1. **SIMPLE MOVING AVERAGE**

The most commonly used type of moving average is the simple moving average (SMA), sometimes referred to as an arithmetic moving average. An SMA is constructed by adding a set of data and then dividing by the number of observations in the period being examined. It can be calculated for any period of time. For example, if you have sales data for a twenty-year period, you can calculate a five-year moving average, a four-year moving average, a three-year moving average and so on.

It is one number that represents a net of certain past numbers. For example, a 20 day moving average is one number that represents all the ‘sales’ for the past 20 days. Thus, it tells how the group of 20 days is behaving, rather than its separate parts.

Where,

t is the time period

Y is the data value

n is the number of observations under examination (moving window)

**Example:**   
Given the following series of daily-prices:  
$10, $11, $12, $16, $17, $19, $20, $22

The SMA calculation for a 7 day period would look like this:  
Summing : $10+$11+$12+$16+$17+$19+$20 = $105  
Dividing : SMA = $105/7 = $15

Here, $15 will be the forecast for 8th day. For the next forecast that is for 9th day, we calculate the moving average for days 2 till day 8. In other words, we drop the price for day 1 and add the price for day 8 to the sum.

SMA can be calculated for various lengths of time. A long moving average period lags behind the actual variable. The average always goes into the same direction as the variable, but takes a bit more time to get moving. This is because a longer time period includes more data observations, and, thus, more information. By including more data in the calculation of the moving average, each day’s data becomes relatively less important in the calculation.

Therefore, a large change in the value on one day does not have a large impact on the longer moving average. This can be of an advantage if this large change is a one-day, irregular outlier in the data. However, if this large move represents the beginning of a significant change in the trend, it takes longer for the underlying trend change to be noticeable. Thus, the longer moving average is slower to pick up trend changes but less likely to falsely indicate a trend change due to a short-term blip in the data.

A shorter moving average has more variability than a longer one. A longer window, compared to a shorter one, will provide more smoothing but will also be slower in signalling any underlying trend changes. Smoothing data removes random variation and shows trend or seasonal or cyclic components (if any). It helps in reducing the effect of random variations or outliers.

Thus, moving averages help in smoothing out shorter fluctuations.

1. **WEIGHTED MOVING AVERAGE**

SMA represents adding most recent period’s value and dropping the earliest period’s value. When calculating the simple moving average, each period’s observation is given equal weights. For example, for a ten-day SMA, the information contained in the value for each of the ten days is given equal importance. However, in certain situations, the most recent observation may have more bearing on the future direction of the variable than the ten-day old observation does. If observations that are more recent contain more relevant information than earlier observations, we want to weight data in favour of the most recent observation. By calculating a weighted moving average, the most recent day’s information is weighted more heavily.

In this method each observation of the moving period is multiplied by its corresponding weight. The total of these numbers are added up and divided by the sum of all the multipliers.

M = Average value

Y = Actual value

W = Weighting factor

n = Number of periods in the weighting group (moving window)

The values can either be linearly weighted or weights can be specified.

**Example:**

The table below shows the prices and quantities that 3 different customers pay for the same product.

| **Customer** | **Shoe Price** | **Shoe Quantity** |
| --- | --- | --- |
| Small Customer | 300 | 20 |
| Medium Customer | 200 | 100 |
| Big Customer | 150 | 225 |

The simple average of the shoe prices would be:

(300+200+150)/ 3 = $216.67

If you look at the numbers, you can see we are selling far more shoes for price less than $200 than we are above $200. Therefore an average of $216.67 does not accurately reflect the real average selling price in the market. It would be more useful to weight those prices based on the quantity purchased.

A weighted average can be calculated like this:

(300∗20+200∗100+150∗225)/ (20+100+225) = $173.19

Since we are selling the vast majority of our shoes between $200 and $150, this number represents an average selling price of our products more accurately than the simple average.

In case of linear weights,

**Example:**

A ten-day linearly weighted moving average multiplies the tenth day observation by 10, the ninth day by 9, the eighth day by 8, and so forth. The total of these multiplied numbers are added up and divided by the sum of all the multipliers. In this case, the sum of all the multipliers (sum of weights) will be 10 + 9 + 8 + 7 + 6 + 5 + 4 + 3 + 2 + 1, or 55. The total of all the multiplied numbers will be divided by 55.

1. **SIMPLE EXPONENTIAL SMOOTHING**

For some analysts, dropping off the earliest trading day’s data that occurs with an SMA or linearly weighted moving average is problematic. If the most recent observation reflects little change in the present value, but the earliest observation, now being omitted, shows considerable change in the present value, the moving average can be unduly influenced by the discarding of the older data.

Although it might seem that such past values wouldn’t hold as much importance as the most recent values, it is still information that may have value. With the above moving averages, this older information, which lies outside of the length of the moving average, is totally being ignored. To address this issue, analysts use the simple exponential average.

In this method, the next forecast is obtained by adjusting the previous forecast in the direction of the previous error by a fractional amount α.

Substituting value of error in the equation and further solving it we get

where,

At-1 is the actual observation of previous period

Ft-1 is the forecast of previous period

An important characteristic of this method is that forecasts made with this model will include a portion of every piece of historical demand. Furthermore, there will be different weights placed on these historical demand values, with older data receiving lower weights. The forecast will be an exponentially weighted (i.e. discounted) moving average with discount factor 1-α.

At first glance this may not be obvious, however, this property is illustrated as follows:

Now doing the backwards calculation

Substituting equation 2 in equation 1

Substituting equation 4 in equation 3

]

is a smoothing factor which dampens the effect of errors, random observations or past actual values. It is basically the percentage of weight given to recent most observation. Higher alpha value means more weightage to recent values and less to older observations. This leads to dampening the effect of fluctuations in the forecasts due to observations in older time periods.

**Example:**

|  |  |  |
| --- | --- | --- |
| YEAR | ACTUAL VALUE | FORECAST(alpha=0.1) |
| 1 | 310 | 300 |
| 2 | 365 | (0.1)(310)+(1-0.1)(300) =301 |
| 3 | 395 | (0.1)(365)+(1-0.1)(301) =307.4 |
| 4 | 415 | 316.16 |

In this illustration we assume that, in the absence of data at startup, we made a guess for the year 1 forecast (300). Then, for each subsequent year (beginning with year 2) we made a forecast using the exponential smoothing model. After the forecast was made, we waited to see what demand unfolded during the year. We then made a forecast for the subsequent year, and so on right through to the forecast for year 4.

The controlling input of the exponential smoothing calculation is known as the smoothing factor [(also called the smoothing constant or alpha (α)]. It essentially represents the weighting applied to the most recent period’s observation. For example, if we want to use 35% as the weighting for the most recent period then our alpha value will be 0.35. As all our weights have to sum up to 100%, the maximum value that alpha can take is 1.

The difference with the exponential smoothing calculation is that instead of us having to also figure out how much weight to apply to each previous period, the smoothing factor is used to automatically do that. For example, if we use 35% as the smoothing factor, the weighting of the most recent period’s observation (period t-1) will be 35%. The weighting of the period before the most recent (period t-2) will be 65% of 35% (65% comes from subtracting 35% from 100%). This equates to 22.75% weighting for that period.

The period before that (period t-3) will be 65% of 65% of 35%, which equates to 14.79%. The period before that (period t-4) will be weighted as 65% of 65% of 65% of 35%, which equates to 9.61%, and so on. And this goes on back through all the previous periods all the way back to the beginning of time (or the point at which you started using exponential smoothing for that particular item).

But the beauty of the exponential smoothing calculation is that rather than having to recalculate the weighting of each previous period, you simply use the output of the exponential smoothing calculation from the previous period to represent all previous periods.

Ft = (At-1 \* 0.35) + (Ft-1 \* 0.65)

This shows that the calculation of forecast for period t involves forecast of t-1and observation of t-1 which will further involves forecast value of t-2, observation of t-2 and so on. In this way previous periods’ values are represented in the calculation without actually going back and recalculating anything.

Simple exponential is easier to calculate as you don’t have to think about what weightings to give to previous periods or how many periods to use to forecast.

1. **DOUBLE EXPONENTIAL SMOOTHING (HOLT’S METHOD)**

A popular smoothing technique commonly used in time series analysis is double exponential smoothing. Basically, it’s an improvement of simple exponential smoothing which does the exponential filter process twice. The SMA models and SES models assume that there is no trend of any kind in the data. The basic idea behind double exponential smoothing is to introduce a term to take into account the possibility of a series exhibiting some form of trend. This is done by the introduction of a second equation with a second constant, β, beta.

Trend is the pattern of gradual change in output or average. It is the general tendency of a series of data points to move in a certain direction over time. Data exhibits a progressively increasing or decreasing trend. At the lowest level, trend can be understood as the rate of change in slope.

where,

at is the level of current time period t

At is the actual value of current time

at-1 is the estimated level of previous period

bt-1 is the estimated trend of previous period

bt is the trend value of current period

Ft+1 is the forecast for next period

Level is defined as the long-term average of the data. It represents a base line with respect to which fluctuations like trends are measured. For forecast, we take the last level component and extrapolate the last trend component out. Last level component is adjusted by the last trend component to reach to a new level.

- this term is nothing but the forecast for the previous period that is t-1.

– this denotes new level of trend for current period t (difference between this period and last period’s estimated level.

The first smoothing equation which estimates level adjusts level for current period directly for the trend of the previous period, bt-1 , by adding it to the last smoothed level value, at-1. This helps to eliminate the lag that can be caused due to trend changes and brings at to the appropriate base of the current value.

The second smoothing equation then updates the trend, which is expressed as the difference between the last two level values.

The initial value of level (a1) can be set to A1 or overall average and the initial value of trend (bt) can be set to the following values.

 Models with small values of β assume that the trend changes only very slowly over time, while models with larger β assume that it is changing more rapidly.

**Example:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **t** | **At** | **at** | **bt** | **Ft+1** |
| 100 | 92 | 90 | 8.5 | 98.5 |
| 101 | 95 | 97.5 | 8.4 | 105.9 |

Suppose we are in time 101 and we use alpha=0.3 and beta=0.1.

a) Forecast demand for t=102

🡪 Find a101

a101 = (0.3)\*(95.0) + (0.7)\*(90.0+8.5)

= 97.5

🡪Find b101

b101 = (0.1)\*(97.5−90.0) +(0.9)\*(8.5) = 8.4

🡪Find F102 = 97.5+8.4 = 105.9

To forecast for p periods ahead with current time period t, the following formula is used

Holt’s models are best suited to shorter-term forecasting say 24 months or lesser due to their assumption that future patterns and trends will resemble current patterns and trends. This is a reasonable assumption in the short term but becomes more tenuous the further you forecast.

1. **TRIPLE EXPONENTIAL SMOOTHING (HOLT-WINTERS)**

The Holt-Winters method is more suitable when the data series to be forecasted show a seasonality pattern. The Holt-Winters method is often used when both trend and seasonality patterns are present in the data series. It incorporates three equations: the first for the level, the second for the trend, and the third for seasonality.

At first, all the three components (level, trend and seasonality) are initialized. This means values of these components at the very first period are determined. Then, with each observation, these components are updated. For example, in January (year t), we update the seasonal component relevant to January (year t-1) while other components are updated with respect to December (year t-1).

In general, there are two Holt-Winters methods, the multiplicative holt- winters method (MHW) method and the additive holt-winters method (AHW) method, depending on whether the seasonality is modelled in multiplicative or additive forms, respectively.

Additive seasonality is the one where the amplitude of the seasonal variation is independent of the level. That is seasonality remains approximately constant over the periods. This is when the absolute difference between two subsequent seasonal indexes is the same. Absolute change is same. For example, difference between January and March values is approximately the same each year.

Multiplicative seasonality is when amplitude and level are connected. Seasonality becomes wider as level increases. It is increasing over time. If level is decreasing then seasonal amplitude would shrink as well. Here, change in data values is measured in % (proportion of change). For example, seasonality in quarter 2 is 20% more than seasonality in quarter 1.

The MHW method is

where

N is the period length (i.e., 12, number of months in a year or 4, number of quarters in a year).

St is the seasonal component for current period t

St-N is the previous seasonal component relevant to period t

m is the m-step-ahead forecast made at current time t (m=1,2,3,4…)

The AHW method is

Alpha, beta and gamma determine the weight of recent data on level, trend and seasonality respectively. These parameters must lie between 0 and 1. Closer the values of these parameters to zero, slower the process of smoothing will be. If they are closer to 1 then it means that smaller fraction of older observations will be considered in the calculation of forecast and thus will dampen the effect of fluctuations in the past on the future. For example, the data you are dealing with is at weekly level and highly volatile. You don’t want level of your series being inflated and deflated on a week by week basis then you would go for a relatively low alpha. Also, data is seasonal on a weekly level and weeks with days like Black Friday affect the seasonality level immensely. Thus you would go for high gamma values to take into account each week’s seasonality.

Higher alpha and beta values will absorb the changes in actual data quickly. The forecast will closely follow the path of ups and downs in the actual data. Lag will be less. Thus if you see that trend is changing frequently and want to capture that in the forecast then a higher beta value would make more sense.

When selecting the smoothing constant subjectively, you use your own experience with this, and similar, series. Also, by specifying the smoothing constant yourself, you tune the forecast to your own beliefs about the future of the series. If you believe that the mechanism generating the series has recently gone through some fundamental changes, use a smoothing constant value of more than 0.5 which will cause distant observations to be ignored. If, however, you think the series is fairly stable and only going through random fluctuations, use a value of less than 0.5.

To select the value of the smoothing constants objectively, you search for values that are best in some sense. The values that give least errors are selected for forecasting.

**Example:**

a0 = 95.25

b0 = 2.47

N = 4

s1-4 = s-3 = 0.7062

s2-4 = s-2 = 1.1114

s3-4 = s-1 = 1.2937

F1 = (95.25 + 2.47) (0.7062)

= 69.0103

Now,

Let α=0.2, β=0.1, γ=0.1 and A0=72

a1 = 0.2(72/0.7062) + 0.8(95.25 + 2.4706) = 98.5673

b1 = 0.1(98.5673 – 95.25) + 0.9(2.4706) = 2.5553

s1 = 0.1(72/98.5673) + 0.9(0.7062) = 0.7086

F2 = (a1 + b1) (s2-4)

= (98.56 + 2.55) (1.11) = 112.3876

For m step ahead forecast at current time period t:

Ft+p = (at + mbt) (st+m-N)

3 periods ahead forecast will be

F1+2 = F3 = (at + 2bt) (s3-4)

= (98.56 + 2\*2.55) (1.2937) = 134.104

The concept is that to predict 3 periods out (months, weeks etc.) you take current level, add 3 times the calculated trend and multiply(or add) by the seasonality of the last corresponding period.

This method is best suited when there is at least 2-3 years of past data with seasonality. This helps in selecting better values of alpha, beta and gamma by visually gathering information on the data trend and seasonality.

1. **AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA)**

One of the most popular and frequently used stochastic time series models is the Autoregressive Integrated Moving Average (ARIMA) model. The basic assumption made to implement this model is that the considered time series is linear and follows a particular known statistical distribution, such as the normal distribution. ARIMA model has subclasses of other models, such as the Autoregressive (AR), Moving Average (MA) and Autoregressive Moving Average (ARMA) models. The popularity of the ARIMA model is mainly due to its flexibility to represent several varieties of time series with simplicity. But the severe limitation of these models is the pre-assumed linear form of the associated time series which becomes inadequate in many practical situations.

ARIMA requires at least 39-40 historical data points. It works best when data exhibits a stable or consistent pattern over time with a minimum amount of outliers and when correlation between past observations is stable.

The first step in applying ARIMA methodology is to check for stationarity. "Stationarity" implies that the series remains at a fairly constant level over time. If a trend exists, as in most economic or business applications, then your data is NOT stationary. The data should also show a constant variance in its fluctuations over time.

If a graphical plot of the data indicates non-stationarity, then you should "difference" the series. Differencing is an excellent way of transforming a non-stationary series to a stationary one. This is done by subtracting the observation in the current period from the previous one. If this transformation is done only once to a series, you say that the data has been "first differenced". This process essentially eliminates the trend if your series is growing at a fairly constant rate. If it is growing at an increasing rate, you can apply the same procedure and difference the data again. Your data would then be "second differenced".

Autocorrelations are numerical values that indicate how a data series is related to itself over time. It measures how strongly observations are correlated to each other at a specified number of periods apart (number of periods apart = lag). For example, autocorrelation at lag 1 measures how values 1 period apart are correlated to one another throughout the series. AC ranges from -1 to +1. A value close to +1 indicates a high positive correlation while a value close to -1 implies a high negative correlation.

 ARIMA models are generally denoted **ARIMA (*p*, *d*, *q*)** where parameters *p*, *d*, and *q* are non-negative integers, *p* is the order (number of time lags) of the autoregressive model, *d* is the degree of differencing (the number of times the data have had past values subtracted), and *q* is the order of the moving-average model.

**p(AR part of ARIMA)** - An AR model with only 1 parameter may be written as

F(t) = a(1) \* A(t-1) + E(t)

where F(t) = time series under investigation

a(1) = the autoregressive parameter of order 1

A(t-1) = the time series lagged 1 period

E(t) = the error term of the model

This simply means that any given value F(t) can be explained by some function of its previous value, A(t-1), plus some unexplainable random error, E(t). If the estimated value of a(1) was .30, then the current value of the series would be related to 30% of its value 1 period ago. Of course, the series could be related to more than just one past value. For example,

F(t) = a(1) \* A(t-1) + a(2) \* A(t-2) + E(t)

This indicates that the current value of the series is a combination of the two immediately preceding values, A(t-1) and A(t-2), plus some random error E(t). Our model is now an autoregressive model of order 2.

**q(MR part)** - Moving average parameters relate what happens in period t only to the random errors that occurred in past time periods, i.e. E(t-1), E(t-2), etc. rather than to A(t-1), A(t-2), A(t-3) as in the autoregressive approaches. A moving average model with one MA term may be written as follows.

F(t) = -b(1) \* E(t-1) + E(t)

The term b(1) is called an MA of order 1. The negative sign in front of the parameter is used for convention only and is usually printed out automatically by most computer programs. The above model simply says that any given value of F(t) is directly related only to the random error in the previous period, E(t-1), and to the current error term, E(t).

**d[I (for "integrated")]** - indicates that the data values have been replaced with the difference between their values and the previous values (and this differencing process may have been performed more than once).

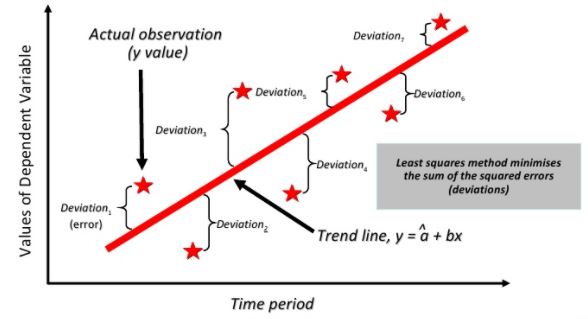
1. **LINEAR REGRESSION**

It attempts to draw a straight line through the historical data points in a fashion that comes as close to the points as possible. (Technically, the approach attempts to reduce the vertical deviations of the points from the trend line, and does this by minimizing the squared values of the deviations of the points from the line).

Ultimately, the statistical formula computes a slope for the trend line (b) and the point where the line crosses the y-axis (a). This results in the straight line equation

Y = a + bX

Where X represents the independent values on the horizontal axis, and Y represents the dependent values on the vertical axis. Independent values are exogenous that is given beforehand. On the basis of these independent X variables, forecast for Y is made. The value of Y is dependent on the value of X.

.

**Example:**

b = 30; a = 295

Y = 295 + 30X

This equation can be used to forecast for any year into the future.

For example: Year 7: Forecast = 295 + 30(7) = 505

It aims at fitting a straight line with minimum errors in the set of points.

Regression is used when there is a correlation between X and Y that is the independent and dependent variables are either positively related or negatively related. This further means if we increase/decrease X by some units then it leads to an increase/decrease in Y as well. In case of a negative relation wherein one increases and other decreases, the coefficient of X will be negative.