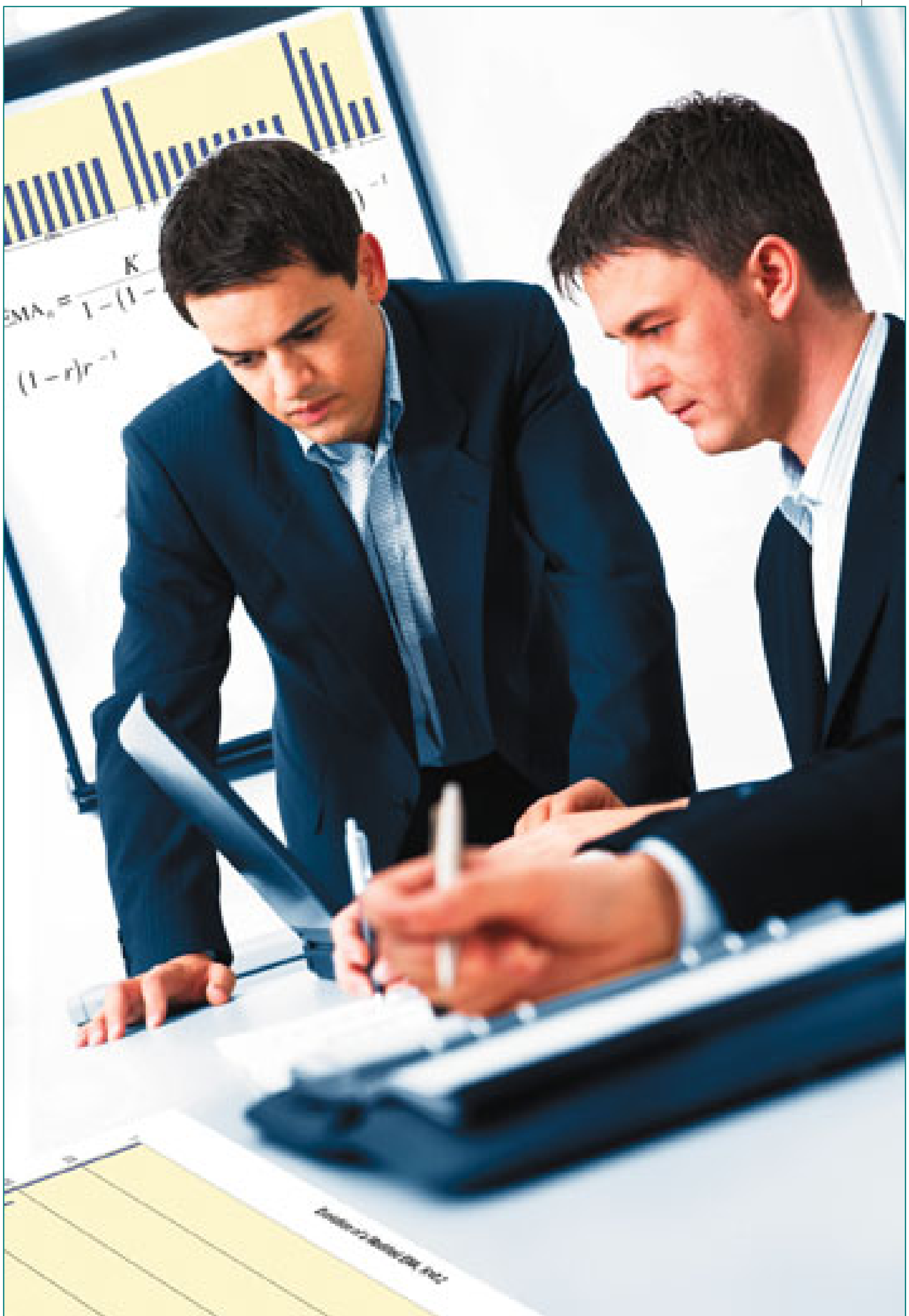
INDICATORS

A Closer Look At A Classic

Examining The EMA

You know how to apply the exponential moving average, but what do you know about its development and behavior? Here's a deeper look at the popular moving average.

by Tim Treloar



The popular exponential moving average (EMA) depends on data to develop its distinctive characteristics, and when it comes to data, more is better. But before you put an EMA to work as a technical indicator, a deeper understanding of its development and behavior is advised. At the very least, a deeper understanding will justify the practical application of it and at most dispel some misconceptions associated with it.

IN THEORY

An EMA is essentially a type of weighted average. The idea of a weighted average is simple: each element in a collection of numerical data is multiplied by a number ranging from zero to 1 (the decimal equivalent of the percentile range zero to 100), called a *weight*, such that the sum of the weights always totals 1, and the weighted elements are then added together.

A simple moving average (MA) is perhaps the simplest example of a weighted average in that all the elements receive exactly the same weight. In an exponential moving average, the weight of each element decreases progressively, usually according to its age and usually by powers of a particular factor. This is done under the premise that recent data is more relevant than older data. The smaller this factor, the faster data devalues, heavily favoring the most recent data — in other words, a short-term EMA. The larger the factor, the slower data devalues, distributing favor more equitably over a longer range of data — a long-term EMA. At some point, the value of old data becomes so small that it can be effectively negligible.

Because the sum of the weights in a collection of data always totals 1, weight must be drawn in some manner from among the other weighted elements in order to make weight available for new additional data. Where in the collection of elements this weight is drawn from has a pronounced effect on the short-term evolution of the weight distribution, but in the long run the weight distribution of all EMAs that use the same parameters will become virtually identical. When represented graphically with bars, the characteristic profile of an EMA weight distribution is a stepped, gradually tapering curve, like steps descending into the base of an escalator.

THE TRADITIONAL EMA

The EMA in its most familiar form is defined recursively as follows:

$$\text{EMA}_i = P_i K + \text{EMA}_{i-1}(1-K) \qquad (1)$$

where:
K = Leading weight value
P_i = Data element value
EMA_i = Current EMA
EMA_{i-1} = Previous EMA
i = Index by chronological order, i = 1, 2, 3,...

Formulas such as equation (1) are known as *recursive* because the resulting output is plugged back into the formula along with the next piece of data to determine the next output. The process can be repeated indefinitely in a recurring pattern. The net result of applying the formula for a prescribed number of pieces of data can be expressed as a single equation:

$$\text{EMA}_n = \text{EMA}_0(1-K)^n + K\sum_{i=1}^{n} P_{n-i+1}(1-K)^{i-1}$$

where n represents the total number of data, excluding the initializing EMA_0 data. The exponents that appear in this form of the formula are the reason for the "exponential" in the EMA's name. All data devalues according to powers of the factor $1 - K$. It may be helpful to think of this factor as a retention factor, as it represents how much value is retained from day to day. The weight of any given piece of data in the i^{th} position in the distribution is given by the expression

$$K(1-K)^{i-1}$$

To better understand how K relates to the perceived length of the EMA, let $i = 1$ to initialize the EMA and let $K = 0.2$ be an arbitrarily selected leading weight. Accordingly, $1 - K = 0.8$ represents the remaining weight:

$$\text{EMA}_1 = 0.2P_1 + 0.8\,\text{EMA}_0$$

The sum of the two weights equals 1 as expected. But EMA_0 represents older data than P_1 and, thus, in exponentially decreasing fashion, its elements should be weighted more lightly. The fact that, in this example, the EMA_0 weight is considerably larger than the P_1 weight suggests that it consists of more than one piece of data.

To address such situations, EMA_0 is typically assigned a simple moving average of sufficient length so that the weight of any one of its elements is less than the leading weight. To determine the number of days in the MA necessary to achieve this, the EMA_0 weight is divided according to the inequality:

$$\frac{1-K}{N} < K, \qquad (2)$$

where N is the number of days in the moving average. Solving for N produces

$$N > \frac{1-K}{K}$$

Using the values in the example above,

$$N > \frac{1-K}{K} = \frac{0.8}{0.2} = 4$$

So the value of N must be at least 5. In order for older data to have less weight than the leading weight of 0.2, the EMA must have at least five days of older data. Since moving averages are usually defined by a particular number of days, the EMA can be defined in terms of N days by solving inequality (2) for K.

$$K > \frac{1}{N+1}$$

Two very simple expressions for K that satisfy this inequality are:

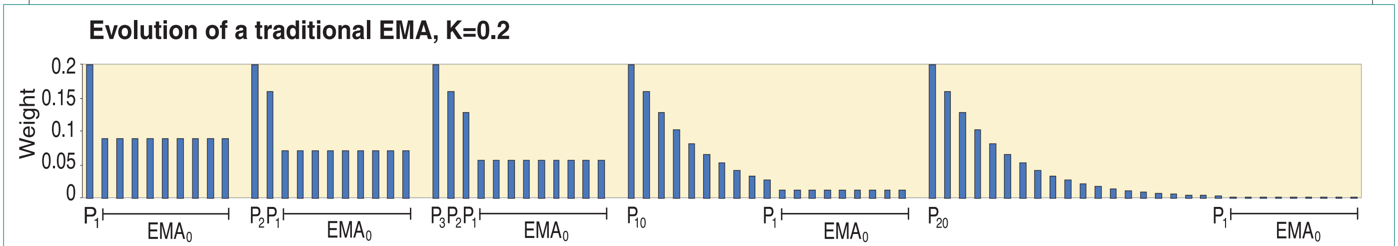


FIGURE 1: TRADITIONAL EMA. When new data is introduced into the calculation, the weight is drawn from the moving average. This reduces the MA height uniformly with each new piece of data.

$$K = \frac{2}{N+1} \quad \text{and} \quad K = \frac{1}{N}$$

The first expression is the most commonly used definition for K. To achieve the leading weight of 0.2 in our example using this definition, the value of N must be raised from the minimum allowable MA of a five-day MA to a nine-day, whereas if the second definition for K is used, the five-day produces the desired leading weight. The weight of each piece of data in the nine-day MA, however, will be considerably less, producing an initial weight distribution that is both longer in length and shorter in overall height but is arguably a better first approximation of the desired weight distribution. It is the moving average from which weight is drawn when new data is introduced, reducing the MA height uniformly with each new piece of data (Figure 1).

So EMAs are named in terms of length of days, presumably in the same way that simple moving averages are, but the implications behind the names are very different. While an N-day simple moving average relies on a steady flux of data into and out of the formula and is calculated using exactly N days of data, exponential moving averages rely on no fixed number of days of data. They can be initialized with as little as a single piece of data and they never discard any of the data they acquire. Further, N exists merely to determine the length of a moving average sufficient to initialize the EMA.

But even the practice of initializing with a MA turns out to be unnecessary. All EMAs have distorted weight distribution profiles early on, whether or not they are initialized with a MA, and no EMA will devalue its data in a smooth, exponential manner until plenty of data is in play.

One thing that can be said about the traditional N-day EMA is that with N as input and with enough data in play to produce a smooth distribution, the cumulative weight of the most recent N pieces of data always accounts for somewhere between 86% and 89% of the total weight. So while there is a definite correlation between the size of N and the length of time required for data to devalue, there is nothing particularly "N-day" about an N-day EMA.

THE MODIFIED EMA

An alternative to the traditional EMA exists. In recursive form,

$$\mathrm{EMA}_i = \frac{K}{1-(1-K)^i} P_i + \frac{\left[1-(1-K)^{i-1}\right](1-K)}{1-(1-K)^i} \mathrm{EMA}_{i-1}$$

where $\mathrm{EMA}_0 = 0$.

And in summation form,

$$\mathrm{EMA}_n = \frac{K}{1-(1-K)^n} \sum_{i=1}^{n} P_{n-i+1}(1-K)^{i-1}$$

Despite the added complexity of these formulas, they have some noteworthy features. Neither formula needs to be initialized with anything more than the first piece of regular data; there are no simple moving averages to be concerned with, and older data always automatically has less value. Because there is no initializing moving average from which to draw weight, it must be drawn from the preceding weights. Proportional amounts are drawn from every preceding weight when new data appears (Figure 2).

In addition, the absence of the MA eliminates the need to establish a minimum number of days N, which in turn implies no inherent dependence on the variable N. While it is true that the choice of N determines the retention rate, it does so indirectly. The retention rate can be defined explicitly by replacing $1 - K$ with a new variable r. Substituting this into equation (3), for instance, yields:

$$\mathrm{EMA}_n = \frac{1-r}{1-r^n} \sum_{i=1}^{n} P_{n-i+1}\, r^{i-1}$$

The choice of r has an advantage beyond simplicity. With r, any retention rate value in the range between zero and 1 can be selected. When N is used to determine the retention rate,

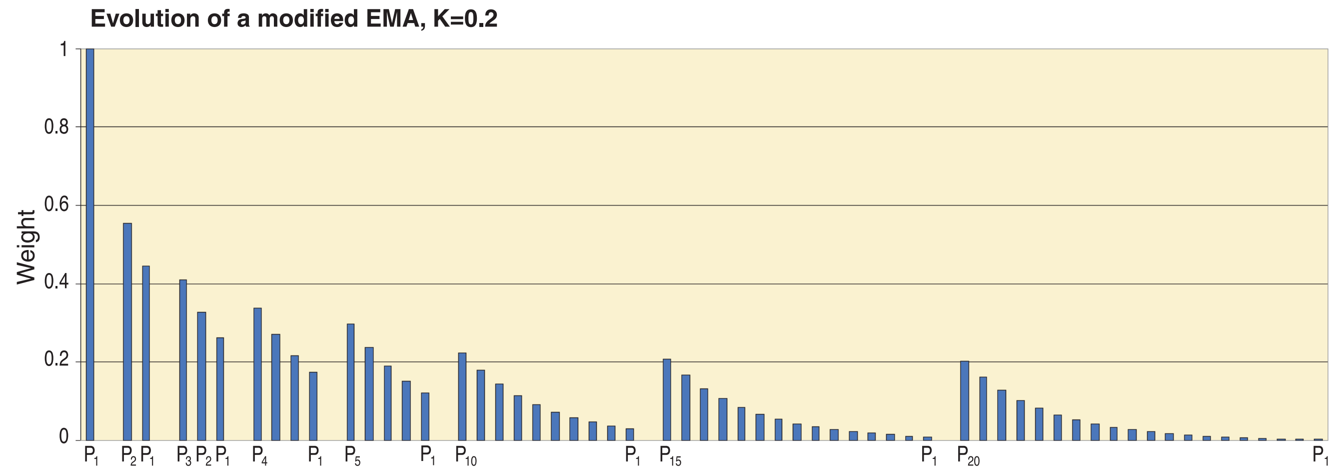


FIGURE 2: MODIFIED EMA. In the modified EMA there is no initiating MA, so proportional amounts are drawn from every preceding weight when new data appears.

many available values will be unselectable because N, representing days, is presumed to be a whole number, and that restriction in turn unnecessarily limits the number of available retention rates.

In the modified EMA, the weight for any piece of data in the i^{th} position is defined as

$$\frac{(1-r)r^{i-1}}{1-r^n}$$

Every weight approaches its own lower limit. The limit value of the i^{th} weight is

$$(1-r)r^{i-1}$$

Technically, the weights never reach their limiting values, but they do get very close to them. How long does it take for weights to get reasonably close to their limits? At low retention rates, newly created weights are much smaller than the weights before them, so large contributions from those preceding weights are not required and the distribution stabilizes in a relatively short time. At high retention rates, newly created weights are only marginally smaller than weights before them, requiring larger contributions from preceding weights, so the distribution takes longer to stabilize. The number of data required to do this will vary with the retention factor and how close to the limiting values the weights are required to be. The formula to determine the minimum number of data n is:

$$n \leq \frac{\ln \varepsilon - \ln(\varepsilon + 1)}{\ln r}$$

where e is the maximum allowable error, as a percentage of the limiting weight. Using the EMA traditional input of N, it generally will take about 2.3 times that amount of data to get the weights to within 1% of their limiting values.

THE *N*-DAY DILEMMA

If an N-day EMA is to have something to do with N days of data, then some meaningful definition must be concocted. One idea is to let the number of days N define the boundary between relevant data and negligible data. But as there is no universally accepted definition for what constitutes relevant or negligible data, there can be no universally accepted boundary.

However, it is possible to let N represent the number of days where a selected cumulative weight is reached. For instance, if we consider data accounting for 95% of the forward weight to be relevant, we can design our formula so that when we call for an N-day EMA, the retention rate is automatically computed so that 95% of the weight always occurs over the most recent N days. The tail of the distribution is still there but accounts for only the remaining 5% of weight. This approach is implemented by defining the retention rate r in terms of both the desired number of days and the desired cumulative weight percentage:

$$r = \sqrt[N]{1-c}$$

CLOSING THOUGHTS

A good EMA should allow us to weight data precisely and devalue it smoothly and predictably over selected time frames. No EMA will be smooth and predictable if it is not primed with enough data for the weight distribution profile to take shape.

That consideration aside, which EMA is superior? The traditional EMA appears computationally simpler and boasts fixed-value forward weights, but it generally requires a moving average to force some semblance of exponential decreasing behavior early on. The modified EMA can be initialized with a single piece of data; it is not subject to unusual initial behavior but appears more computationally demanding and features limit-approaching forward weights. Another question posed is whether to use a traditional N-day input, a retention rate r directly, or whether to consider a variation of r that takes into account a selected cumulative weight over N days.

In the end, it may come down to user preferences. As is true of many technical indicators, the power to choose and adjust parameters is largely what gives the EMA such utility, flexibility, and appeal.

Tim Treloar has experience as a commodities daytrader and position trader and holds degrees in both mathematics and education. He has taught mathematics for 10 years and also does some consulting work.

S&C

620098