Exponential Smoothing

Exponential smoothing schemes weight past observations using exponentially decreasing weights.

This is a very popular scheme to produce a smoothed Time Series. Whereas in Single Moving Averages the past observations are weighted equally, Exponential Smoothing assigns exponentially decreasing weights as the observation get older.
In other words, recent observations are given relatively more weight in forecasting than the older observations.
In the case of moving averages, the weights assigned to the observations are the same and are equal to 1/N. In exponential smoothing, however, there are one or more smoothing parameters to be determined (or estimated) and these choices determine the weights assigned to the observations.
Single, double and triple Exponential Smoothing will be described in this section.

Single Exponential Smoothing

This smoothing scheme begins by setting S_2 to y_1, where S_i stands for smoothed observation or Exponentially Weighted Moving Average (EWMA), and y stands for the original observation. The subscripts refer to the time periods, 1, 2, ..., n. For the third period, $S_3 = \alpha\, y_2 + (1-\alpha)\, S_2$; and so on. There is no S_1; the smoothed series starts with the smoothed version of the second observation.
For any time period t, the smoothed value St 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 is the basic equation of exponential smoothing and the constant or parameter α is called the smoothing constant.

Why is it called "Exponential"?
Expand basic equation
Let us expand the basic equation by first substituting for S_{t-1} in the basic equation to obtain

$$S_t = \alpha\, y_{t-1} + (1-\alpha)\, [\, \alpha\, y_{t-2} + (1-\alpha)\, S_{t-2}\,]$$
$$= \alpha\, y_{t-1} + \alpha\, (1-\alpha)\, y_{t-2} + (1-\alpha)^2\, S_{t-2}$$

By substituting for S_{t-2}, then for S_{t-3}, and so forth, until we reach S_2 (which is just y_1), it can be shown that the expanding equation can be written as:

$$S_t = \alpha \sum_{i=1}^{t-2} (1-\alpha)^{i-1} y_{t-i} + (1-\alpha)^{t-2} S_2, \quad t \geq 2$$

For example, the expanded equation for the smoothed value S_5 *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:

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

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

Example for α = .3

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 α 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

We choose the best value for α so the value which results in the smallest MSE.

Example

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

Time	y_t	**S (α=.1)**	**Error**	**Error squared**
1	71			
2	70	71	-1.00	1.00
3	69	70.9	-1.90	3.61
4	68	70.71	-2.71	7.34
5	64	70.44	-6.44	41.47
6	65	69.80	-4.80	23.04
7	72	69.32	2.68	7.18
8	78	69.58	8.42	70.90
9	75	70.43	4.57	20.88
10	75	70.88	4.12	16.97
11	75	71.29	3.71	13.76
12	70	71.67	-1.67	2.79

The sum of the squared errors (SSE) = 208.94. The mean of the squared errors (MSE) is the SSE /11 = 19.0.

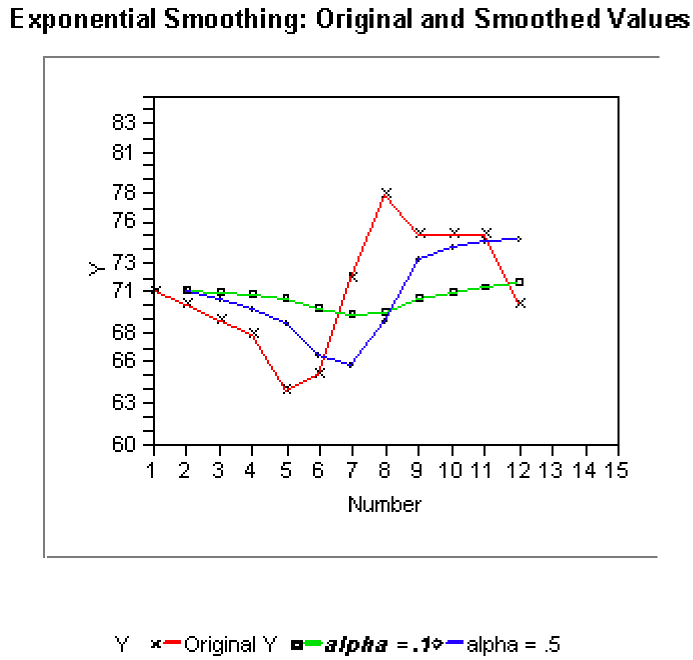
Calculate for different values of α

The MSE was again calculated for α = .5 and turned out to be 16.29, so in this case we would prefer 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. We determine the best initial choice for α and then search between $\alpha - \Delta$ and $\alpha + \Delta$. We could repeat this perhaps one more time to find the best α to 3 decimal places.

Nonlinear optimizers can be used

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 the value of α that minimizes the MSE.

Sample plot showing smoothed data for 2 values of α



Double Exponential Smoothing

Double exponential smoothing uses two constants and is better at handling trends

As was previously observed, Single Smoothing does not excel in following the data when there is a trend. This situation can be improved by the introduction of a second equation with a second constant, γ, which must be chosen in conjunction with α.

Here are the two equations associated with Double Exponential Smoothing:

$$S_t = \alpha y_t + (1-\alpha)(S_{t-1} + b_{t-1}) \qquad 0 \le \alpha \le 1$$

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)\, b_{t-1} \qquad 0 \le \gamma \le 1$$

Note that the current value of the series is used to calculate its smoothed value replacement in double exponential smoothing.

Initial Values

Several methods to choose the initial values

As in the case for single smoothing, there are a variety of schemes to set initial values for S_t and b_t in double smoothing.

S_1 is in general set to y_1. Here are three suggestions for b_1:

$$b_1 = y_2 - y_1$$

$$b_1 = [(y_2 - y_1) + (y_3 - y_2) + (y_4 - y_3)]/3$$

$$b_1 = (y_n - y_1)/(n - 1)$$

Comments

Meaning of the smoothing equations

The first smoothing equation adjusts S_t directly for the trend of the previous period, b_{t-1}, by adding it to the last smoothed value, S_{t-1}. This helps to eliminate the lag and brings S_t 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 values. The equation is similar to the basic form of single smoothing, but here applied to the updating of the trend.

Non-linear optimization techniques can be used

The values for α and γ can be obtained via non-linear optimization techniques, such as the Marquardt Algorithm.

Triple Exponential Smoothing

What happens if the data show trend **and** seasonality?

To handle seasonality, we have to add a third parameter

In this case double smoothing will not work. We now introduce a third equation to take care of seasonality (sometimes called periodicity). The resulting set of equations is called the "Holt-Winters" (HW) method after the names of the inventors.

The basic equations for their method are given by:

$$S_t = \alpha \frac{y_t}{I_{t-L}} + (1-\alpha)(S_{t-1} + b_{t-1})$$ OVERALL SMOOTHING

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1}$$ TREND SMOOTHING

$$I_t = \beta \frac{y_t}{S_t} + (1-\beta) I_{t-L}$$ SEASONAL SMOOTHING

$$F_{t+m} = (S_t + m b_t) I_{t-L+m}$$ FORECAST

where

- y is the observation
- S is the smoothed observation
- b is the trend factor
- I is the seasonal index
- F is the forecast at m periods ahead
- t is an index denoting a time period

and α, β, and γ are constants that must be estimated in such a way that the MSE of the error is minimized. This is best left to a good software package.

Complete season needed

To initialize the HW method we need at least one complete season's data to determine initial estimates of the seasonal indices I_{t-L}.

L periods in a season

A complete season's data consists of L periods. And we need to estimate the trend factor from one period to the next. To accomplish this, it is advisable to use two complete seasons; that is, $2L$ periods.

Initial values for the trend factor

How to get initial estimates for trend and seasonality parameters

The general formula to estimate the initial trend is given by

$$b = \frac{1}{L}\left(\frac{y_{L+1} - y_1}{L} + \frac{y_{L+2} - y_2}{L} + \cdots + \frac{y_{L+L} - y_L}{L}\right)$$

Initial values for the Seasonal Indices

As we will see in the example, we work with data that consist of 6 years with 4 periods (that is, 4 quarters) per year. Then

Step 1: compute yearly averages

Step 1: Compute the averages of each of the 6 years

$$A_p = \frac{\sum_{i=1}^{4} y_i}{4} \qquad p = 1, 2, \cdots, 6$$

Step 2: divide by yearly averages

Step 2: Divide the observations by the appropriate yearly mean

1	2	3	4	5	6
y_1/A_1	y_5/A_2	y_9/A_3	y_{13}/A_4	y_{17}/A_5	y_{21}/A_6
y_2/A_1	y_6/A_2	y_{10}/A_3	y_{14}/A_4	y_{18}/A_5	y_{22}/A_6
y_3/A_1	y_7/A_2	y_{11}/A_3	y_{15}/A_4	y_{19}/A_5	y_{23}/A_6
y_4/A_1	y_8/A_2	y_{12}/A_3	y_{16}/A_4	y_{20}/A_5	y_{24}/A_6

Step 3: form seasonal indices

Step 3: Now the seasonal indices are formed by computing the average of each row. Thus the initial seasonal indices (symbolically) are:

$$I_1 = (y_1/A_1 + y_5/A_2 + y_9/A_3 + y_{13}/A_4 + y_{17}/A_5 + y_{21}/A_6)/6$$
$$I_2 = (y_2/A_1 + y_6/A_2 + y_{10}/A_3 + y_{14}/A_4 + y_{18}/A_5 + y_{22}/A_6)/6$$
$$I_3 = (y_3/A_1 + y_7/A_2 + y_{11}/A_3 + y_{15}/A_4 + y_{19}/A_5 + y_{22}/A_6)/6$$
$$I_4 = (y_4/A_1 + y_8/A_2 + y_{12}/A_3 + y_{16}/A_4 + y_{20}/A_5 + y_{24}/A_6)/6$$

We now know the algebra behind the computation of the initial estimates.

The next page contains an example of triple exponential smoothing.

The case of the Zero Coefficients

Zero coefficients for trend and seasonality parameters

Sometimes it happens that a computer program for triple exponential smoothing outputs a final coefficient for trend (γ) or for seasonality (β) of zero. Or worse, both are outputted as zero!

Does this indicate that there is no trend and/or no seasonality?

Of course not! It only means that the initial values for trend and/or seasonality were right on the money. No updating was necessary in order to arrive at the lowest possible MSE. We should inspect the updating formulas to verify this.

Source: http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc43.htm