. Ehlers

. SMA

From [1], pp xxx-xxx.

An n-day SMA is formed by adding the samples over n days and dividing by n. Another way to view an SMA is as average of the data within an n-sample moving window. The average is usially plotted at the right-hand side of the window which causes a lag.

The average price is always the price at the center of the window expressed at $\frac{n-1}{2}$ **[WHY?]**.

If we increase SMA window to incude full-period of sinusoid, the SMA value will be always equal to zero because there are as many points above the mean as there are below it. **[this is continuous case. sampling?]** The same happens when SMA window includes an integer number of full-periods of sinusoid.
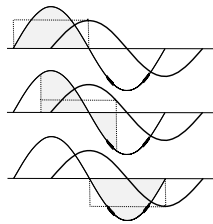

Figure . *svg*

SMA coefficients form rectangle across the width of the filter, resulting a center of gravity being 1/2 across the window.

. Another view

A simple moving average of length n is just the mean of the last n observations of x_1, x_2, x_3, ... , x_k, where x_k is the most recent value:

$$m_k = \frac{1}{n} \sum_{i=k-n+1}^{k} x_i.$$

In the beginning, when $k < n$, this is a cumulative moving average:

$$m_k = \frac{1}{k} \sum_{i=1}^{k} x_i.$$

Actually the last equation is the particular case of the first one for $k = n$. The recurrence for the first equation is

$$m_k = \frac{1}{n}\left[x_k + \sum_{i=k-n}^{k-1} x_i - x_{k-n}\right] = m_{k-1} + \frac{x_k - x_{k-n}}{n}.$$

Similarly for the second equation

$$m_k = \frac{1}{k-1}\left[x_k + \sum_{i=1}^{k-1} x_i\right]\frac{k-1}{k} = m_{k-1} + \frac{x_k - m_{k-1}}{k}.$$

Note that both can be rewritten as

$$m_k = m_{k-1} + \frac{\delta_k - \delta_{k-n}}{N_k}. \qquad ()$$

where

$$\begin{cases} \delta_l = x_l - m_{k-1} & l > 0 \\ \delta_l = 0 & l \le 0 \end{cases} \qquad ()$$

and

$$\begin{cases} N_k = n & , k \ge n \\ N_k = k & , k < n \end{cases} \qquad ()$$

. EMA

The exponential smoothing is defined by the following recurrent relation

$$\begin{cases} m_1 = x_1 & k = 1 \\ m_k = \alpha x_k + (1-\alpha)m_{k-1} = \alpha x_k + \lambda m_{k-1} & k > 1 \end{cases}$$

where $\lambda = 1 - \alpha$. This recurrence also defines a weighted average

$$m_k = \sum_{i=1}^{k} w_{k,i} x_i$$

with the following weights

$$w_{k,i} = \begin{cases} \lambda^k & i = 1 \\ \alpha\lambda^{k-i} & i > 1 \end{cases}$$

The weights decrease exponentially with the decay factor $\nu = -ln(\lambda)$, which gave the name for this smoothing:

$$\begin{cases} \dfrac{w_{k,i}}{w_{k,i+1}} = \lambda = e^{-\nu} & i = k-1, \ldots, 2 \\ \dfrac{w_{k,1}}{w_{k,2}} = \dfrac{\lambda}{\alpha} & i = 1 \end{cases} \qquad ()$$

The weight $w_{k,1}$ in the previous equation does not follow the same exponential rule. For $\lambda > \frac{1}{2}$ it is even not the smallest one. The pure exponential average can be defined as

$$m_k = \frac{S_k}{N_k}$$

where

$$S_k = \sum_{i=1}^{k} \lambda^{k-i} x_i$$

$$N_k = \sum_{i=1}^{k} \lambda^{k-i}$$

The above equations can be written as recurrent relations:

$$\begin{aligned} S_k &= x_k + \lambda S_{k-1} \\ N_k &= 1 + \lambda N_{k-1} \end{aligned} \tag{}$$

Using these equations, it is easy to derive a recurrence for the mean value

$$\begin{aligned} m_k &= \frac{x_k + \lambda S_{k-1}}{N_k} \\ &= \frac{x_k}{N_k} + \frac{\lambda N_{k-1}}{N_k} m_{k-1} \\ &= m_{k-1} + \frac{x_k}{N_k} + \left(\frac{\lambda N_{k-1}}{N_k} - 1 \right) m_{k-1} \\ &= m_{k-1} + \frac{x_k}{N_k} - \frac{m_{k-1}}{N_k} \end{aligned}$$

Finally,

$$m_k = m_{k-1} + \frac{\delta_k}{N_k} \tag{}$$

where

$$\delta_k = x_k - m_{k-1} \tag{}$$

The equations ??? and ??? are remarkably similar to ??? - ???. The smoothing length is well defined in the simple average. It is the norm N_k ???. The norm ??? is the generalization of ??? and defines the effective smoothing length for the exponential average:

$$N_k = \frac{1 - \lambda^k}{1 - \lambda} = \frac{1 - \lambda^k}{\alpha}$$

This norm accounts for 100 % of all weights. At sufficiently large k

$$\frac{1}{N_k} \to \frac{1}{N_\infty} = \alpha$$

and $\frac{\lambda^{k-i}}{N_k} \approx w_{k,i}$???. Therefore both definitions of exponential smoothing coincide for large k. Actually the difference between the two definitions tends to zero as λ^k.

The definition ??? seems to be natural, however historically the smoothing period P for the exponential smoothing ??? is defined as

$$\alpha = \frac{2}{P+1}$$

These two definitions related as

$$N_k = \frac{P+1}{2}\left[1 - \left(\frac{P-1}{P+1}\right)^k\right] \approx \frac{P+1}{2}$$

In RiskMetrics [3] the effective averaging length L is defined as

$$\frac{N_L}{N_\infty} = 0.999 = 1 - \epsilon$$

Therefore

$$1 - \lambda^L = 1 - \epsilon$$
$$\lambda^L = \epsilon$$

or

$$\lambda = \epsilon^{\frac{1}{L}}$$

This length is related to the natural length as

$$L = \frac{ln(\epsilon)}{ln(\lambda)} = \frac{ln(\epsilon)}{ln(1 - \frac{1}{N_\infty})}$$

For $N_\infty \gg 1$

$$L \approx 6.9\, N_\infty$$

In particular, for $\lambda = 0.94$ [3], the above definitions give the following values: $L = 112, P = 33, N_\infty = 17$.

A couple of specific cases are also worth mentioning. For $N_\infty = P = 1\ \lambda = 0$ - no averaging; and for $L = 1\ \lambda = 0.001$.

The values of all definitions for selected λ are collected in the table below.

λ	L	P	N_∞	Comment
0	0	1	1	No averaging
0.001	1	1.002	1.001	
0.5	10	3	2	

λ	**L**	**P**	N_∞	**Comment**
0.75	24	7	4	
0.875	52	15	8	
0.94	112	33	17	RiskMetrix

Table .

. WMA

From [1], pp xxx-xxx.

Close to SMA (xxx), but multiplier coefficients are not constant across the window width but are linearly weighted.

$$WMA_N = \frac{\sum_{i=0}^{N-1}(N-i)P_i}{\sum_{i=0}^{N-1}(N-i)}$$

The sum of the data and coefficient products are divided by the sum of coefficients to normalize the averaging process.

The major advantage of WMA is the reduced lag which results from the most recent data being the most heavily weighted. WMA coefficients form triangle across the width of the filter, resulting a center of gravity being 1/3 across the window.

. ZEMA

Was introduced by John Ehlers in [1], pp 167-175.

Zema is a suboptimal "almost" zero-lag filter based on a simplification of a Kalman filter with a constant gain. The idea of Zema comes from the tracking filters that use a liner model to estimate the position of a military target.

Imagine a gunner shooting at an enemy target. He estimates the angle of his gun and shoots. The forward shoot soldiers radio back how much deviation there was from the target. The gunner computes the incremental change required for his new gun angle from the deviation.

This model constitutes using the previous estimate plus a constant times the difference between the last real position and the last estimate

$$x_{estim} = x_{estim}[1] + \alpha(x_{real} - x_{estim}).$$

We can improve the estimate of position by adding an estimate of the velocity to the last known position:

$$x_{estim} = x_{estim} + \alpha(x_{real} + g*v - x_{estim}[1])$$

where v is an estimate of velocity, g is a gain factor. The velocity estimate can be an EMA of the rate of change of position,

$$v_{estim} = v_{estim}[1] + \beta(v - v_{estim}[1]).$$

The equation ??? looks like an EMA:

$$ema(x) = \alpha x + (1 - \alpha)ema(x)[1] = ema(x)[1] + \alpha(x - ema(x)[1]).$$

Considering a special case $\beta = 1$, Ehlers uses a momentum to approximate the velocity

. Hilbert transformer

From [1], pp xxx-xxx.

When data are sampled at a sample frequency f_s, that sampling frequency acts like a radio carrier signal. That is, the real data being sampled and heterodyned into upper and lower sidebands of the sampling frequency. Mathematically, heterodyning is multiplying two frequencies.

So, if we have a baseband data frequency of f_b, the phase of two signals is

$$\frac{1}{2}cos(2\pi f_s t)cos(2\pi f_b t) = cos(2\pi(f_s - f_b)t) + cos(2\pi(f_s + f_b)t)$$

The lower sideband can be considered as a negative frequency relative to f_s, and the upper sideband can be considered as a positive frequency. Furthermore, every harmonic of f_s exists having its own lower and upper sidebands containing the base signal.

The sample data spectrum can be pictured as

[svg]

The baseband range of frequencies is limited to $f_s/2$ following the Nyquist sampling criteria. As we are talking over complex functions, the sampled spectrum can extend below zero.

An interesting observation is that that either the upper or lower sideband of any harmonic of f_s can be processed with exactly the same result because the same information resides in all sidebands. Therefore, the frequency selection for processing is a matter of convenience and is usually the baseband because demodulation of the zero frequency harmonic is not required.

Analytic signals can be defined as a special case of a complex function without imaginary values that have only positive or negative frequencies, but not both. Analytic signal can be split into the sum of two complex signals that are odd and even functions around zero. These two functions may be sin and cos. Recall, $cos(wt) = cos(-wt)$ and $sin(wt) = -sin(-wt)$. Also, by Euler's equations $cos(wt) = \frac{1}{2}(e^{jwt} + e^{-jwt})$ and $sin(wt) = \frac{1}{2j}(e^{jwt} - e^{-jwt})$.

In DSP parlance, an analytical signal is a sum of inPhase and Quadrature components. InPhase and Quadrature components can be created from the analytical signal with Hilbert transform. It can be shown, [HOW?] that the Hilbert transform

$$H(e^{jw}) = \sum_{n=-\infty}^{\infty} C_n e^{jwn}, C_n = \frac{sin^2(\frac{\pi}{2}n)}{\frac{\pi}{2}n}, n \neq 0, C_0 = 0.$$

From this formula, Ehlers concludes that $C_n = \frac{1}{n}$ for odd values of n and $C_n > 0$ for $n > 0, C_n < 0$ for $n < 0$.

The $\frac{\pi}{2}$ factor in C_n can be ignored because each of C_n coefficient is divided by the sum of C_n to produce a normalized amplitude response.

Truncating the filter at $n = 7$,

$$Q = \frac{\frac{1}{7}P + \frac{1}{5}P[2] + \frac{1}{3}P[4] + P[6] - P[8] - \frac{1}{3}P[10] - \frac{1}{5}P[12] - \frac{1}{7}P[14]}{1 + \frac{1}{3} + \frac{1}{5} + \frac{1}{7}}, I = P[7]$$

where inPhase is a value at the center of the filter. The lag of the filter is $14/2 = 7$ bars.

Truncating at $n = 3$, we get

$$Q = \frac{\frac{1}{3}P + P[2] - P[4] - \frac{1}{3}P[6]}{1 + \frac{1}{3}} = 0.25P + 0.75P[2] - 0.75P[4] - 0.25P[6], I = P[3].$$

The lag is $6/2 = 3$ bars.

But this truncation produces not so good amplitude response. Then Ehlers says that the response can be improved adjusting filter coefficients "by trial and error", which results in

$$Q = 0.0962P + 0.5769P[2] - 0.5769P[4] - 0.0962P[6], I = P[3].$$

[coefficients are not ~ 1/n, is this a Hilbert transform then?]

. References

[1]   Ehlers, John F. (2001). *Rocket Science for Traders: Digital Signal Processing Applications*. Wiley, 2001. ISBN:9780471405672 online offline

[2]   Ehlers, John F. (2004). *Cybernetic Analysis for Stocks and Futures: Cutting-Edge DSP Technology to Improve Your Trading*. Wiley, 2004. ISBN:9780471463078 online offline

[3]   Ehlers, John F. (2013). *Cycle Analytics for Traders + Downloadable Software: Advanced Technical Trading Concepts*. John Wiley & Sons, 2013. ISBN:9781118728604 online offline

[4]   Ehlers, John F. (1992). *MESA and Trading Market Cycles*. Wiley, 1992. ISBN:9780471549437 online offline

[5]   Ehlers, John F. and Way, Rick (2010, November). *Zero Lag (Well, Almost)*. Technical Analysis of Stocks & Commodities, vol. 28 (11), pp. 30-35. ISSN:0738-3355 online offline