A General Harmonic Rule Controller for Run-to-Run Process Control

Fangyi He, Kaibo Wang, and Wei Jiang

***Abstract*—The existence of initial bias in parameter estimation is an important issue in controlling short-run processes in semiconductor manufacturing. Harmonic rule has been widely used in machine setup adjustment problems. This paper generalizes the harmonic rule to a new controller called general harmonic rule (GHR) controller in run-to-run process control. The stability and optimality of the GHR controller is discussed for a wide range of stochastic disturbances. A numerical study is performed to compare the sensitivity of the GHR controller, the exponentially weighted moving average (EWMA) controller and the variable EWMA controller. It is shown that the GHR controller is more robust than the EWMA controller when the process parameters are estimated with uncertainty.**

***Index Terms*—Automatic process control, EWMA, robust control, worst case.**

I. Introduction

MANY semiconductor manufacturing processes are suffering from sudden component failures, initial setup bias, gradual wear of components or aging effects. To produce conforming products, feedback controllers are needed for such process to generate control actions and maintain output on target.

The following model has been used extensively in the literature to represent diverse semiconductor processes (see, e.g., Tsung *et al.* [23], Tsung and Apley [22], Apley and Kim [1]):

$$Y_t = \alpha + \beta \cdot x_{t-1} + N_t, \tag{1}$$

where x_{t-1} denotes the process input recipe at the end of run $t-1$ (beginning of run t) and N_t, which may not be white noise (WN), denotes the process disturbance that accounts for the variability in the process. The parameter α is called the offset or intercept and the parameter β is called the process gain or slope. It should be noted that the disturbance N_t, models different types of process faults illustrated above. Initial bias, sudden process shifts and gradual process drift can all be expressed by this model.

An IMA(1,1) model is a commonly used structure in semiconductor process control (Box *et al.* [2], Box and Kramer [4], Janakiram and Keats [12], Montgomery *et al.* [14], Tsung *et al.* [23], Tsung and Apley [22], Vander Wiel *et al.* [26], Vander Wiel [25], Del Castillo and Hurwitz [7], Box and Luceño [3], Luceño [13], Chen and Guo [5], Apley and Kim [1]). An IMA(1,1) disturbance model can characterize the behavior of a non-stationary process, which is

$$N_t = N_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}, \quad 0 < \theta < 1 \tag{2}$$

where ε_t is an independently identically distributed (i.i.d.) sequence of random noise with mean 0 and variance σ_ε^2, i.e., $\{\varepsilon_t\} \sim \mathrm{WN}(0, \sigma_\varepsilon^2)$. θ is the IMA parameter. In order to verify whether a noise sequence follows an IMA(1,1) model, one may take one-step lagged difference of the sequence, and fit an MA(1) model to the differences.

Among others, the exponentially weighted moving average (EWMA) controller has been widely used to compensate process deviations or faults. To adjust process output, Y_{t+1}, to its target value τ, the EWMA controller suggests

$$x_t = \frac{\tau - \alpha}{\beta}.$$

Sachs *et al.* [24] assume that an estimate b of the process gain β is available prior to the beginning of the control session. However, to account for the process disturbance N_t, the intercept α is estimated recursively using an EWMA equation. In this controller, the estimate of α at run t is denoted by a_t. This estimate is then updated using the following EWMA equation,

$$a_t = \omega(Y_t - bx_{t-1}) + (1-\omega)a_{t-1},$$

where ω is a parameter $(0 \leq \omega \leq 1)$ that gives more weight to the most recently observed forecast error of the quality characteristic the closer it is to 1. The control law follows

$$x_t = \frac{\tau - a_t}{b}.$$

If the process gain is estimated accurately, i.e., $b = \beta$, and the process output at time 0 is on target, i.e., $\alpha + \beta x_0 = \tau$, the EWMA controller is the minimum MSE (MMSE) controller when the EWMA parameter ω is set equal to $1-\theta$ (Box *et al.* [2], Sachs *et al.* [24]).

Manuscript received November 09, 2007; revised January 16, 2009. Current version published May 06, 2009. The work of K. Wang was supported by the National Natural Science Foundation of China (NSFC) under Grant 70802034. The work of W. Jiang was supported by The Hong Kong University of Science and Technology under Grant HKUST-DAG08/09.EG10.

F. He is with the School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: fangyi.he@stevents.edu).

K. Wang is with the Department of Industrial Engineering, Tsinghua University, Beijing 100084, China (e-mail: kbwang@tsinghua.edu.cn).

W. Jiang is with the Department of Industrial Engineering and Logistics Management, The Hong Kong University of Science and Technology, Kowloon, Hong Kong (e-mail: jiangwei08@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSM.2009.2017627

However, two critical constrains are confronting the semiconductor manufacturing practice. First, estimates of process parameter are never accurate; parameter estimation uncertainties always exist. Therefore, controllers have to be robust to inaccurate parameter values and guarantee process stability and control performance in all circumstances. Second, short-run processes leave narrow space for process adjustment and call for quick actions to improve transient performance. As a result, when initial setup bias or fault exists, controllers that respond slowly to such shifts or faults are not capable in new manufacturing scenarios.

Patel and Jenkins [16] presented a scheme to change the EWMA weight adaptively to compensate step and drift disturbances. Chen *et al.* [6] demonstrated that an ARI(3,1) model is more suitable for the metal sputter deposition process under investigation. To handle this special type of disturbance, the authors developed an extended Kalman filter controller. Wu *et al.* [28] investigated the effect of metrology delay on control performance. The authors also suggested alternatives to compensate the adverse effect caused by metrology delay when the underlying process exhibits nonstationary or highly autoregressive disturbance. Nonetheless, none of these works focuses on compensation of initial bias and short-run processes.

Short-run processes are becoming more and more prevalent due to high-degree customization, frequent process maintenance and the mix-product trend in semiconductor manufacturing (see, e.g., Sullo and Vandeveen [18], Pan [15], Tsiamyrtzis and Hawkins [21]). For example, in a wafer preparing process, all facilities have to be adjusted everyday or even several times per day to meet technical specifications of different orders. An etching process has to be re-configured frequently when a new batch of wafers with varying critical dimensions arrives. In recent years, production lines designed with fast-switch capability are frequently seen. When processes are run with small batches, the short-term performance of a controller will become a critical concern.

However, the EWMA controller's short-term performance may be deteriorated due to the following reasons. First, the process may have large initial setup bias, i.e., $|\alpha+\beta x_0-\tau| \gg 0$; Second, the effect of inaccurately estimated process gain parameter, i.e., $b \neq \beta$, is extrusive during the initial stage; Third, the disturbance sequence may be wrongly identified. When N_t is not an IMA(1,1) process, and even when N_t follows an IMA(1,1) process but the EWMA parameter ω is not set to $1-\theta$ because of inaccurate estimate of θ, the performance of the EWMA controller becomes unpredictable.

Tseng *et al.* [20] proposed a variable EWMA (VEWMA) controller, which allows ω to vary at each step. That is, a_t is updated as follows:

$$a_t = \omega_t(Y_t - bx_{t-1}) + (1-\omega_t)a_{t-1}$$

where $\omega_t = \omega_0+\delta^t$, δ is a discount factor. The authors show that the VEWMA can compensate initial bias faster than the EWMA controller. However, in order to setup the optimal VEWMA controller, the amount of initial bias has to be known in advanced. The performance of the VEWMA controller may deteriorate if this parameter is not estimated accurately.

In this paper, we propose a new control scheme called the general harmonic rule (GHR) controller which has higher robustness when parameter estimation uncertainties exist and better short-term performance than EWMA controller when the process output has a large initial bias. We show that the GHR controller is optimal when N_t is white noise and $b = \beta$. In comparison with the EWMA controller's optimality, the GHR controller's optimality is derived without any assumption on the process initial condition, i.e., the initial bias is taken as an unknown value instead of 0. We also investigate the sensitivity of both GHR and EWMA controllers and show that the GHR controller is more robust than the EWMA controller, especially when θ is overestimated. The sensitivity analysis also extends to imperfect estimate of the process gain β, i.e., when β is overestimated offline ($|b| > |\beta|$ and $b \cdot \beta > 0$), the GHR controller has much better performance than that of EWMA controller. This result is very significant since in practice people tend to use an overestimated β rather than an underestimated one in the EWMA controller due to its stability condition, i.e., $0 < \omega\beta/b < 2$ (Ingolfsson and Sachs [11], Sachs *et al.* [24]). The stability condition guarantees the control system away from bursting if β is not underestimated too much.

The rest of this paper is organized as follows. Section II presents the motivation and problem descriptions. Section III discusses the stability and optimality of the GHR controller. Section IV provides a sensitivity analysis of both GHR controller and EWMA controller. A numerical example is presented in Sections V and VI concludes the paper.

II. The GHR Controller

The setup adjustment problem was first studied by Grubbs [9]. Suppose the measurements Y_t represent some quality characteristic of the items as they are produced at discrete points in time $t = 1, 2, \ldots$. Grubbs [9] proposed a method for the adjustment of the machine in order to bring the process back to target if at start-up it was off-target by d units, where d is an unknown value. In this section, we will consider the process model (1) used in run-to-run control. Since both α and β are unknown and need to be estimated from the offline procedure, the initial bias of the process output is an unknown value. Using the similar derivation of Grubbs', we shall in the following derive the optimal controller based on unknown initial bias.

Assume that the disturbance N_t in model (1) follows a family of processes which could be represented by

$$N_t = \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \cdots \tag{3}$$

where $\varepsilon_t \sim \mathrm{WN}(0, \sigma_\varepsilon^2)$ for $t > 0$ and $\varepsilon_t = 0$ for all $j \leq 0$. It includes many popular random processes. For example, N_t is a white noise if $\psi_j = 0$ for $j \geq 1$, an IMA(1,1) with parameter θ if $\psi_j = 1-\theta$ for $j \geq 1$, an ARMA(1,1) with parameters (ϕ, θ) if $\psi_j = \phi^{j-1}(\phi-\theta)$ for $j \geq 1$, and an ARIMA(1,1,1) with parameters (ϕ, θ) if $\psi_j = 1+\phi-\theta/1-\phi(1-\phi^j)$ for $j \geq 1$.

At any time t, denote the process mean conditional on the past outputs as μ_t, i.e., $\mu_t = \mathrm{E}(Y_t|\, Y_1^{t-1})$, where Y_1^{t-1} stands for the observations $\{Y_{t-1}, Y_{t-2}, \ldots, Y_1\}$. Set Y_1^{t-1} to $\varnothing$ when $t \leq 1$. Let d represent the deviation of μ_1 from the target value τ. We call d the initial bias since it comes from the inaccurate estimate of the process parameters at time 0. d can be represented as

$$d = \mu_1 - \tau = (\alpha + \beta x_0) - \tau \tag{4}$$

where $x_0 = (\tau - a_0)/b$. Then the true value for the first output is $Y_1 = \tau + d + \varepsilon_1$. Note that τ is known and Y_1 is observable, so we can measure $Y_1 - \tau$ directly. However, d is unknown and cannot be determined since α and β are unkown. In the following derivation, without loss of generality, assume $\tau = 0$ so that Y_t represents the output's deviation from target at time t.

After producing the first item and observing the first deviation Y_1, we can adjust the last period's input x_0 by $k_1 Y_1$ before making the second item. That is,

$$x_1 = x_0 - k_1 Y_1 = x_0 - k_1(d + \varepsilon_1). \tag{5}$$

From (1), (3), (4) and (5), we know that, after the first adjustment,

$$Y_2 = (1 - \beta k_1)\, d + (1 - \beta k_1)\, \varepsilon_1 + (\psi_1 - 1)\, \varepsilon_1 + \varepsilon_2.$$

It follows that $\mu_2 = (1 - \beta k_1)\, d + (1 - \beta k_1)\, \varepsilon_1 + (\psi_1 - 1)\, \varepsilon_1$. Similarly, an adjustment can be made on the last period's input, i.e.,

$$x_2 = x_1 - k_2 Y_2$$

and

$$\begin{aligned}\mu_3 = {} & (1 - \beta k_1)(1 - \beta k_2) d \\ & + \left[(1 - \beta k_1)(1 - \beta k_2) + (1 - \beta k_2)(\psi_1 - 1) \right. \\ & \left. + (\psi_2 - \psi_1)\right] \varepsilon_1 + \left[(1 - \beta k_2) + (\psi_1 - 1)\right] \varepsilon_2. \end{aligned} \tag{6}$$

Continuing this iteration, by making the corrections $k_3 Y_3$, $k_4 Y_4, \ldots$, etc., in general, the process mean at time t conditional on the past observable deviations follows:

$$\begin{aligned}\mu_t = d \prod_{i=1}^{t-1} (1 - \beta k_i) + \sum_{i=1}^{t-1} \varepsilon_i \Bigg[& \prod_{j=i}^{t-1} (1 - \beta k_j) \\ & + \sum_{r=i+1}^{t} \prod_{j=r}^{t-1} (1 - \beta k_j)(\psi_{r-i} - \psi_{r-i-1}) \Bigg] \end{aligned} \tag{7}$$

where we specify that $\psi_0 = 1$ and $\prod_{j=t}^{t-1}(1 - \beta k_j) = 1$. Our aim is to determine the adjustment coefficients k_i $(i = 1, 2, \ldots, t-1)$, which solve the following optimization problem:

$$\begin{aligned} \min_{k_i, i=1,2,\ldots,t-1} \quad & \mathrm{MSE}(\mu_t) \\ \text{s.t.} \quad & \mathrm{E}(\mu_t) = 0. \end{aligned} \tag{8}$$

Theorem 1 gives a recursive form to determine k_{t-1} given $k_{t-2}, k_{t-3}, \ldots, k_1$ for any time t.

Theorem 1: If $k_1, k_2, \ldots, k_{t-1}$ solve the problem (8), then

$$k_{t-1} = \frac{1}{\beta} \cdot \frac{1 + (t-1) \cdot \psi_1 - \psi_{t-1} - M_{t-2}(k_1, k_2, \ldots, k_{t-2})}{t - 1 - M_{t-2}(k_1, k_2, \ldots, k_{t-2})} \tag{9}$$

where

$$\begin{aligned} & M_{t-2}(k_1, k_2, \ldots, k_{t-2}) \\ & = \sum_{i=1}^{t-3} \sum_{r=i+2}^{t-1} \prod_{j=r}^{t-2} (1 - \beta k_j)(\psi_{r-i} - \psi_{r-i-1}) \\ & \quad + \psi_1 \sum_{i=1}^{t-2} \prod_{j=i+1}^{t-2} (1 - \beta k_j) & (10) \\ & = (1 - \beta k_{t-2}) \cdot M_{t-3}(k_1, k_2, \ldots, k_{t-3}) + \psi_{t-2}. & (11) \end{aligned}$$

The proof of Theorem 1 can be found in Appendix A. Note that it is obvious from (10) that $M_0 = 0$. Theorem 1 gives us a way to obtain the adjustment k_t at any time t based on all the adjustments before time t.

According to Theorem 1, we can design a new controller as follows. Suppose the process to be controlled can be modeled as (1) with the disturbance (3), the new control rule is

$$x_t = x_{t-1} - k_t Y_t \tag{12}$$

where

$$k_t = \frac{1}{b} \cdot \frac{1 + t \cdot \psi_1 - \psi_t - M_{t-1}}{t - M_{t-1}} \tag{13}$$

and

$$M_t = (1 - b \cdot k_t) \cdot M_{t-1} + \psi_t. \tag{14}$$

At time 0, $x_0 = -a_0/b$ and $M_0 = 0$, where a_0 and b are the off-line estimates of the process parameters α and β, respectively. It can be seen from (13) that when all $\psi_j = 0 (j \geq 1)$ and $b = 1$, k_t reduces to $1/t$, which is exactly the harmonic rule in the process setup adjustment problem derived by Grubbs [9]. We call the new controller General Harmonic Rule (GHR) controller and ψ_j $(j \geq 1)$ are called the GHR parameters. In particular, for the IMA(1,1) disturbance with parameter θ, $\psi_j = 1 - \theta$ for all $j \geq 1$.

III. Stability and Optimality

For any control scheme to be practical, a fundamental requirement is that the process should achieve long-term stability. Although this paper focuses on the performance of short-run manufacturing processes, it is worthwhile to investigate stability conditions for the GHR controller. A process $\{Y_t\}$ is said to be asymptotically stable if

$$\lim_{t \to \infty} \mathrm{E}(Y_t) = \tau \quad \text{and} \quad \lim_{t \to \infty} \mathrm{Var}(Y_t) < \infty. \tag{15}$$

The stability of a process ensures that the mean of the process output converges to the desired target, while its asymptotic variance remains bounded. The following theorem gives the stability condition for the GHR controller when the disturbance follows IMA(1,1) process.

Theorem 2: The GHR controller defined in (12)–(14) is asymptotically stable if $0 < \beta/b < 2$ when the disturbance N_t is an IMA(1,1) process.

The proof of Theorem 2 can be found in Appendix B. This condition implies that if b is of the same sign and larger (in absolute value) than β, then stability is guaranteed. If b is of the same sign and smaller (in absolute value) than β, it must be larger (in absolute value) than $\beta/2$ for stability. If b and β have different signs, the process will always be unstable. The following theorem gives the conditions under which the GHR controller is optimal.

Theorem 3: If $b = \beta$ and N_t is white noise, i.e., $\psi_j = 0$ for all $j \geq 1$, the GHR controller is optimal.

The proof of Theorem 3 can be found in Appendix C. Our proof is very similar to that of Grubbs [9], but from a controller's aspect. In comparison with the EWMA controller's optimality, the GHR controller is derived from no assumption of a_0. That is, if the process disturbance is white noise and $b = \beta$, the GHR controller is the optimal controller no matter what the process initial status is.

IV. Sensitivity Analysis

In this section, we take the Deep Reactive Ion Etching (DRIE) process in semiconductor manufacturing as an example to study the performance of the newly proposed GHR controller. The DRIE process is an important step for forming desired patterns on wafers in micro/nano-scale fabrication; it involves complex chemical-mechanical reactions. Wafers to be etched are loaded into a chamber. The system first releases etching plasma into the chamber to generate trenches subject to designed mask patterns; then in the deposition step, different gases are introduced into the chamber to generate a protective film on the sidewalls. The etching and deposition steps repeat alternately until the preset processing time is reached or the end-point detection module confirms the correct etching depth. A more detailed illustration of the etching process is referred to Wang and Tsung [27]. DRIE has been successfully used in producing photonic crystals, magnetic nanostructures and MEMS resonators (STS [17]).

One of the key quality characteristics produced by the DRIE process is the etched profile. As is shown in Fig. 1. An ideal profile should be vertical, having an angle of 90° against the horizontal line ($\tau = 90$). The etch/deposition time ratio is usually adjusted to compensate over-etched or under-etched wafers to generate vertical sidewalls. In an analysis presented by Wang and Tsung [27], the authors studied the DRIE process and suggested that the slope of the produced profiles can be modeled by:

$$Y_t = 91.7 - 1.8x_{t-1} + N_t,$$

where N_t is an IMA(1,1) time series with $\theta = 0.6$. The parameters are estimated from a designed experiment.

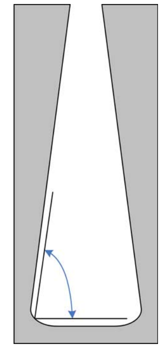


Fig. 1. Illustration of an etched profile.

As each production run is very time-consuming and it is not practical to study the performance of the proposed controller by adjusting the real process, we treat the above Equation as the true process model and study the performance of the GHR controller. The EWMA and VEWMA controllers are also set up for the same simulated process for comparison. Our comparison is divided into the following two parts, when β is known and unknown.

A. β is Known

In many setup adjustment literature, the process gain β is usually assumed known and set to 1 (Grubbs [9], Triestsch [19] and Del Castillo *et al.* [8]). In this section, we first assume β is known or could be accurately estimated offline, i.e., $b = \beta$, and investigate the performance of both GHR and EWMA controllers given different offline estimate of α (i.e., a_0). We will first assume the disturbance follows an IMA model and then takes a more general ARIMA(1,1,1) model.

1) N_t is an IMA(1,1) Process: The IMA(1,1)model has been widely adopted to characterize disturbances in diverse applications. The conventional EWMA controller has been proved to be optimal to compensate such a disturbance series (Box *et al.* [2]). However, certain conditions must be satisfied in order for the EWMA controller to achieve its optimal performance. First, the process parameters α and β are both known or accurately estimated offline, i.e., $a_0 = \alpha$ and $b = \beta$; Second, the EWMA parameter ω should be set to $1 - \theta$ which implies that the IMA parameter θ should be known or accurately estimated offline. In short-run production scenarios, however, it is often impossible to obtain accurate estimate of these parameters. In the following simulations, we take $\alpha = 91.7$, $\beta = -1.8$, $\theta = 0.6$ and $\sigma_\varepsilon^2 = 1$ as the true parameters of the DRIE process. The EWMA, VEWMA and GHR controllers are then set up to adjust the process under different hypothetical settings.

Three cases are studied in the following. In case 1, we assume θ is accurately estimated (i.e., $\hat{\theta} = \theta$); in case 2, we assume θ is overestimated (i.e., $\hat{\theta} > \theta$); and in case 3, we assume θ is underestimated (i.e., $\hat{\theta} < \theta$). In each case, the EWMA parameter ω and the GHR parameters ψ_j $(j \geq 1)$ are all set to $1 - \hat{\theta}$. For the VEWMA controller, the initial value of the smoothing parameter, ω_0, is set to $1 - \hat{\theta}$. The choice of the discount factor in the VEWMA controller depends on the initial process bias, which is unknown in real scenarios. Therefore, we choose a moderate

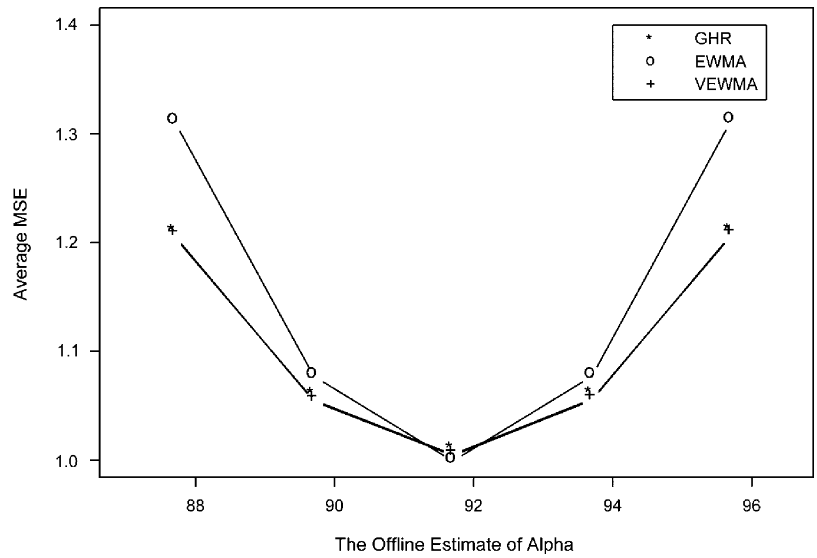
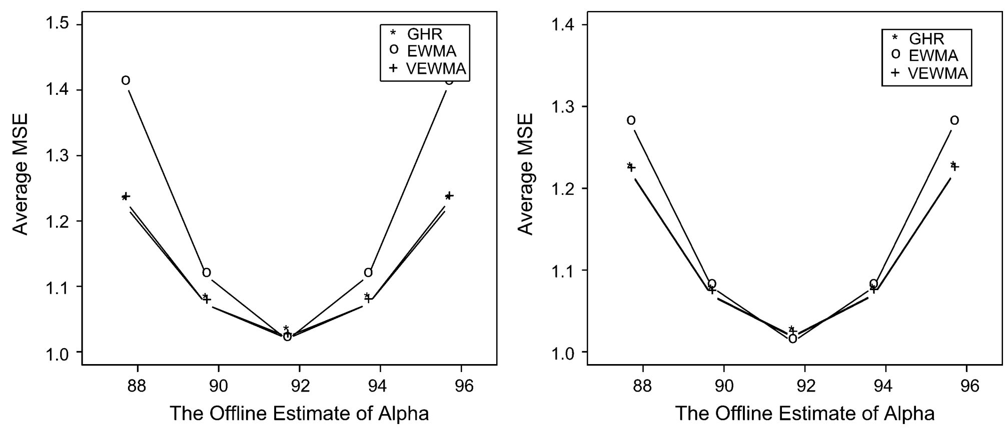


Fig. 2. N_t is IMA(1,1) process and θ is known ($\hat{\theta} = 0.6$).



(a) $\hat{\theta} = 0.7$ (b) $\hat{\theta} = 0.5$

Fig. 3. AMSE of the EWMA and GHR Controllers when θ is Over- and Under-estimated.

TABLE I
THE AMSE WHEN N_t IS IMA(1,1)

$\theta = 0.6$	$\hat{\theta} = 0.6$				
a_0	87.7	89.7	91.7	93.7	95.7
The GHR Controller	1.205	1.055	1.005	1.056	1.206
The EWMA Controller	1.311	1.077	0.999	1.077	1.312
The VEWMA Controller	1.206	1.054	1.004	1.055	1.207

All the SEAMSE's are smaller than 0.002

TABLE II
THE AMSE WHEN N_t IS IMA(1,1)

$\theta = 0.6$	$\hat{\theta} = 0.7$				
a_0	87.7	89.7	91.7	93.7	95.7
The GHR Controller	1.222	1.072	1.023	1.073	1.223
The EWMA Controller	1.410	1.116	1.018	1.116	1.410
The VEWMA Controller	1.230	1.072	1.020	1.073	1.231

All the SEAMSE's are smaller than 0.002

TABLE III
THE AMSE WHEN N_t IS IMA(1,1)

$\theta = 0.6$	$\hat{\theta} = 0.5$				
a_0	87.7	89.7	91.7	93.7	95.7
The GHR Controller	1.218	1.068	1.018	1.069	1.219
The EWMA Controller	1.279	1.079	1.012	1.079	1.279
The VEWMA Controller	1.219	1.069	1.019	1.070	1.220

All the SEAMSE's are smaller than 0.002

value, $\delta = 0.5$, in the following simulations. The estimated intercept, a_0, is varying within a neighborhood of α.

We did 10 000 simulations for each case. In each simulation, we run 80 steps for Y_t. The MSE of $Y_1, Y_2, \ldots, Y_{80}$ is computed. The Average MSE (AMSE) of the 10 000 simulations is reported in Tables I–III. The simulation results are graphically shown in Figs. 2 and 3(a) and 3(b). The standard error in the AMSE (SEAMSE) is reported as well, which is defined as

$$\text{SEAMSE} = \frac{\text{SDMSE}}{\sqrt{\text{No. of replicates}}}$$

where SDMSE means the standard deviation of mean square errors, and the number of replicates used in the simulation is

10 000. Essentially, SEAMSE measures the significance of the difference of AMSE values among all controllers.

Both Table I and Fig. 2 show the results when $\hat{\theta} = \theta$. It is seen that although the EWMA controller is an MMSE controller when $a_0 = \alpha$, which achieves the lowest MSE, the GHR controller is more robust than the EWMA controller when a_0 is deviated from α. The VEWMA controller performs in between; it is superior to the GHR controller when a_0 is close to α, while is inferior to the GHR controller when a_0 deviates far away from α. The AMSE of the GHR controller increases much slower than the EWMA controller does when a_0 deviates from its true value α.

Similar patterns are observed when θ is either overestimated or underestimated. When θ is overestimated, as $\hat{\theta} = 0.7$ in case 2, Fig. 3(a) and Table II show that the AMSE of Y_t using the GHR controller is almost consistently smaller than that using the EWMA controller for any a_0 values and the margin could be as big as 13.3% when $a_0 \neq \alpha$. The VEMWA controller shows its capability in compensating initial bias, as Tseng *et al.* [20] demonstrated. When θ is over estimated, the VEWMA controller is slightly better than the GHR controller or equally good when a_0 is close to α, while it becomes less favored when the difference between a_0 and α becomes large.

This result has very important implications in practice. As Hunter [10] suggested, the EWMA parameter ω is usually set to a value between 0.1 and 0.3. This fact implies that $\hat{\theta}$ is ordinarily assumed to be between 0.7 and 0.9, although θ could be smaller than 0.7 in many cases. When θ is overestimated using the rule of thumb, the advantage of the GHR controller over the EWMA controller becomes more significant.

Table III and Fig. 3(b) show the results when θ is underestimated ($\hat{\theta} = 0.5$). Again, the robustness of the GHR controller is proved by its slow increasing trend when a_0 deviates from α. Now, the VEWMA controller is slightly inferior to the GHR controller for all tested a_0 values.

2) N_t is an ARIMA(1,1,1) Process: Even though the process we investigate has an IMA(1,1) distrubance series, it is meaningful to study the performance of the GHR controller under other types of disturbance series. In the following simulation, we assume that the disturbance N_t follows an general ARIMA(1,1,1) process with parameters ϕ and $\tilde{\theta}$, while it is misidentified as IMA(1,1) process with parameter θ. Appendix D shows that the estimate of IMA parameter θ would be around $|\phi - \tilde{\theta}|/\sqrt{1-\phi^2}$. We consider two cases in the following simulations. In case 1, ϕ is set to 0.2 and $\tilde{\theta}$ is set to 0.6; In case 2, ϕ is set to 0.6 and $\tilde{\theta}$ is set to 0.2. If these ARIMA(1,1,1) processes are misidentified as IMA(1,1) processes, $\hat{\theta}$ is around 0.4 in case 1 and around 0.5 in case 2. Similar to the above experiment, we replicated 10,000 simulations for each case. In each simulation, we run 80 steps for Y_t and computed the AMSE of the 10,000 simulations as shown in Tables IV and V.

Fig. 4 further plots the AMSE versus a_0 for case 1 and case 2 respectively. Once again, the GHR controller has more robust performance in both cases. The inaccurate offline estimate of α has less impact on the GHR controller than on the EWMA controller. The VEWMA controller and the GHR controller shows similar performance. The VEWMA controller's AMSE is generally larger when θ is underestimated, while smaller when θ is overestimated.

TABLE IV
THE AMSE WHEN N_t IS ARIMA(1,1,1)

$\tilde{\theta} = 0.6$	$\phi = 0.2$		$\hat{\theta} = 0.4$		
a_0	87.7	89.7	91.7	93.7	95.7
The GHR Controller	1.210	1.060	1.010	1.061	1.211
The EWMA Controller	1.245	1.067	1.008	1.067	1.246
The VEWMA Controller	1.214	1.063	1.012	1.063	1.215

All the SEAMSE's for the GHR controller are smaller than 0.00002
All the SEAMSE's for the EWMA and VEWMA controller are smaller than 0.002

TABLE V
THE AMSE WHEN N_t IS ARIMA(1,1,1)

$\tilde{\theta} = 0.2$	$\phi = 0.6$		$\hat{\theta} = 0.5$		
a_0	87.7	89.7	91.7	93.7	95.7
The GHR Controller	2.840	2.690	2.641	2.691	2.841
The EWMA Controller	2.935	2.736	2.669	2.736	2.936
The VEWMA Controller	2.831	2.681	2.631	2.682	2.832

All the SEAMSE's for the GHR controller are smaller than 0.00002
All the SEAMSE's for the EWMA and VEWMA controller are smaller than 0.008

B. β is Unknown

In most run-to-run production processes, offline sample is usually not large enough to guarantee accuracy of the parameter estimates, especially the process gain β. When β is unknown and has to be estimated by offline experiments, it is important to investigate the performance of the GHR controller when all parameters are estimated with uncertainties.

In the following study, the process is still assumed to following (1) with N_t being an IMA(1,1) series. True process parameters are chosen as $\alpha = 91.7$, $\beta = -1.8$, $\theta = 0.6$ and $\sigma_\varepsilon^2 = 1$. Similar to the case when β is known, we analyze three cases here: θ is known, θ is overestimated, and θ is underestimated. In each case, the EWMA parameter ω and the GHR parameters ψ_j $(j \geq 1)$ are all set to $1 - \hat{\theta}$. The discount factor of the VEWMA controller is still set to 0.5. Estimated parameters, a_0 and b, are allowed to vary with their respective neighborhood areas. The simulation results are reported in Tables VI–VIII. Fig. 5 shows the contour plots of AMSE for the GHR and EWMA controller respectively when θ is known. It is easy to see that, for both controllers, the AMSE changes slowly along the line $b = \beta/(\alpha - \tau) \cdot (a_0 - \tau)$ and achieves the minimum when $a_0 = \alpha$ and $b = \beta$. Comparing with the GHR controller, the contour curve of the EWMA controller is more dense, which indicates that the AMSE value increases more quickly when $b/(a_0 - \tau)$ deviates from $\beta/(\alpha - \tau)$ due to poor offline estimates of α and β. That is, the GHR controller is more robust than the EWMA controller when offline parameter estimates are not accurate, especially when $|b|$ is bigger than $|\beta|$. The VEWMA controller outperforms the GHR controller only when the estimated values are close to their respective true values. This tends to be an advantage of the GHR controller when β is overestimated in practice as we discussed before for the controller's stability.

Since using small values of ω in the EWMA controller is a rule of thumb in practice (Hunter [10]), Figs. 6 and 7 further present contour plots of the AMSE for the three controllers when θ is overestimated and underestimated. Although all

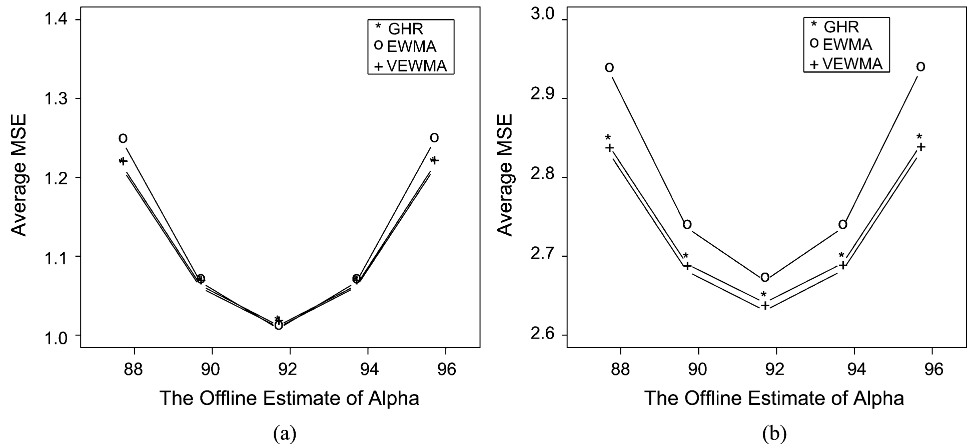


Fig. 4. AMSE of the EWMA and GHR Controllers when N_t is an ARIMA(1,1,1) process but is misidentified as an IMA(1,1) process. (a) ARIMA with $\phi = 0.2$ and $\theta = 0.6$. (b) ARIMA with $\phi = 0.6$ and $\theta = 0.2$.

TABLE VI
THE AMSE WHEN θ IS KNOWN ($\hat{\theta} = 0.6$)

The GHR Controller					
b	$a_0 = 89.7$	90.7	91.7	92.7	93.7
-2.1	1.057	1.023	1.009	1.013	1.036
-2.0	1.055	1.020	1.006	1.013	1.040
-1.9	1.054	1.018	1.005	1.014	1.046
-1.8	1.055	1.018	1.005	1.018	1.056
-1.7	1.058	1.019	1.008	1.025	1.070
-1.6	1.064	1.022	1.013	1.035	1.089
-1.5	1.074	1.028	1.020	1.050	1.117
The EWMA Controller					
b	$a_0 = 89.7$	90.7	91.7	92.7	93.7
-2.1	1.089	1.031	1.006	1.013	1.052
-2.0	1.084	1.026	1.002	1.013	1.058
-1.9	1.080	1.021	1.000	1.015	1.066
-1.8	1.077	1.018	0.999	1.018	1.077
-1.7	1.076	1.017	1.000	1.025	1.092
-1.6	1.077	1.017	1.003	1.035	1.111
-1.5	1.080	1.020	1.010	1.048	1.136
The VEWMA Controller					
b	$a_0 = 89.7$	90.7	91.7	92.7	93.7
-2.1	1.058	1.023	1.008	1.012	1.036
-2.0	1.055	1.020	1.005	1.012	1.040
-1.9	1.054	1.017	1.004	1.013	1.046
-1.8	1.054	1.017	1.004	1.017	1.055
-1.7	1.057	1.017	1.006	1.023	1.068
-1.6	1.062	1.020	1.011	1.033	1.086
-1.5	1.070	1.026	1.018	1.047	1.112

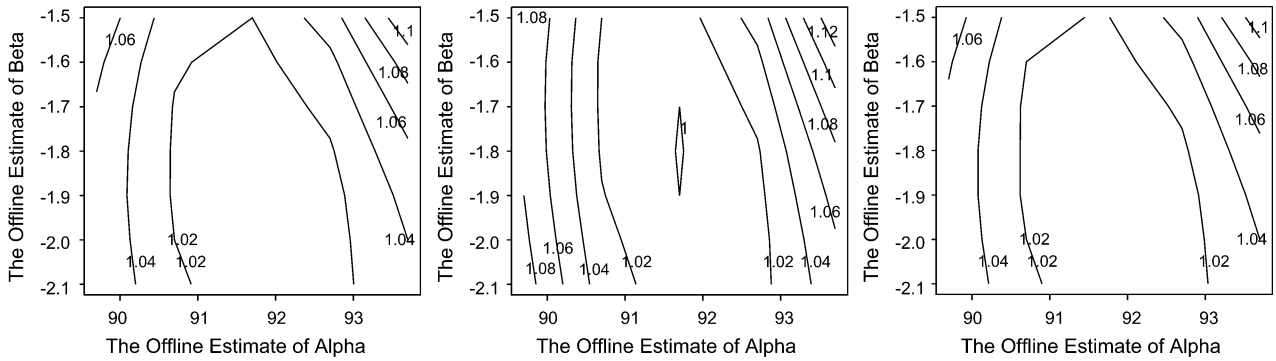
All the SEAMSE's for the three controllers are smaller than 0.002

TABLE VII
THE AMSE WHEN θ IS OVERESTIMATED ($\hat{\theta} = 0.7$)

The GHR Controller					
b	$a_0 = 89.7$	90.7	91.7	92.7	93.7
-2.1	1.093	1.059	1.045	1.049	1.072
-2.0	1.085	1.050	1.036	1.043	1.070
-1.9	1.078	1.042	1.029	1.038	1.070
-1.8	1.072	1.035	1.023	1.035	1.073
-1.7	1.068	1.029	1.018	1.035	1.080
-1.6	1.067	1.024	1.015	1.037	1.092
-1.5	1.067	1.022	1.014	1.043	1.111
The EWMA Controller					
b	$a_0 = 89.7$	90.7	91.7	92.7	93.7
-2.1	1.150	1.077	1.045	1.054	1.104
-2.0	1.138	1.065	1.035	1.048	1.105
-1.9	1.126	1.053	1.026	1.045	1.109
-1.8	1.116	1.042	1.018	1.043	1.116
-1.7	1.106	1.033	1.012	1.043	1.126
-1.6	1.098	1.024	1.007	1.046	1.141
-1.5	1.091	1.017	1.004	1.052	1.161
The VEWMA Controller					
-2.1	1.097	1.060	1.043	1.048	1.074
-2.0	1.088	1.050	1.035	1.042	1.071
-1.9	1.080	1.041	1.027	1.037	1.071
-1.8	1.072	1.033	1.020	1.033	1.073
-1.7	1.067	1.026	1.014	1.032	1.078
-1.6	1.062	1.020	1.011	1.033	1.087
-1.5	1.060	1.016	1.009	1.037	1.102

All the SEAMSE's for the three controllers are smaller than 0.002

controllers' performance are worse than that in Fig. 5, the EWMA controller's performance deteriorates more significantly; the VEWMA controller is also inferior to the GHR controller for most tested cases when θ is underestimated. The GHR controller is seen to be the most robust against parameter deviations, especially when $|b|$ is greater than $|\beta|$.



(a) GHR Controller (b) EWMA Controller (c) VEWMA Controller

Fig. 5. Contour plot of AMSE when N_t is IMA(1,1) process and θ is known.

TABLE VIII
THE AMSE WHEN θ IS UNDERESTIMATED ($\hat{\theta} = 0.5$)

			The GHR Controller		
b	$a_0 = 89.7$	90.7	91.7	92.7	93.7
-2.1	1.053	1.019	1.005	1.009	1.032
-2.0	1.056	1.021	1.007	1.014	1.041
-1.9	1.061	1.025	1.012	1.021	1.053
-1.8	1.068	1.031	1.018	1.031	1.069
-1.7	1.078	1.039	1.028	1.045	1.090
-1.6	1.092	1.050	1.041	1.063	1.117
-1.5	1.111	1.066	1.058	1.087	1.155
			The EWMA Controller		
b	$a_0 = 89.7$	90.7	91.7	92.7	93.7
-2.1	1.071	1.022	1.001	1.007	1.040
-2.0	1.072	1.023	1.003	1.012	1.050
-1.9	1.074	1.025	1.006	1.019	1.063
-1.8	1.079	1.029	1.012	1.029	1.079
-1.7	1.085	1.035	1.020	1.042	1.099
-1.6	1.095	1.044	1.032	1.059	1.125
-1.5	1.109	1.057	1.048	1.081	1.158
			The VEWMA Controller		
-2.1	1.053	1.020	1.005	1.009	1.032
-2.0	1.057	1.022	1.008	1.015	1.041
-1.9	1.062	1.026	1.013	1.022	1.054
-1.8	1.069	1.032	1.019	1.032	1.070
-1.7	1.080	1.040	1.029	1.046	1.091
-1.6	1.094	1.052	1.042	1.064	1.119
-1.5	1.113	1.068	1.060	1.089	1.156

All the SEAMSE's for the three controllers are smaller than 0.002

V. AN ILLUSTRATIVE EXAMPLE WITH INITIAL BIAS

The above section has studied the performance and robustness of the GHR controller. In this section, we use a specific parameter setting and study the impact of initial estimation bias on the newly proposed controller.

The process is again assumed to follow (1) and the disturbance N_t is an IMA(1,1) time series. The true parameter settings are $\alpha = 91.7, \beta = -1.8$ and $\theta = 0.6$, while the estimated values are assumed to be $a_0 = 85$, $b = -3.0$ and $\hat{\theta} = 0.8$, respectively. That is, initial bias exist in estimating all the parameters in the model. The target value $\tau = 90$.

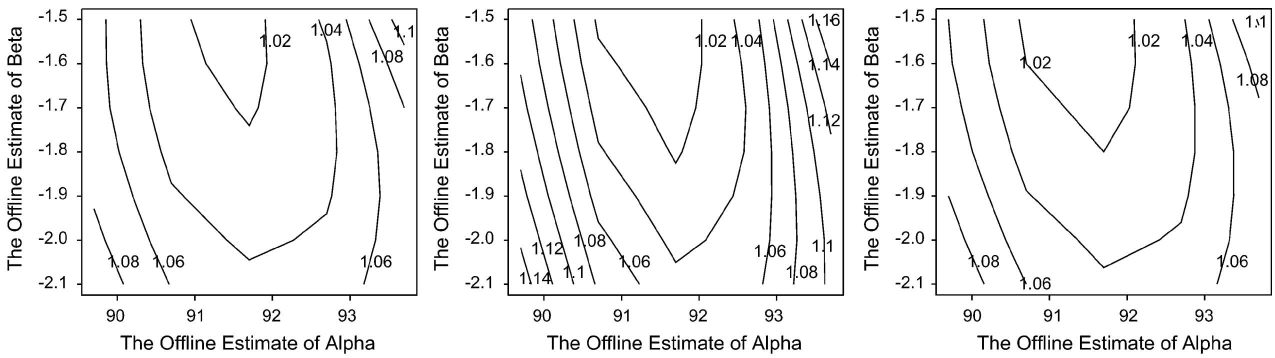
Fig. 8 shows the process output Y_t against time t when the three controllers are applied. The curves show that, the EWMA controller requires a moderately large number of runs to bring the process output to the target when the process has large initial bias $(\alpha + \beta(\tau - a_0)/b - \tau = 4.7)$. The VEWMA controller is more efficient than the EWMA controller in removing initial bias. However, the GHR controller is the fastest in bringing the process output back to target. After the first 30 runs, the process output Y_t are almost the same using both controllers. Although this is just a single realization of the control process, it demonstrates the effectiveness of the GHR controller when the initial bias is significant.

To better reveal the difference between the VEWMA and GHR controllers, we further investigated the mean response of a process with $\alpha = 91.7$ and $\beta = -1.8$. We consider different cases when estimated α and β, denoted by a_0 and b, are biased. Fig. 9 shows the mean responses of a process controlled by the VEWMA and GHR controllers under different initial bias. All the cases have an equal initial bias $d = 2$. It is seen from (a) that when only α is biased, GHR can bring the output back to target immediately, while VEWMA consumes several more steps to approach the target. When β is also biased, in case (b), GHR and VEWMA performs closely, while in case (c), GHR outperforms VEWMA in compensating the initial bias.

To summarize, when the disturbance sequence follows an IMA model and parameters are accurately estimated, the EWMA controller is always suggested. This is supported by both theory and simulation results. While when the disturbance model is not IMA or parameters cannot be estimated accurately, either VEWMA or GHR is suggested. More specifically, as the GHR controller is more robust in most cases, GHR is more favored when initial bias or estimation uncertainties is large.

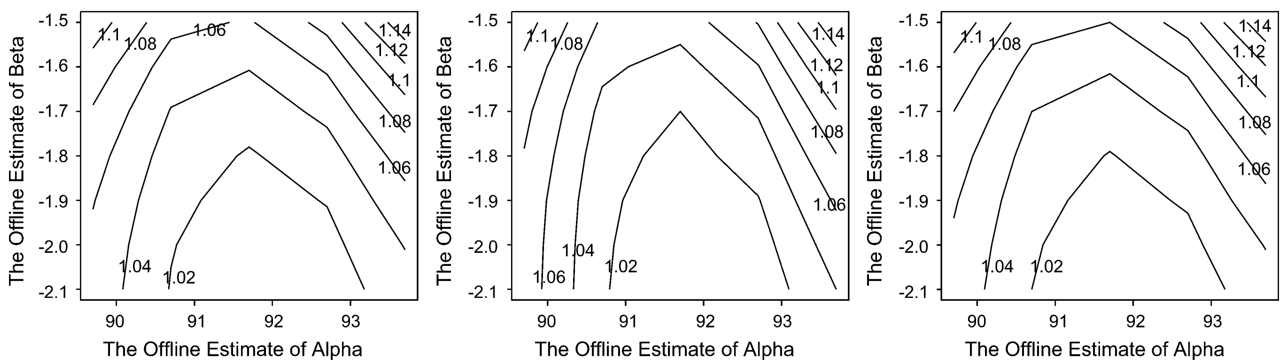
VI. CONCLUSION

In short-run production processes, the performance of an EWMA controller critically depends on offline estimates of



(a) GHR Controller (b) EWMA Controller (c) VEWMA Controller

Fig. 6. Contour plot of AMSE when N_t is IMA(1,1) process and θ is overestimated.



(a) GHR Controller (b) EWMA Controller (c) VEWMA Controller

Fig. 7. Contour plot of AMSE when N_t is IMA(1,1) process and θ is underestimated.

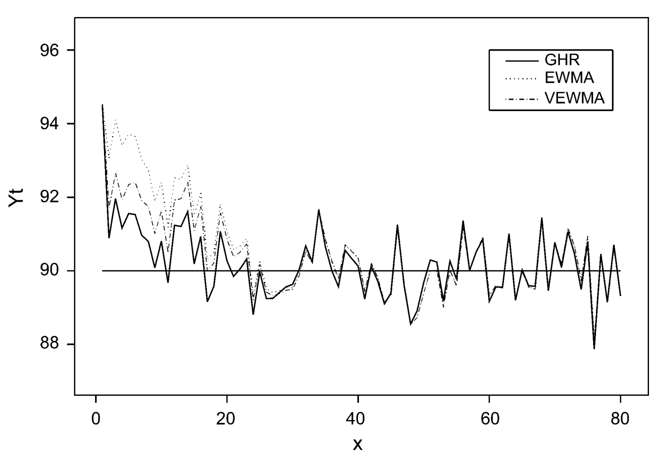
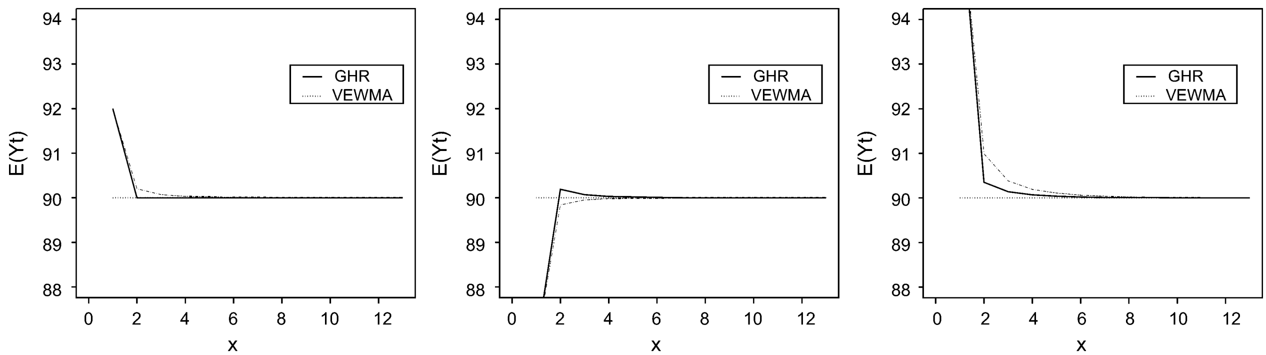


Fig. 8. The output Y_t against time t.

process parameters. This paper proposes a new controller based on the harmonic rule used in machine setup adjustment problems. The sensitivity of the new controller is compared with the EWMA and VEWMA controllers under different scenarios of the disturbance parameter estimate, the process offset and gains estimate, as well as the misidentification of the disturbance. The short-run performance of the GHR controller is shown better than that of the EWMA controller when the offline estimates of the process or disturbance parameters are inaccurate. Even though the VEWMA can compensate initial bias faster than the EWMA controller, it is inferior to the GHR controller when estimated parameter values deviate far from their true settings, especially when θ is underestimated. The stability and optimality conditions are also derived for the new controller. The sensitivity analysis indicates that the new GHR controller is more robust than the EWMA controller under model misspecification and parameter estimation uncertainties.

It should be noted that dislike the EWMA controller, which is optimal for IMA(1,1) disturbance series only, the GHR controller can be applied to processes with any general disturbance models that follow (3).

Initial bias in parameter estimation can seriously deteriorate control performance, especially in short-run processes. The GHR controller assumes the initial bias is an unknown but fixed value in this research. In many practices, due to frequent process setup, initial bias may be better modelled as a random variable rather than a fixed value. The GHR controller can be extended to take random initial bias into considerations, which should be a topic for further research. This work focuses on single-input-single-output processes only. In many semiconductor manufacturing scenarios, a process may have multiple correlated inputs and outputs. Extending the GHR controller to multivariate processes is another important topic for future research



(a) $a_0 = 89.7$, $b = -1.8$ (b) $a_0 = 89.716$, $b = -1.7$ (c) $a_0 = 89.68$, $b = -1.9$

Fig. 9. Mean responses of a process controlled by GHR and VEWMA controllers when initial bias exists.

APPENDIX A

Since

$$\mathrm{MSE}(\mu_t) = \mathrm{Var}(\mu_t) + [\mathrm{E}(\mu_t)]^2$$

the problem (8) is equivalent to

$$\min_{k_i} \quad \mathrm{Var}(\mu_t) \qquad \text{s.t.} \quad \mathrm{E}(\mu_t) = 0. \tag{16}$$

From (7), we get that

$$\mathrm{E}(\mu_t) = d\prod_{i=1}^{t-1}(1-\beta k_i) \tag{17}$$

and

$$\mathrm{Var}(\mu_t) = \sigma_\varepsilon^2 \sum_{i=1}^{t-1}\left[\prod_{j=i}^{t-1}(1-\beta k_j) + \sum_{r=i+1}^{t}\prod_{j=r}^{t-1}(1-\beta k_j)(\psi_{r-i}-\psi_{r-i-1})\right]^2. \tag{18}$$

Denote λ as the Lagrange multiplier, we set

$$f(k_i) = \sigma_\varepsilon^2 \sum_{i=1}^{t-1}\left[\prod_{j=i}^{t-1}(1-\beta k_j) + \sum_{r=i+1}^{t} \times \prod_{j=r}^{t-1}(1-\beta k_j)(\psi_{r-i}-\psi_{r-i-1})\right]^2 + \lambda\prod_{i=1}^{t-1}(1-\beta k_i)$$

and equate $\partial/\partial k_i f(k_i)$ to zero. Since

$$\prod_{i=1}^{t-1}(1-\beta k_i) = 1 - \sum_{i=1}^{t-1}\beta k_i \prod_{j=i+1}^{t-1}(1-\beta k_j) \tag{19}$$

we find that

$$2\sigma_\varepsilon^2\left[(1-\beta k_i)\prod_{j=i+1}^{t-1}(1-\beta k_j) + \sum_{r=i+1}^{t}\prod_{j=r}^{t-1}(1-\beta k_j)(\psi_{r-i}-\psi_{r-i-1})\right] \times \beta\prod_{j=i+1}^{t-1}(1-\beta k_j) + \lambda\beta\prod_{j=i+1}^{t-1}(1-\beta k_j) = 0. \tag{20}$$

That is,

$$\beta k_i \prod_{j=i+1}^{t-1}(1-\beta k_j) = \frac{\lambda}{2\sigma_\varepsilon^2} + \sum_{r=i+2}^{t}\prod_{j=r}^{t-1} \times(1-\beta k_j)(\psi_{r-i}-\psi_{r-i-1}) + \psi_1\prod_{j=i+1}^{t-1}(1-\beta k_j) \tag{21}$$

where $i = 1, 2, \ldots, t-1$ and we used $\psi_0 = 1$.

From the side condition $\mathrm{E}(\mu_t) = 0$ and (17), (19), we know that

$$\sum_{i=1}^{t-1}\beta k_i \prod_{j=i+1}^{t-1}(1-\beta k_j) = 1. \tag{22}$$

So from (21) and (22), we get (23), found at the bottom of the page. Since (21) is right for any $i = 1, 2, \ldots, t-1$, let us set $i = t-1$. Then we can get that

$$k_{t-1} = \frac{1}{\beta}\cdot\left(\frac{\lambda}{2\sigma_\varepsilon^2} + \psi_1\right). \tag{24}$$

From (23) and (24), we get

$$k_{t-1} = \frac{1}{\beta}\cdot\frac{1 + (t-1)\cdot\psi_1 - \psi_{t-1} - M_{t-2}(k_1, k_2, \ldots, k_{t-2})}{t-1-M_{t-2}(k_1, k_2, \ldots, k_{t-2})} \tag{25}$$

$$\lambda = 2\sigma_\varepsilon^2 \frac{1 - \psi_1\sum_{i=1}^{t-1}\prod_{j=i+1}^{t-1}(1-\beta k_j) - \sum_{i=1}^{t-1}\sum_{r=i+2}^{t}\prod_{j=r}^{t-1}(1-\beta k_j)(\psi_{r-i}-\psi_{r-i-1})}{t-1} \tag{23}$$

where

$$M_{t-2}(k_1,k_2,\ldots,k_{t-2}) = \sum_{i=1}^{t-3}\sum_{r=i+2}^{t-1}\prod_{j=r}^{t-2}(1-\beta k_j) \times(\psi_{r-i}-\psi_{r-i-1}) + \psi_1\sum_{i=1}^{t-2}\prod_{j=i+1}^{t-2}(1-\beta k_j). \quad (26)$$

It is not difficult to get the recursive version of M_{t-2} that

$$M_{t-2}(k_1,k_2,\ldots,k_{t-2}) = (1-\beta k_{t-2})\cdot M_{t-3}(k_1,k_2,\ldots,k_{t-3}) + \psi_{t-2}.$$

APPENDIX B

We only prove the case that β and b are positive. Note that $\psi_j = 1-\theta$ for all $j \geq 1$ when N_t is IMA(1,1) process. According to the GHR controller's definition, it's easy to get $E(Y_t) = E(\mu_t) = d\prod_{i=1}^{t-1}(1-\beta k_i)$ from (7). So Y_t is asymptotically stable if and only if there exists a $T > 0$ such that $|1-\beta k_j| < 1$ for all $j > T$. We only need to prove if $b > \beta/2$, then

$$0 < |1-\beta k_{t+1}| < 1$$

is satisfied whenever

$$0 < |1-\beta k_t| < 1 \quad (27)$$

is satisfied.

Conditional on $b > \beta/2$, from (13) and (27) we can get

$$M_{t-1} < \frac{\beta}{2b-\beta}\left[\frac{2b}{\beta}t - 1 - (t-1)(1-\theta)\right]. \quad (28)$$

Because of (27), it's easy to get

$$1-\frac{2b}{\beta} < 1 - bk_t < 1. \quad (29)$$

Then from (14) and (29), we know

$$M_t < M_{t-1} + (1-\theta). \quad (30)$$

So under the conditions (28) and (30), we know

$$\begin{aligned} M_t &< \frac{\beta}{2b-\beta}\left[\frac{2b}{\beta}t - 1 - (t-1)(1-\theta)\right] + (1-\theta) \\ &< \frac{\beta}{2b-\beta}\left[\frac{2b}{\beta}(t+1) - 1 - t(1-\theta)\right]. \end{aligned} \quad (31)$$

From (13), (31) and $b > \beta/2$, we can get

$$0 < k_{t+1} < \frac{2}{\beta}$$

that is, $|1-\beta k_{t+1}| < 1$. So $\lim_{t\to\infty} \mathrm{E}(Y_t) = 0$.

Since we have proved that $|1-\beta k_j| < 1$ for all $j \geq 1$ under the condition $b > \beta/2$, we can find a real number q such that $|1-\beta k_j| \leq q < 1$. Then

$$\begin{aligned} \mathrm{Var}(Y_t) &= \mathrm{Var}(\mu_t) + \mathrm{Var}(\varepsilon_t) \\ &= \sigma_\varepsilon^2 \sum_{i=1}^{t-1}\left[\prod_{j=i}^{t-1}(1-\beta k_j)\right]^2 + \sigma_\varepsilon^2 \\ &= \sigma_\varepsilon^2 \sum_{i=1}^{t-1}\left[\prod_{j=1}^{i}(1-\beta k_{t-j})^2\right] + \sigma_\varepsilon^2 \\ &\leq \sigma_\varepsilon^2 \sum_{i=1}^{t-1} q^{2i} + \sigma_\varepsilon^2. \end{aligned} \quad (32)$$

So $\lim_{t\to\infty} \mathrm{Var}(Y_t) < \infty$.

APPENDIX C

When $\psi_j = 0\,(j \geq 1)$ and $b = \beta$, the GHR controller will degenerate into the form that

$$x_t = x_{t-1} - k_t\cdot Y_t \quad \text{and} \quad k_t = \frac{1}{\beta}\cdot\frac{1}{t} \qquad (t \geq 1). \quad (33)$$

In order to prove the GHR controller be optimal, we need to prove the following fact: if

$$k_j = \frac{1}{\beta}\cdot\frac{1}{j} \quad (1 \leq j \leq t-1)$$

are the solutions of the problem (8), then $k_1, k_2, \ldots, k_{t-1}, k_t$ are the solutions of the problem

$$\begin{aligned} \min_{k_i,\ i=1,2,\ldots,t} \quad & \mathrm{MSE}(\mu_{t+1}) \\ \text{s.t.} \quad & \mathrm{E}(\mu_{t+1}) = 0 \end{aligned} \quad (34)$$

where

$$k_t = \frac{1}{\beta}\cdot\frac{1}{t}.$$

From (18) and $\psi_0 = 1$, we can derive that

$$\begin{aligned} \mathrm{Var}(\mu_t) = \sigma_\varepsilon^2 \cdot \sum_{i=1}^{t-1}\Bigg[&\prod_{j=i}^{t-1}(1-\beta k_j) \\ &+ \prod_{j=i+1}^{t-1}(1-\beta k_j)(\psi_1 - 1) \\ &+ \sum_{r=i+2}^{t}\prod_{j=r}^{t-1}(1-\beta k_j)(\psi_{r-i}-\psi_{r-i-1})\Bigg]^2. \end{aligned}$$

When $\psi_j = 0\ (j \geq 1)$,

$$\mathrm{Var}(\mu_t) = \sigma_\varepsilon^2 \cdot \sum_{i=1}^{t-1}\left[-\beta k_i \cdot \prod_{j=i+1}^{t-1}(1-\beta k_j)\right]^2.$$

Put $\beta k_j = 1/j\ (1 \le j \le t-1)$ into the equation above, we get

$$\mathrm{Var}(\mu_t) = \sigma_\varepsilon^2 \cdot \frac{1}{t-1}. \tag{35}$$

From (7), it's easy to derive that

$$\mu_{t+1} = (1-\beta k_t)(\mu_t + \varepsilon_t) + (\psi_1 - 1)\varepsilon_t + \sum_{j=1}^{t-1}(\psi_{j+1} - \psi_j)\varepsilon_{t-j}.$$

When $\psi_j = 0\ (j \ge 1)$, we can derive from the above equation that

$$\begin{aligned}\mathrm{Var}(\mu_{t+1}) &= (1-\beta k_t)^2 \cdot \mathrm{Var}(\mu_t) + \sigma_\varepsilon^2 \cdot \beta^2 k_t^2 \\ &= \left[\mathrm{Var}(\mu_t) + \sigma_\varepsilon^2\right] \cdot \left[\beta k_t - \frac{\mathrm{Var}(\mu_t)}{\mathrm{Var}(\mu_t) + \sigma_\varepsilon^2}\right]^2 \\ &\quad + \frac{\sigma_\varepsilon^2 \cdot \mathrm{Var}(\mu_t)}{\mathrm{Var}(\mu_t) + \sigma_\varepsilon^2}. \end{aligned} \tag{36}$$

So in order to minimize $\mathrm{Var}(\mu_{t+1})$, we need to let

$$\beta k_t = \frac{\mathrm{Var}(\mu_t)}{\mathrm{Var}(\mu_t) + \sigma_\varepsilon^2}. \tag{37}$$

Then put (35) into (37), we get

$$k_t = \frac{1}{\beta} \cdot \frac{1}{t}$$

thus we finish our proof.

APPENDIX D

Suppose N_t is ARIMA(1,1,1) process with parameter ϕ and $\tilde{\theta}$, i.e.,

$$\begin{aligned} N_t - N_{t-1} &= \frac{1-\tilde{\theta}\mathcal{B}}{1-\phi\mathcal{B}} \cdot \varepsilon_t \\ &= (1 + \phi\mathcal{B} + \phi^2\mathcal{B}^2 + \cdots)(1-\tilde{\theta}\mathcal{B})\varepsilon_t \\ &= \varepsilon_t + (\phi - \tilde{\theta}) \cdot \varepsilon_{t-1} + (\phi - \tilde{\theta}) \cdot \sum_{i=1}^{t-2} \phi^{t-1-i} \cdot \varepsilon_i. \end{aligned} \tag{38}$$

Then

$$\mathrm{Var}(N_t - N_{t-1}) = \sigma_\varepsilon^2 + (\phi - \tilde{\theta})^2 \cdot \sigma_\varepsilon^2 + (\phi - \tilde{\theta})^2 \cdot \sigma_\varepsilon^2 \cdot \sum_{i=1}^{t-2} \phi^{2(t-1-i)}. \tag{39}$$

If we have infinite sample size, i.e., $t \to \infty$, then

$$\mathrm{Var}(N_t - N_{t-1}) = \sigma_\varepsilon^2 \cdot \left[1 + (\phi - \tilde{\theta})^2 + (\phi - \tilde{\theta})^2 \cdot \frac{\phi^2}{1-\phi^2}\right]. \tag{40}$$

Now, N_t is misidentified as IMA(1,1) process with parameter θ. Using moment estimation, we make

$$\sigma_\varepsilon^2 \cdot (1 + \hat{\theta}^2) = \sigma_\varepsilon^2 \cdot \left[1 + (\phi - \tilde{\theta})^2 + (\phi - \tilde{\theta})^2 \cdot \frac{\phi^2}{1-\phi^2}\right]. \tag{41}$$

That is,

$$\hat{\theta} = \frac{|\phi - \tilde{\theta}|}{\sqrt{1-\phi^2}}. \tag{42}$$

ACKNOWLEDGMENT

The authors are grateful to the three referees and the Associate Editor for many helpful suggestions that have significantly improved the quality of this paper.

REFERENCES

[1] D. W. Apley and J. Kim, "Cautious control of industrial process variability with uncertain input and disturbance model parameters," *Technometrics*, vol. 46, no. 2, pp. 188–199, May 2004.

[2] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 1994.

[3] G. E. P. Box and A. Luceño, *Statistical Control by Monitoring and Feedback Adjustment*. New York: Wiley, 1997.

[4] G. E. P. Box and T. Kramer, "Statistical process monitoring and feedback adjustment-a discussion," *Technometrics*, vol. 34, pp. 251–267, 1992.

[5] A. Chen and R. Guo, "Age-based double EWMA controller and its application to CMP processes," *IEEE Trans. Semicond. Manuf.*, vol. 14, no. 1, pp. 11–19, Feb. 2001.

[6] J. H. Chen, T. W. Kuo, and A. C. Lee, "Run-by-run process control of metal sputter deposition: Combining time series and extended kalman filter," *IEEE Trans. Semicond. Manuf.*, vol. 20, no. 3, pp. 278–285, Aug. 2007.

[7] E. Del Castillo and A. M. Hurwitz, "Run-to-Run process control: Literature review and extensions," *J. Quality Technol.*, vol. 29, no. 2, pp. 184–196, 1997.

[8] E. Del Castillo, R. Pan, and B. M. Colosimo, "A unifying view of some process adjustment methods," *J. Quality Technol.*, vol. 35, no. 3, Jul. 2003.

[9] F. E. Grubbs, "An optimum procedure for setting machines or adjusting processes," *Indust. Quality Control, July, 1957. Reprinted in J. Quality Technol.*, vol. 15, pp. 186–189, 1983.

[10] J. S. Hunter, "The exponentially weighted moving average," *J. Quality Technol.*, vol. 18, pp. 203–210, 1986.

[11] A. Ingolfsson and E. Sachs, "Stability and sensitivity of an EWMA controller," *J. Quality Technol.*, vol. 25, no. 4, pp. 271–287, Oct. 1993.

[12] M. Janakiram and J. B. Keats, "Combining SPC and EPC in a hybrid industry," *J. Quality Technol.*, vol. 30, pp. 189–200, 1998.

[13] A. Luceño, "Performance of discrete feedback adjustment schemes with dead band, under stationary versus nonstationary stochastic disturbances," *Technometrics*, vol. 40, pp. 223–233, 1998.

[14] D. C. Montgomery, J. B. Keats, G. C. Runger, and W. S. Messina, "Integrating statistical process control and engineering process control," *J. Quality Technol.*, vol. 26, no. 2, pp. 79–87, 1994.

[15] R. Pan, "Statistical Process Adjustment Methods for Quality Control in Short-Run Manufacturing," Ph.D. dissertation, Pennsylvania State Univ., , 2002.

[16] N. S. Patel and S. T. Jenkins, "Adaptive optimization of run-to-run controllers: The EWMA example," *IEEE Trans. Semicond. Manuf.*, vol. 13, no. 1, pp. 97–107, Feb. 2000.

[17] STS Innovations STS, STS in Nanotechnology, 2006.

[18] P. Sullo and M. Vandeven, "Optimal adjustment strategies for a process with run-to-run variation and 0–1 quality loss," *IIE Trans.*, vol. 31, no. 12, pp. 1135–1145, 1999.

[19] D. Trietsch, "The harmonic rule for process setup adjustment with quadratic loss," *J. Quality Technol.*, vol. 30, no. 1, pp. 75–84, 1998.

[20] S.-T. Tseng, A.-B. Yeh, F. Tsung, and Y.-Y. Chan, "A study of variable EWMA controller," *IEEE Trans. Semicond. Manuf.*, vol. 16, no. 4, pp. 633–643, Nov. 2003.

[21] P. Tsiamyrtzis and D. M. Hawkins, "A bayesian scheme to detect changes in the mean of a short-run process," *Technometrics*, vol. 47, no. 4, pp. 446–456, 2005.

[22] F. Tsung and D. W. Apley, "The dynamic T^2 chart for monitoring feedback-controlled processes," *IIE Trans.*, vol. 34, pp. 1043–1053, 2002.

[23] F. Tsung, H. Wu, and V. N. Nair, "On the efficiency and robustness of discrete proportional-integral control schemes," *Technometrics*, vol. 40, no. 3, pp. 214–222, Aug. 1998.

[24] E. Sachs, A. Hu, and A. Ingolfsson, "Run by run process control: Combining SPC and feedback control," *IEEE Trans. Semicond. Manuf.*, vol. 8, no. 1, pp. 26–43, Feb. 1995.

[25] S. A. Vander Wiel, "Monitoring processes that wander using integrated moving average models," *Technometrics*, vol. 38, pp. 139–151, 1996.

[26] S. A. Vander Wiel, W. T. Tucker, F. W. Faltin, and N. Doganaksoy, "Algorithmic statistical process control: Concepts and an application," *Technometrics*, vol. 34, pp. 286–297, 1992.

[27] K. Wang and F. Tsung, "Run-to-Run process adjustment using categorical observations," *J. Quality Technol.*, vol. 39, no. 4, pp. 312–325, 2007.

[28] M. F. Wu, C. H. Lin, D. S. H. Wong, S. S. Jang, and S. T. Tseng, "Performance analysis of EWMA controllers subject to metrology delay," *IEEE Trans. Semicond. Manuf.*, vol. 21, no. 3, pp. 413–425, Aug. 2008.

Fangyi He received the B.S. and M.S. degrees in statistics from University of Science and Technology of China (USTC), and the Ph.D. degree in industrial engineering from the School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ.

After his graduation, he joined the group of Equity and Derivative of BNP Paribas, New York, working as a quantitative researcher. His research interests include statistical process control, engineering process control, time series analysis, applied probability and statistical inference.

Dr. He is a member of American Statistical Association (ASA).

Kaibo Wang received the B. Eng. and M. Eng. degrees in mechatronics from Xi'an Jiaotong University, Xi'an, China, and the Ph.D. degree in industrial engineering and engineering management from Hong Kong University of Science and Technology (HKUST), Hong Kong.

He is currently an Assistant Professor in the Department of Industrial Engineering at Tsinghua University. He is on the Editorial Board for *Journal of the Chinese Institute of Industrial Engineers* (JCIIE). He has published papers on *Journal of Quality Technology*, *Quality and Reliability Engineering International*, *International Journal of Production Research* and others. His research interests include quality management, statistical process control, and run-to-run process control.

Wei Jiang received the B.Sc. and M.Sc. degrees from Xi'an Jiaotong University, Xi'an, China, in 1989 and 1992, respectively, and the Ph.D degree from Hong Kong University of Science and Technology (HKUST), Hong Kong, in 2000.

He has been working in AT&T labs for four years as a senior technical staff member and appointed as Assistant Professor in 2003 and Associated Professors in 2008 at Stevens Institute of Technology. He is now a Visiting Associate Professor in the Hong Kong University of Science and Technology. His research focuses on fault detection and activity monitoring using statistical and data mining methods. His current research interests include data mining and enterprise intelligence, statistical and data mining methods for quality control and management, as well as logistics management and financial applications.

Dr. Wei Jiang has been awarded a NSF project on business activity monitoring and is a recipient of NSF CAREER award on information quality management. He has published many research papers on quality journals including *IIE Transactions*, *Technometrics*, *Journal of Quality Technology*, etc., and is now serving as an associate editor of *Journal of Statistical Computation and Simulation*. He is the past chair of the Data Mining Section of the Institute for Operations Research and Management Sciences.