# Time series

Time series refers to a sequence of data points collected and recorded at regular intervals over a period of time. These data points could represent various metrics such as temperature readings, stock prices, sales figures, or any other measurable quantity. It is widely used in various fields including finance, economics, meteorology and any other for analysing trends, making forecasts, and understanding underlying patterns.

## Components of Time Series

There are three components of time series, Seasonal, Trend and residuals. Visualizing the components of time series is known as decomposition.

**Seasonality**: Patterns that repeat at a regular interval such as daily, weekly, monthly or yearly once.

**Trend**: The long-term movement or direction of the data, indicating whether is increasing, decreasing or remaining stable over time. It is also referred to as level.

**Noise**: Random fluctuations or irregularities that make the data deviate from the underlying pattern is called noise. They represent information that we cannot model or predict.

Forecasting:

*Forecasting* is predicting the future using historical data and knowledge of future events that might affect our forecasts.

***How time series forecasting is different from***

***other regression tasks***

***Time series have an order*** time series have an order, and we cannot change that order when modeling. In time series forecasting, we express future values as a function of past values. Therefore, we must keep the data in order, so as to not violate this relationship. Other regression tasks in machine learning often do not have an order. time series are indexed by time, and that order must be kept. Otherwise, you would be training your model with future information that it would not have at prediction time. This is called *look-ahead bias* in more formal terms. The resulting model would therefore not be reliable and would most probably perform poorly when you make future forecasts.

***Time series sometimes do not have features*** It is possible to forecast time series without the use of features other than the time series itself. In regression task, we are used to having dataset with many column each representing a potential predictor for out target variable. However, with time series, it is quite common to be given a simple dataset with a time column and a value at that point in time. Without any other features, we must learn ways of using past values of the time series to forecast future values.

Baseline model:

A *baseline model* is a trivial solution to our problem. It often uses heuristics, or simple

statistics, to generate predictions. The baseline model is the simplest solution you can

think of—it should not require any training, and the cost of implementation should

be very low.

Time series forecasting starts with a baseline model that serves as a benchmark

for comparison with more complex models.

 A baseline model is a trivial solution to our forecasting problem because it only

uses heuristics, or simple statistics, such as the mean.

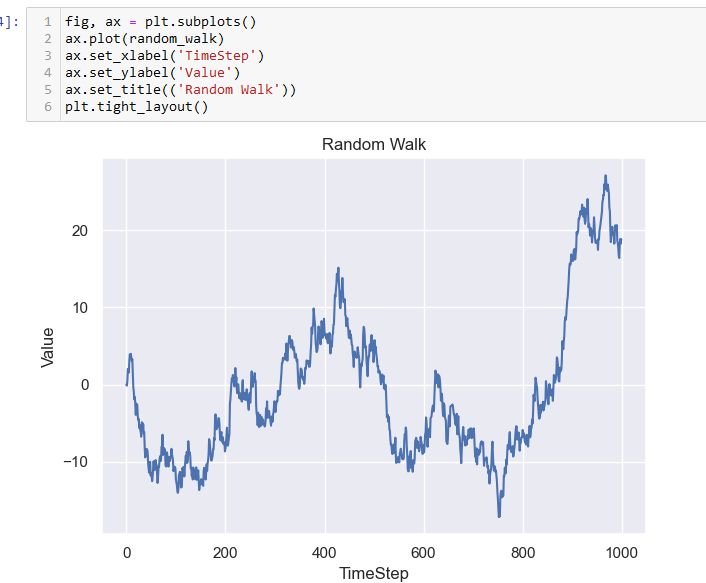
 MAPE stands for *mean absolute percentage error*, and it is an intuitive measure of

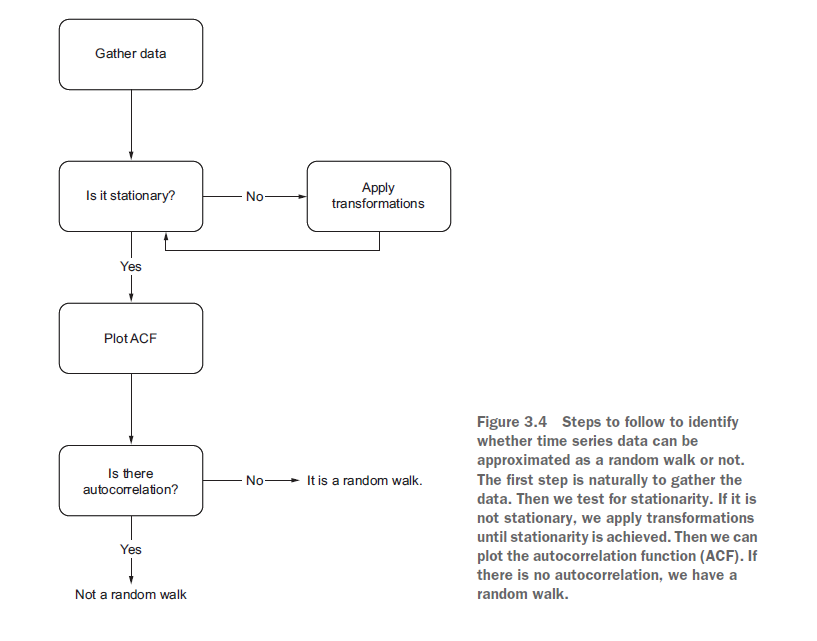
how much a predicted value deviates from the actual value.

***The random walk process***

A *random walk* is a process in which there is an equal chance of going up or down by a random number. This is usually observed in financial and economic data, like the daily closing price of GOOGL. Random walks often expose long periods where a positive or negative trend can be observed. They are also often accompanied by sudden changes in direction. In a random walk process, we say that the present value is a function of the value at the previous timestep , a constant C, and a random number ϵ*t*, also termed *white noise*. Here, ϵ*t* is the realization of the standard normal distribution, which has a variance of 1 and a mean of 0. Therefore, we can mathematically express a random walk with the following equation, where is the value at the present time , is a constant, is the value at the previous timestep , and ϵ*t* is a random number.

In the context of time series, a *random walk* is defined as a series whose first difference is stationary and uncorrelated.





***Stationarity***

A *stationary time series* is one whose statistical properties do not change over time. In

other words, it has a constant mean, variance, and autocorrelation, and these properties

are independent of time.

We can view stationarity as an assumption that can make our lives easier when forecasting. Of course, we will rarely see a stationary time series in its original state because we are often interested in forecasting processes with a trend or with seasonal cycles. This is when models like ARIMA and SARIMA come into play.

Transformation

A transformation is simply a mathematical manipulation of the data that stabilizes its mean and variance, thus making it stationary.

The simplest transformation one can apply is differencing. This transformation

helps stabilize the mean, which in turn removes or reduces the trend and seasonality

effects. Differencing involves calculating the series of changes from one timestep to

another. To accomplish that, we simply subtract the value of the previous timestep *yt*–1

from the value in the present *yt* to obtain the differenced value *y't*.

***Testing for stationarity***

Once a transformation is applied to a time series, we need to test for stationarity to determine if we need to apply another transformation to make the time series stationary, or if we need to transform it at all. A common test is the augmented Dickey-Fuller (ADF) test.

Augmented Dickey-Fuller (ADF) test

The augmented Dickey-Fuller (ADF) test helps us determine if a time series is stationary

by testing for the presence of a unit root. If a unit root is present, the time series

is not stationary.

The null hypothesis states that a unit root is present, meaning that our time series

is not stationary.

Once we have a stationary series, we must determine whether there is autocorrelation or

not. Remember that a random walk is a series whose first difference is stationary and

uncorrelated. The ADF test takes care of the stationarity portion, but we’ll need to use

the autocorrelation function to determine if the series is correlated or not.

***The autocorrelation function***

Once a process is stationary, plotting the autocorrelation function (ACF) is a great way

to understand what type of process you are analyzing. In this case, we will use it to

determine if we are studying a random walk or not.

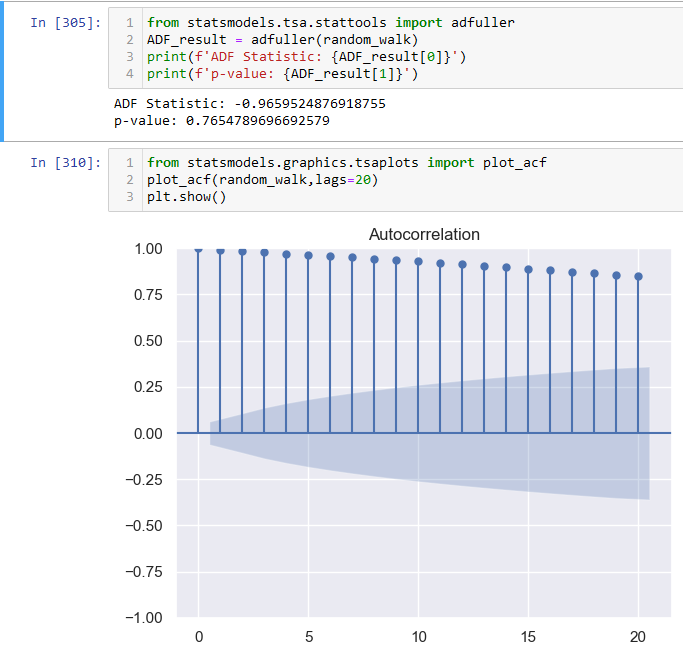
We know that correlation measures the extent of a linear relationship between two

variables. Autocorrelation therefore measures the linear relationship between lagged

values of a time series. Thus, the ACF reveals how the correlation between any two values

changes as the lag increases. Here, the lag is simply the number of timesteps separating

two values.



***Defining a moving average process***

A *moving average process*, or the moving average (MA) model, states that the current

value is linearly dependent on the current and past error terms. The error terms are

assumed to be mutually independent and normally distributed, just like white noise.

A moving average model is denoted as , where is the order. The model

expresses the present value as a linear combination of the mean of the series , the

present error term , and past error terms . The magnitude of the impact of past

errors on the present value is quantified using a coefficient denoted as . Mathematically,

we express a general moving average process of order *q*

In a moving average (MA) process, the current value depends linearly on the mean of

the series, the current error term, and past error terms.

The moving average model is denoted as MA(*q*), where *q* is the order.

The larger *q* is, the more past error

terms affect the present value. Therefore, it is important to determine the order of the

moving average process in order to fit the appropriate model

ACF plot is used to determine the order for moving average *q*

Forecasting using the MA(***q***) model

When using an MA(*q*) model, forecasting beyond *q* steps into the future will simply

return the mean, because there are no error terms to estimate beyond *q* steps. We

can use rolling forecasts to predict up to *q* steps at a time in order avoid predicting

only the mean of the series.

***Defining the autoregressive process***

An autoregressive process establishes that the output variable depends linearly on its

own previous values. In other words, it is a regression of the variable against itself.

An *autoregressive process* is denoted as an process, where is the order. In

such a process, the present value is a linear combination of a constant C, the present

error term , which is also white noise, and the past values of the series . The magnitude

of the influence of the past values on the present value is denoted as , which

represents the coefficients of the model. Mathematically, we express a general

AR(*p*) model as:

Autoregressive process

An autoregressive process is a regression of a variable against itself. In a time

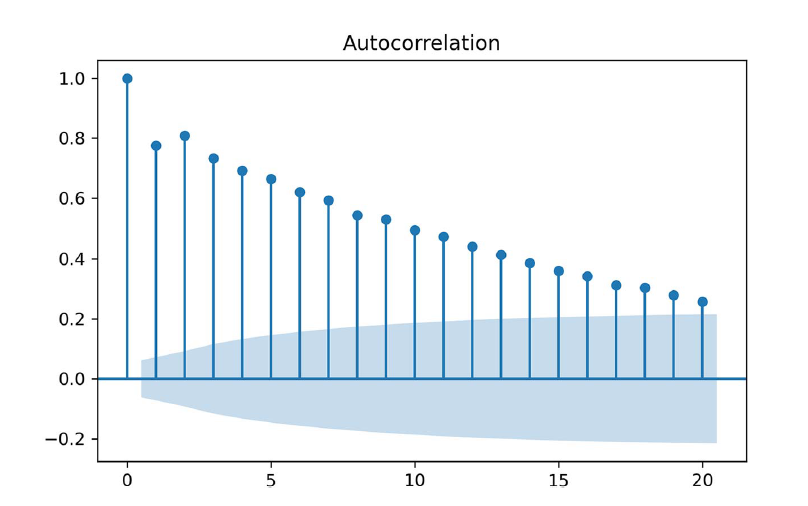
series, this means that the present value is linearly dependent on its past values.

When the ACF plot of a stationary process exhibits a pattern of exponential decay,

we probably have an autoregressive process in play, and we must find another way to

identify the order *p* of the AR(*p*) process. Specifically, we must turn our attention to

the *partial autocorrelation function* (PACF) plot.



Partial autocorrelation

Partial autocorrelation measures the correlation between lagged values in a time

series when we remove the influence of correlated lagged values in between. We can

plot the partial autocorrelation function to determine the order of a stationary AR(*p*)

process. The coefficients will be non-significant after lag *p*.

***Autoregressive Moving average process***

Even after using the ACF and PACF plots

to determine the orders *q* and *p*, respectively, if both plots will show

either a slowly decaying or sinusoidal pattern. Thus, we will define a general modeling

procedure that will allow us to model such complex time series. This procedure involves

model selection using the *Akaike information criterion* (AIC), which will determine the

optimal combination of *p* and *q* for our series.

The *autoregressive moving average process* is a combination of the autoregressive process

and the moving average process. It states that the present value is linearly dependent

on its own previous values and a constant, just like in an autoregressive process, as well

as on the mean of the series, the current error term, and past error terms, like in a

moving average process.

The autoregressive moving average process is denoted as ARMA(*p*,*q*), where *p* is

the order of the autoregressive portion, and *q* is the order of the moving average portion.

Mathematically, the ARMA(*p*,*q*) process is expressed as a linear combination of a

constant C, the past values of the series *yt–p*, the mean of the series μ, past error terms

ϵ*t–q*, and the current error term ϵ*t*, as shown below

|  |
| --- |
|  |

If your process is stationary and both the ACF and PACF plots show a decaying or sinusoidal

pattern, then it is a stationary ARMA(*p*,*q*) process.

Akaike information criterion (AIC)

The Akaike information criterion (AIC) is a measure of the quality of a model in relation to other models. It is used for model selection. The AIC is a function of the number of parameters *k* in a model and the maximum

value of the likelihood function :

The lower the value of the AIC, the better the model. Selecting according to the AIC

allows us to keep a balance between the complexity of a model and its goodness of

fit to the data.

***Understanding residual analysis***

Using the AIC as a model selection criterion, we found that an ARMA(1,1)

model is the best model relative to all others. This brings us to the last steps before forecasting, which is residual analysis, the Q-Q plot show a straight line, and are the residuals uncorrelated? The residuals of a model are simply the difference between the predicted values and the actual values. In a perfect situation the residuals of a model are white noise. This indicates that the model has captured all predictive information,and there is only a random fluctuation left that cannot be modelled. Thus, the residuals

must be uncorrelated and have a normal distribution in order for us to conclude

that we have a good model for making forecasts.

QUALITATIVE ANALYSIS: STUDYING THE Q-Q PLOT

The first step in residual analysis is the study of the *quantile-quantile plot* (Q-Q plot).

The Q-Q plot is a graphical tool for verifying our hypothesis that the model’s residuals

are normally distributed.

|  |  |
| --- | --- |
|  |  |

A Q-Q plot is a plot of the quantiles of two distributions against each other. In time

series forecasting, we plot the distribution of our residuals on the *y*-axis against the

theoretical normal distribution on the *x*-axis.

This graphical tool allows to us to assess the goodness of fit of our model. If the distribution

of our residuals is similar to a normal distribution, we will see a straight line

lying on *y* = *x*. This means that our model is a good fit, because the residuals are

similar to white noise.

On the other hand, if the distribution of our residuals is different from a normal distribution,

we will see a curved line. We can then conclude that our model is not a good

fit, since the residuals’ distribution is not close to a normal distribution, and therefore

the residuals are not similar to white noise.

QUANTITATIVE ANALYSIS: APPLYING THE LJUNG-BOX TEST

The Ljung-Box test is a statistical test that determines whether the autocorrelation of

a group of data is significantly different from 0.

In time series forecasting, we apply the Ljung-Box test on the model’s residuals to

test whether they are similar to white noise. The null hypothesis states that the data

is independently distributed, meaning that there is no autocorrelation. If the p-value

is larger than 0.05, we cannot reject the null hypothesis, meaning that the residuals

are independently distributed. Therefore, there is no autocorrelation, the residuals

are similar to white noise, and the model can be used for forecasting.

If the p-value is less than 0.05, we reject the null hypothesis, meaning that our residuals

are not independently distributed and are correlated. The model cannot be used

for forecasting.

Autoregressive integrated moving average model

An *autoregressive integrated moving average* (ARIMA) process is the combination of

the AR(*p*) and MA(*q*) processes, but in terms of the differenced series.

It is denoted as ARIMA(*p*,*d*,*q*), where *p* is the order of the AR(*p*) process, *d* is the order

of integration, and *q* is the order of the MA(*q*) process.

Integration is the reverse of differencing, and the order of integration *d* is equal to the

number of times the series has been differenced to be rendered stationary.

The general equation of the ARIMA(*p*,*d*,*q*) process is

A time series that can be rendered stationary by applying differencing is said to be an

*integrated* series. In the presence of a non-stationary integrated time series, we can use

the ARIMA(*p*,*d*,*q*) model to produce forecasts. in simple terms, the ARIMA model is simply an ARMA model that can be

applied on non-stationary time series. Whereas the ARMA(*p*,*q*) model requires the

series to be stationary before fitting an ARMA(*p*,*q*) model, the ARIMA(*p*,*d*,*q*) model

can be used on non-stationary series. We must simply find the order of integration *d*,

which corresponds to the minimum number of times a series must be differenced to

become stationary. We know that the order of integration *d* is simply the

minimum number of times a series must be differenced to become stationary. Therefore,

if a series is stationary after being differenced once, then *d* = 1. If it is stationary

after being differenced twice, then *d* = 2. In my experience, a time series rarely needs

to be differenced more than twice to become stationary.

Seasonal autoregressive integrated moving average (SARIMA) model

The *seasonal autoregressive integrated moving average* (SARIMA) model adds seasonal

parameters to the ARIMA(*p*,*d*,*q*) model.

It is denoted as SARIMA(*p*,*d*,*q*)(*P*,*D*,*Q*)*m*, where *P* is the order of the seasonal AR(*P*)

process, *D* is the seasonal order of integration, *Q* is the order of the seasonal MA(*Q*)

process, and *m* is the frequency, or the number of observations per seasonal cycle.

Note that a SARIMA(*p*,*d*,*q*)(0,0,0)*m* model is equivalent to an ARIMA(*p*,*d*,*q*) model.

Key Characteristics:

**Temporal Ordering**: The data points are ordered chronologically, with each observation corresponding to a specific period in time.

**Seasonality**: Patterns that repeat at a regular interval such as daily, weekly, monthly or yearly once.

**Trend**: The long-term movement or direction of the data, indicating whether is increasing, decreasing or remaining stable over time. It is also referred to as level.

**Noise**: Random fluctuations or irregularities that make the data deviate from the underlying pattern is called noise.

**Upward Trend:** An upward trend refers to a pattern in time series data where the values tend to increase over time. This could indicate growth, expansion or improvement in the variable being measured.

**Downward Trend:** A downward trend is the opposite of an upward trend, where the values in the time series decrease over time. This could indicate decline, contraction or deterioration in the variable being measured.

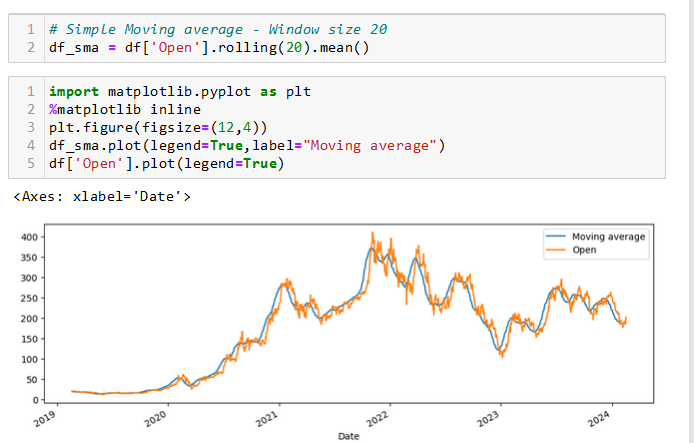
**Stationary Data**: Stationary data refers to time series data where the statistical properties such as mean, variance, and auto-correlation remain constant over time. In stationary data, there are no long-term trends or seasonality. Stationary data is easier to analyze and model because its properties do not change over time.

**Cyclic Data:** Cyclic data refers to patterns in time series data that repeat at irregular intervals, usually over a longer time span than seasonality. These cycles are not necessarily of fixed duration and can be influenced by various factor such as economic cycles, business cycles or natural phenomena. Unlike seasonality, which occurs at regular intervals, cyclic patterns may have varying durations and magnitudes.

## Simple Moving average

Formula

Explanation: The simple moving average calculates the average of a specified number of data points over a given period. It is a straight forward method for smoothing out fluctuations in data and identifying trends. The formula sumps up the values of the data points over chosen period and divides by the number of points to compute the average.



## Cumulative moving average

Formula

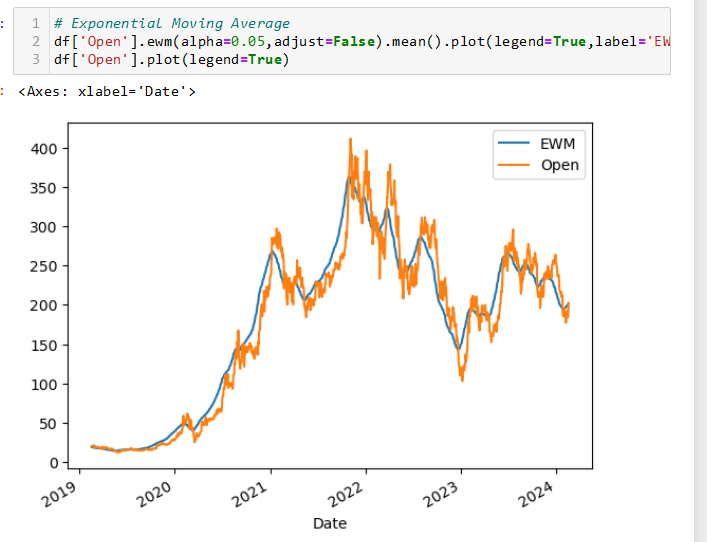
Explanation: The cumulative moving average calculates the average of all data points up to a specific point in time. Unlike SMA, which considers a fixed window of data points, the CMA includes all the data points up to the current observation. Each new observation adjusts the average, incorporating the latest data point.



## Exponential Weighted Moving Average

Formula =

Explanation: The exponential weighted moving average assigns exponentially decreasing weights to older data points. The parameter determines the smoothing factor, with higher values giving more weight to recent observations. This method provides more responsiveness to recent changes while still incorporating past data.



Exponential weighted moving average

Moving average Formula -ACF (MA)

Auto regressive model -PACF (AR)