The Weighted Moving Average Technique

Chapter · June 2010

DOI: 10.1002/9780470400531.eorms0964

1 author:

Marcus B. Perry
University of Alabama

THE WEIGHTED MOVING AVERAGE TECHNIQUE

MARCUS B. PERRY
Department of Information Systems, Statistics and Management Science, The University of Alabama, Tuscaloosa, Alabama

INTRODUCTION

The w-term moving average (MA) is most often applied to time series data as a means to smooth out seasonal variation or irregular fluctuations in the data, permitting the analyst to more easily identify the structural patterns. It is also used as a means for computing short-term forecasts of time series. Its application in practice is widespread across many disciplines. For example, in an engineering application, one might use the MA in efforts to control the quality of a product or process. Further, in a business application, the MA might be used to forecast future sales demand or employee turnover. Depending on the application, different types of MAs can be computed (e.g., simple, weighted, and cumulative). Further, MAs can be centered by placing each w-term average in the middle of its corresponding w-length time interval. If w is odd, then by this placement, the MA is centered. On the other hand, if w is even, a 2-term moving average is required to center the original MA. Centered MAs are more typical when the primary objective of the analysis is seasonal adjustment of the data.

In this tutorial, emphasis is placed on the *weighted* moving average (WMA), where each w-term average is placed at the upper bound of its corresponding w-length time interval. In particular, consider the n-length time series $\{x_i\}$, then its w-term WMA is defined as

$$z_i = \sum_{j=i-w+1}^{i} \omega_j x_j, \tag{1}$$

for $i \in [w, w+1, \dots, n]$, where the ω_j's are weights and the x_j's are a time sequence of (possibly autocorrelated) random variables. Notice that the number of observations in the series $\{z_i\}$ is reduced to $n - w + 1$; that is, $w - 1$ terms are lost in the WMA. As noted by one of the reviewers, the right-hand side of Equation (1) defines a convex hull for the data points, and thus z_i is a point in the convex hull of data up to the present. The WMA defined in Equation (1) is useful for computing one-step-ahead (i.e., short-term) forecasts or predictions. Further, it lends nicely to the development of process monitoring algorithms useful in, say, statistical quality control applications.

The WMA in Equation (1) can be written in vector notation as

$$\begin{aligned} z_i &= (\omega_{i-w+1}, \omega_{i-w+2}, \dots, \omega_i) \begin{pmatrix} x_{i-w+1} \\ x_{i-w+2} \\ \vdots \\ x_i \end{pmatrix} \\ &= \boldsymbol{\omega}_i{}' \mathbf{x}_i, \end{aligned} \tag{2}$$

where $\boldsymbol{\omega}_i$ and $\mathbf{x}_i$ are both $w \times 1$ and $\sum_{k=1}^{w} \omega_{i-w+k} = 1$. If the ω_i's are constant for all i, then z_i can be written as

$$z_i = (\omega_1, \omega_2, \cdots, \omega_w) \begin{pmatrix} x_{i-w+1} \\ x_{i-w+2} \\ \vdots \\ x_i \end{pmatrix} = \boldsymbol{\omega}' \mathbf{x}_i, \tag{3}$$

where $\sum_{k=1}^{w} \omega_k = 1$, and whereby applying larger weights to the more recent observations yields the additional constraint $\omega_1 \le \omega_2 \le \cdots \le \omega_w$. Consider the special case where $\omega_k = w^{-1}$ for all k, then this produces the w-term *simple* MA given by

$$z_i = w^{-1} \sum_{j=i-w+1}^{i} x_j, \tag{4}$$

which is just the arithmetic average of the w terms.

Wiley Encyclopedia of Operations Research and Management Science, edited by James J. Cochran

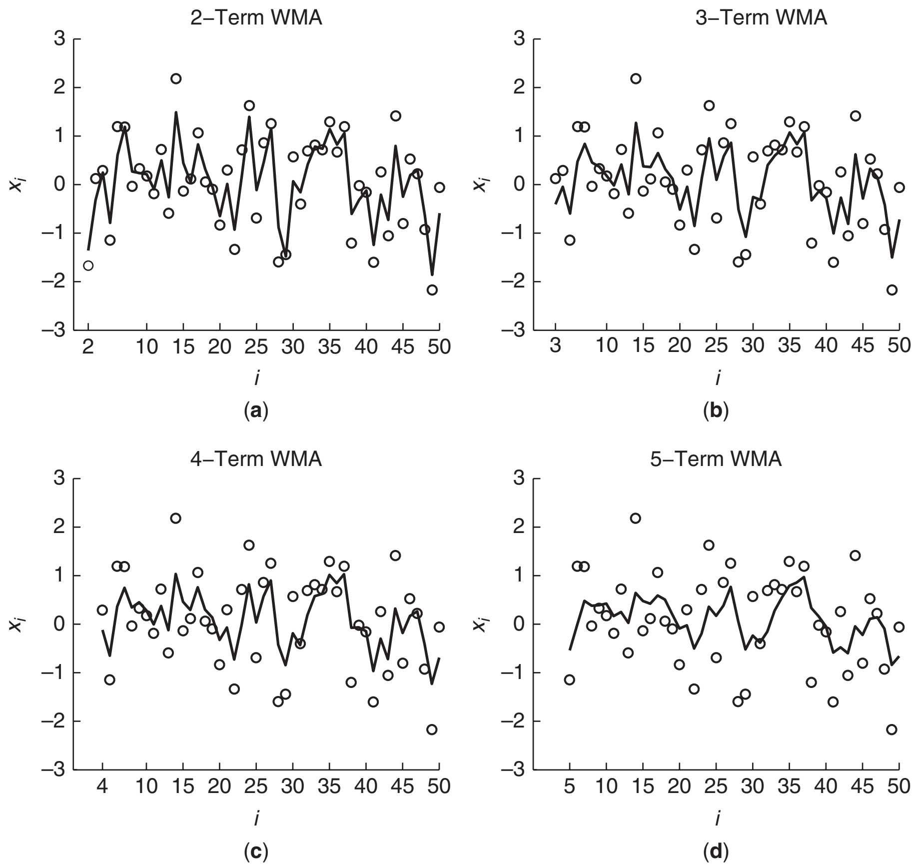

Figure 1. Examples of weighted moving averages.

To illustrate the computation of a w-term *simple* MA, consider $w = 3$ and the series

$$\{x_i\} = \begin{bmatrix} 1.3 & 2.5 & 4.1 & 2.9 & 1.6 \end{bmatrix}.$$

Since $n = 5$ and $w = 3$, the new series $\{z_i\}$ will contain three observations (i.e., z_3, z_4, z_5) computed, respectively, as

$$z_3 = (1.3 + 2.5 + 4.1)/3 = 2.63$$

$$z_4 = (2.5 + 4.1 + 2.9)/3 = 3.17$$

$$z_5 = (4.1 + 2.9 + 1.6)/3 = 2.87,$$

and thus the new series $\{z_i\}$ (for $i = 3, 4, 5$) is given as[1]

$$\{z_i\} = \begin{bmatrix} \cdot & \cdot & 2.63 & 3.17 & 2.87 \end{bmatrix}.$$

To illustrate further, consider the scatter plots of $n = 50$ independently generated standard normal random deviates shown in Fig. 1. For each plot, the horizontal axis denotes time and the vertical axis denotes the observed value of X at time i. The solid line in Fig. 1a shows a 2-term WMA with $\omega_1 = 0.25$ and $\omega_2 = 0.75$. Similarly, the solid line in Fig 1b shows a 3-term WMA with $\omega_1 = 0.15$, $\omega_2 = 0.25$, and $\omega_3 = 0.60$. A 4-term WMA with $\omega_1 = 0.10$, $\omega_2 = 0.15$, $\omega_3 = 0.25$, and $\omega_4 = 0.50$ is given in Fig. 1c, while a 5-term WMA with $\omega_1 = 0.10$, $\omega_2 = 0.15$, $\omega_3 = 0.20$, $\omega_4 = 0.25$, and $\omega_5 = 0.30$ is shown in Fig. 1d.

It seems apparent from Fig. 1 that as the number of terms in the MA increases, so does the amount by which $\{x_i\}$ is smoothed. The implication is that the variance of z_i reduces with the addition of more terms; however, at the expense of bias. This is quite

[1]Note that the centered MA for this case is given as $\{z_i\} = [\cdot \quad 2.63 \quad 3.17 \quad 2.87 \quad \cdot]$.

intuitive and in-line with the well-known variance-bias trade-off. This is discussed in more detail in a later section.

In the next section several properties of z_i are evaluated, including the mean, variance, and the variance of the one-step-ahead forecast errors. In subsequent sections, some common applications of the WMA in practice are discussed.

PROPERTIES

Let $\mathbf{x}_i = (x_{i-w+1}, x_{i-w+2}, \ldots, x_i)'$ be a $w \times 1$ random vector with mean $\boldsymbol{\mu}_i$ and finite autocovariance matrix $\boldsymbol{\Sigma} = \sigma_x^2\mathbf{R}$, where $\mathbf{R}$ denotes the $w \times w$ autocorrelation matrix of $\mathbf{x}_i$ and σ_x^2 denotes the variance of the original process $\{x_i\}$. Note the special case of the uncorrelated process when $\mathbf{R} = \mathbf{I}$. Note further that $\boldsymbol{\mu}_i$ is $w \times 1$ and is given explicitly as[2]

$$\boldsymbol{\mu}_i = \begin{pmatrix} \mu_{i-w+1} & \mu_{i-w+2} & \cdots & \mu_i \end{pmatrix}' \tag{5}$$

and $\mathbf{R}$ has the form[3]

$$\mathbf{R} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{w-1} \\ \rho_{-1} & 1 & \rho_1 & \cdots & \rho_{w-2} \\ \rho_{-2} & \rho_{-1} & 1 & \cdots & \rho_{w-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{-w+1} & \rho_{-w+2} & \rho_{-w+3} & \cdots & 1 \end{pmatrix}, \tag{6}$$

where $\rho_m = \rho_{-m}$ denotes the lag m autocorrelation of $\mathbf{x}_i$.

Applying the expectation operator to z_i we obtain

$$E(z_i) = \boldsymbol{\omega}'\boldsymbol{\mu}_i \tag{7}$$

and by applying the variance operator to z_i we find

$$\sigma_z^2 = \sigma_x^2(\boldsymbol{\omega}'\mathbf{R}\boldsymbol{\omega}), \tag{8}$$

where for uncorrelated processes

$$\sigma_z^2 = \sigma_x^2 \sum_{k=1}^{w} \omega_k^2. \tag{9}$$

Suppose that one uses z_i to forecast x_{i+1}, then applying the variance operator to $z_i - x_{i+1}$ yields the variance of the one-step-ahead forecast errors (i.e., prediction errors), or

$$\sigma_{\text{pred}}^2 = \sigma_x^2 \left(1 + \boldsymbol{\omega}'\mathbf{R}\boldsymbol{\omega} - 2\sum_{k=1}^{w} \omega_k \rho_{w-k+1}\right), \tag{10}$$

where for the uncorrelated process $\rho_m = 0$ $(m > 0)$ and Equation (10) reduces to

$$\sigma_{\text{pred}}^2 = \sigma_x^2 + \sigma_z^2, \tag{11}$$

which is the sum of the variances of the original process $\{x_i\}$ and the WMA process $\{z_i\}$.

In much of what follows, references will be made to different time series models for autocorrelated processes. These models consist of the class of stationary and invertible autoregressive moving average (ARMA) models discussed in Box *et al.* [1]. Additional references to these models include Abraham and Ledolter [2] and Pankratz [3]. The notation ARMA(p, q) is often used to denote the ARMA model of order p and q; that is, we have a model with p autoregressive terms and q moving average terms. The notation is shortened further when the model contains only autoregressive or moving average terms. That is, a model containing p autoregressive terms and no moving average terms is denoted by AR(p). Similarly, a model containing q moving average terms and no autoregressive terms is denoted by MA(q). This notation will be used in subsequent sections of this tutorial.

3-Term WMA Applied to AR(1) Process

Suppose that the process under consideration is well characterized by a first-order autoregressive model, or

$$x_i = \mu + \phi(x_{i-1} - \mu) + \epsilon_i. \tag{12}$$

[2] Note that the original process $\{x_i\}$ need not be mean stationary.

[3] For covariance stationary processes, $\text{Var}(\mathbf{x}_i) = \sigma_x^2\mathbf{R}$ for all $i \Rightarrow$ covariance is only a function of lag.

Suppose further that $\sigma_\epsilon^2 = 1$ and $\phi = 0.50$, then it can be shown that $\sigma_x^2 = (1-\phi^2)^{-1} = 1.3333$. Also, for a 3-term WMA, we have the correlation matirx

$$\mathbf{R} = \begin{pmatrix} 1.00 & 0.50 & 0.25 \\ 0.50 & 1.00 & 0.50 \\ 0.25 & 0.50 & 1.00 \end{pmatrix} \quad (13)$$

so that for $\omega = [0.15, 0.25, 0.60]'$ we obtain for all i

$$E(z_i) = \mu \quad (14)$$

$$\sigma_z^2 = 0.9033, \quad (15)$$

and the variance of the one-step-ahead forecast errors obtained from Equation (10) is given by[4]

$$\sigma_{\text{pred}}^2 = 1.2200. \quad (16)$$

To put this example into practical context, suppose the process in Equation (12) well models a sales demand process and one is interested in monitoring for changes in mean sales demand over time. To accomplish this, one can plot z_i for each i and compare to the limits $\mu_0 \pm L\sigma_z$, where μ_0 is some hypothesized *expected* value of sales demand, $\sigma_z = \sqrt{0.9033}$ and L is a constant expressed in standard deviation units.[5] If any point exceeds these limits, one can conclude that the mean sales demand has changed, and thus plans might be devised at this point to increase or decrease production for the next period.

5-Term WMA Applied to ARMA(1,1) Process

Suppose now that the process under consideration is well modeled by the ARMA(1,1) model with $\phi = 0.75$ and $\theta = -0.35$, or

$$x_i = \mu + 0.75\,(x_{i-1} - \mu) + 0.35\epsilon_{i-1} + \epsilon_i. \quad (17)$$

Suppose again that $\sigma_\epsilon^2 = 1$, then it can be shown that

$$\sigma_x^2 = \frac{1+\theta^2 - 2\phi\theta}{(1-\phi^2)} = 3.7657 \quad (18)$$

and for a 5-term WMA we have the correlation matrix

$$\mathbf{R} = \begin{pmatrix} 1.0000 & 0.8429 & 0.6322 & 0.4742 & 0.3556 \\ 0.8429 & 1.0000 & 0.8429 & 0.6322 & 0.4742 \\ 0.6322 & 0.8429 & 1.0000 & 0.8429 & 0.6322 \\ 0.4742 & 0.6322 & 0.8429 & 1.0000 & 0.8429 \\ 0.3556 & 0.4742 & 0.6322 & 0.8429 & 1.0000 \end{pmatrix}, \quad (19)$$

so that for $\omega = [0.10, 0.15, 0.20, 0.25, 0.30]'$ we obtain for all i

$$E\,(z_i) = \mu \quad (20)$$

$$\sigma_z^2 = 2.8163 \quad (21)$$

and the variance of the one-step-ahead forecast errors obtained from Equation (10) is given by

$$\sigma_{\text{pred}}^2 = 2.1703. \quad (22)$$

Notice that $\sigma_{\text{pred}}^2 < \sigma_z^2$ for this process,[6] and thus for our sales demand monitoring example described above, one might consider monitoring the one-step-ahead forecasts $x_i - z_{i-1}$ in place of z_i to gain a reduction in variance of the charting statistic. That is, the quantity $x_i - z_{i-1}$ would be computed and compared to the limits $\pm L\sqrt{2.1703}$.

Figure 2 shows a single realization of the ARMA$(0.75, -0.35)$ process with two 5-term WMAs superimposed. The WMA shown as the solid line has weights $\omega_1 = 0.10$, $\omega_2 = 0.15$, $\omega_3 = 0.20$, $\omega_4 = 0.25$, and $\omega_5 = 0.30$, while the WMA shown as the dotted line has weights $\omega_k = 0.20$ for $k = 1, 2, \ldots, 5$, which is the 5-term *simple* MA. Note that for this

[4]Note also that $E(z_i - x_{i+1}) = 0$.

[5]L is often chosen on the basis of an acceptable type-I error.

[6]There is more uncertainty associated with $\{z_i\}$ than with $\{x_i - z_{i-1}\}$ due to strong positive autocorrelation. In general, $\sigma_{\text{pred}}^2 < \sigma_z^2$ when $\sum_{k=1}^{w} \omega_k \rho_{w-k+1} > \frac{1}{2}$.

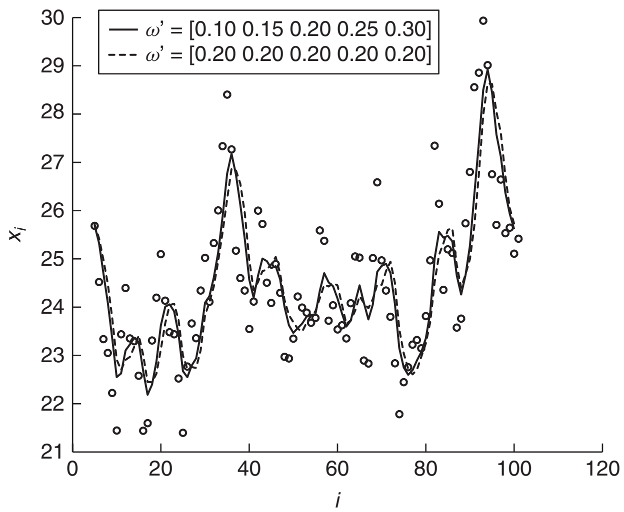

Figure 2. Single realization of ARMA$(0.75, -0.35)$ process with two 5-term WMAs.

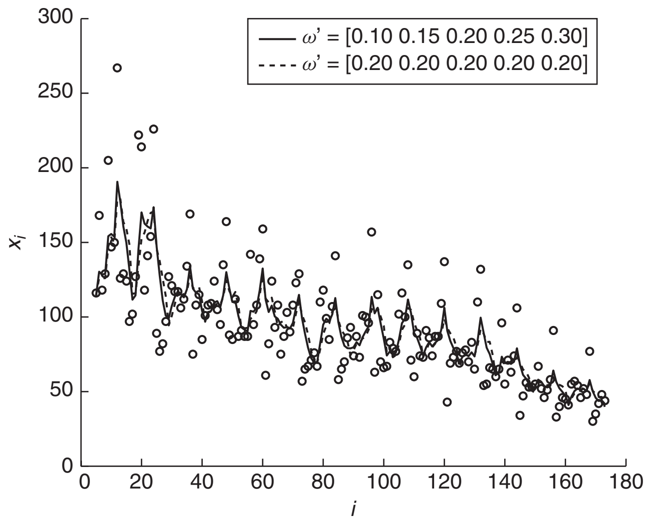

Figure 3. Monthly rose wine sales data with a 5-term simple MA and a 5-term WMA superimposed. Total number of observations is $n = 173$ and $w = 4$ terms are lost.

realization, the two MAs track closely, with only marginal differences. In the next section some of the more common applications of the WMA in practice are discussed.

APPLICATIONS OF THE WEIGHTED MOVING AVERAGE

One of the most common applications of the WMA is generating one-step-ahead forecasts of time series. For example, consider Fig. 3, where monthly Australian sales of rose wine is plotted over 173 months from approximately January 1980 to April 1995. The data and its original source can be obtained from the Time Series Data Library.[7] Superimposed on this data are a 5-term WMA[8] (solid line) and a 5-term simple MA (dotted line). Notice that this

[7]The Time Series Data Library is accessible at http://www.robjhyndman.com/TSDL/.

[8]Note that $\omega = [0.10, 0.15, 0.20, 0.25, 0.30]$ for this application.

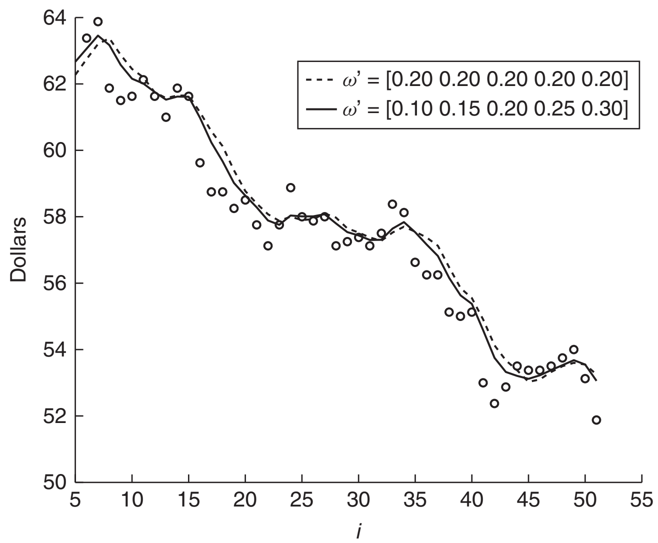

Figure 4. Weekly closing price of AT&T common shares. Total number of observations is $n = 51$ and $w = 4$ terms are lost.

process has some significant seasonal effects evidenced by the overall trend and cyclical-type fluctuations in the MAs. The forecast for x_{174} is then given by $\hat{x}_{174} = z_{173}$, which for the simple MA and WMA cases we have $\hat{x}_{174} = 39.80$ and $\hat{x}_{174} = 41.85$, respectively. The actual sales in the 174th month was $x_{174} = 45.00$, thus, for this realization, the WMA provides a more accurate forecast than that provided by the simple MA.

Consider a second example taken from Pankrats [3] involving the weekly closing price of AT&T common shares. The data set consists of $n = 52$ observations taken over each week in 1979. Figure 4 shows the data up through the 51st week along with 5-term simple and WMAs superimposed. If the WMA is used to forecast closing price in the 52nd week, then $\hat{x}_{52} = 53.05$, and the forecast using the simple MA is $\hat{x}_{52} = 53.25$. The actual closing price at week 52 was $x_{52} = 52.25$. Thus, again we see that for this realization, applying higher weights to more recent observations provides a more accurate forecast (albeit marginal in this case).

We can improve on the *accuracy* of the forecast by reducing the number of terms in the MA; however, this will result in an increase in σ_z^2 (i.e., a decrease in precision). For example, Fig. 5 shows three different WMAs with $w = 2, 3,$ and 5, respectively. Notice that the 2-term WMA forecast for x_{52} is $\hat{x}_{52} = 52.19$, which is very close to the actual value of $x_{52} = 52.25$. However, the amount by which z_i fluctuates when $w = 2$ is noticeably greater if, say, $w = 5$.

This reduction in σ_z^2 as w increases is easily shown for the uncorrelated process. If the observations are uncorrelated, then Equation (8) can be written as

$$\sigma_z^2 = \sigma_x^2(\boldsymbol{\omega}'\boldsymbol{\omega}) = \sigma_x^2 \sum_{k=1}^{w} \omega_k^2 \tag{23}$$

and for the simple MA $\omega_k = w^{-1}$ for all $k = 1, 2, \ldots, w$ and

$$\sigma_z^2 = \frac{\sigma_x^2}{w} \tag{24}$$

so that as $w \to \infty$ then $\sigma_z^2 \to 0$.

The above examples demonstrate how the WMA can be applied to generate one-step-ahead forecasts of a time series. Typically, in these applications, the WMA is useful when the forecast horizon is short-term [4]. Other methods are more frequently used for computing mid-to-long-term forecasts.

Another common application of the WMA is in the area of quality control. In particular, the WMA is often used for monitoring critical process quality characteristics over time

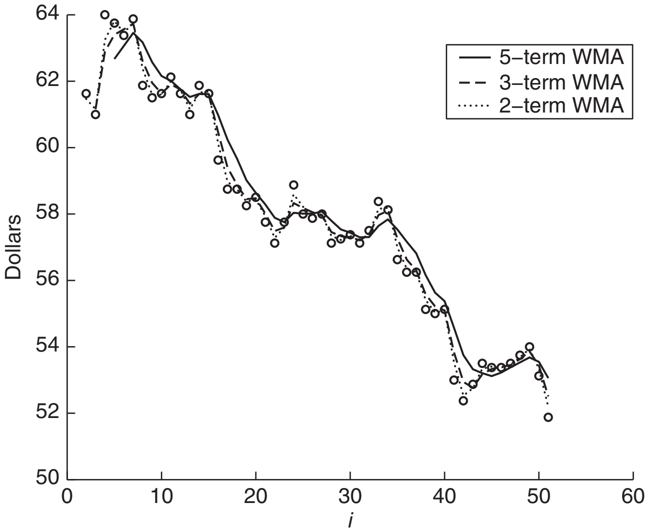

Figure 5. Three different WMAs of weekly closing price of AT&T common shares. Total number of observations is $n = 51$.

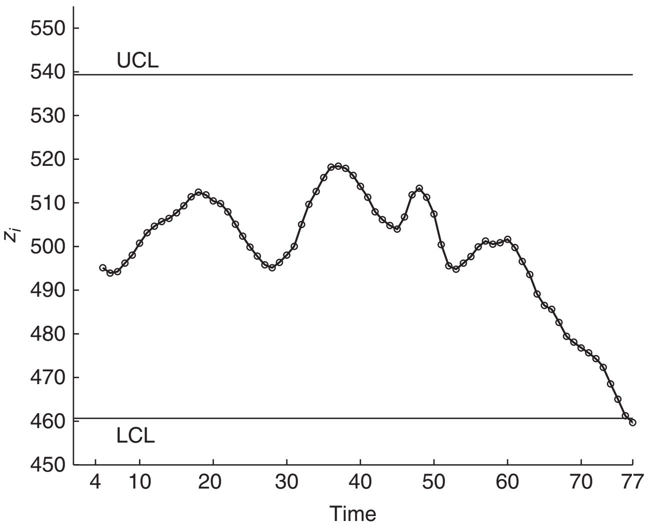

Figure 6. 5-term *two-sigma* WMA control chart applied to simulated ARMA$(0.95, -0.65)$ process. Note the control chart signaled on the 77th observation, suggesting a decrease in the expected yield of the process.

[5,6]. For example, suppose that the process under consideration is a chemical process where the output being monitored is yield. Suppose further that the process is highly autocorrelated and its *expected* behavior is well modeled by the ARMA$(0.95, -0.65)$ process, or

$$x_i = 500 + 0.95(x_{i-1} - 500) + 0.65\epsilon_{i-1} + \epsilon_i \tag{25}$$

for all i with $\sigma_\epsilon^2 = 15.00$. Thus, $\sigma_x^2 = 408.8462$ and for a 5-term WMA,[9] we find

$$\sigma_z^2 = 387.0393. \tag{26}$$

Suppose it is of interest to detect changes in the expected yield of the process, say from $\mu_i = 500$ to $\mu_i = \mu_1$, where $\mu_1 \neq 500$.

[9] $\omega' = [0.10, 0.15, 0.20, 0.25, 0, 30]$.

The WMA can be used as a basis for constructing a *control chart*.[10] That is, we could compute z_i (at each time i) and compare to the upper and lower control limits $500 \pm L\sqrt{\sigma_z^2} = 500 \pm 19.6733L$, where L is a constant. If z_i exceeds either of these limits, then a signal is issued and a special-cause investigation is initiated in efforts to uncover the underlying cause of the process change. The constant L is expressed in standard deviation units and is set to achieve an acceptable false alarm rate. If we set $L = 2$, then Fig. 6 shows a *two-sigma* WMA control chart applied to a simulated realization of the ARMA(0.95, −0.65) process with $\sigma_\epsilon^2 = 15.00$. Initially, the process was simulated so that $\mu_i = 500$ for all $i = 1, 2, \ldots, 50$. However, for $i > 50$ the process was simulated so that $\mu_i = 485$. Notice that $z_{77} <$ LCL, which suggest that the expected yield of the process has decreased.[11] Since high yield is desirable, production would cease at this point and a special-cause search is initiated. When using $\{z_i\}$ as a control charting statistics, smaller shifts in the process mean are guarded against most effectively with larger w; however, at the expense of quick response to large shifts [7]. Therefore, if the detection of smaller-sized shifts is of interest, w should be set sufficiently large. For larger-sized shifts, w should be set sufficiently small. Sparks [5] provides further discussion on strategies for choosing $\boldsymbol{\omega}$ and w when using the WMA for process monitoring.

The above illustrations are not the only applications of the WMA; however, they are common applications of the WMA in the management science and quality engineering disciplines. In general, application of the WMA is widespread throughout several disciplines, including health, finance, criminal justice, sports, industry, demography, and many more.

[10]Control charts are often used to distinguish between common-cause and special-cause process variability. Common-cause variation is deemed “unavoidable” and special-cause variation “avoidable.”

[11]Note that the control chart required $77 - 50 = 27$ additional samples after the change in order to detect the simulated decrease.

SUMMARY

In this tutorial, the WMA was discussed. As previously noted, the WMA is often used for smoothing irregular fluctuations (i.e., noise) in a time series to permit the data analyst to better reveal the trend-cycle patterns over time. Additionally, the WMA is frequently used to compute short-term forecasts of time series (e.g., sale and stocks). Some common applications of the WMA in the management science and quality engineering disciplines were also discussed. Further, the expected value, variance, and one-step-ahead prediction variance of the WMA were evaluated. For more information on the WMA and its application, see Bowerman *et al*. [8], Abraham and Ledolter [2], and Makridakis *et al*. [9].

REFERENCES

1. Box GEP, Jenkins GM, Reinsel RC. Time series analysis: forecasting and control. New Jersey: Prentice Hall; 1983.
2. Abraham B, Ledolter J. Statistical methods for forecasting. New York: John Wiley & Sons, Inc.; 1983.
3. Pankrats A. Forecasting with univariate Box-Jenkins models: concepts and cases. New York: John Wiley & Sons, Inc.; 1983.
4. Mentzer JT, Cox JE Jr. Familiarity, application and performance of sales forecasting techniques. J Forecast 1984;3:27–36.
5. Sparks R. Weighted moving averages: an efficient plan for monitoring specific location shifts. Int J Prod Res 2004;42(12):2521–2528.
6. Nugent T, Baykal-Gursoy M, Gursoy K. Monitoring k-step-ahead controlled processes. Qual Reliab Eng Int 2005;21:63–80.
7. Montgomery DC. Introduction to statistical quality control. New York: John Wiley & Sons, Inc.; 2005.
8. Bowerman BL, O'Connell RT, Koehler AB. Forecasting, time series, and regression. Belmont, CA: Duxbury; 2005.
9. Makridakis S, Wheelwright SC, Hyndman RJ. Forecasting: methods and applications. New York: John Wiley & Sons, Inc.; 1998.