**ABC of Time Series Analysis**

Let me start off with a saying, “Time is of the essence”. I know you guys will agree with this because up until now, whatever we learnt, we never took time into consideration. Time is one of the most important factors which needs to be considered. We cannot just go ahead with the prediction and not consider time as an aspect. Well, of course we can go ahead and predict but then the very essence time brings in will be left behind. There might be questions that you might not answer like how time impacts our prediction, why somethings happen at this particular time and not in other times. So, you see, time is very important.

Let’s take an example, say Cola consumption in India. You might notice that cola consumption will be higher during the months of April, May, June and July. By now, you would have figured out the reason. Yes, you are absolutely correct, as these months are the hottest in India. Cola consumption is lower during winter. How do I know this, well we considered time as an aspect? So, let’s move ahead.

**Introduction to Time Series Analysis:**

Time Series data is an ordered sequence of values that collectively represents how a system, a process or even behaviour changes over time.

As wiki would have it, A time series is a series of data points indexed (or listed or graphed) in time order. Most commonly, a time series is a sequence taken at successive equally spaced points in time. Thus, it is a sequence of discrete-time data. Examples of time series are heights of ocean tides, counts of sunspots, and the daily closing value of the Dow Jones Industrial Average.

So, we can conclude that time series is concerned with analysis of data over time such as daily, weekly, monthly or yearly.

Goal of time series analysis is forecasting future values based on previously observed value. We can in short say that the past values influence present and future values.

There are so many things we can do with time series out of which below are the few I am mentioning for you to be aware.

1. Analyse historical trends.
2. We can look at the state of the system at any point in time and try to visualise it to predict.
3. We can do real-time monitoring.
4. Troubleshoot problem as they appear (Example being COVID cases).
5. Identify and Fixing problem even before they appear.

Okay, so we come to a question as in when not to use time series?

1. When the values are constant over the period of time. Understand this, if we have values which is constant with the given amount of time, would it be feasible to use time series? Nah, I guess not so. The values are just not moving ahead. It is a flat line. We cannot analyse anything from that set of data. We would need to use some other feature in order to come up with any prediction and time series is not one of the models that we can use in this case.
2. Values can be represented using a known function. Well, in this case as well we cannot use time series.

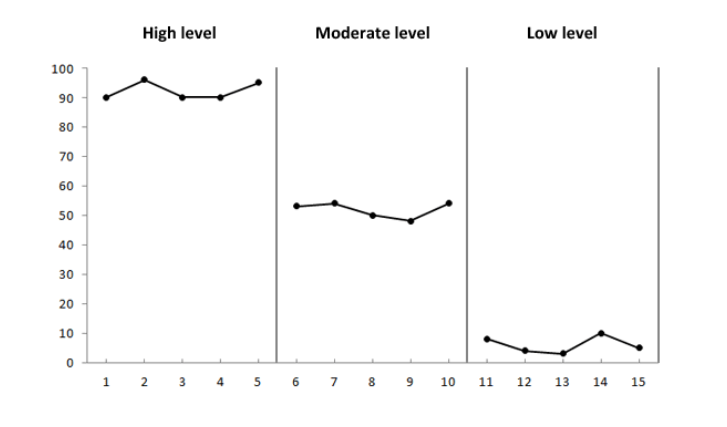
**Components of Time Series:**

Okay, let’s move further into the deeper side of time series. So, we have several components of time series namely, Level, Random Error, Trend, Cyclical and Seasonal. We will talk about each one of them. But first, let me tell you this, each time series will have Level and Random Error but Trend, Cyclical and Seasonal components are optional. Not all of them will have it.

So, moving ahead, let’s discuss each component one by one.

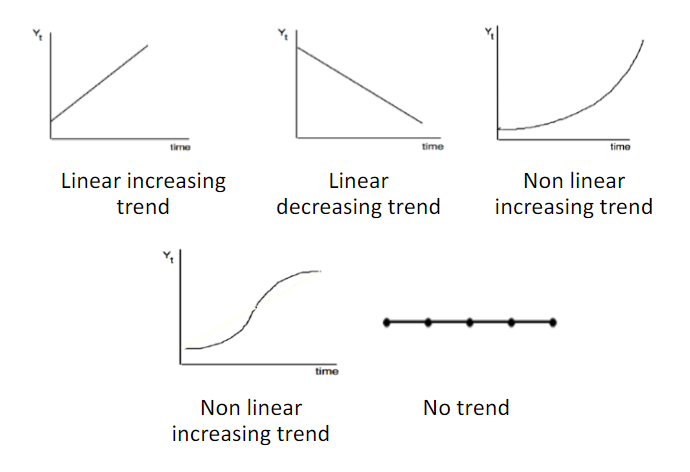
1. **Level:**

Level component of time series is the position of the data set taken from Y-axis. Or, you can say, Level is the average value in the series.



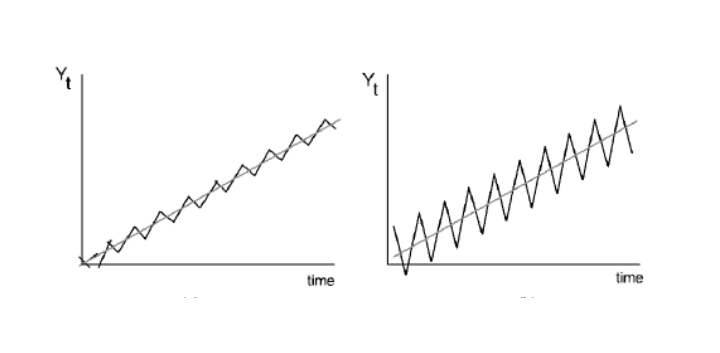
1. **Trend:**

As the name itself clarifies, the trend component is a long-term gradual increase or decrease of data indicating direction of data as in where it is heading and what we can get from it. It can be defined as increasing or decreasing value in the series.



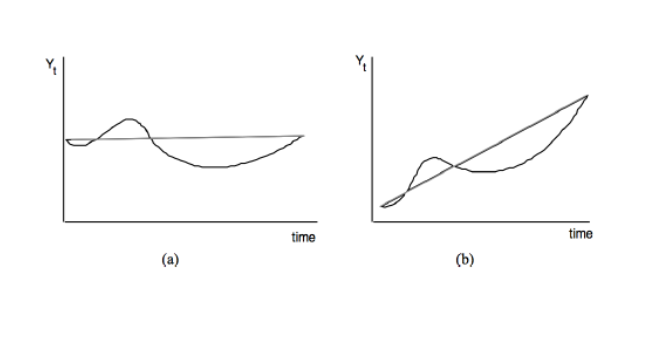
1. **Seasonality:**

Seasonality is repeating short term cycle in the series. The pattern is repeated within the time frame of the year. The cola consumption example was a seasonal one as there is seasonal increase in consumption during Summer.



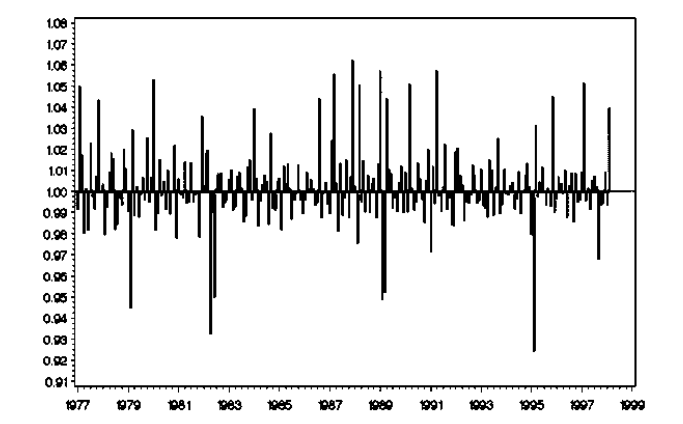
1. **Cyclical:**

A gradual long term up and down potentially irregular swings of the data. The pattern repeats irregularly over time.



1. **Random Error:**

A random increase or decrease of the data for the specific period of time. These are the kind of fluctuations that are neither systematic nor predictable.



**Decomposing a Time Series:**

Decomposing a time series can be thought of as a system which helps break time series into various components. You can think of a series as a whole i.e combination of time series components. Why do we decompose a time series? Why is there a need for doing so? Let’s find out next.

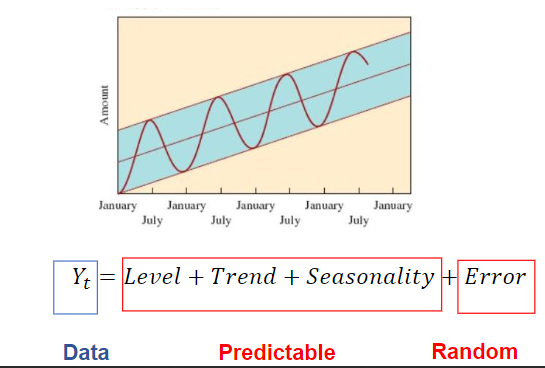
We decompose time series in order to extract vital patterns of time series data which are often difficult to detect through visual inspection. Also, it gives us the understanding of which of the components contribute the most or the least to the variance of time series data. And, once we get that understanding, we can either retain or remove such component depending on the contribution it will make towards modelling the time series data.

**How to decompose a time series?**

There are two models with the help of which we can decompose a time series.

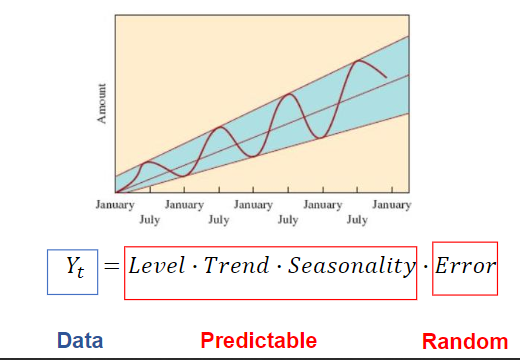
1. Additive Model:

Additive model is used when the variance of the time series does not change over different values of time series or has a normal distribution with constant variance over time.



1. Multiplicative Model:

In the multiplicative model, the original time series is expressed as the product of trend, seasonal and irregular components. So, if the seasonal variation is increasing over time, we use this model of decomposition.



**Moving Average:**

As wiki would have it, in statistics, a moving average (rolling average or running average) is a calculation to analyse data points by creating a series of averages of different subsets of the full data set. It is also called a moving mean or rolling mean and is a type of finite impulse response filter. Variations include: simple, and cumulative, or weighted forms.

Given a series of numbers and a fixed subset size, the first element of the moving average is obtained by taking the average of the initial fixed subset of the number series. Then the subset is modified by "shifting forward"; that is, excluding the first number of the series and including the next value in the subset.

Moving Average is a very important statistical component when it comes to time series. Let’s now move ahead and discuss the metrics used.

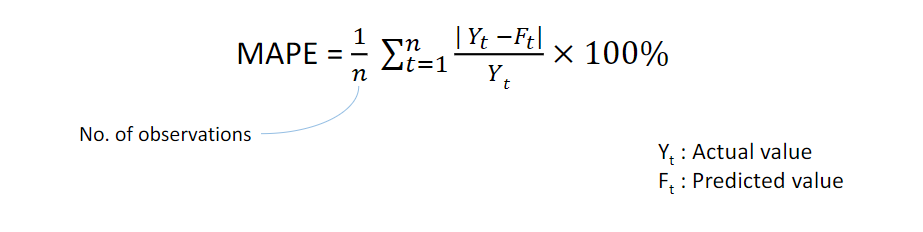
**Evaluation Metrics for Time Series Analysis:**

Up until now, I did not say a thing about input variables. But now I will say it. For the time series, Input variable is the same as output variable. Time series are almost similar to Regression model but with only difference being the input and output variables.

The Metrics that are commonly used in case of time series are as follows:

1. Mean Absolute Percentage Error

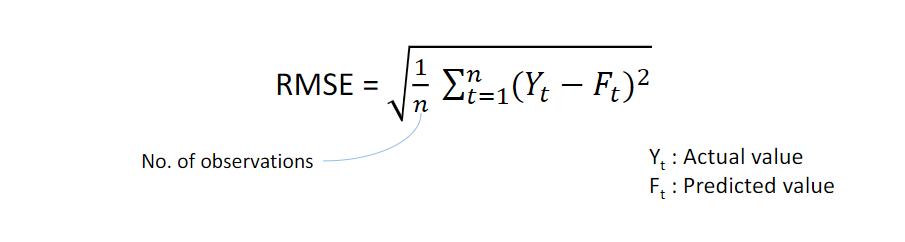
It is a measure of prediction accuracy of a forecasting method in statistics, for example in trend estimation, also used as a loss function for regression problems in machine learning. It is given by formula:



MAPE expresses the average error in percentage terms and is easy to interpret.

1. Root Mean Square Error:

The root-mean-square deviation (RMSD) or root-mean-square error (RMSE) is a frequently used measure of the differences between values (sample or population values) predicted by a model or an estimator and the values observed. Lower RMSE implies better prediction. However, it also depends on the scale of the data.



**Stationary vs Non-Stationary time series:**

A stationary (time) series is one whose statistical properties such as the mean, variance and autocorrelation are all constant over time. Hence, a non-stationary series is one whose statistical properties change over time.

In order to work with Time Series, you must have a Stationary Dataset. If in case you have Non-Stationary dataset, it must first be converted to Stationary dataset (for example by trend removal), so that further statistical analysis can be done on the de-trended stationary data. This is so because let’s say for example, if the series we have is consistently increasing over time, then the sample mean and variance will grow with the size of the sample, and they will always underestimate the mean and variance in future periods. The usual mean of “de-trending” a series is to fit a regression line and then subtract it from the original data.

To be honest, most of the statistical forecasting methods are based on the assumption that the time series dataset is stationary. And also, a stationary series is easy to predict.

**How to make any Time Series Stationary:**

In order to make any time series stationary, we perform differencing. Okay, so what is differencing now? Let’s see.

Differencing is a method of transforming a non-stationary time series into a stationary one. This is an important step in preparing data to be used in an ARIMA model. So, how does it work?

The first differencing value is the difference between the current time period and the previous time period. If these values fail to revolve around a constant mean and variance then we find the second differencing using the values of the first differencing. We repeat this until we get a stationary series.

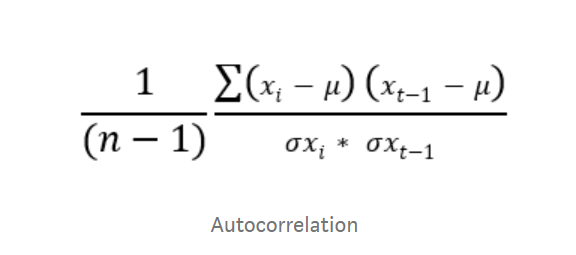
**What is Correlation?**

We are well aware about this but still, I wanted to bring it up. Correlation denotes association between two variables. For example, let’s say how Salary hike is correlated with Expense. We can say that more the salary, more the expense. We assumed this, but that’s what correlation is. How both variables are associated with each other. Let’s move ahead.

**What is Autocorrelation?**

According to wiki, Autocorrelation, also known as serial correlation, is the correlation of a signal with a delayed copy of itself as a function of delay. Informally, it is the similarity between observations as a function of the time lag between them. That’s some heavy definition. Okay, let me cut it short and simple. Autocorrelation is correlation with itself. I mean, how the variable is related or associated with itself. It is also called serial or lagged correlation.

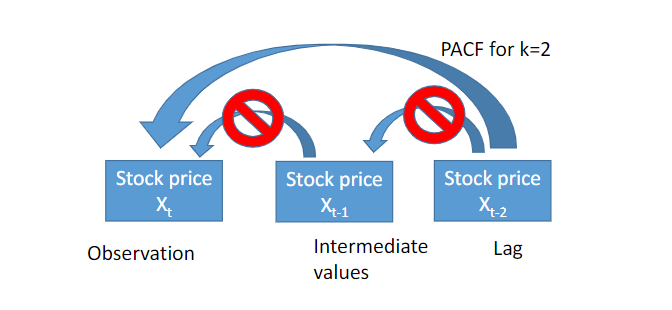
Autocorrelation can be defined as the correlation between itself and the other values of same variable(features) (in our case correlation between (Xt and Xt-1) (Xt and Xt-2). etc…) and it is denoted as ρ.



**Partial Auto Correlation:**

The partial autocorrelation at lag k is the correlation that results after removing the effect of any correlations due to the terms at shorter lags. This is some high stuff. Okay, in normal terms, PACF is conditional Correlation. It is just like an interviewer asking you What was your contribution to making of the whole project.

In autocorrelation we find correlation between present(xt) and next (xt-1) values. Partial Autocorrelation is finding the correlation between present(xt) random lags value (xt-h) so, the correlation in the middle values like (xt-1) (xt-2) (xt-3) …. (xt-(h-1)) will not be taken into account.



Before moving to discuss on models, there is one more thing I would like you to know.

**White Noise:**

A time series is white noise if the variables are independent and identically distributed with a mean of zero. This means that all variables have the same variance (sigma^2) and each value has a zero correlation with all other values in the series. Okay, and why does it matter to us? Let’s see.

White noise is very important concept in time series. The reason being

1. Predictability:

If at all there is white noise in your time series, that means by definition it is random. What I mean is that the things are happening at random and you cannot make predictions or reason with it. Still did not get it. Let’s understand with an example.

We are under attack from a virus. Let’s just assume one day that government says, there is no virus at all. And one day they say, these many people got infected. I meant, there is no fix series of things occurring, virus is coming and going at random. Would you be able to make prediction? Wouldn’t it be very hard? This is what White noise does. Predicting things becomes harder.

1. Model Diagnostics:

The series of errors from a time series forecast model should ideally be white noise. When forecast errors are white noise, it means that all of the signal information in the time series has been harnessed by the model in order to make predictions. All that is left is the random fluctuations that cannot be modelled.

**Forecasting Time Series data using multiple Models:**

1. Auto Regression Model (AR):

An autoregression (AR) model predicts future behaviour based on past behaviour. It’s used for forecasting when there is some correlation between values in a time series and the values that precede and succeed them. You only use past data to model the behaviour, hence the name autoregressive. The process is basically a linear regression of the data in the current series against one or more past values in the same series.

In an AR model, the value of the outcome variable (Y) at some point t in time is like “regular” linear regression directly related to the predictor variable (X). Where simple linear regression and AR models differ is that Y is dependent on X and previous values for Y.

1. Moving Average Model (MA)

According to Wiki, in time series analysis, the moving-average model (MA model), also known as moving-average process, is a common approach for modelling univariate time series. The moving-average model specifies that the output variable depends linearly on the current and various past values of a stochastic (imperfectly predictable) term.

1. ARIMA Model:

ARIMA Model is a combination of AR and MA model. ARIMA, short for ‘Auto Regressive Integrated Moving Average’ is actually a class of models that ‘explains’ a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values.

The model comprises of 3 components:

p is the order of AR term.

q is the order of MA term.

d is the number of differencing.

We can fit ARIMA Model on data using the value of (p,d,q) obtained and make prediction.

Phew!! That was vast. I tried to cut this short but I am sorry for this article to be this long. I couldn’t stop writing. Anyways, that’s it for this topic. Let me know if any improvement needed. Also, you can have a look at my Jupyter notebook which is places at GitHub.

Links:

<https://github.com/njain5/datasciencecode>

<https://www.linkedin.com/in/nikhiljain93/>

References:

<https://www.wikipedia.org/>

<https://towardsdatascience.com/>

<https://www.analyticsvidhya.com/>

<https://www.machinelearningplus.com/>