6.4.3.1. Single Exponential Smoothing

Exponential smoothing weights past observations with exponentially decreasing weights to forecast future values

This smoothing scheme begins by setting S_0 to y_1, where S stands for smoothed observation or EWMA, and y for the observation The subscripts refer to the time periods, 1, 2, ..., n. For the second period, $S_2 = \alpha y_2 + (1-\alpha) S_1$ and so on.

For any time period t, the smoothed value S_t is found by computing

$$S_t = \alpha y_{t-1} + (1-\alpha) S_{t-1} \qquad 0 < \alpha \le 1 \qquad t \ge 3$$

This equation is due to Roberts (1959) and is called the **basic equation of exponential smoothing** and the constant or parameter α is called the **smoothing constant**.

Setting the first EWMA

The first forecast is very important

The initial EWMA plays an important role in computing all the subsequent EWMA's. Setting S_0 to y_1 is one method of initialization. Another way is to set it to the target of the process.
Still another possibility would be to average the first four or five observations.

It can also be shown that the smaller the value of α, the more important becomes the selection of the initial EWMA The user would be wise to try a few methods, (assuming that the software has them available) before finalizing the settings.

Why is it called "Exponential"?

Let us expand the basic equation by first substituting for S_{t-1} in the basic equation to obtain

$$\begin{aligned} S_t &= \alpha y_t + (1-\alpha)[\alpha y_{t-1} + (1-\alpha) S_{t-2}] \\ &= \alpha y_t + \alpha (1-\alpha) y_{t-1} + (1-\alpha)^2 S_{t-2} \end{aligned}$$

By substituting for S_{t-2}, then for S_{t-3}, and so forth, until we substitute for S_0 it can be shown that the expanding equation can be written as:

$$S_t = \alpha \sum_{i=0}^{t-1} (1-\alpha)^i y_{t-i} + (1-\alpha)^t S_0$$

For example, the expanded equation for the smoothed value S_4 is:

$$S_5 = \alpha \left[(1-\alpha)^0 y_{5-1} + (1-\alpha)^1 y_{5-2} + (1-\alpha)^2 y_{5-3} \right] + (1-\alpha)^3 S_2$$

This illustrates the exponential behavior. The weights, $\alpha (1-\alpha)^t$ decrease geometrically, and their sum is unity as shown below, using a property of geometric series:

$$\alpha \sum_{i=0}^{t-1} (1-\alpha)^i = \alpha \left[\frac{1-(1-\alpha)^t}{1-(1-\alpha)} \right] = 1-(1-\alpha)^t$$

From the last formula we can see that the summation term shows that the contribution to the smoothed value S_t becomes less at each consecutive time period.

Let α = .3. Observe that the weights $\alpha(1-\alpha)^t$ decrease exponentially (geometrically) with time.

	Value	Weight
last	y_1	.2100
	y_2	.1470
	y_3	.1029
	y_4	.0720

What is the "best" value for α?

How do you choose the weight parameter?

The speed at which the older responses are dampened (smoothed) is a function of the value of α. When α is close to 1, dampening is quick and when a is close to 0, dampening is slow. This is illustrated in the table below:

---------------> towards past observations

α	$(1-\alpha)$	$(1-\alpha)^2$	$(1-\alpha)^3$	$(1-\alpha)^4$
.9	.1	.01	.001	.0001
.5	.5	.25	.125	.0625
.1	.9	.81	.729	.6561

The best value for α is that value which results in the smallest MSE.

Let us illustrate this principle with an example. Consider the following data set consisting of 12 observation taken over time:

Example

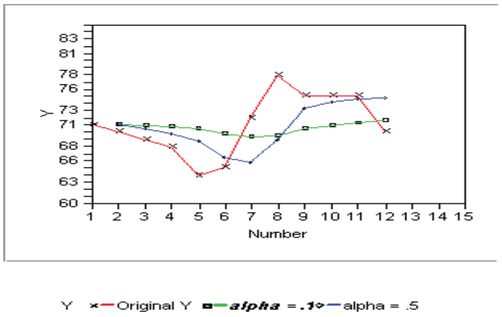
Time	y_t	$S(\alpha=.1)$	Error	Error sq
1	71	71.0	0	
2	70	70.9	−.90	0.81
3	69	70.7	−1.71	2.92
4	68	70.4	−2.44	5.95
5	64	69.8	−5.80	33.58
6	65	69.3	−4.32	18.62
7	72	69.6	2.42	5.84
8	78	70.4	7.57	57.37
9	75	70.9	4.12	16.95
10	75	71.3	3.71	13.73
11	75	71.7	3.34	11.12
12	70	71.5	−1.50	2.25

The sum of the squared errors (SSE) =169.143 The mean of the squared errors (MSE) is the SSE /12 = 14.095

The MSE was again calculated for α = .5 and turned out to be3.78 so in this case we would select an α of 5. Can we do better? We could apply the proven trial and error method. This is an iterative procedure beginning with a range of α between .1 and .9. The we find the smallest value for α and search between α - Δ to α + Δ. We could repeat this maybe one more time to find the best α to 3 decimals.

But there are better search methods, such as the Marquardt procedure. This is a nonlinear optimizer that minimizes the sum of squares of residuals. In general, most well designed statistical software programs should be able to find that value for α that minimizes the MSE.

Exponential Smoothing: Original and Smoothed Values



NIST SEMATECH HOME TOOLS & AIDS SEARCH BACK NEXT