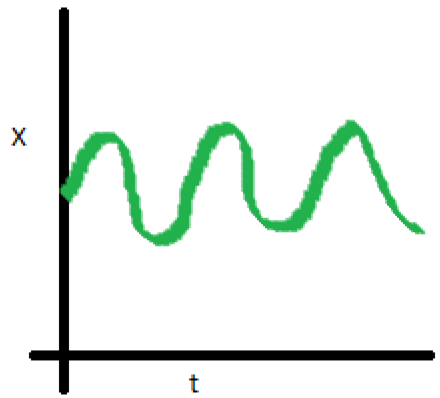
Time Series

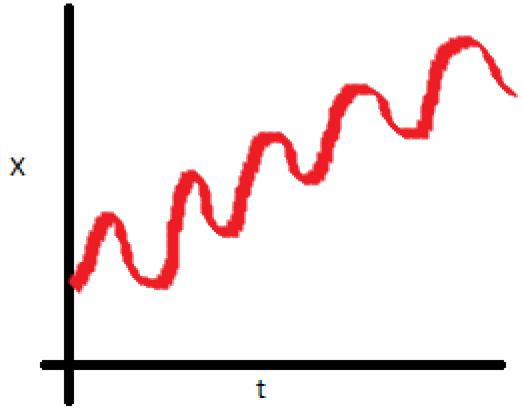
Sonal Ghanshani

Stationary Series

- The mean of the series should not be a function of time rather should be a constant. The image below has the left hand graph satisfying the condition whereas the graph in red has a time dependent mean.



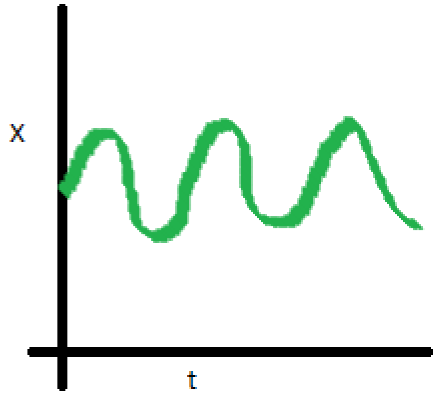
Stationary series



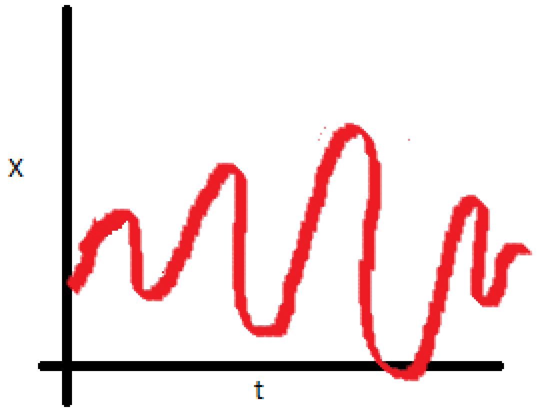
Non-Stationary series

Stationary Series

- The variance of the series should not a be a function of time. This property is known as homoscedasticity. Following graph depicts what is and what is not a stationary series. (Notice the varying spread of distribution in the right hand graph)



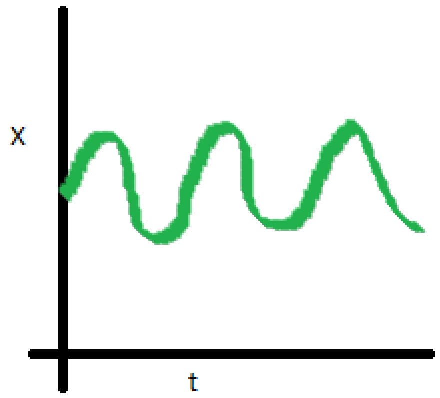
Stationary series



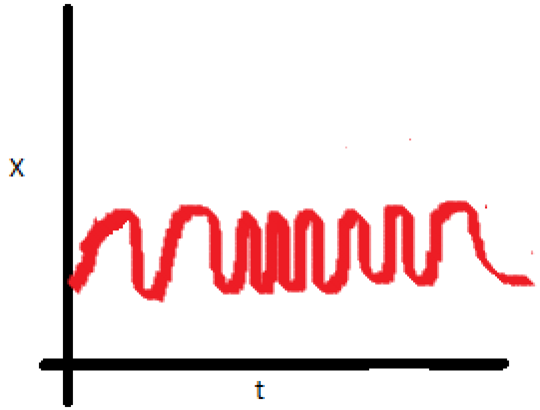
Non-Stationary series

Stationary Series

- The covariance of the i th term and the (i + m) th term should not be a function of time. In the following graph, you will notice the spread becomes closer as the time increases. Hence, the covariance is not constant with time for the 'red series'.



Stationary series



Non-Stationary series

Why do I care about ‘stationarity’ of a time series?

- Unless your time series is stationary, you cannot build a time series model. In cases where the stationary criterion are violated, the first requisite becomes to stationaries the time series and then try stochastic models to predict this time series. There are multiple ways of bringing this stationarity. Some of them are Detrending, Differencing etc.

Auto-Regressive Time Series Model

- The current GDP of a country say x(t) is dependent on the last year's GDP i.e. x(t – 1). The hypothesis being that the total cost of production of products & services in a country in a fiscal year (known as GDP) is dependent on the set up of manufacturing plants / services in the previous year and the newly set up industries / plants / services in the current year. But the primary component of the GDP is the former one.
- Hence, we can formally write the equation of GDP as:

x(t) = alpha * x(t – 1) + error (t)

- This equation is known as AR(1) formulation. The numeral one (1) denotes that the next instance is solely dependent on the previous instance. The alpha is a coefficient which we seek so as to minimize the error function.

Moving Average Time Series Model

- A manufacturer produces a certain type of bag, which was readily available in the market. Being a competitive market, the sale of the bag stood at zero for many days. So, one day he did some experiment with the design and produced a different type of bag. This type of bag was not available anywhere in the market. Thus, he was able to sell the entire stock of 1000 bags (lets call this as x(t)). The demand got so high that the bag ran out of stock. As a result, some 100 odd customers couldn’t purchase this bag. Lets call this gap as the error at that time point. With time, the bag had lost its woo factor. But still few customers were left who went empty handed the previous day. Following is a simple formulation to depict the scenario :

x(t) = beta * error(t-1) + error (t)

In MA model, noise / shock quickly vanishes with time. The AR model has a much lasting effect of the shock.

Difference between AR and MA models

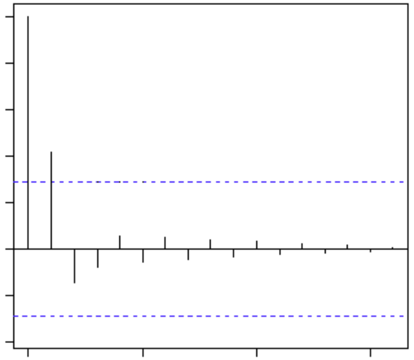
- The primary difference between an AR and MA model is based on the correlation between time series objects at different time points.
- The correlation between x(t) and x(t-n) for n > order of MA is always zero. This directly flows from the fact that covariance between x(t) and x(t-n) is zero for MA models.
- However, the correlation of x(t) and x(t-n) gradually declines with n becoming larger in the AR model.
- This difference gets exploited irrespective of having the AR model or MA model.
- The correlation plot can give us the order of MA and AR model.

Exploiting ACF and PACF plots

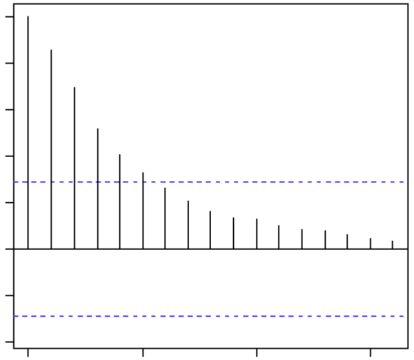
- Once we have got the stationary time series, we must answer two primary questions:

Q1. Is it an AR or MA process?

- Can be answered using Auto – correlation Function / ACF. ACF is a plot of total correlation between different lag functions.
- In a moving average series of lag n, we will not get any correlation between x(t) and x(t – n -1) . Hence, the total correlation chart cuts off at nth lag. So it becomes simple to find the lag for a MA series.



ACF



PACF

- Clearly, the graph above has a cut off on ACF curve after 2nd lag which means this is mostly a MA(2) process.

Formula

$$\hat{r}_k = \frac{\sum_{t=k+1}^{n-k} (x_{t-k} - \overline{x})(x_t - \overline{x})}{\sum_{t=1}^{n} (x_t - \overline{x})^2}$$

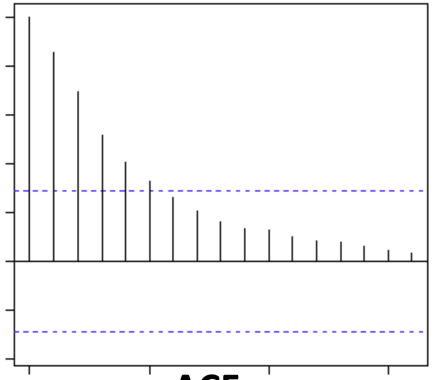
Notation

Term	Description
k	lag; $k = 1, 2, \ldots$
x_t	value of x at row t
$\overline{x}$	mean of x
n	number of observations in the series

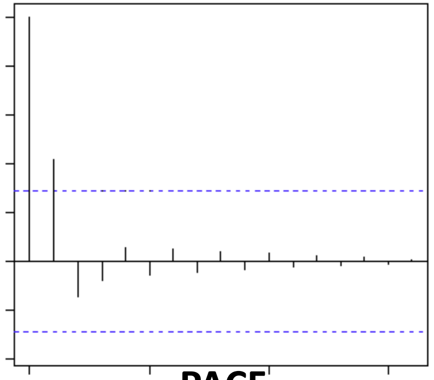
Exploiting ACF and PACF plots

Q2. What order of AR or MA process do we need to use?

- For an AR series this correlation will gradually go down without any cut off value.
- If we find out the partial correlation of each lag, it will cut off after the degree of AR series. For instance,if we have a AR(1) series, if we exclude the effect of 1st lag (x (t-1)), our 2nd lag (x (t-2)) is independent of x(t). Hence, the partial correlation function (PACF) will drop sharply after the 1st lag.



ACF



PACF

- The blue line above shows significantly different values than zero. Clearly, the graph above has a cut off on PACF curve after 2nd lag which means this is mostly an AR(2) process.