

# Design and Performance Analysis of the Exponentially Weighted Moving Average Mean Estimate for Processes Subject to Random Step Changes

Argon CHEN

Graduate Institute of Industrial Engineering  
National Taiwan University  
Taipei, Taiwan 106

E. A. ELSAYED

Department of Industrial Engineering  
Rutgers University  
Piscataway, NJ 08854

The exponentially weighted moving average (EWMA) is a well-known and popular statistic used for smoothing and forecasting time series and as a process mean estimator, due to its simplicity and ability to capture nonstationarity. The EWMA statistic has been shown to be an optimal mean estimator for a certain disturbance process and an effective estimator for various other processes. In this article we focus on a practical disturbance process—relatively small random step changes that are difficult to distinguish from white noise and usually overlooked by practitioners. We propose an optimal EWMA parameter for step-change disturbance processes, as well as methodologies to identify and estimate the process models. The EWMA estimator's performance is then evaluated analytically. We demonstrate that a well-designed EWMA control scheme can effectively reduce the process variation even for processes subject to infrequent, small step changes. A semiconductor process example illustrates the design and analysis.

KEY WORDS: Exponentially weighted moving average; Mean estimator; Run-to-run control; Step change.

## 1. INTRODUCTION

Early applications of the exponentially weighted moving average (EWMA) appear in economics and inventory control and forecasting (Muth 1960; Cox 1961; Brown 1962) and in quality and process control (Roberts 1959; Hunter 1986; MacGregor 1988; Lucas and Saccucci 1990; Box and Kramer 1992). Most recently, the EWMA controller has received special attention from the semiconductor industry and is the core of most run-to-run controllers for semiconductor fabrication processes (Ingolfsson and Sachs 1993; Del Castillo and Hurwitz 1997).

Process control seeks to minimize variations and ensure consistent product quality. For example, in a semiconductor oxidation process, a layer of silicon dioxide ( $\text{SiO}_2$ ) is grown in a furnace; thickness is considered an important quality characteristic. Because semiconductor manufacturing usually involves more than 200 process steps, variations in the  $\text{SiO}_2$  layer thickness will impact subsequent processes and consequently affect the final quality of integrated circuit chips. A run-to-run feedback process control mechanism (Sachs, Hu, and Ingolfsson 1995; Boning et al. 1995) has been proposed to reduce the layer variation by comparing the actual thickness grown with its target value, then using the difference to adjust the process accordingly.

An important step of process control is to estimate the process mean. The EWMA estimator is optimal when the process mean follows a first-order integrated moving average [IMA(1, 1)] model (Box and Jenkins 1970). In fact, the EWMA estimator can be implemented as a widely used

proportional-integral-derivative (PID) controller (Åström and Hägglund 1995). Although the estimator performs well for various other processes, few studies (Ingolfsson and Sachs 1993; Yashchin 1995) have analytically showed the estimator's wide applicability.

In this article we demonstrate the EWMA estimator's performance for a stochastic process other than the IMA(1, 1) process, in particular one whose mean is subject to random step changes. We refer to this type of variation as *step-change disturbances*. Different batches of raw materials (e.g., quality differences of different raw wafers, oxygen densities) in the oxidation process often produce step changes in the process. Sudden changes in environmental conditions and machine parts replacement may also be causes. These causes of step-change disturbances are usually inevitable and difficult to eliminate; they differ from variation due to process mean drifts caused by, for example, process deterioration or cumulative step changes.

We investigate the step-change disturbance process whose mean step changes are too small to be distinguished from white noise. The disturbance process can be viewed as a collection of segmented processes in which each segment has its own *local* mean with white-noise variations and the mean

of each segment varies about an *overall* grand mean. When the variation of local means is not significantly larger than white noise, the process appears to be varying only about the overall mean. As a result, the process, when left without controls, exhibits a larger variation than does the underlying white noise. In the past, such variations have been monitored through incoming quality control (IQC) and statistical process control (SPC). Recently, control strategies using the EWMA statistic, such as run-to-run controllers in semiconductor manufacturing (Moyne, Del Castillo, and Hurwitz 2000), have been proposed to further reduce the quality deviation resulting from small step changes. In this article we propose a disturbance process model characterizing such a process and then use the EWMA statistic to estimate the process mean. Because the EWMA estimator can be implemented as a PID controller, the methodology will also provide an analytical basis for tuning PID controllers. Thus we propose designing and analyzing such a known controller for a practical disturbance process seldom discussed in the literature.

In Section 2 we describe a control system using the EWMA mean estimator and explain the association between the EWMA estimator and the PID controller. In Section 3 we present the practical random step-change disturbance model. In Section 4 we develop a design methodology that optimizes the EWMA estimator's performance for a given step-change process. In Section 5 we demonstrate, with examples, how a process model with step-change disturbances can be identified and estimated. In Section 6 we evaluate and discuss the EWMA estimator's performance. Finally, we discuss practical issues in Section 7.

## 2. CONTROL SYSTEM USING EXPONENTIALLY WEIGHTED MOVING AVERAGE ESTIMATORS

Figure 1 depicts a system whose output is controlled by adjusting a variable to compensate for the process disturbance to reduce the output variation. In this system,  $U_t$  represents the control action of the manipulated variable at time  $t$ ,  $Y_t$  is the system output in the absence of disturbances, and  $X_t$  is the disturbance added to  $Y_t$ . The observed system output  $\delta_t$  is the sum of  $Y_t$  and  $X_t$ .

Revisiting the oxidation process,  $U_t$  could be the temperature setting of the oxidation furnace that is related approximately linearly to the oxidation layer's growing rate.  $Y_t$  could be the nominal thickness corresponding to the temperature setting  $U_t$ , and  $X_t$  in this case would be the disturbance that causes deviations in the actual thickness from the nominal thickness. Let  $T$  be the target of the system output. The purpose of the control system in Figure 1 is to estimate  $X_t$  (by  $\hat{X}_t$ ) and then adjust  $U_t$  such that  $Y_t = T - \hat{X}_t$ . With a good estimate of  $X_t$  ( $\hat{X}_t \approx X_t$ ), the system output  $\delta_t$  will approach the target  $T$  ( $\delta_t = T - \hat{X}_t + X_t \approx T$ ).

If the disturbance model is white noise, (i.e.,  $X_t = \varepsilon_t$ , where  $\varepsilon_t$  are normal with mean 0 and constant variance  $\sigma^2$ ), then the best control strategy is no control, because white noise is not predictable. Any attempt to estimate and control white noise will result in overcontrol and thus a larger variation of the system output (Deming 1982). In many applications, a disturbance sequence may be nonstationary (i.e., nonconstant

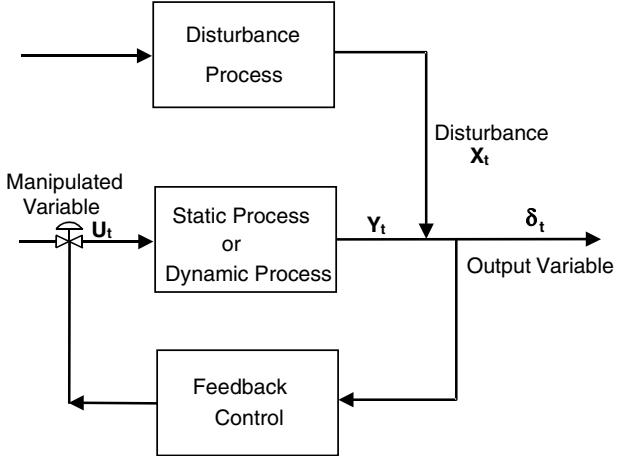


Figure 1. Process and Disturbance Models.

mean). A simple, but very useful nonstationary disturbance model is the IMA(1, 1) model,

$$X_t - X_{t-1} = \varepsilon_t - \theta \varepsilon_{t-1}, \quad (1)$$

where  $\theta$  is a parameter ( $-1 \leq \theta \leq 1$ ) that determines the degree to which the current disturbance is related to the past disturbances. Under this model, we can estimate  $X_t$ , if  $\theta$  is known, by

$$\hat{X}_t = (1 - \theta)X_{t-1} + \theta\hat{X}_{t-1}. \quad (2)$$

This estimate can be expressed recursively as an EWMA,

$$\begin{aligned} \hat{X}_t = \lambda & [X_{t-1} + (1 - \lambda)X_{t-2} + (1 - \lambda)^2X_{t-3} + \dots \\ & + (1 - \lambda)^{t-2}X_1] + (1 - \lambda)^{t-1}X_0, \end{aligned} \quad (3)$$

where we let  $\hat{X}_0 = X_0$  and  $\lambda = 1 - \theta$ .  $\hat{X}_t$  is then used to adjust  $U_t$  to compensate for deviations from the target.

Box and Jenkins (1970) first established the relationships between the IMA(1, 1) model and a PID controller. A similar proof was also provided by MacGregor (1987, 1988). Box and Kramer (1992) again derived this relationship and reported extensive discussions about it by researchers and practitioners. In short, for a second-order dynamic process, the EWMA estimates of disturbances in (3) result in PID adjustments on  $U_t$ . If the disturbance process follows (1), then the PID adjustments can reduce the output variation to the white-noise variation  $\sigma^2$  and is known as a *minimum-variance controller* (Box and Luceño 1997).

## 3. STEP-CHANGE DISTURBANCE PROCESS MODEL

As indicated in (1), the IMA disturbance model is a linear process model in which the present disturbance is linearly correlated to the earlier disturbances. The IMA model can be difficult to interpret physically. Here we propose another kind of disturbance process model that we call a *step-change* model. This model is derived directly from realistic situations in which the process mean is subject to occasionally occurring step changes caused by variations in the physical conditions. For example, a semiconductor fabrication process

is very sensitive to small changes in environmental temperature and raw material qualities, which can be modeled as random step changes.

Let  $X_t$  denote the disturbance normally distributed with mean  $\mu_t$  and variance  $\sigma^2$  at time  $t$ , that is,

$$X_t = \mu_t + \varepsilon_t$$

where  $\mu_0 \sim f(\xi, \tau^2)$ , and

$$\mu_t = \begin{cases} \mu_{t-1} & \text{with probability } 1-p, \quad t=1, 2, \dots \\ \sim f(\xi, \tau^2) & \text{with probability } p \end{cases} \quad (4)$$

The constant probability ( $p$ ) represents the frequency of changes to the process and an assumption that the step-change occurrence is independent of the prior history of the process. The foregoing model is similar to that of Chernoff and Zacks (1964) but differs in that their model deals with accumulated step changes, rather than randomly and independently distributed step changes. Harrison and Stevens (1971, 1976) used multiprocess state-space models with Bayesian forecasting techniques to forecast the time series subject to random accumulated step changes and drift. Yashchin (1995) showed that the EWMA statistic is an optimum estimator among weighted-average estimation procedures for processes subject to one sudden change in the mean level. In this article we study the design and performance analysis of the EWMA mean estimator for processes subject to independent step changes with random size and occurrence.

#### 4. DESIGN OF THE EXPONENTIALLY WEIGHTED MOVING AVERAGE ESTIMATOR FOR STEP-CHANGE DISTURBANCES

When a control mechanism is present as in Figure 1, the objective of the feedback control scheme is to adjust the manipulated variable such that the system output is as close to the target value as possible. When the system is subject to occasional mean step changes, the objective is to estimate the true mean of the disturbance so that the variation of the system can be reduced by the control mechanism. The EWMA estimator in (2) is now used to estimate  $\mu_t$ ,

$$\hat{\mu}_t = \lambda X_{t-1} + (1 - \lambda) \hat{\mu}_{t-1}, \quad (5)$$

where  $\hat{\mu}_0 = X_0$  is an estimate of the overall mean  $\xi$  from historic data. This estimator may not be ideal for estimating the mean in the presence of step changes because it takes into account all past data, whereas the data before the most recent step change is actually irrelevant. Nonetheless, the advantages of the EWMA estimator include its simple computation and PID-controller implementation. To choose an appropriate value for  $\lambda$  in (5), we rely on the following observation.

*Observation 1.* If random variables  $X_1, X_2, \dots, X_t, \dots$  follow the distribution laws given in (4), then

- (a)  $E(\mu_t) = \xi$ ;
- (b)  $E(\mu_t^2) = \tau^2 + \xi^2$  and  $\text{var}(\mu_t) = \tau^2$ ;
- (c)  $E(\mu_t \mu_{t+k}) = (1-p)^k \tau^2 + \xi^2$  and  $\text{cov}(\mu_t, \mu_{t+k}) = (1-p)^k \tau^2$ ; and

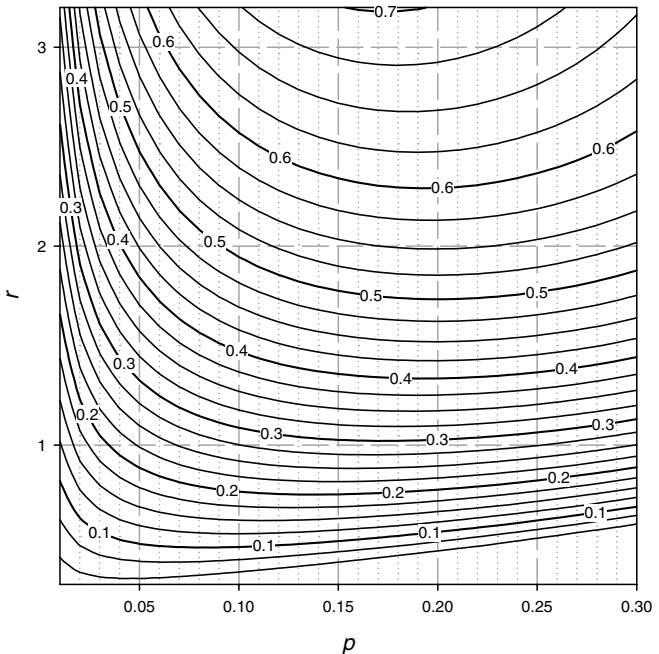


Figure 2. Contours of  $\lambda^*$  for Various Values of  $r = \tau/\sigma$  and  $p$ .

$$(d) \quad \lim_{t \rightarrow \infty} \text{Risk}(\mu_t, \hat{\mu}_t) = \lim_{t \rightarrow \infty} E[L(\mu_t, \hat{\mu}_t)] \\ = \frac{\lambda}{2-\lambda} \sigma^2 + \frac{2p}{(2-\lambda)[1-(1-p)(1-\lambda)^2]} \tau^2, \quad (6)$$

where  $L(\mu_t, \hat{\mu}_t) = (\mu_t - \hat{\mu}_t)^2$  and  $0 < p < 1$ .

The proof of Observation 1 is given in Appendix A. Equation (6) is a convex function of  $\lambda$  for  $0 \leq \lambda \leq 1$ , so a closed-form solution for  $\lambda^*$  that minimizes eq. (6), given the standard deviation ratio  $r = \tau/\sigma$  and the step-change occurrence probability  $p$ , is

$$\lambda^* = \frac{-p(1+r^2) + r\sqrt{p^2r^2 - p^2 + 2p}}{(1-p)}. \quad (7)$$

Figure 2 shows the contours of  $\lambda^*$  for various values of  $r (= \tau/\sigma)$  and  $p$ . As expected,  $\lambda^*$  increases as the size of the step changes increases or as the step change occurs more frequently. When the step-change occurrence is less frequent ( $p < .02$ ), the value of  $\lambda^*$  (between 0 and .3) is determined mainly by the occurrence frequency, and the impact of the step-change size on  $\lambda^*$  will not be as significant. Similarly, the effect of the step-change occurrence frequency on  $\lambda^*$  (with a value between 0 and .2) decreases with smaller step-change size ( $\tau/\sigma < 1.0$ ).

Determination of  $\lambda^*$  from (7) or Figure 2 requires prior knowledge of the step-change occurrence probability ( $p$ ) and the ratio of the natural process variation to the step-change variation ( $r$ ). The next section presents procedures to identify the step-change disturbance process and to estimate  $p$  and  $r$ .

#### 5. IDENTIFICATION AND PARAMETER ESTIMATION OF THE STEP-CHANGE DISTURBANCE PROCESS

The identification of process dynamics in a control system has been discussed by, for example, authors in a special issue of *Automatica* (1981), Åström and Wittenmark (1984) and

Davis and Vinter (1985). Here we focus only on the identification of the step-change disturbances. When we assume, without loss of generality, that  $Y_t = 0$  and that the system described in Figure 1 is not controlled, then the process output is determined solely by the disturbance; that is,  $\delta_t = X_t$ .

Box and Jenkins (1970) identified disturbance (or time series) models through graphical representations of autocorrelation and partial autocorrelation functions. The methodology has been well received by practitioners and has been implemented as a standard procedure in most major statistical software packages, including SAS, SPSS, and Statgraphics, so we incorporate autocorrelation functions in our proposal for identifying the step-change disturbances. From Observation 1(a)–(c), the process defined in (4) constitutes a stationary stochastic process. The autocorrelation function (ACF) at lag  $k$ ,  $\rho_k$ , is estimated by the sample autocorrelation function (SACF),

$$\hat{\rho}_k = \frac{c_k}{c_0},$$

where  $c_k = \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X})$ ,  $n$  is the sample size, and  $\bar{X}$  is the sample mean. From (4), the SACFs will probably not be significantly different from 0 if the sample size is too small to contain any step changes.

*Example.* Thickness data of the  $\text{SiO}_2$  layer were sampled after each run of an uncontrolled oxidation process. The run-to-run thickness sample means were collected from December 12, 1996 to January 29, 1997 in a Taiwan Semiconductor Manufacturing Company (TSMC) facility. We have intentionally normalized and scaled the data because they are proprietary. Figure 3 shows the normalized and scaled thickness sample means. Figure 4 illustrates the SACFs calculated using 30 thickness sample-mean observations (from the 15th to the 44th) taken from a total of 175 observations in Figure 3. No significant correlation is seen.

The significance of autocorrelations is observed only when the sample size is enlarged to cover several step changes in the disturbances. The following observation describes the behavior of the ACFs when step changes are present.

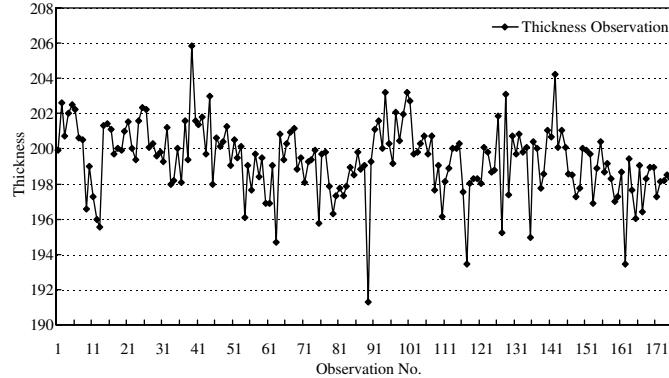


Figure 3. Thickness Sample Mean Observations From TSMC's Oxidation Process.

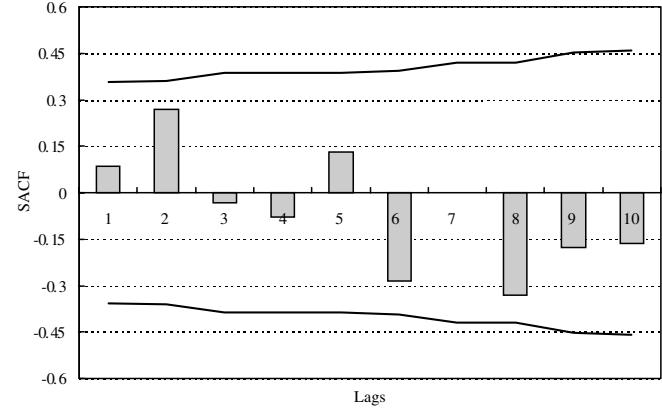


Figure 4. SACF Calculated From a Small Oxidation Thickness Sample (■ SACF; — two standard error limits).

*Observation 2.* If random variables  $X_1, X_2, \dots$  follow the distribution laws given in (4), then

$$\begin{aligned} \rho_k(X_t) &= \frac{E[(X_t - \xi)(X_{t+k} - \xi)]}{\sqrt{E(X_t - \xi)^2 E(X_{t+k} - \xi)^2}} \\ &= \frac{r^2}{r^2 + \sigma^2} (1-p)^k = \frac{r^2}{r^2 + 1} (1-p)^k. \end{aligned} \quad (8)$$

The proof is given in Appendix B. This exponential decay of the theoretical autocorrelation function would not be observed in SACFs until the sample includes more than one step-change occurrence. The drastic change in the sample autocorrelation function from a small to a larger sample size is a unique property of the step-change disturbance model and is not seen in linear disturbance models such as the IMA model. Equation (8) suggests that we can estimate  $p$  by fitting the nonlinear regression model,

$$\hat{\rho}_k = c \times (1 - \hat{p})^k, \quad k = 1, 2, 3, \dots,$$

and then estimate  $r$  by  $[c/(1-c)]^{1/2}$ . When all observed SACFs are nonnegative, we can fit the linear regression model  $\log \hat{\rho}_k = A + Bk$  and estimate  $r$  and  $p$  by  $[\exp A/(1 - \exp A)]^{1/2}$  and  $1 - \exp B$ .

These estimates  $\hat{p}$  and  $\hat{r}$  provide sufficient knowledge about the disturbance process to design an optimized EWMA mean estimator and are simpler to compute than those for linear disturbance processes. Note that knowledge of the times when new raw material batches are introduced to the process or when possible step-change incidents occur can be used to estimate  $p$  more efficiently.

Further exploration of the proposed estimation procedure's stability and its impact on the EWMA estimator may be conducted in a future study. In this article we use an example to illustrate the procedure and evaluate the resulting EWMA mean estimator.

*Example (Continued).* Figure 5 shows the SACFs and the fitted model calculated from all 175 thickness observations in Figure 3. The fitted model is  $\hat{\rho}_k = .437(1 - .253)^k$ , yielding  $\hat{p} = .253$  and  $\hat{r} = (.437/.563)^{1/2} = .88$ , whence  $\lambda^* = .225$  by (7) or Figure 2.

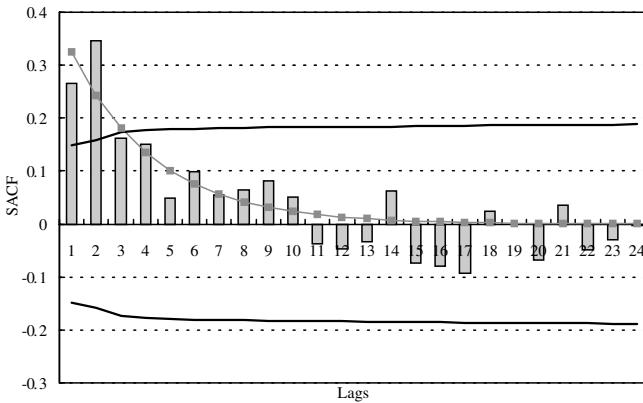


Figure 5. SACF From the Oxidation Thickness Sample Versus Fitted ACF From the Step-Change Disturbance Model (■ SACF; — fitted ACF; — two standard error limits).

Another property of the ACFs for step-change disturbances can be observed from the difference of two consecutive disturbances. Let  $Z_t = X_t - X_{t-1}$ .

*Observation 3.* If random variables  $X_1, X_2, \dots$  follow the distribution laws given in (4), then

$$\rho_1(Z_t) = \frac{E(Z_t Z_{t+1})}{\sqrt{E(Z_t^2) E(Z_{t+1}^2)}} = -\frac{1}{2} \left[ 1 - \frac{p(1-p)\tau^2}{\sigma^2 + p\tau^2} \right].$$

When  $X_t$  is white noise (i.e.,  $p = 0$ ), the ACF of  $Z_t$  equals  $-1/2$  at lag 1 and 0 at lags greater than 1. Observation 3 (see App. C) shows that the ACF of  $Z_t$  for disturbances  $X_t$  following (6) behave similarly with  $\rho_1(Z_t)$  close to  $-1/2$  and correlation functions at lags greater than 1 equal to 0. In other words, although the ACF pattern of  $X_t$  can vary greatly with different sample sizes, the ACF pattern of  $Z_t$  will look quite consistently similar to white noise. This property of  $Z_t$  also can be useful for identifying step-change disturbances. Figure 6 shows the SACF of  $Z_t$  calculated from the  $\text{SiO}_2$  thickness observations.

An estimate of the white-noise variance ( $\sigma^2$ ), the process lower-bound variance that a controller can achieve, compared to the uncontrolled process sample variance  $s^2$ , indicates possible room for improvement in process quality. To estimate  $\sigma$ ,

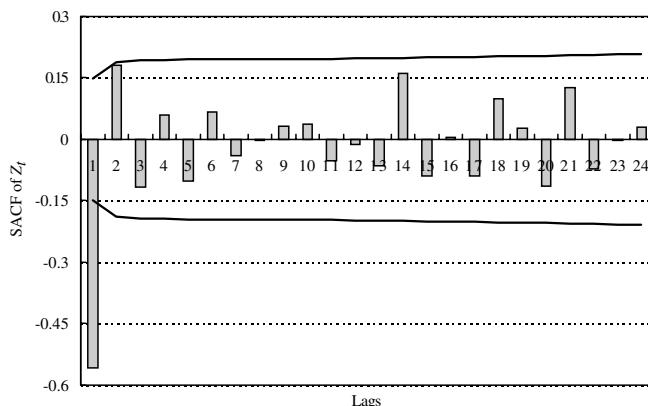


Figure 6. SACF of  $Z_t$  From the Oxidation Thickness Sample (■ SACF; — two standard error limits).

we use the average moving range. If  $X_1$  and  $X_2$  are normal with means  $\mu_1$  and  $\mu_2$  and constant variance  $\sigma^2$ , where the means,  $\mu_1$  and  $\mu_2$ , are iid normal with constant mean and variance  $\tau^2$ , then (see also Vardeman and Jobe 1999)

$$E(|X_1 - X_2|) = \frac{2\sigma}{\sqrt{\pi}} \sqrt{1 + \frac{\tau^2}{\sigma^2}}. \quad (9)$$

*Observation 4.* If random variables  $X_1, X_2, \dots$  follow the distribution laws given in (4) and  $f(\xi, \tau^2) = N(\xi, \tau^2)$ , then

$$E\left(\frac{\overline{MR}}{1.128}\right) = \sigma \left[ 1 - p + p \sqrt{1 + \frac{\tau^2}{\sigma^2}} \right], \quad (10)$$

where

$$\overline{MR} = \frac{\sum_{i=1}^{n-1} |X_i - X_{i+1}|}{n-1}.$$

Observation 4 is a direct result of (4) and (9). Using this observation and the previously estimated  $p$  and  $\tau/\sigma$ , we can estimate  $\sigma$ ,

$$\hat{\sigma} = \frac{\overline{MR}}{1.128[1 - \hat{p} + \hat{p}\sqrt{1 + \hat{r}^2}]} \quad (11)$$

The white noise standard deviation for the oxidation layer thickness is estimated using (11),

$$\hat{\sigma} = \frac{\overline{MR}}{1.128[1 - .253 + .253\sqrt{1 + (.88)^2}]} = 1.443,$$

or  $\hat{\sigma}^2 = 2.081$ . Because sample variance ( $s^2$ ) of thickness observations is 4.100, the potential for quality variation reduction is almost 50%.

Before any control action is taken, it is important to ensure that the controller does not overcontrol and thus introduce larger variations in the process. The graphical display of SACFs can serve as a guide. When the SACFs all appear to be within the range of two standard errors, this indicates that the step changes either occur very rarely or are so small that the process engineer may ignore them and choose not to control the process at all. A larger sample size may be needed to introduce at least one step change in the process so that significant autocorrelations can be observed. For example, step-change disturbances with  $p = .01$  demand at least several hundred observations to enable detection of significant autocorrelations. Many production processes, including the oxidation process, already have sample data taken for SPC charting. The existing historic data for control charting may be sufficient for analysis and design of the controller. For example, the thickness data in oxidation processing were originally collected for SPC charting, but are shown here to have a great potential for further improvement by run-to-run control (Moyne et al. 2000).

## 6. PERFORMANCE OF THE EXPONENTIALLY WEIGHTED MOVING AVERAGE ESTIMATOR

Figure 7(a) and (b) show pure white noise disturbances and disturbances with step changes ( $p = .02$ ,  $\tau/\sigma = 1$ ). Comparison of the two figures reveals the difficulty of detecting by eye the step changes in the disturbance process. If observers of the figures are not informed of the true models, then the two disturbance models may look very similar, because the

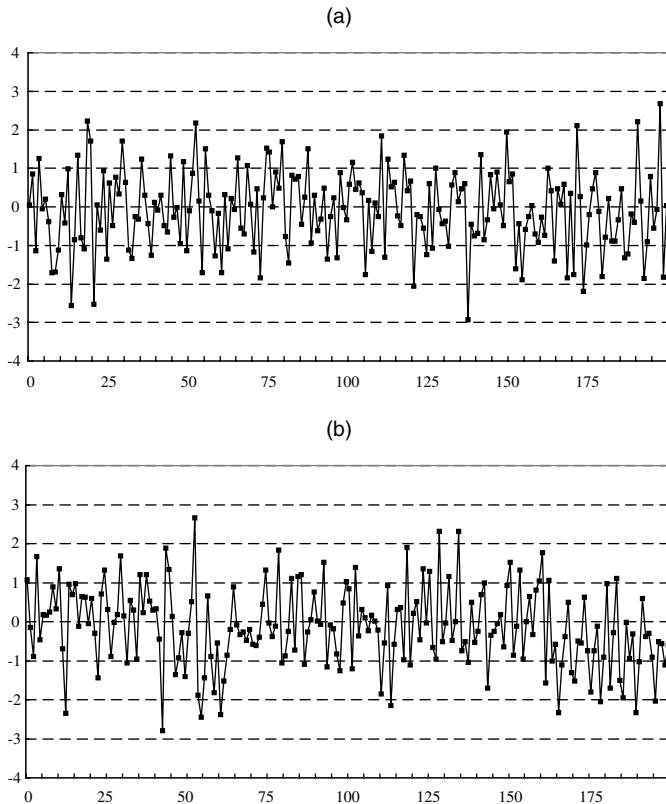


Figure 7. (a) Simulated 200 White-Noise Observations and (b) Simulated 200 Step-Change Disturbance Observations With  $\tau/\sigma = 1.0$  and  $p = .02$ .

disturbances in both figures vary about the center line. The only difference seems to be the larger variation in Figure 7(b). Even if the step changes were observed, it would be still difficult to identify the points of changes and the new means after changes.

Researchers and practitioners have discussed whether a process as in Figure 7(b) should be adjusted through a feedback control loop or simply monitored by SPC charts (Box and Kramer 1992). In semiconductor manufacturing, the run-to-run thickness data of oxidation layers are used mainly for SPC charting, and corrective actions are taken only for out-of-control incidents. In this section we show that the EWMA control actions are quite effective even for cases with rare and small step changes.

An advantage of the EWMA mean estimator is that it does not require the identification of changepoints. The estimator filters out the noise and gives a clearer picture of how the true mean level varies. Figure 8 shows the disturbances with the means estimated by the EWMA estimator with  $\lambda^* = .17$  found by (7) or Figure 2 for  $p = .02$  and  $\tau/\sigma = 1$ .

This figure indicates several mean level changes. However, the true mean level is distorted by the fluctuation of the mean estimates. As discussed earlier, other estimators may exist that can further smooth the fluctuation of the estimated means (Yashchin 1995; Chen and Elsayed 2000). Yashchin (1995) proposed some alternatives to the EWMA procedure for processes with one fixed-size mean step change. The results, however, show that the EWMA statistic still outperforms the alternatives for step-change sizes smaller than  $2\sigma$ . Thus, the

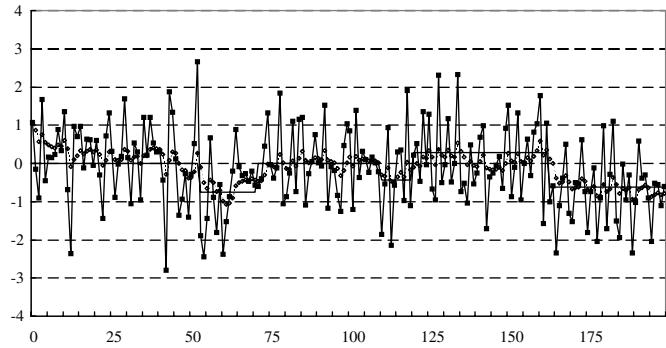


Figure 8. Estimated Means Using EWMA for  $\tau/\sigma = 1.0$  and  $p = .02$   
 (— true mean; —■— observation; ···◇··· EWMA estimate).

EWMA mean estimator seems to be a natural choice for processes with small step changes ( $<2\sigma$ ). Another factor affecting the EWMA estimator's performance, but rarely discussed in the literature, is the frequency of step-change occurrence. In this section we analyze the performance of the EWMA estimator, or, equivalently, the PID controller, for random step changes using mean squared error (MSE).

For a process without control (i.e., with an open control loop), its MSE, denoted as  $MSE_o$ , can be calculated as

$$\begin{aligned} MSE_o(t) &= E(\delta_t - T)^2 = E(Y_t + X_t - T)^2 = E(X_t - \xi)^2 \\ &= E(\mu_t^2) + E(\varepsilon_t^2) - \xi^2 = \sigma^2 + \tau^2. \quad (12) \end{aligned}$$

This result is based on Observation 1(b) and an assumption that  $Y_t$  has been previously set to offset the overall disturbance mean  $\xi$ ; that is,  $Y_t = T - \xi$ . When the feedback controller is implemented as proposed, its MSE, denoted by  $MSE_c$ , becomes

$$\begin{aligned} MSE_c(t) &= E(\delta_t - T)^2 = E(-\hat{\mu}_t + \mu_t + \varepsilon_t)^2 \\ &= E(\mu_t - \hat{\mu}_t)^2 + \sigma^2 \end{aligned}$$

as  $t \rightarrow \infty$ , and

$$MSE_c = \lim_{t \rightarrow \infty} MSE_c(t) = \sigma^2 + \frac{\lambda}{2-\lambda}\sigma^2 + \frac{2p}{(2-\lambda)[1-(1-p)(1-\lambda)]}\tau^2, \quad (13)$$

a direct result from Observation 1(d). Because the MSE measure is directly proportional to Taguchi's definition of quality loss (see Taguchi, Elsayed, and Hsiang 1989), we evaluate the EWMA controller's performance by calculating the percent MSE improvement,

$$\text{MSE improvement\%}(I\%) = \frac{MSE_o - MSE_c}{MSE_o} \times 100\%. \quad (14)$$

Using (12), (13), and (14), Figure 9 shows the contours of  $I\%$  given  $\lambda^*$  for various values of  $r (= \tau/\sigma)$  and  $p$ , from which we estimate the reduction in the oxidation process MSE as about 7%. Applying the designed EWMA ( $\lambda^* = .225$ ) estimator to the oxidation data in Figure 3, the actual MSE reduction is 8.42% (from 4.1 to 3.755). Figure 10 shows the estimated mean of the  $\text{SiO}_2$  thickness using the designed EWMA estimator.

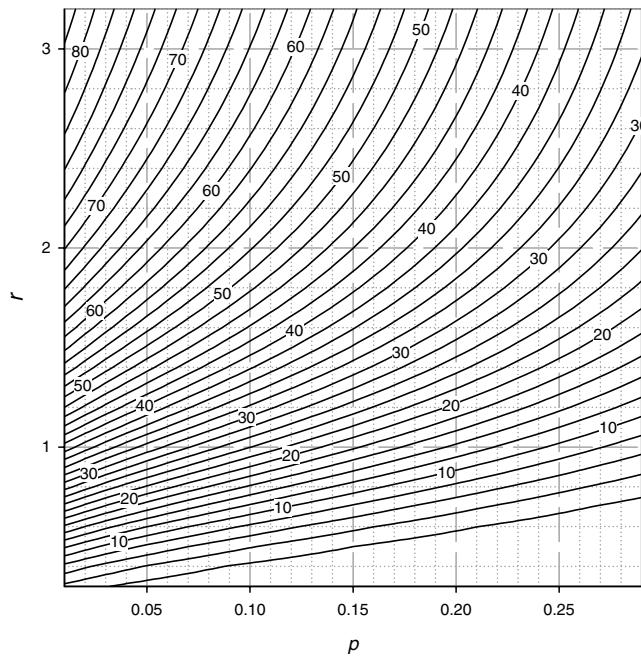


Figure 9. Contours of  $I\%$  for Various Values of  $r = \tau/\sigma$  and  $p$ .

Although the research focus of this article is on understanding step-change disturbances and the corresponding EWMA design and performance, we compare the effectiveness of the proposed procedure to other design procedures, including least squares (LS) fitting. The design quality of EWMA estimators results directly from the efficiency of disturbance model estimation. Both deserve more rigorous studies in future research. In the final section, we briefly address the problem of LS fitting.

In addition to the percent MSE improvement, we also compare the performances of the EWMA and the ideal minimum-variance controllers via the measure

Control capability% ( $C\%$ )

$$\begin{aligned} &= \frac{MSE_o - MSE_c}{MSE_o - \sigma^2} \times 100\% \\ &= \left( 1 - \frac{2p}{(2-\lambda)[1-(1-p)(1-\lambda)]} - \frac{\lambda}{(2-\lambda)r^2} \right) \times 100\% \end{aligned}$$

With  $\lambda^*$  specified by (7), Figure 11 shows the contours of  $C\%$  for various values of  $r$  and  $p$ .

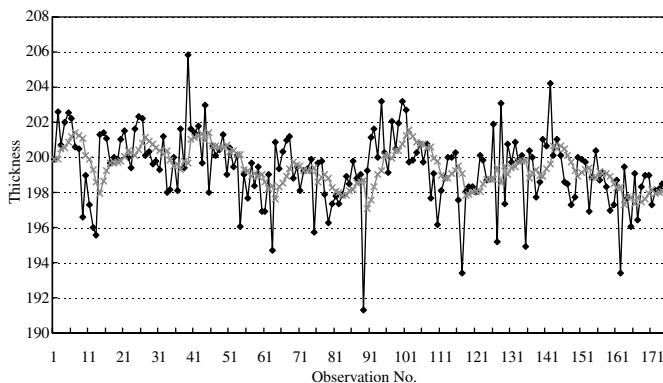


Figure 10. EWMA Estimated Thickness ( $\times$ ) Versus Actual Thickness Observations (—◆—).

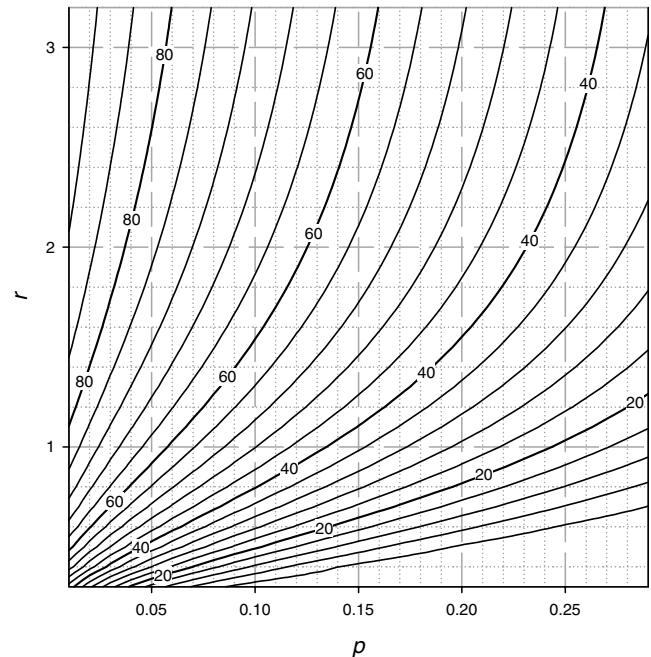


Figure 11. Contours of  $C\%$  for Various Values of  $r = \tau/\sigma$  and  $p$ .

To illustrate the EWMA estimator's overall performance, we consider six cases obtained by all combinations of  $\tau/\sigma = 1.0$  or  $.5$  and  $p = .1, .02$ , or  $.01$ . Figure 9 shows that the EWMA controller can reduce the MSE by 25% to 40% for cases with  $r = 1.0$ . When the step change probability  $p$  is  $.1$  (i.e., the historic observations quickly become irrelevant to the current observation), the EWMA controller remarkably improves the MSE by 25%. In this case, the EWMA controller is 50% as good as the minimum-variance controller. Cases with  $r = 1.0$  and  $p = .02$  or  $.01$  are the most likely cases in practice. The step-change occurrence rate (1 every 50 to 100) is reasonable, and the step-change sizes are difficult to detect by SPC charts. Both cases have MSE improvement exceeding 35%, and the controller is about 75% as good as an ideal controller.

For cases with  $r = .5$ , the EWMA controller improves the MSE by 5% to 14%. In these cases, the step changes are relatively small ( $\tau/\sigma = .5$ ), and the room for improvement is very limited. The case with  $r = .5$  and  $p = .1$  is an extreme case in which the EWMA controller can achieve only 5% improvement. Any cases with  $r < .5$  and  $p > 0$  are considered not cost-effective for control actions. For cases with  $p = .02$  or  $.01$ , the EWMA controller impressively reduces the MSE by more than 10% and is 58% as capable as an ideal controller.

Overall, a controller based on the EWMA estimator can improve the MSE up to 60% for disturbances with less frequent (e.g.,  $p < .05$ ) and relatively large (e.g.,  $\tau/\sigma > 2$ ) mean step changes and is almost as capable as an ideal controller (more than 80%). The controller becomes less effective (20% or less of an ideal controller's capability), when the mean step changes occur very frequently (e.g.,  $p > .1$ ) and are relatively small (e.g.,  $\tau/\sigma < .5$ ).

## 7. DISCUSSION AND CONCLUDING REMARKS

Several issues could be potentially critical to the applicability of the proposed procedure.

## 7.1 Assumptions of the Step-Change Disturbance Model in Practice

Recently, many high-tech manufacturing processes, such as in semiconductor fabrication, are becoming more and more complex, involving many raw materials and equipment parts, where changes in quality may be overlooked. In practice, the quality of the incoming materials and parts should be tightly controlled by means of incoming material quality control (IQC) as a part of the total quality control (TQC) system, but variations in the raw material quality are still inevitable. The model of step-change disturbances will describe a large class of disturbances actually observed.

An important assumption of the proposed disturbance model is the constant step-change occurrence rate. This assumption is equivalent to assuming the Poisson arrival of step changes. Although more case studies may be needed to validate its general applicability, the assumption seems theoretically rational because of the numerous possible step-change causes present in the process and the slim chance for each cause to actually occur. (The Poisson distribution is the limiting distribution of binomial distribution).

## 7.2 Practical Issues of Step-Change Disturbance Identification

We have proposed procedures to identify and estimate the step-change disturbance model. Among these procedures, the insignificance of SACFs with a small sample of data (20–100 sample observations) is distinct from other linear disturbance models and could mislead practitioners into thinking that the process is already in control. We suggest that at least 300–500 sample observations (3–5 step-change occurrences on average for disturbances with  $p = .01$ ) should be collected to calculate the SACFs before making a decision on whether or not a further control is needed. This recommendation is based on the fact that a minimum of 10% MSE reduction is almost warranted if the SACFs are found to be significant for the suggested sample size. In the case of short-run production, which is increasingly important in modern high-mix production settings, sample observations from different production recipes may need to be normalized to the same scale to form a sufficiently large sample size for analysis.

In addition to graphical displays of SACFs, a contrast between the sample variance  $s^2$  and the estimated white-noise variance  $\hat{\sigma}^2$  also manifests the special nature of step-change disturbances and the need to further tighten the system variation by means of process adjustment, such as EWMA run-to-run controllers. A simple statistical hypothesis test

proposed by Kamat (1953) is based on a similar contrast. The test statistic is

$$W = \frac{\sum_{i=1}^{n-1} \frac{|X_i - X_{i+1}|}{n-1}}{\sqrt{\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n}}} = \frac{\bar{MR}}{s}. \quad (15)$$

For randomly occurring small step changes, let the null hypothesis be  $H_0: p = 0$ . From (10) and (D.1), it can be seen that for  $p > 0$  and  $\tau^2 > 0$ , even just small positive values, the denominator of the test statistic (15) becomes significantly larger than the numerator and results in the rejection of the null hypothesis. (Critical values can be found in Kamat 1953). This test is powerful in detecting hard-to-notice irregular processes, such as small step changes, but it has received little attention from practitioners. In Appendix D we also propose a procedure for estimating  $\tau$ .

## 7.3 Issues of EWMA Parameter Design

Without understanding the nature of the process, one can always fit the EWMA statistic [or the IMA(1, 1) model] to the observed sample data to find a LS-fitted  $\lambda$ . In the oxidation thickness example, the LS-fitted  $\lambda$  is .215 with minimum  $MSE = 3.754$ . This result is very close to that obtained from estimating parameters of a step-change disturbance process. For a given set of observed data, the LS-fitted  $\lambda$  minimizes the MSE, but it is optimal only for this particular set of sample data and cannot guarantee an optimum EWMA design ( $\lambda^*$ ) for the entire disturbance process. If one has little interest in the disturbance nature and a sufficiently large sample of observations is readily collected, then LS fitting will always be an easy and quick approach to finding a good  $\lambda$  value. With relatively small sample size and smaller step changes, the performance of the LS fitting procedure would be in question. A complete future study is needed to compare the LS fitting procedure and the proposed step-change estimation procedure.

To illustrate the possible ineffectiveness of fitting  $\lambda$  by LS for certain step-change disturbances under limited sample data collection, we simulated 10 processes each with two types of step-change disturbances,  $p = .02$  and  $r = .5$ , and  $p = .02$  and  $r = 1.0$ . We test two sample sizes, 1,000 and 500, which have an average of 20 and 10 step-change occurrences. Table 1 shows the quadratic estimation loss (QL) for the LS-fitting and step-change estimation procedures. The results show that the simple LS-fitting procedure was sufficient to achieve a good EWMA design without knowing the disturbance nature for disturbances with larger shifts. But in cases of small step-change size ( $r = .5$ ), the step-change estimation procedure outperforms the LS-fitting procedure,

Table 1. Comparison of Two EWMA Design Procedures

|                       | $p = .02$            | Sample size 1,000                      | Sample size 500                        |
|-----------------------|----------------------|--|--|
| $r = .5$              | LS fitting           | $\bar{\lambda} = .0815$ ; QL = .000594 | $\bar{\lambda} = .0850$ ; QL = .001234 |
| $(\lambda^* = .0761)$ | Step-change estimate | $\bar{\lambda} = .0730$ ; QL = .000371 | $\bar{\lambda} = .0720$ ; QL = .000803 |
| $r = 1.0$             | LS fitting           | $\bar{\lambda} = .1451$ ; QL = .000751 | $\bar{\lambda} = .1447$ ; QL = .002036 |
| $(\lambda^* = .1633)$ | Step-change estimate | $\bar{\lambda} = .1404$ ; QL = .001046 | $\bar{\lambda} = .1404$ ; QL = .002497 |

NOTE:  $\bar{\lambda}$  is the average estimated value from 10 simulated datasets; QL is the quadratic estimation loss estimated from 10 simulated datasets.

indicating that there are opportunities for improvement in such situations.

#### APPENDIX A: PROOF OF OBSERVATION 1

(a)

$$\begin{aligned} E(\mu_t) &= E(\mu_t | \text{step change at } t) \times p \\ &\quad + E(\mu_t | \text{no step change at } t) \times (1-p) \\ &= \xi p + E(\mu_{t-1})(1-p). \end{aligned}$$

Recursively, we obtain

$$\begin{aligned} E(\mu_t) &= \xi p + \xi p(1-p) + \xi p(1-p)^2 + \dots \\ &\quad + \xi P(1-p)^{t-1} + E(\mu_0)(1-p)^t = \xi, \end{aligned}$$

because  $E(\mu_0) = \xi$ .

(b)

$$\begin{aligned} E(\mu_t^2) &= E(\mu_t^2 | \text{step change at } t) \times p \\ &\quad + E(\mu_t^2 | \text{no step change at } t) \times (1-p) \\ &= (\tau^2 + \xi^2)p + E(\mu_{t-1}^2)(1-p). \end{aligned}$$

Recursively, we obtain

$$\begin{aligned} E(\mu_t^2) &= (\tau^2 + \xi^2)p + (\tau^2 + \xi^2)p(1-p) + \dots \\ &\quad + (\tau^2 + \xi^2)p(1-p)^{t-1} + E(\mu_0^2)(1-p)^t, \\ &= \tau^2 + \xi^2 \end{aligned}$$

because  $E(\mu_0^2) = \tau^2 + \xi^2$ . Thus

$$\text{var}(\mu_t) = E(\mu_t^2) - [E(\mu_t)]^2 = \tau^2.$$

(c)

$$\begin{aligned} E(\mu_t \mu_{t+k}) &= E(\mu_t^2 | \text{one or more step changes in } [t, t+k]) \\ &\quad \times [1 - (1-p)^k] + E(\mu_t \mu_{t+k} | \text{no step changes} \\ &\quad \text{in } [t, t+k])(1-p)^k \\ &= E(\mu_t)^2[1 - (1-p)^2] + E(\mu_t^2)(1-p)^k \\ &= \xi^2[1 - (1-p)^k] + (\tau^2 + \xi^2)(1-p)^k \\ &= \xi^2 + \tau^2(1-p)^k. \end{aligned}$$

Thus, from (a),

$$\begin{aligned} \text{cov}(\mu_t, \mu_{t+k}) &= E[(\mu_t - E(\mu_t))(\mu_{t+k} - E(\mu_{t+k}))] \\ &= (1-p)^k \tau^2 \end{aligned}$$

(d)

$$\begin{aligned} \lim_{t \rightarrow \infty} E(\mu_t - \hat{\mu}_t)^2 &= \lim_{t \rightarrow \infty} [\text{var}(\mu_t - \hat{\mu}_t) + (E(\mu_t) - E(\hat{\mu}_t))^2] = \lim_{t \rightarrow \infty} \text{var}(\mu_t - \hat{\mu}_t) \\ &= \lim_{t \rightarrow \infty} \text{var}(\mu_t - \lambda X_{t-1} - \lambda(1-\lambda)X_{t-2} - \lambda(1-\lambda)^2X_{t-3} - \dots) \\ &= \lim_{t \rightarrow \infty} \text{var}(\mu_t - \lambda\mu_{t-1} - \lambda(1-\lambda)\mu_{t-2} - \dots) + \frac{\lambda^2}{1-(1-\lambda)^2}\sigma^2, \end{aligned}$$

because  $\mu_t$  and  $\varepsilon_t$  are independent of each other. From (b) and (c), we have shown that  $\text{var}(\mu_t) = \tau^2$  and  $\text{cov}(\mu_t, \mu_{t+k}) = (1-p)^k \tau^2$ . Thus

$$\begin{aligned} \lim_{t \rightarrow \infty} \text{var}(\mu_t - \lambda\mu_{t-1} - \lambda(1-\lambda)\mu_{t-2} - \lambda(1-\lambda)^2\mu_{t-3} - \dots) &= \lim_{t \rightarrow \infty} \text{var}(\mu_t) + \lambda^2 \lim_{t \rightarrow \infty} \text{var}(\mu_{t-1}) + \lambda^2(1-\lambda)^2 \lim_{t \rightarrow \infty} \text{var}(\mu_{t-2}) \\ &\quad + \dots - 2 \lim_{t \rightarrow \infty} \sum_{i=1}^t \lambda(1-\lambda)^{i-1} \text{cov}(\mu_t, \mu_{t-i}) \\ &\quad + 2 \lim_{t \rightarrow \infty} \sum_{i=1}^{t-1} \sum_{j=i+1}^t \lambda^2(1-\lambda)^{i+j-2} \text{cov}(\mu_{t-i}, \mu_{t-j}) \\ &= \lim_{t \rightarrow \infty} \text{var}(\mu_t) + \lambda^2 \lim_{t \rightarrow \infty} \text{var}(\mu_{t-1}) + \lambda^2(1-\lambda)^2 \lim_{t \rightarrow \infty} \text{var}(\mu_{t-2}) \\ &\quad + \dots - 2 \lim_{t \rightarrow \infty} \sum_{i=1}^t \lambda(1-\lambda)^{i-1}(1-p)^i \tau^2 \\ &\quad + 2 \lim_{t \rightarrow \infty} \sum_{i=1}^{t-1} \sum_{j=i+1}^t \lambda^2(1-\lambda)^{i+j-2}(1-p)^{j-i} \tau^2 \\ &= \frac{2p}{(2-\lambda)[1 - (1-\lambda)(1-p)]} \tau^2. \end{aligned}$$

Finally,

$$\lim_{t \rightarrow \infty} E(\mu_t - \hat{\mu}_t)^2 = \frac{\lambda}{2-\lambda} \sigma^2 + \frac{2p}{(2-\lambda)[1 - (1-\lambda)(1-p)]} \tau^2.$$

#### APPENDIX B: PROOF OF OBSERVATION 2

It can be shown that

$$\begin{aligned} \frac{E[(X_t - \xi)(X_{t+k} - \xi)]}{\sqrt{E(X_t - \xi)^2 E(X_{t+k} - \xi)^2}} &= \frac{E(X_t X_{t+k}) - \xi E(X_t) - \xi E(X_{t+k}) + \xi^2}{\sqrt{E(X_t^2 - 2\xi X_t + \xi^2) E(X_{t+k}^2 - 2\xi X_{t+k} + \xi^2)}}. \end{aligned}$$

Because  $X_t = \mu_t + \varepsilon_t$ ,  $X_{t+k} = \mu_{t+k} + \varepsilon_{t+k}$ , and  $\mu_t$  is independent of  $\varepsilon_t$  for any  $t$ , from Observation 1(a)–(c) we immediately obtain

$$\begin{aligned} E[(X_t - \xi)(X_{t+k} - \xi)] &= E[(\mu_t - \xi)(\mu_{t+k} - \xi)] \\ &= (1-p)^k \tau^2 \quad (\text{B.1}) \end{aligned}$$

and

$$\begin{aligned} \sqrt{E(X_t - \xi)^2 E(X_{t+k} - \xi)^2} &= \sqrt{[\sigma^2 + E(\mu_t^2) - 2\xi E(\mu_t) + \xi^2][\sigma^2 + E(\mu_{t+k}^2) - 2\xi E(\mu_{t+k}) + \xi^2]} \\ &= \sigma^2 + \tau^2. \quad (\text{B.2}) \end{aligned}$$

We finally obtain

$$\rho = \frac{(\text{B.1})}{(\text{B.2})} = \frac{(1-p)^k \tau^2}{\sigma^2 + \tau^2}.$$

## APPENDIX C: PROOF OF OBSERVATION 3

It can be easily shown that

$$\begin{aligned} & \frac{E(Z_t Z_{t+1})}{\sqrt{E(Z_t^2)E(Z_{t+1}^2)}} \\ &= \frac{E(X_{t+1}X_t) + E(X_t X_{t-1}) - E(X_{t+1}X_{t-1}) - E(X_t^2)}{\sqrt{E(X_{t+1}^2 - 2X_{t+1}X_t + X_t^2)E(X_t^2 - 2X_t X_{t-1} + X_{t-1}^2)}}. \end{aligned}$$

Because  $X_i = \mu_i + \varepsilon_i$ , from Observation 1(a)–(c) we obtain

$$\begin{aligned} E(Z_t Z_{t-1}) &= E(\mu_{t+1}\mu_t) + E(\mu_t\mu_{t-1}) - E(\mu_{t+1}\mu_{t-1}) \\ &- \sigma^2 - E(\mu_t^2) = -p^2\tau^2 - \sigma^2 \quad (\text{C.1}) \end{aligned}$$

and

$$\begin{aligned} & \sqrt{E(Z_t^2)E(Z_{t+1}^2)} \\ &= \sqrt{[2\sigma^2 + E(\mu_{t+1} - \mu_t)^2][2\sigma^2 + E(\mu_t - \mu_{t-1})^2]} \\ &= 2\sigma^2 + 2p\tau^2. \quad (\text{C.2}) \end{aligned}$$

Thus

$$\frac{(\text{C.1})}{(\text{C.2})} = \frac{-p^2\tau^2 - \sigma^2}{2\sigma^2 + 2p\tau^2} = -\frac{1}{2} \left[ 1 - \frac{p(1-p)\tau^2}{\sigma^2 + p\tau^2} \right].$$

APPENDIX D: PROCEDURE FOR ESTIMATING  $\tau$ 

To estimate  $\tau$ , the following observation is needed.

*Observation D.1.* If random variables  $X_1, X_2, \dots$  follow the distribution laws given in (4), then

(a)

$$E(s^2) = \sigma^2 + \tau^2 - \frac{2\tau^2}{n(n-1)} \sum_{i=1}^{n-1} i(1-p)^{n-i}$$

(b)

$$E(s^2) = \begin{cases} \sigma^2, & p = 0 \\ \sigma^2 + \tau^2 - \frac{2\tau^2}{n(n-1)} \frac{(1-p)(np-1+(1-p)^n)}{p^2}, & 0 < p < 1 \\ \sigma^2 + \tau^2, & p = 1, \end{cases}$$

where

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \quad \text{and} \quad \bar{X} = \frac{\sum_{i=1}^n X_i}{n}.$$

*Proof.* (a) It can be easily shown that

$$E(s^2) = \sigma^2 + \frac{1}{n-1} E \left( \sum_{i=1}^n (\mu_i - \bar{\mu})^2 \right) \quad (\text{D.1})$$

where

$$\begin{aligned} E \left( \sum_{i=1}^n (\mu_i - \bar{\mu})^2 \right) &= \sum_{i=1}^n E(\mu_i^2) - nE(\bar{\mu}^2) \\ &= (n-1)\tau^2 - \frac{2\tau^2}{n} \sum_{i=1}^{n-1} i(1-p)^{n-i}. \quad (\text{D.2}) \end{aligned}$$

Substituting (D.2) into (D.1), we obtain

$$E(s^2) = \sigma^2 + \tau^2 - \frac{2\tau^2}{n(n-1)} \sum_{i=1}^{n-1} i(1-p)^{n-i}.$$

(b) The results can be established directly from (a).

Because we have already estimated  $p$  and  $\sigma$ , we can then estimate  $\tau$  by the foregoing Observation,

$$\hat{\tau}^2 = \frac{n(n-1)\hat{p}^2}{n(n-1)\hat{p}^2 - 2[(1-\hat{p})(n\hat{p}-1+(1-\hat{p})^n)]}(s^2 - \hat{\sigma}^2).$$

## ACKNOWLEDGEMENTS

The authors thank TSMC for providing the valuable data and Karen Kafadar and the anonymous associate editor and referees for their valuable comments, which have greatly improved the article. This research was funded in part by NSC research project 89-2416-H-002-044.

[Received September 1999. Revised May 2000.]

## REFERENCES

- Åström, K. L., and Hägglund, T. (1995), *PID Controllers* (2nd ed.),—: Instrument Society of America.
- Åström, K. L., and Wittenmark, B. (1984), *Computer Controlled Systems*, Englewood Cliffs, NJ: Prentice Hall.
- Automatica (1981), Special Issue on Identifications and System Parameter Estimation, 17.
- Boning, D., Hurwitz, A., Moyne, J., Moyne, W., Smith, T., Taylor, J., and Telfeyan, R. (1996), “Run-by-Run Control of Chemical Mechanical Polishing,” *IEEE Transactions on Components, Packaging and Manufacturing Technology*, Part C, 19, 307–314.
- Box, G. E., and Jenkins, G. M. (1968), “Some Recent Advances in Forecasting and Control,” *Applied Statistics*, 17, 91–109.
- (1970), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.
- Box, G. E., and Kramer, T. (1992), “Statistical Process Monitoring and Feedback Adjustment—A Discussion,” *Technometrics*, 34, 251–285.
- Box, G. E., and Luceño, A. (1997), *Statistical Control by Monitoring and Feedback Adjustment*, New York: Wiley.
- Chen, A., and Elsayed, E. A. (2000), “An Alternative Mean Estimator for Processes Monitored by SPC Charts,” *International Journal of Production Research*.
- Chernoff, H., and Zacks, S. (1964), “Estimating the Current Mean of a Normal Distribution Which is Subject to Change in Time,” *Annals of Mathematical Statistics*, 35, 999–1018.
- Cox, D. R. (1961), “Prediction by Exponentially Weighted Moving Averages and Related Methods,” *Journal of the Royal Statistical Society, Ser. B*, 23, 414–422.
- Davis, M. H. A., and Vinter, R. B. (1985), *Stochastic Modelling and Control*, New York: Chapman and Hall.
- Deming, W. E. (1982), *Out of Crisis*, Cambridge, MA: MIT Center for Advanced Engineering Study.
- Del Castillo, E., and Hurwitz, A. (1997), “Run-to-Run Process Control: Literature Review and Extensions,” *Journal of Quality Technology*, 29, 184–196.
- Harrison, P. J., and Stevens, C. F. (1971), “A Bayesian Approach to Short-term Forecasting,” *Operational Research Quarterly*, 22, 341–362.
- (1976), “Bayesian Forecasting,” *Journal of the Royal Statistical Society, Ser. B*, 38, 205–247.
- Hunter, J. S. (1986), “The Exponential Weighted Moving Average,” *Journal of Quality Technology*, 18, 203–210.
- Ingolfsson, A., and Sachs, E. (1993), “Stability and Sensitivity of an EWMA Controller,” *Journal of Quality Technology*, 25, 271–287.
- Kamat, A. R. (1953), “On the Mean Successive Difference and Its Ratio to the Root Mean Square,” *Biometrika*, 40, 116–127.

- Lucas, J. M., and Saccucci, M. S. (1992), "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements," *Technometrics*, 32, 1–29.
- MacGregor, J. F. (1987), "Interfaces Between Process Control and Online Statistical Process Control," *Computing and Systems Technology Division Communications*, 10, 9–20.
- (1988), "On Line Statistical Process Control," *Chemical Engineering Progress*, October, 21–31.
- Moyné, J., Del Castillo, E., and Hurwitz, A. M. (2000), *Run to Run Control in Semiconductor Manufacturing*, CRC Press.
- Roberts, S. W. (1959), "Control Chart Tests Based on Geometric Moving Averages," *Technometrics*, 42, 97–101.
- Sachs, E., Hu, A., and Ingolfsson, A. (1995), "Run by Run Process Control: Combining SPC and Feedback Control," *IEEE Transactions in Semiconductor Manufacturing*, 8, 26–43.
- Taguchi, G., Elsayed, E. A., and Hsiang, T. (1989), *Quality Engineering in Production Systems*, New York: McGraw-Hill.
- Yashchin, E. (1995), "Estimating the Current Mean of a Process Subject to Abrupt Changes," *Technometrics*, 37, 311–323.