



Exponential smoothing

Exponential smoothing or **exponential moving average (EMA)** is a rule of thumb technique for smoothing time series data using the exponential window function. Whereas in the simple moving average the past observations are weighted equally, exponential functions are used to assign exponentially decreasing weights over time. It is an easily learned and easily applied procedure for making some determination based on prior assumptions by the user, such as seasonality. Exponential smoothing is often used for analysis of time-series data.

Exponential smoothing is one of many window functions commonly applied to smooth data in signal processing, acting as low-pass filters to remove high-frequency noise. This method is preceded by Poisson's use of recursive exponential window functions in convolutions from the 19th century, as well as Kolmogorov and Zurbenko's use of recursive moving averages from their studies of turbulence in the 1940s.

The raw data sequence is often represented by $\{x_t\}$ beginning at time $t = 0$, and the output of the exponential smoothing algorithm is commonly written as $\{s_t\}$, which may be regarded as a best estimate of what the next value of x will be. When the sequence of observations begins at time $t = 0$, the simplest form of exponential smoothing is given by the following formulas:^[1]

$$\begin{aligned}s_0 &= x_0 \\ s_t &= \alpha x_t + (1 - \alpha)s_{t-1}, \quad t > 0\end{aligned}$$

where α is the *smoothing factor*, and $0 < \alpha < 1$. If s_{t-1} is substituted into s_t continuously so that the formula of s_t is fully expressed in terms of $\{x_t\}$, then exponentially decaying weighting factors on each raw data x_t is revealed, showing how exponential smoothing is named.

The simple exponential smoothing is not able to predict what would be observed at $t + m$ based on the raw data up to t , while the double exponential smoothing and triple exponential smoothing can be used for the prediction due to the presence of b_t as the sequence of best estimates of the linear trend.

Basic (simple) exponential smoothing

The use of the exponential window function is first attributed to Poisson^[2] as an extension of a numerical analysis technique from the 17th century, and later adopted by the signal processing community in the 1940s. Here, exponential smoothing is the application of the exponential, or Poisson, window function. Exponential smoothing was suggested in the statistical literature without citation to previous work by Robert Goodell Brown in 1956,^[3] and expanded by Charles C. Holt in 1957.^[4] The formulation below, which is the one commonly used, is attributed to Brown and is known as "Brown's simple exponential smoothing".^[5] All the methods of Holt, Winters, and Brown may be seen as a simple application of recursive filtering, first found in the 1940s^[2] to convert finite impulse response (FIR) filters to infinite impulse response filters.

The simplest form of exponential smoothing is given by the formula:

$$s_t = \alpha x_t + (1 - \alpha)s_{t-1},$$

where α is the *smoothing factor*, with $0 \leq \alpha \leq 1$. In other words, the smoothed statistic s_t is a simple weighted average of the current observation x_t and the previous smoothed statistic s_{t-1} . Simple exponential smoothing is easily applied, and it produces a smoothed statistic as soon as two observations are available. The term *smoothing factor* applied to α here is something of a misnomer, as larger values of α actually reduce the level of smoothing, and in the limiting case with $\alpha = 1$ the smoothing output series is just the current observation. Values of α close to 1 have less of a smoothing effect and give greater weight to recent changes in the data, while values of α closer to 0 have a greater smoothing effect and are less responsive to recent changes. In the limiting case with $\alpha = 0$, the output series is just flat or a constant as the observation x_0 at the beginning of the smoothening process $t = 0$.

The method for choosing α must be decided by the modeler. Sometimes the statistician's judgment is used to choose an appropriate factor. Alternatively, a statistical technique may be used to *optimize* the value of α . For example, the method of least squares might be used to determine the value of α for which the sum of the quantities $(s_t - x_{t+1})^2$ is minimized.^[6]

Unlike some other smoothing methods, such as the simple moving average, this technique does not require any minimum number of observations to be made before it begins to produce results. In practice, however, a "good average" will not be achieved until several samples have been averaged together; for example, a constant signal will take approximately $3/\alpha$ stages to reach 95% of the actual value. To accurately reconstruct the original signal without information loss, all stages of the exponential moving average must also be available, because older samples decay in weight exponentially. This is in contrast to a simple moving average, in which some samples can be skipped without as much loss of information due to the constant weighting of samples within the average. If a known number of samples will be missed, one can adjust a weighted average for this as well, by giving equal weight to the new sample and all those to be skipped.

This simple form of exponential smoothing is also known as an exponentially weighted moving average (EWMA). Technically it can also be classified as an autoregressive integrated moving average (ARIMA) (0,1,1) model with no constant term.^[7]

Time constant

The time constant of an exponential moving average is the amount of time for the smoothed response of a unit step function to reach $1 - 1/e \approx 63.2\%$ of the original signal. The relationship between this time constant, τ , and the smoothing factor, α , is given by the following formula:

$$\alpha = 1 - e^{-\Delta T/\tau}, \text{ thus } \tau = -\frac{\Delta T}{\ln(1 - \alpha)}$$

where ΔT is the sampling time interval of the discrete time implementation. If the sampling time is fast compared to the time constant ($\Delta T \ll \tau$) then, by using the Taylor expansion of the exponential function,

$$\alpha \approx \frac{\Delta T}{\tau}, \text{ thus } \tau \approx \frac{\Delta T}{\alpha}$$

Choosing the initial smoothed value

Note that in the definition above, s_0 (the initial output of the exponential smoothing algorithm) is being initialized to x_0 (the initial raw data or observation). Because exponential smoothing requires that, at each stage, we have the previous forecast s_{t-1} , it is not obvious how to get the method started. We could assume that the initial forecast is equal to the initial value of demand; however, this approach has a serious drawback. Exponential smoothing puts substantial weight on past observations, so the initial value of demand will have an unreasonably large effect on early forecasts. This problem can be overcome by allowing the process to evolve for a reasonable number of periods (10 or more) and using the average of the demand during those periods as the initial forecast. There are many other ways of setting this initial value, but it is important to note that the smaller the value of α , the more sensitive your forecast will be on the selection of this initial smoother value s_0 .^{[8][9]}

Optimization

For every exponential smoothing method, we also need to choose the value for the smoothing parameters. For simple exponential smoothing, there is only one smoothing parameter (α), but for the methods that follow there are usually more than one smoothing parameter.

There are cases where the smoothing parameters may be chosen in a subjective manner – the forecaster specifies the value of the smoothing parameters based on previous experience. However, a more robust and objective way to obtain values of the unknown parameters included in any exponential smoothing method is to estimate them from the observed data.

The unknown parameters and the initial values for any exponential smoothing method can be estimated by minimizing the sum of squared errors (SSE). The errors are specified as $e_t = y_t - \hat{y}_{t|t-1}$ for $t = 1, \dots, T$ (the one-step-ahead within-sample forecast errors) where y_t and $\hat{y}_{t|t-1}$ are a variable to be predicted at t and a variable as the prediction result at t (based on the previous data or prediction), respectively. Hence, we find the values of the unknown parameters and the initial values that minimize

$$\text{SSE} = \sum_{t=1}^T (y_t - \hat{y}_{t|t-1})^2 = \sum_{t=1}^T e_t^2 \text{ [10]}$$

Unlike the regression case (where we have formulae to directly compute the regression coefficients which minimize the SSE) this involves a non-linear minimization problem, and we need to use an optimization tool to perform this.

"Exponential" naming

The name *exponential smoothing* is attributed to the use of the exponential function as the filter impulse response in the convolution.

By direct substitution of the defining equation for simple exponential smoothing back into itself we find that

$$\begin{aligned}
s_t &= \alpha x_t + (1 - \alpha)s_{t-1} \\
&= \alpha x_t + \alpha(1 - \alpha)x_{t-1} + (1 - \alpha)^2 s_{t-2} \\
&= \alpha [x_t + (1 - \alpha)x_{t-1} + (1 - \alpha)^2 x_{t-2} + (1 - \alpha)^3 x_{t-3} + \cdots + (1 - \alpha)^{t-1} x_1] + (1 - \alpha)^t x_0.
\end{aligned}$$

In other words, as time passes the smoothed statistic s_t becomes the weighted average of a greater and greater number of the past observations $s_{t-1}, \dots, s_{t-n}, \dots$, and the weights assigned to previous observations are proportional to the terms of the geometric progression

$$1, (1 - \alpha), (1 - \alpha)^2, \dots, (1 - \alpha)^n, \dots$$

A geometric progression is the discrete version of an exponential function, so this is where the name for this smoothing method originated according to Statistics lore.

Comparison with moving average

Exponential smoothing and moving average have similar defects of introducing a lag relative to the input data. While this can be corrected by shifting the result by half the window length for a symmetrical kernel, such as a moving average or gaussian. This approach is not possible for exponential smoothing since it is an IIR filter and therefore has an asymmetric kernel and frequency-dependent group delay. This means each constituent frequency is shifted by a different amount and therefore, there is no single number of samples that can be used to shift the output signal to account for the lag.

Both filters also both have roughly the same distribution of forecast error when $\alpha = 2/(k + 1)$ where k is the number of past data points in consideration of moving average. They differ in that exponential smoothing takes into account all past data, whereas moving average only takes into account k past data points. Computationally speaking, they also differ in that moving average requires that the past k data points, or the data point at lag $k + 1$ plus the most recent forecast value, to be kept, whereas exponential smoothing only needs the most recent forecast value to be kept.^[11]

In the signal processing literature, the use of non-causal (symmetric) filters is commonplace, and the exponential window function is broadly used in this fashion, but a different terminology is used: exponential smoothing is equivalent to a first-order infinite-impulse response (IIR) filter and moving average is equivalent to a finite impulse response filter with equal weighting factors.

Double exponential smoothing (Holt linear)

Simple exponential smoothing does not do well when there is a trend in the data.^[1] In such situations, several methods were devised under the name "double exponential smoothing" or "second-order exponential smoothing," which is the recursive application of an exponential filter twice, thus being termed "double exponential smoothing". 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 slope component is itself updated via exponential smoothing.

One method works as follows:^[12]

Again, the raw data sequence of observations is represented by x_t , beginning at time $t = 0$. We use s_t to represent the smoothed value for time t , and b_t is our best estimate of the trend at time t . The output of the algorithm is now written as F_{t+m} , an estimate of the value of x_{t+m} at time $m > 0$ based on the raw data up to time t . Double exponential smoothing is given by the formulas

$$\begin{aligned}s_0 &= x_0 \\ b_0 &= x_1 - x_0\end{aligned}$$

and for $t > 0$ by

$$\begin{aligned}s_t &= \alpha x_t + (1 - \alpha)(s_{t-1} + b_{t-1}) \\ b_t &= \beta(s_t - s_{t-1}) + (1 - \beta)b_{t-1}\end{aligned}$$

where α ($0 \leq \alpha \leq 1$) is the *data smoothing factor*, and β ($0 \leq \beta \leq 1$) is the *trend smoothing factor*.

To forecast beyond x_t is given by the following approximation:

$$F_{t+m} = s_t + m \cdot b_t.$$

Setting the initial value b is a matter of preference. An option other than the one listed above is $\frac{x_n - x_0}{n}$ for some n .

Note that F_0 is undefined (there is no estimation for time 0), and according to the definition $F_1 = s_0 + b_0$, which is well defined, thus further values can be evaluated.

A second method, referred to as either Brown's linear exponential smoothing (LES) or Brown's double exponential smoothing, has only one smoothing factor, α :^[13]

$$\begin{aligned}s'_0 &= x_0 \\ s''_0 &= x_0 \\ s'_t &= \alpha x_t + (1 - \alpha)s'_{t-1} \\ s''_t &= \alpha s'_t + (1 - \alpha)s''_{t-1} \\ F_{t+m} &= a_t + mb_t,\end{aligned}$$

where a_t , the estimated level at time t , and b_t , the estimated trend at time t , are given by

$$\begin{aligned}a_t &= 2s'_t - s''_t \\ b_t &= \frac{\alpha}{1 - \alpha}(s'_t - s''_t).\end{aligned}$$

Triple exponential smoothing (Holt–Winters)

Triple exponential smoothing applies exponential smoothing three times, which is commonly used when there are three high frequency signals to be removed from a time series under study. There are different types of seasonality: 'multiplicative' and 'additive' in nature, much like addition and multiplication are basic operations in mathematics.

If every month of December we sell 10,000 more apartments than we do in November the seasonality is *additive* in nature. However, if we sell 10% more apartments in the summer months than we do in the winter months the seasonality is *multiplicative* in nature. Multiplicative seasonality can be represented as a constant factor, not an absolute amount.^[14]

Triple exponential smoothing was first suggested by Holt's student, Peter Winters, in 1960 after reading a signal processing book from the 1940s on exponential smoothing.^[15] Holt's novel idea was to repeat filtering an odd number of times greater than 1 and less than 5, which was popular with scholars of previous eras.^[15] While recursive filtering had been used previously, it was applied twice and four times to coincide with the Hadamard conjecture, while triple application required more than double the operations of singular convolution. The use of a triple application is considered a rule of thumb technique, rather than one based on theoretical foundations and has often been over-emphasized by practitioners. Suppose we have a sequence of observations x_t , beginning at time $t = 0$ with a cycle of seasonal change of length L .

The method calculates a trend line for the data as well as seasonal indices that weight the values in the trend line based on where that time point falls in the cycle of length L .

Let s_t represent the smoothed value of the constant part for time t , b_t is the sequence of best estimates of the linear trend that are superimposed on the seasonal changes, and c_t is the sequence of seasonal correction factors. We wish to estimate c_t at every time $t \bmod L$ in the cycle that the observations take on. As a rule of thumb, a minimum of two full seasons (or $2L$ periods) of historical data is needed to initialize a set of seasonal factors.

The output of the algorithm is again written as F_{t+m} , an estimate of the value of x_{t+m} at time $t + m > 0$ based on the raw data up to time t . Triple exponential smoothing with multiplicative seasonality is given by the formulas^[1]

$$s_0 = x_0$$

$$s_t = \alpha \frac{x_t}{c_{t-L}} + (1 - \alpha)(s_{t-1} + b_{t-1})$$

$$b_t = \beta(s_t - s_{t-1}) + (1 - \beta)b_{t-1}$$

$$c_t = \gamma \frac{x_t}{s_t} + (1 - \gamma)c_{t-L}$$

$$F_{t+m} = (s_t + mb_t)c_{t-L+1+(m-1) \bmod L},$$

where α ($0 \leq \alpha \leq 1$) is the *data smoothing factor*, β ($0 \leq \beta \leq 1$) is the *trend smoothing factor*, and γ ($0 \leq \gamma \leq 1$) is the *seasonal change smoothing factor*.

The general formula for the initial trend estimate b is

$$b_0 = \frac{1}{L} \left(\frac{x_{L+1} - x_1}{L} + \frac{x_{L+2} - x_2}{L} + \cdots + \frac{x_{L+L} - x_L}{L} \right).$$

Setting the initial estimates for the seasonal indices c_i for $i = 1, 2, \dots, L$ is a bit more involved. If N is the number of complete cycles present in your data, then

$$c_i = \frac{1}{N} \sum_{j=1}^N \frac{x_{L(j-1)+i}}{A_j} \quad \text{for } i = 1, 2, \dots, L$$

where

$$A_j = \frac{\sum_{k=1}^L x_{L(j-1)+k}}{L} \quad \text{for } j = 1, 2, \dots, N.$$

Note that A_j is the average value of x in the j^{th} cycle of your data.

This results in

$$c_i = \frac{1}{N} \sum_{j=1}^N \frac{x_{L(j-1)+i}}{\frac{1}{L} \sum_{k=1}^L x_{L(j-1)+k}}$$

Triple exponential smoothing with additive seasonality is given by

$$\begin{aligned} s_0 &= x_0 \\ s_t &= \alpha(x_t - c_{t-L}) + (1 - \alpha)(s_{t-1} + b_{t-1}) \\ b_t &= \beta(s_t - s_{t-1}) + (1 - \beta)b_{t-1} \\ c_t &= \gamma(x_t - s_{t-1} - b_{t-1}) + (1 - \gamma)c_{t-L} \\ F_{t+m} &= s_t + mb_t + c_{t-L+1+(m-1) \bmod L}. \end{aligned}$$

Implementations in statistics packages

- R: the HoltWinters function in the stats package^[16] and ets function in the forecast package^[17] (a more complete implementation, generally resulting in a better performance^[18]).
- Python: the holtwinters module of the statsmodels package allow for simple, double and triple exponential smoothing.
- IBM SPSS includes Simple, Simple Seasonal, Holt's Linear Trend, Brown's Linear Trend, Damped Trend, Winters' Additive, and Winters' Multiplicative in the Time-Series modeling procedure within its Statistics and Modeler statistical packages. The default Expert Modeler feature evaluates all seven exponential smoothing models and ARIMA models with a range of nonseasonal and seasonal p , d , and q values, and selects the model with the lowest Bayesian Information Criterion statistic.
- Stata: tssmooth command^[19]
- LibreOffice 5.2^[20]
- Microsoft Excel 2016^[21]

See also

- Autoregressive moving average model (ARMA)
- Errors and residuals in statistics
- Moving average

Notes

1. "NIST/SEMATECH e-Handbook of Statistical Methods" (<http://www.itl.nist.gov/div898/handbook/>). NIST. Retrieved 23 May 2010.
2. Oppenheim, Alan V.; Schaffer, Ronald W. (1975). *Digital Signal Processing*. Prentice Hall. p. 5. ISBN 0-13-214635-5.
3. Brown, Robert G. (1956). *Exponential Smoothing for Predicting Demand* (<http://legacy.library.ucsf.edu/tid/dae94e00;jsessionid=104A0CEFFA31ADC2FA5E0558F69B3E1D.tobacco03>). Cambridge, Massachusetts: Arthur D. Little Inc. p. 15.
4. Holt, Charles C. (1957). "Forecasting Trends and Seasonal by Exponentially Weighted Averages". *Office of Naval Research Memorandum*. **52**. reprinted in Holt, Charles C. (January–March 2004). "Forecasting Trends and Seasonal by Exponentially Weighted Averages". *International Journal of Forecasting*. **20** (1): 5–10. doi:10.1016/j.ijforecast.2003.09.015 (<https://doi.org/10.1016%2Fj.ijforecast.2003.09.015>).
5. Brown, Robert Goodell (1963). *Smoothing Forecasting and Prediction of Discrete Time Series* (<http://babel.hathitrust.org/cgi/pt?id=mdp.39015004514728;view=1up;seq=1>). Englewood Cliffs, NJ: Prentice-Hall.
6. "NIST/SEMATECH e-Handbook of Statistical Methods, 6.4.3.1. Single Exponential Smoothing" (<http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc431.htm>). NIST. Retrieved 5 July 2017.
7. Nau, Robert. "Averaging and Exponential Smoothing Models" (<http://www.duke.edu/~rnau/411avg.htm>). Retrieved 26 July 2010.
8. "Production and Operations Analysis" Nahmias. 2009.
9. Čisar, P., & Čisar, S. M. (2011). "Optimization methods of EWMA statistics." *Acta Polytechnica Hungarica*, 8(5), 73–87. Page 78.
10. 7.1 Simple exponential smoothing | *Forecasting: Principles and Practice* (<https://www.otexts.org/fpp/7/1>).
11. Nahmias, Steven; Olsen, Tava Lennon. *Production and Operations Analysis* (7th ed.). Waveland Press. p. 53 (<https://books.google.com/books?id=SIsoBgAAQBAJ&pg=PA53>). ISBN 9781478628248.
12. "6.4.3.3. Double Exponential Smoothing" (<http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc433.htm>). *itl.nist.gov*. Retrieved 25 September 2011.
13. "Averaging and Exponential Smoothing Models" (<http://www.duke.edu/~rnau/411avg.htm>). *duke.edu*. Retrieved 25 September 2011.
14. Kalehar, Prajakta S. "Time series Forecasting using Holt–Winters Exponential Smoothing" (http://www.itb.ac.in/~praj/acads/seminar/04329008_ExponentialSmoothing.pdf) (PDF). Retrieved 23 June 2014.
15. Winters, P. R. (April 1960). "Forecasting Sales by Exponentially Weighted Moving Averages". *Management Science*. **6** (3): 324–342. doi:10.1287/mnsc.6.3.324 (<https://doi.org/10.1287%2Fmnsc.6.3.324>).
16. "R: Holt–Winters Filtering" (<https://stat.ethz.ch/R-manual/R-patched/library/stats/html/HoltWinters.html>). *stat.ethz.ch*. Retrieved 5 June 2016.
17. "ets {forecast} | inside-R | A Community Site for R" (<https://web.archive.org/web/20160716153135/http://www.inside-r.org/packages/cran/forecast/docs/ets>). *inside-r.org*. Archived from the original (<http://www.inside-r.org/packages/cran/forecast/docs/ets>) on 16 July 2016. Retrieved 5 June 2016.
18. "Comparing HoltWinters() and ets()" (<http://robjhyndman.com/hyndsight/estimation2/>). *Hyndsight*. 29 May 2011. Retrieved 5 June 2016.
19. tssmooth (<https://www.stata.com/help.cgi?tssmooth>) in Stata manual
20. "LibreOffice 5.2: Release Notes – the Document Foundation Wiki" (https://wiki.documentfoundation.org/ReleaseNotes/5.2#New_spreadsheet_functions).

21. "Excel 2016 Forecasting Functions | Real Statistics Using Excel" (<http://www.real-statistics.com/time-series-analysis/basic-time-series-forecasting/excel-2016-forecasting-functions/>).

External links

- [Lecture notes on exponential smoothing \(Robert Nau, Duke University\) \(http://people.duke.edu/~rna/411avg.htm\)](http://people.duke.edu/~rna/411avg.htm)
- [Data Smoothing \(http://demonstrations.wolfram.com/DataSmoothing/\)](http://demonstrations.wolfram.com/DataSmoothing/) by Jon McLoone, [The Wolfram Demonstrations Project](#)
- [The Holt–Winters Approach to Exponential Smoothing: 50 Years Old and Going Strong \(https://foresight.forecasters.org/product/foresight-issue-19/\)](https://foresight.forecasters.org/product/foresight-issue-19/) by Paul Goodwin (2010) [Foresight: The International Journal of Applied Forecasting](#)
- [Algorithms for Unevenly Spaced Time Series: Moving Averages and Other Rolling Operators \(https://web.archive.org/web/20150623211648/http://www.eckner.com/papers/ts_alg.pdf\)](https://web.archive.org/web/20150623211648/http://www.eckner.com/papers/ts_alg.pdf) by Andreas Eckner

Retrieved from "https://en.wikipedia.org/w/index.php?title=Exponential_smoothing&oldid=1304582081"