R_{t-1} + \varphi_2 R_{t-2} + \varepsilon_t$
- AR(3): $R_t = \mu + \varphi_1 R_{t-1} + \varphi_2 R_{t-2} + \varphi_3 R_{t-3} + \varepsilon_t$

As example: Suppose R_t is a time series of stock price

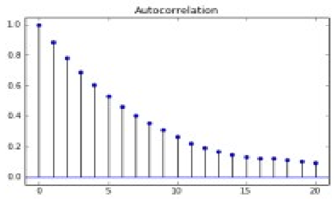
- if phi is -ve, then a positive return last peroid, at time t-1 , implies that this peroid's return is more likely to be negative i.e mean reversion
- if phi is +ve, then a positive return last peroid, at time t-1 , implies that this peroid's return is expected to be positive i.e momentum

To simulate AR model, we need to reverse the sign of phi (Below phi is 0.9)

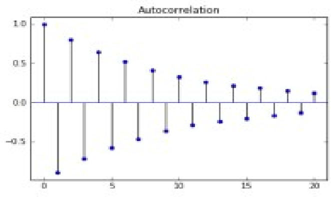
```
from statsmodels.tsa.arima_process import ArmaProcess
ar = np.array([1, -0.9])
ma = np.array([1])
AR_object = ArmaProcess(ar, ma)
simulated_data = AR_object.generate_sample(nsample=1000)
plt.plot(simulated_data)
```

6.1)Comparison of AR(1) Autocorrelation Functions:

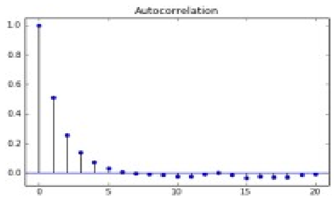
- $\phi = 0.9$



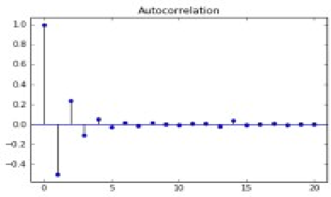
- $\phi = -0.9$



- $\phi = 0.5$



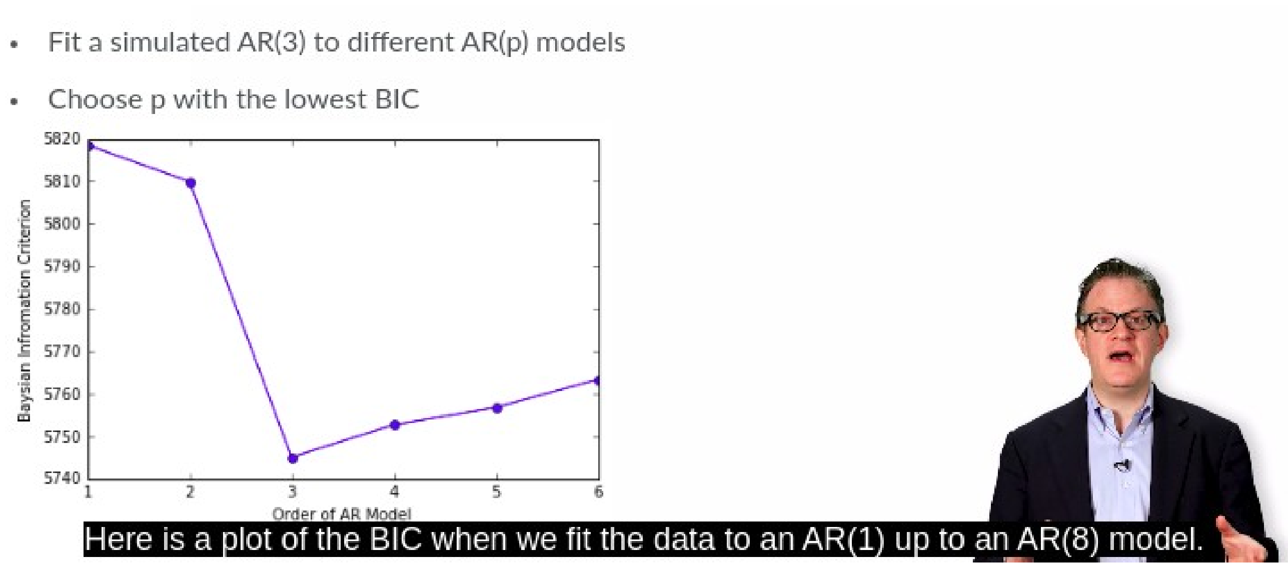
- $\phi = -0.5$



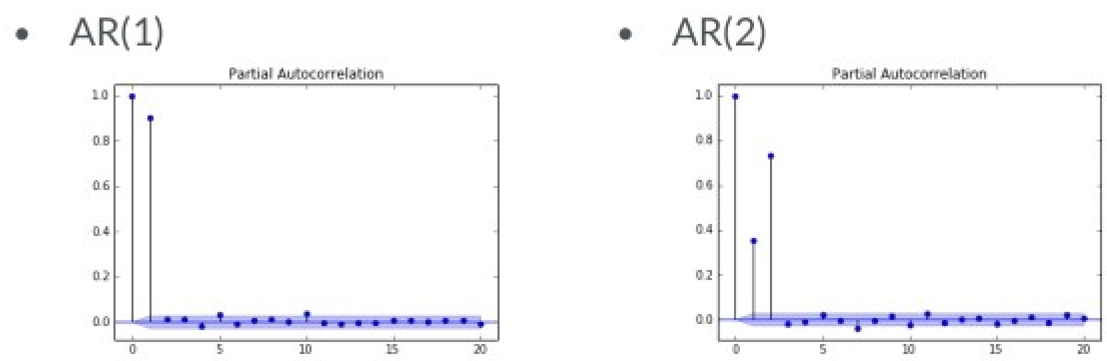
- The autocorrelation function decays exponentially for an AR time series at a rate of the AR parameter. For example, if the AR parameter, ϕ=+0.9, the first-lag autocorrelation will be 0.9, the second-lag will be (0.9)2=0.81, the third-lag will be (0.9)3=0.729, etc.
- When phi is negative, the ACF still decays exponentially, but the sign of the ACF reverse a each lag for a example for negative AR parameter, say -0.9, the decay will flip signs, so the first-lag autocorrelation will be -0.9, the second-lag will be (−0.9)2=0.81, the third-lag will be (−0.9)3=−0.729, etc.

6.2)Identifying the Order of an AR Model

- The order of an AR(p) model will usually be unknown
- One useful tool to identify the order of an AR model is to look at the Partial Autocorrelation Function (PACF).
- Two techniques to determine order
    - Partial Autocorrelation Function
    - Information criteria: adjusts goodness-of- fir for number of parameters
    - Two popular adjusted goodness-of-fit measures
        - AIC (Akaike Information Criterion)
        - BIC (Bayesian Information Criterion)



The PACF measure the incremental benefits of adding another lag



The number of non-zero partial autocorrelations gives the order of the AR model.

- At AR(1), PACF have significant lag-1 and zeros after that. So AR order is 1
- At AR(2), PACF have significant lag-1 & lag-2 and zeros after that. So AR order is 2

Note: ACF is used to identify order of MA term, and PACF for AR. There is a thumb rule that for MA, the lag where ACF shuts off suddenly is the order of MA and similarly for PACF and AR.

7)MA Model (Moving Average):

Higher Order MA Models

- MA(1)

$$R_t = \mu + \epsilon_t - \theta_1\, \epsilon_{t-1}$$

- MA(2)

$$R_t = \mu + \epsilon_t - \theta_1\, \epsilon_{t-1} - \theta_2\, \epsilon_{t-2}$$

- MA(3)

$$R_t = \mu + \epsilon_t - \theta_1\, \epsilon_{t-1} - \theta_2\, \epsilon_{t-2} - \theta_3\, \epsilon_{t-3}$$

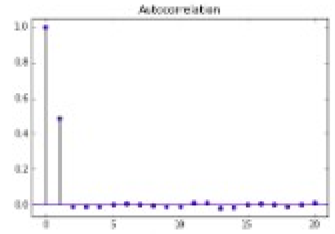
If MA parameter is 0, then the process is white noice

Suppose a R time series of stock return

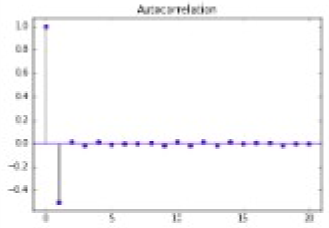
If phi is -ve, then a positive return last peroid,represented by epsilon t-1 ,would have caused last peroid’s return is more likely to be negative i.e mean reversion

If phi is -ve, then a positive return last peroid,represented by epsilon t-1 ,would have caused last peroid’s return is more likely to be negative i.e mean reversion

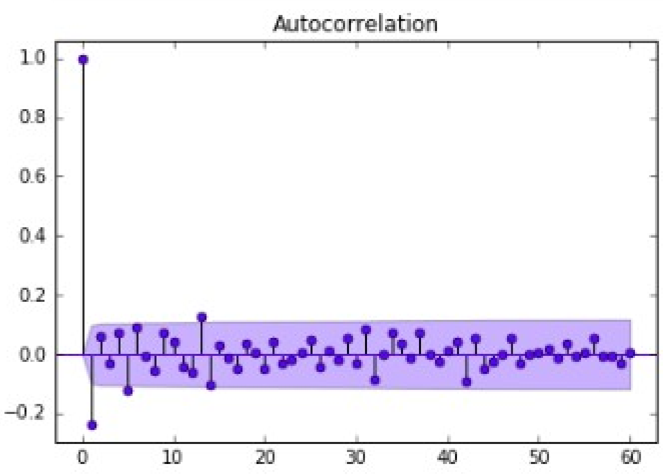
- $\theta = 0.9$



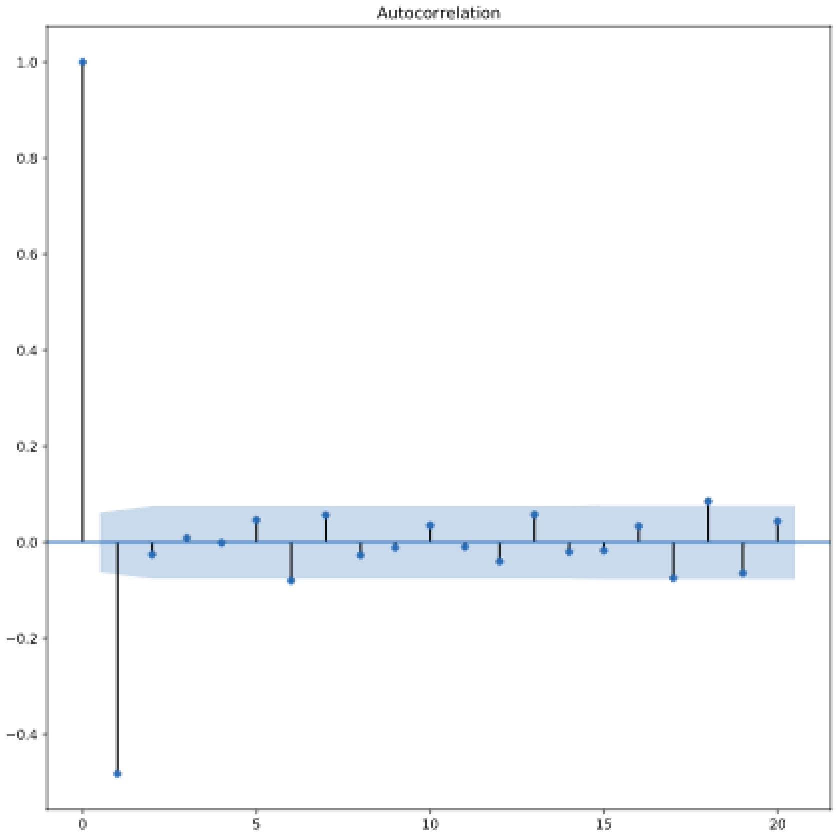
- $\theta = -0.9$



When theta is +ve, the lag-1 autocorrelation is +ve and When theta is -ve, the lag-1 autocorrelation is -ve



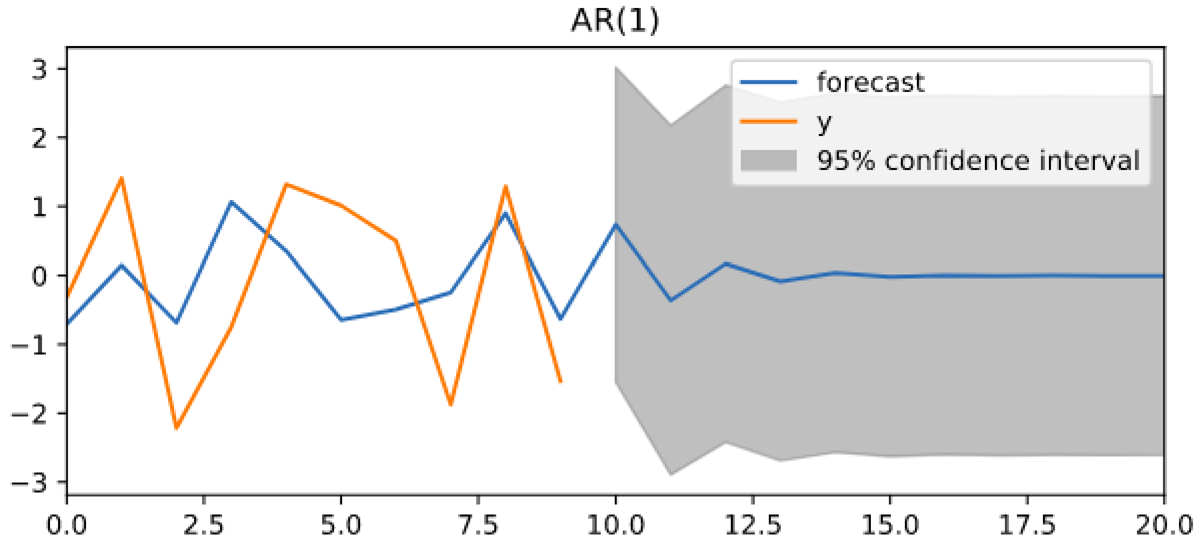
The bid/ask bounce induces a negative lag-1 autocorrelation, ut no autocorrelation beyond lag-1

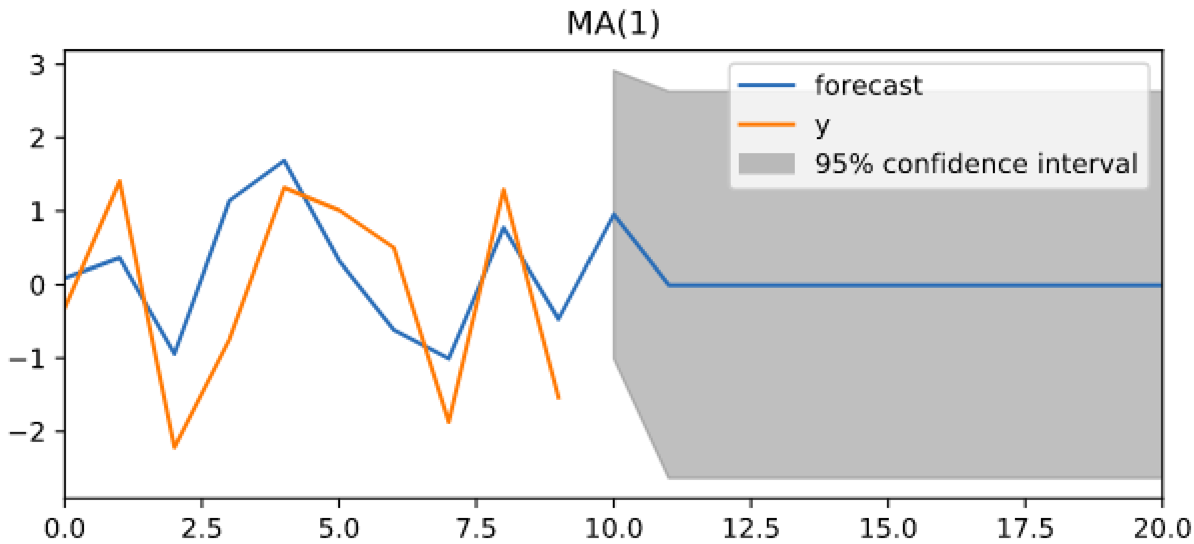


The lag-1 autocorrelation for an MA(1) model is not θ, but rather θ/(1+θ2). For example, if the MA parameter, θ, is = +0.9, the first-lag autocorrelation will be 0.9/(1+(0.9)2)=0.497, and the autocorrelation at all other lags will be zero. If the MA parameter, θ, is -0.9, the first-lag autocorrelation will be −0.9/(1+(−0.9)2)=−0.497.

Note: One thing to note is that with an MA(1) model, unlike an AR model, all forcast beyond the one-step ahead forcast will be the same.

* One big difference you will see between out-of-sample forecasts with an MA(1) model and an AR(1) model is that the MA(1) forecasts more than one period in the future are simply the mean of the sample. Below shows example





8)ARMA Model:

An ARMA model is the combination of AR and MA model.

- ARMA(1,1) model:

$$R_t = \mu + \phi R_{t-1} + \epsilon_t + \theta \epsilon_{t-1}$$

In ARMA(1,1) model, which has the familiar AR(1) and MA(1) components

9)Cointegration:

If the prices of two different assets both follow random walk, it is possible that a linear combination of them is not a random walk.

If its true, then even though P and Q are not forecastable because they are random walk but there linear combination is forecastable, and we say that P and Q are cointegrated.

- Two series, P_t and Q_t can be random walks
- But the linear combination $P_t - c\, Q_t$ may not be a random walk!
- If that's true
  - $P_t - c\, Q_t$ is forecastable
  - P_t and Q_t are said to be cointegrated

Example:

Analogy: Dog on a Leash

- $P_t =$ Owner
- $Q_t =$ Dog
- Both series look like a random walk
- Difference, or distance between them, looks mean reverting
  - If dog falls too far behind, it gets pulled forward
  - If dog gets too far ahead, it gets pulled back

9.1)What Types of Series are Cointegrated?

- Economic substitutes
  - Heating Oil and Natural Gas
  - Platinum and Palladium
  - Corn and Wheat
  - Corn and Sugar
  - ...
  - Bitcoin and Ethereum?
- How about competitors?

- Coke and Pepsi?
- Apple and Blackberry? No! Leash broke and dog ran away

9.2)Two Steps to Test for Cointegration:

- Regress P t on Q t and get slope c
- Run Augmented Dickey-Fuller test on P t − c Q t to test for random walk
- Alternatively, can use coint function in statsmodels that combines both steps

```
from statsmodels.tsa.stattools import coint
coint(P,Q)
```