Moving averages

Rob J Hyndman

November 8, 2009

A moving average is a time series constructed by taking averages of several sequential values of another time series. It is a type of mathematical convolution. If we represent the original time series by $y_1, \ldots, y_n$, then a **two-sided moving average** of the time series is given by

$$z_t = \frac{1}{2k+1} \sum_{j=-k}^{k} y_{t+j}, \qquad t = k+1, k+2, \ldots, n-k.$$

Thus $z_{k+1}, \ldots, z_{n-k}$ forms a new time series which is based on averages of the original time series, $\{y_t\}$. Similarly, a **one-sided moving average** of $\{y_t\}$ is given by

$$z_t = \frac{1}{k+1} \sum_{j=0}^{k} y_{t-j}, \qquad t = k+1, k+2, \ldots, n.$$

More generally, weighted averages may also be used. Moving averages are also called running means or rolling averages. They are a special case of "filtering", which is a general process that takes one time series and transforms it into another time series.

The term "moving average" is used to describe this procedure because each average is computed by dropping the oldest observation and including the next observation. The averaging "moves" through the time series until z_t is computed at each observation for which all elements of the average are available.

Note that in the above examples, the number of data points in each average remains constant. Variations on moving averages allow the number of points in each average to change. For example, in a cumulative average, each value of the new series is equal to the sum of all previous values.

Moving averages are used in two main ways: Two-sided (weighted) moving averages are used to "smooth" a time series in order to estimate or highlight the underlying trend; one-sided (weighted) moving averages are used as simple forecasting methods for time series. While moving averages are very simple methods, they are often building blocks for more complicated methods of time series smoothing, decomposition and forecasting.

1 Smoothing using two-sided moving averages

It is common for a time series to consist of a smooth underlying trend observed with error:

$$y_t = f(t) + \varepsilon_t,$$

where $f(t)$ is a smooth and continuous function of t and $\{\varepsilon_t\}$ is a zero-mean error series. The estimation of $f(t)$ is known as smoothing, and a two-sided moving average is one way of doing so:

$$\hat{f}(t) = \frac{1}{2k+1} \sum_{j=-k}^{k} y_{t+j}, \qquad t = k+1, k+2, \ldots, n-k.$$

The idea behind using moving averages for smoothing is that observations which are nearby in time are also likely to be close in value. So taking an average of the points near an observation will provide a reasonable estimate of the trend at that observation. The average eliminates some of the randomness in the data, leaving a smooth trend component.

Moving averages do not allow estimates of $f(t)$ near the ends of the time series (in the first k and last k periods). This can cause difficulties when the trend estimate is used for forecasting or analysing the most recent data.

Each average consists of $2k+1$ observations. Sometimes this is known as a $(2k+1)$ MA smoother. The larger the value of k, the flatter and smoother the estimate of $f(t)$ will be. A smooth estimate is usually desirable, but a flat estimate is biased, especially near the peaks and troughs in $f(t)$. When $\{\varepsilon_t\}$ is a white noise series (i.e., independent and identically distributed with zero mean and variance σ^2), the bias is given by $\mathrm{E}[\hat{f}(x)] - f(x) \approx \frac{1}{6}f''(x)k(k+1)$ and the variance by $\mathrm{V}[\hat{f}(x)] \approx \sigma^2/(2k+1)$. So there is a trade-off between increasing bias (with large k) and increasing variance (with small k).

2   Centered moving averages

The simple moving average described above requires an odd number of observations to be included in each average. This ensures that the average is centered at the middle of the data values being averaged. But suppose we wish to calculate a moving average with an even number of observations. For example, to calculate a 4-term moving average, the trend at time t could be calculated as

$$\hat{f}(t-0.5) = (y_{t-2} + y_{t-1} + y_t + y_{t+1})/4$$
$$\text{or} \quad \hat{f}(t+0.5) = (y_{t-1} + y_t + y_{t+1} + y_{t+2})/4.$$

That is, we could include two terms on the left and one on the right of the observation, or one term on the left and two terms on the right, and neither of these is centered on t. If we now take the average of these two moving averages, we obtain something centered at time t:

$$\hat{f}(t) = \tfrac{1}{2}\left[(y_{t-2} + y_{t-1} + y_t + y_{t+1})/4\right] + \tfrac{1}{2}\left[(y_{t-1} + y_t + y_{t+1} + y_{t+2})/4\right]$$
$$= \tfrac{1}{8}y_{t-2} + \tfrac{1}{4}y_{t-1} + \tfrac{1}{4}y_t + \tfrac{1}{4}y_{t+1}\tfrac{1}{8}y_{t+2}.$$

So a 4 MA followed by a 2 MA gives a **centered moving average**, sometimes written as 2×4 MA. This is also a weighted moving average of order 5, where the weights for each period are unequal. In general, a $2 \times m$ MA smoother is equivalent to a weighted MA of order $m+1$ with weights $1/m$ for all observations except for the first and last observations in the average, which have weights $1/(2m)$.

Centered moving averages are examples of how a moving average can itself be smoothed by another moving average. Together, the smoother is known as a *double moving average*. In fact, any combination of moving averages can be used together to form a double moving average. For example, a 3×3 moving average is a 3 MA of a 3 MA.

3   Moving averages with seasonal data

If the centered 4 MA was used with quarterly data, each quarter would be given equal weight. The weight for the quarter at the ends of the moving average is split between the two years. It is this property that makes 2×4 MA very useful for estimating a trend in the presence of quarterly seasonality. The seasonal variation will be averaged out exactly when the moving average is computed. A slightly longer or a slightly shorter moving average will still retain some seasonal

variation. An alternative to a 2×4 MA for quarterly data is a 2×8 MA or 2×12 MA, which will also give equal weights to all quarters and produce a smoother fit than the 2×4 MA. Other moving averages tend to be contaminated by the seasonal variation.

More generally, a $2 \times (km)$ MA can be used with data with seasonality of length m where k is a small positive integer (usually 1 or 2). For example, a 2×24 MA may be used for estimating a trend in monthly seasonal data (where $m = 12$).

4 Weighted moving averages

A weighted k-point moving average can be written as

$$\hat{f}(t) = \sum_{j=-k}^{k} a_j y_{t+j}.$$

For the weighted moving average to work properly, it is important that the weights sum to one and that they are symmetric, that is $a_j = a_{-j}$. However, we do not require that the weights are between 0 and 1. The advantage of weighted averages is that the resulting trend estimate is much smoother. Instead of observations entering and leaving the average abruptly, they can be slowly downweighted. There are many schemes for selecting appropriate weights. Kendall et al. (1983, chapter 46) give details.

Some sets of weights are widely used and have been named after their proposers. For example, Spencer (1904) proposed a $5 \times 4 \times 4$ MA followed by a weighted 5-term moving average with weights $a_0 = 1$, $a_1 = a_{-1} = 3/4$, and $a_2 = a_{-2} = -3/4$. These values are not chosen arbitrarily, but because the resulting combination of moving averages can be shown to have desirable mathematical properties. In this case, any cubic polynomial will be undistorted by the averaging process. It can be shown that Spencer's MA is equivalent to the 15-point weighted moving average whose weights are $-.009$, $-.019$, $-.016$, $.009$, $.066$, $.144$, $.209$, $.231$, $.209$, $.144$, $.066$, $.009$, $-.016$, $-.019$, and $-.009$. Another Spencer's MA that is commonly used is the 21-point weighted moving average. Henderson's weighted moving averages are also widely used, especially as part of seasonal adjustment methods (Ladiray & Quenneville 2001). The set of weights is known as the *weight function*. Table 1 shows some common weight functions. These are all symmetric, so $a_{-j} = a_j$.

Table 1: Weight functions a_j for some common weighted moving averages.

Name	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}
3 MA	.333	.333										
5 MA	.200	.200	.200									
2×12 MA	.083	.083	.083	.083	.083	.083	.042					
3×3 MA	.333	.222	.111									
3×5 MA	.200	.200	.133	.067								
S15 MA	.231	.209	.144	.066	.009	−.016	−.019	−.009				
S21 MA	.171	.163	.134	.037	.051	.017	−.006	−.014	−.014	−.009	−.003	
H5 MA	.558	.294	−.073									
H9 MA	.330	.267	.119	−.010	−.041							
H13 MA	.240	.214	.147	.066	.000	−.028	−.019					
H23 MA	.148	.138	.122	.097	.068	.039	.013	−.005	−.015	−.016	−.011	−.004

S = Spencer's weighted moving average
H = Henderson's weighted moving average

Weighted moving averages are equivalent to kernel regression when the weights are obtained from a kernel function. For example, we may choose weights using the quartic function

$$Q(j,k) = \begin{cases} \left\{1 - [j/(k+1)]^2\right\}^2 & \text{for } -k \leq j \leq k; \\ 0 & \text{otherwise.} \end{cases}$$

Then a_j is set to $Q(j,k)$ and scaled so the weights sum to one. That is,

$$a_j = \frac{Q(j,k)}{\sum\limits_{i=-k}^{k} Q(i,k)}.$$

5 Forecasting using one-sided moving averages

A simple forecasting method is to simply average the last few observed values of a time series. Thus

$$\hat{y}_{t+h|t} = \frac{1}{k+1} \sum_{j=0}^{k} y_{t-j}$$

provides a forecast of y_{t+h} given the data up to time t.

As with smoothing, the more observations included in the moving average, the greater the smoothing effect. A forecaster must choose the number of periods $(k+1)$ in a moving average. When $k = 0$, the forecast is simply equal to the value of the last observation. This is sometimes known as a "naïve" forecast.

An extremely common variation on the one-sided moving average is the exponentially weighted moving average. This is a weighted average where the weights decrease exponentially. It can be written as

$$\hat{y}_{t+h|t} = \sum_{j=0}^{t-1} a_j y_{t-j},$$

where $a_j = \lambda(1 - \lambda)^j$. Then, for large t, the weights will approximately sum to one. An exponentially weighted moving average is the basis of simple exponential smoothing. It is also used in some process control methods.

6 Moving average processes

A related idea is the moving average process, which is a time series model that can be written as

$$y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q},$$

where $\{e_t\}$ is a white noise series. Thus, the observed series $\{y_t\}$, is a weighted moving average of the unobserved $\{e_t\}$ series. This is a special case of an Autoregressive Moving Average (or ARMA) model and is discussed in more detail on page ??. An important difference between this moving average and those considered previously is that here the moving average series is directly observed, and the coefficients $\theta_1, \ldots, \theta_q$ must be estimated from the data.

References

Kendall, M. G., Stuart, A. & Ord, J. K. (1983), *Kendall's advanced theory of statistics. Vol. 3*, Hodder Arnold, London.

Ladiray, D. & Quenneville, B. (2001), *Seasonal adjustment with the X-11 method*, Vol. 158 of *Lecture notes in statistics*, Springer-Verlag.

Makridakis, S., Wheelwright, S. C. & Hyndman, R. J. (1998), *Forecasting: methods and applications*, 3rd edn, John Wiley & Sons, New York.

Spencer, J. (1904), 'On the graduation of the rates of sickness and mortality', *Journal of the Institute of Actuaries* **38**, 334–343.