Exponential smoothing

From Wikipedia, the free encyclopedia

Exponential smoothing is a rule of thumb technique for smoothing time series data. 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 used for analysis of financial time-series data as well as the field of signal processing.

Exponential smoothing is commonly applied to smooth data, as many window functions are in signal processing, acting as low-pass filters to remove high frequency noise. This method is proceeded 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 formulas:[1]

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

where α is the *smoothing factor*, and $0 < \alpha < 1$.

Contents

Background

Window functions

The simple moving average (SMA)

Intuitively, the simplest way to smooth a time series is to calculate a simple, or unweighted, moving average. This is known as using a rectangular or "boxcar" window function. The smoothed statistic s_t is then just the mean of the last k observations:

$$s_t = \frac{1}{k} \sum_{n=0}^{k-1} x_{t-n} = \frac{x_t + x_{t-1} + x_{t-2} + \cdots + x_{t-k+1}}{k} = s_{t-1} + \frac{x_t - x_{t-k}}{k},$$

where the choice of an integer $k > 1$ is arbitrary. A small value of k will have less of a smoothing effect and be more responsive to recent changes in the data, while a larger k will have a greater smoothing effect, and produce a more pronounced lag in the smoothed sequence. One disadvantage of this technique is that it cannot be used on the first $k-1$ terms of the time series without the addition of values created by some other means. This means effectively extrapolating outside the existing data, and the validity of this section would therefore be questionable and not a direct representation of the data. However, as long as the time series contains at least k values, this has no effect on forecasts of future values.

It also introduces a phase shift into the data of half the window length. For example, if the data were all the same except for one high data point, the peak in the "smoothed" data would appear half a window length later than when it actually occurred. Where the phase of the result is important, this can be simply corrected by shifting the resulting series back by half the window length.

A major drawback of the SMA is that it lets through a significant amount of the signal shorter than the window length. Worse, it actually inverts it. This can lead to unexpected artifacts, such as peaks in the "smoothed" result appearing where there were troughs in the data. It also leads to the result being less "smooth" than expected since some of the higher frequencies are not properly removed.

The weighted moving average

A slightly more intricate method for smoothing a raw time series $\{x_t\}$ is to calculate a weighted moving average by first choosing a set of weighting factors

$$\{w_1, w_2, \ldots, w_k\} \text{ such that } \sum_{n=1}^{k} w_n = 1$$

and then using these weights to calculate the smoothed statistics $\{s_t\}$:

$$s_t = \sum_{n=1}^{k} w_n x_{t+1-n} = w_1 x_t + w_2 x_{t-1} + \cdots + w_k x_{t-k+1}.$$

In practice the weighting factors are often chosen to give more weight to the most recent terms in the time series and less weight to older data. Notice that this technique has the same disadvantage as the simple moving average technique (i.e., it cannot be used until at least k observations have been made), and that it entails a more complicated calculation at each step of the smoothing procedure. In addition to this disadvantage, if the data from each stage of the averaging is not available for analysis, it may be difficult if not impossible to reconstruct a changing signal accurately (because older samples may be given less weight). If the number of stages missed is known however, the weighting of values in the average can be adjusted to give equal weight to all missed samples to avoid this issue.

Basic 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 first suggested in the statistical literature without citation to previous work by Robert Goodell Brown in 1956,[3] and then 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 FIR filters to IIR filters.

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

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

where α is the *smoothing factor*, and 0 < α < 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}. 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 α = 1 the output series is just the current observation. Simple exponential smoothing is easily applied, and it produces a smoothed statistic as soon as two observations are available.

Values of α close to one have less of a smoothing effect and give greater weight to recent changes in the data, while values of α closer to zero have a greater smoothing effect and are less responsive to recent changes. There is no formally correct procedure for choosing α. 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/α* 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 set 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 formula:

$$\alpha = 1 - e^{\frac{-\Delta T}{\tau}}$$

Where ΔT is the sampling time interval of the discrete time implementation. If the sampling time is fast compared to the time constant then

$$\alpha \approx \frac{\Delta T}{\tau}$$

Choosing the initial smoothed value

Note that in the above definition s_1 is being initialized to x_0. Because exponential smoothing requires that at each stage we have the previous forecast, 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_1.[8]

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 is 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 for 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 SSE. The errors are specified as $e_t = y_t - \hat{y}_{t|t-1}$ for t=1,...,T (the one-step-ahead within-sample forecast errors). Hence we find the values of the unknown parameters and the initial values that minimize

$$SSE = \sum_{t=1}^{T}(y_t - \hat{y}_{t|t-1})^2 = \sum_{t=1}^{T} e_t^2 \ ^{[9]}$$

Unlike the regression case (where we have formulas that return the values of 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.

Why is it "exponential"?

The name 'exponential smoothing' is attributed to the use of the exponential window function during convolution. It is no longer attributed to Holt, Winters & Brown.

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

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

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 x_{t-n}, and the weights assigned to previous observations are in general proportional to the terms of the geometric progression $\{1, (1-\alpha), (1-\alpha)^2, (1-\alpha)^3, ...\}$. 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, it is unclear how appropriate this would be for exponential smoothing. They also both have roughly the same distribution of forecast error when $\alpha = 2/(k+1)$. 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.[10]

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 or IIR filter and moving average is equivalent to a Finite Impulse Response or FIR filter with equal weighting factors.

Double exponential smoothing

Simple exponential smoothing does not do well when there is a trend in the data, which is inconvenient.[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". This nomenclature is similar to quadruple exponential smoothing, which also references its recursion depth.[11] 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, sometimes referred to as "Holt-Winters double exponential smoothing"[12] works as follows:[13]

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 at time $t+m$, $m>0$ based on the raw data up to time t. Double exponential smoothing is given by the formulas

$$s_1 = x_1$$
$$b_1 = x_1 - x_0$$

And for t > 2 by

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

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

To forecast beyond x_t

$$F_{t+m} = s_t + mb_t$$

Setting the initial value b_0 is a matter of preference. An option other than the one listed above is $(x_n - x_0)/n$ for some $n > 1$.

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 works as follows.[14]

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

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

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

Triple exponential smoothing

Triple exponential smoothing takes into account seasonal changes as well as trends (all of which are trends). Seasonality is defined to be the tendency of time-series data to exhibit behavior that repeats itself every L periods, much like any periodic function. The term season is used to represent the period of time before behavior begins to repeat itself. 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. Can be represented by an 'absolute' increase. 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. [15]

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.[16] Holt's novel idea was to repeat filtering an odd number of times (ignoring 1) . 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.

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.

$\{s_t\}$ represents the smoothed value of the constant part for time t. $\{b_t\}$ represents the sequence of best estimates of the linear trend that are superimposed on the seasonal changes. $\{c_t\}$ is the sequence of seasonal correction factors. c_t is the expected proportion of the predicted trend at any 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 at time $t+m$, $m>0$ 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 is the *data smoothing factor*, $0 < \alpha < 1$, β is the *trend smoothing factor*, $0 < \beta < 1$, and γ is the *seasonal change smoothing factor*, $0 < \gamma < 1$.

The general formula for the initial trend estimate b_0 is:

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

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

$$c_i = \frac{1}{N}\sum_{j=1}^{N} \frac{x_{L(j-1)+i}}{A_j} \quad \forall i \qquad = 1, 2, \ldots, L$$

where

$$A_j = \frac{\sum_{i=1}^{L} x_{L(j-1)+i}}{L} \quad \forall j \qquad = 1, 2, \ldots, N$$

Note that A_j is the average value of x in the jth cycle of your data.

Triple exponential smoothing with additive seasonality is given by:

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

Implementations in statistics packages

- R: the HoltWinters function in the stats package[17] and ets function in the forecast package[18] (a more complete implementation, generally resulting in a better performance[19]).
- 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.
- LibreOffice 5.2[20]
- Microsoft Excel 2016[21]

See also

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

Notes

1. "NIST/SEMATECH e-Handbook of Statistical Methods" (http://www.itl.nist.gov/div898/handbook/). NIST. Retrieved 2010-05-23.
2. Oppenheim, Alan V.; Schafer, 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" (http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V92-4BJVV07-3&_user=1535420&_coverDate=03%2F31%2F2004&_rdoc=1&_fmt=high&_orig=gateway&_origin=gateway&_sort=d&_docanchor=&view=c&_searchStrId=1755927165&_rerunOrigin=google&_acct=C000053610&_version=1&_urlVersion=0&_userid=1535420&md5=0e881c06b0512cbf976a92268d48edc4&searchtype=a). *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 2017-07-05.
7. "Averaging and Exponential Smoothing Models" (http://www.duke.edu/~rnau/411avg.htm). Retrieved 26 July 2010.
8. "Production and Operations Analysis" Nahmias. 2009.
9. https://www.otexts.org/fpp/7/1
10. Nahmias, Steven. *Production and Operations Analysis* (6th ed.). ISBN 0-07-337785-6.
11. "Model: Second-Order Exponential Smoothing" (http://help.sap.com/saphelp_45b/helpdata/en/7d/c27a14454011d182b40000e829fbfe/content.htm). SAP AG. Retrieved 23 January 2013.
12. Prajakta S. Kalekar. "Time series Forecasting using Holt-Winters Exponential Smoothing" (https://labs.omniti.com/people/jesus/papers/holtwinters.pdf) (PDF).
13. "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.
14. "Averaging and Exponential Smoothing Models" (http://www.duke.edu/~rnau/411avg.htm). *duke.edu*. Retrieved 25 September 2011.
15. Kalehar, Prajakta S. "Time series Forecasting using Holt-Winters Exponential Smoothing" (http://www.it.iitb.ac.in/~praj/acads/seminar/04329008_ExponentialSmoothing.pdf) (PDF). Retrieved 23 June 2014.
16. Winters, P. R. (April 1960). "Forecasting Sales by Exponentially Weighted Moving Averages" (http://mansci.journal.informs.org/cgi/content/abstract/6/3/324). *Management Science*. **6** (3): 324–342. doi:10.1287/mnsc.6.3.324 (https://doi.org/10.1287%2Fmnsc.6.3.324).
17. "R: Holt-Winters Filtering" (https://stat.ethz.ch/R-manual/R-patched/library/stats/html/HoltWinters.html). *stat.ethz.ch*. Retrieved 2016-06-05.
18. "ets {forecast} | inside-R | A Community Site for R" (http://www.inside-r.org/packages/cran/forecast/docs/ets). *www.inside-r.org*. Retrieved 2016-06-05.
19. "Comparing HoltWinters() and ets()" (http://robjhyndman.com/hyndsight/estimation2/). *Hyndsight*. 2011-05-29. Retrieved 2016-06-05.
20. https://wiki.documentfoundation.org/ReleaseNotes/5.2#New_spreadsheet_functions
21. 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)
- Data Smoothing by Jon McLoone, The Wolfram Demonstrations Project
- The Holt-Winters Approach to Exponential Smoothing: 50 Years Old and Going Strong by Paul Goodwin (2010) Foresight: The International Journal of Applied Forecasting
- Algorithms for Unevenly Spaced Time Series: Moving Averages and Other Rolling Operators by Andreas Eckner