

This article was downloaded by: [Tsinghua University]

On: 25 March 2015, At: 07:27

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## IIE Transactions

Publication details, including instructions for authors and subscription information:  
<http://www.tandfonline.com/loi/uiie20>

### Optimal setup adjustment and control of a process under ARMA disturbances

Fangyi He<sup>a</sup>, Huiliang Xie<sup>b</sup> & Kaibo Wang<sup>c</sup>

<sup>a</sup> School of Finance, Southwestern University of Finance and Economics, Chengdu, Sichuan, 611130, People's Republic of China, E-mail:

<sup>b</sup> Division of Biostatistics, University of Miami, Miami, FL 33136, USA

<sup>c</sup> Department of Industrial Engineering, Tsinghua University, Beijing, 100084, People's Republic of China

Accepted author version posted online: 11 Jun 2014. Published online: 31 Oct 2014.



[Click for updates](#)

To cite this article: Fangyi He, Huiliang Xie & Kaibo Wang (2015) Optimal setup adjustment and control of a process under ARMA disturbances, IIE Transactions, 47:3, 230-244, DOI: [10.1080/0740817X.2014.928959](https://doi.org/10.1080/0740817X.2014.928959)

To link to this article: <http://dx.doi.org/10.1080/0740817X.2014.928959>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

# Optimal setup adjustment and control of a process under ARMA disturbances

FANGYI HE<sup>1,\*</sup>, HUILIANG XIE<sup>2</sup> and KAIBO WANG<sup>3</sup>

<sup>1</sup>*School of Finance, Southwestern University of Finance and Economics, Chengdu, Sichuan, 611130, People's Republic of China, E-mail: fangyi\_he@swufe.edu.cn*

<sup>2</sup>*Division of Biostatistics, University of Miami, Miami, FL 33136, USA*

<sup>3</sup>*Department of Industrial Engineering, Tsinghua University, Beijing, 100084, People's Republic of China*

Received October 2011 and accepted March 2013

Process adjustment uses information from past runs to adjust settings for the next run and bring the output to its target. The efficiency of a control algorithm depends on the nature of the disturbance and dynamics of the process. This article develops a control algorithm when the disturbance is a general ARMA( $p, q$ ) process, in the presence of measurement error and adjustment error together with a random initial bias. Its optimality property is established and the stability conditions are derived. It is shown that the popular Exponentially Weighted Moving Average (EWMA) controller is a special case of the proposed controller. In addition, Monte Carlo simulations are conducted to study the finite sample behavior of the proposed controller and compare it with the proportional–integral–derivative controller when the disturbance is an ARMA(1,1) process and with the EWMA controller when the disturbance is an IMA(1,1) process. The ARMA controller is also implemented to control an ARMA(2,1) disturbance and its performance is compared with the other two controllers. All of the results reflect the new controller's superiority when multiple sources of uncertainty exist or a general ARMA( $p, q$ ) disturbance is incurred.

**Keywords:** Process adjustment, state-space model, optimal control, ARMA, LQG

## 1. Introduction

In engineering and manufacturing, quality engineering is extensively involved in developing systems to ensure that products or services meet customers' requirements. Despite efforts to remove the causes of variation such as inaccurate testing methods, fluctuations in raw materials, and differences in operators, processes may still not be fully brought to a satisfactory state of stability. For example, an incorrect machine setup can result in an offset in the quality characteristic of the parts produced in the batch of product made subsequent to the setup; many semiconductor manufacturing processes can suffer from sudden component failures, gradual wear of components, or aging effects. To produce conforming products, a scheme of process adjustment or regulation is usually necessary for such processes to generate control actions and maintain the output on target. A recent review on statistical process adjustment can be found in Del Castillo (2006).

To motivate the type of adjustment problems considered in this article, we start by presenting the following classic Single-Input Single-Output (SISO) process:

$$y_t = f(x_{t-1}) + n_t, \quad (1)$$

where  $y_t$  is a continuous process output that needs to be adjusted to its target value  $\tau$ ,  $x_{t-1}$  denotes the process input recipe at the end of run  $t - 1$  (beginning of run  $t$ ), and  $n_t$  denotes the process disturbance that accounts for the variability in the process. The objective of a statistical process adjustment rule is to bring  $y_t$  as close to  $\tau$  as possible by adjusting  $x_t$  after each run. The efficiency of a control algorithm depends on the nature of the disturbance—i.e.,  $n_t$ —and the dynamics of the process; i.e.,  $f(\cdot)$ . Special forms of  $f(\cdot)$  have been widely studied in the literature; see, for example, Del Castillo and Hurwitz (1997), Wang and Tsung (2008), Jin and Tsung (2009), and Lin and Wang (2011).

For the case that  $f(x) = \alpha + x$  and  $n_t$  is White Noise (WN), Grubbs (1954) originally studied the setup adjustment problem. He designed an adjustment policy, which was called a “harmonic rule” by Trietsch (1998), to tune the process  $y_t$  close to the target value if at the startup  $y_t$  was off target by  $d$  units, where  $d$  is assumed to be an unknown value. In the same paper, Grubbs also presented a second, “extended rule,” which was designed for the case

\*Corresponding authors

Color versions of one or more of the figures in the article can be found online at [www.tandfonline.com/uiie](http://www.tandfonline.com/uiie).

that  $d$  is allowed to be a random variable with a prior distribution. Under the assumption that  $n_t$  follows a normal distribution, a general formulation and unification of these two control rules and others for setup adjustment problems can be found in Del Castillo *et al.* (2003). These authors showed how the previous setup adjustment rules are all cases of Linear Quadratic Gaussian (LQG) control. Instead of assuming process parameters are all known and considering only off-target costs, Lian and Del Castillo (2006) gave a solution to the setup adjustment problem using a dynamic programming formulation for the case that the measurement error variance is unknown and there are also fixed adjustment costs in additions to quadratic off-target costs. Lian *et al.* (2006) designed a new Sequential Monte Carlo (SMC) adjustment method to solve the setup adjustment problem under the scenario that the initial bias mean, variance, and measurement error variance are all unknown. Lian and Del Castillo (2007) considered a similar problem except for assuming no initial bias but adjustment error incurred when adjusting  $x_t$ . And the variance of the adjustment error was also assumed to be unknown in that paper. The authors designed an adaptive deadband control rule using a Bayesian approach based on SMC methods.

In addition to setup problems, various Run-to-Run (R2R) controllers have been developed in semiconductor manufacturing with respect to Equation (1) when  $f(x) = \alpha + \beta x$ . A popular controller in industrial practice is the Exponentially Weighted Moving Average (EWMA) controller, which was proposed in Ingolfsson and Sachs (1993). Variants of the EWMA controller include the double EWMA controller investigated in Butler and Stefani (1994), Del Castillo (1999), and Tseng *et al.* (2002) and the variable EWMA controller proposed in Tseng *et al.* (2003) and Tseng *et al.* (2007). When  $f(x) = \alpha + \beta x$  and  $n_t$  is an IMA(1,1) process that can be represented by

$$n_t = n_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}, \quad |\theta| < 1 \quad \text{and} \quad \varepsilon_t \sim \text{WN}, \quad (2)$$

the EWMA controller enjoys the Minimum Mean Square Error (MMSE) property when all of the parameters ( $\alpha$ ,  $\beta$ ,  $\theta$ ) are known and the initial bias  $d$  is zero (Box *et al.*, 1994). However, several factors may prevent the EWMA controller from achieving its optimal status. First, process parameters are seldom known and parameter estimation uncertainties always exist. As a result, the initial bias  $d$  rarely has a value of zero. Robustness of the EWMA controller when the parameters are estimated instead of known was investigated in Ingolfsson and Sachs (1993) and He *et al.* (2009). Second, process disturbance may not follow an IMA model; for example, a more general ARMA model may fit the disturbance better because of certain inertia effects. Third, multiple error sources, such as measurement error in observing the process output  $y_t$  and adjustment error in changing settings of controllable factors  $x_t$ , may exist. These errors are not considered in constructing the EWMA controller. In fact, the EWMA

controller was developed based on the single source of error model. The only error considered in EWMA controller is  $\varepsilon_t$  in Equation (2), hereafter referred to as the intrinsic error within the disturbance. The intrinsic error cannot be eliminated by any control algorithm.

Another widely used feedback controller is a Proportional–Integral–Derivative (PID) controller, which involves three separate constant parameters: the proportional, the integral, and the derivative values, denoted by  $k_P$ ,  $k_I$ , and  $k_D$ . Heuristically, these values can be interpreted in terms of time:  $k_P$  depends on the present error,  $k_I$  on the accumulation of past errors, and  $k_D$  is a prediction of future errors based on current rate of change. PID controllers are widely used in automated process control. Tsung and Shi (1999) developed a methodology to choose  $k_P$ ,  $k_I$ , and  $k_D$  when the process disturbance is an ARMA(1,1) process. However, they did not consider the measurement error, the adjustment error, and the initial bias of the process. For a general ARMA( $p$ ,  $q$ ) disturbance, they also did not provide an implemental method.

In this article, we build a more general framework to consider the problem that the process disturbance is assumed an ARMA( $p$ ,  $q$ ) disturbance. The ARMA model has been widely used to describe process dynamics; for example, the ARMA(1,1) model in Tsung and Shi (1999); ARMA(2,1) model in Hong and Macgregor (1975), MacGregor (1976), Pandit and Wu (1977) and Jiang *et al.* (2000); and the ARMA(3,2) model in Jiang *et al.* (2000). In addition to the intrinsic error, the new framework considers the measurement error and the adjustment error as well as allowing an initial bias with a prior distribution. Some setup adjustment problems and some R2R control problems can be included in the new framework. We derive an optimal control algorithm, the ARMA controller, based on the multiple source of error framework. Since an ARMA approach can be used to model any weakly stationary time series process as implied by the Wold representation theorem (Wold, 1948), the proposed ARMA controller can be applied to a wide range of disturbances.

The rest of this article is organized as follows. Section 2 gives the state-space framework of the model we consider. We first build the framework for a SISO system and then quickly extend the results to a Multiple-Input Multiple-Output (MIMO) system. Section 3 develops the ARMA controller based on Bayesian rule and Kalman filter theory and establishes its optimality under the LQG formulation without considering the adjustment cost. Section 4 presents the ARMA controller's stability conditions. Relations between the ARMA controller and the EWMA controller are examined in Section 5. Section 6 provides a performance analysis of ARMA controller, PID controller, and EWMA controller under multiple types of uncertainty. Concluding remarks and discussion of future work are made in Section 7.

The notation used in this article is as follows. Scalars are denoted by lowercase letters and column vectors by

lowercase letters in bold font. Matrices are denoted by upper case letters. For a column vector  $\mathbf{v}$ ,  $\mathbf{v}^\#$  denotes the number of components in  $\mathbf{v}$ . For an unknown parameter  $\beta$  or vector  $\boldsymbol{\beta}$ ,  $\hat{\beta}$  or  $\hat{\boldsymbol{\beta}}$  denotes its estimator, respectively.  $\mathcal{B}$  is the back-shift operator; i.e.,  $\mathcal{B}n_t = n_{t-1}$ ;  $\mathbf{0}$  denotes a vector or matrix with all components 0s.  $\mathbf{y}_1^t$  is the history of the output process from 1 to  $t$ ; i.e.,  $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t\}$ . We let

$$\text{diag}(A_1, A_2, \dots, A_s) = \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_s \end{pmatrix}.$$

## 2. Problem description

In this section, we will first consider a SISO system with a general ARMA( $p, q$ ) disturbance, involving measurement errors, adjustment errors, as well as an initial bias with a prior distribution. The SISO system's state-space framework will be developed. Next we quickly extend the results to a MIMO system.

### 2.1. Model setup for a SISO system

Suppose a simple process to be controlled,  $y_{it}$ , which corresponds to the  $i$ th component of a complex system, can be expressed as

$$y_{it} = g_{it} + n_{it} + e_{it}, \quad (3)$$

where  $g_{it}$  is the state of  $y_{it}$  at time  $t$ ;  $n_{it}$  is an ARMA( $p, q$ ) process satisfying

$$\phi_i(\mathcal{B})n_{it} = \theta_i(\mathcal{B})\varepsilon_{it}, \quad t = 0, \pm 1, \dots,$$

where  $\phi_i(z) = 1 - \phi_{i1}z - \phi_{i2}z^2 - \dots - \phi_{ip}z^p$  and  $\theta_i(z) = 1 + \theta_{i1}z + \theta_{i2}z^2 + \dots + \theta_{iq}z^q$ .  $\{\varepsilon_{it}\} \sim \text{WN}(0, \sigma_{i\varepsilon}^2)$  and  $\sigma_{i\varepsilon}^2$  is known. We call  $e_{it}$  the measurement error for  $y_{it}$  with  $\{e_{it}\} \sim \text{WN}(0, \sigma_{ie}^2)$ , where  $\sigma_{ie}^2$  is known. Neither  $g_{it}$  nor  $n_{it}$  can be measured or observed directly. At time  $t = 0$ , the process initial bias  $d$  (i.e.,  $g_{i0}$ ) is assumed to be a random variable with mean and variance  $\mu_{id}$  and  $\sigma_{id}^2$ , respectively. At time  $t \geq 1$ , suppose we need to make an adjustment of magnitude  $\Delta g_{it}$  to  $g_{it}$  to bring the process output to target in the next run. That is,

$$g_{i(t+1)} = g_{it} + \Delta g_{it} + w_{it}, \quad (4)$$

where  $w_{it}$  is a WN process called the adjustment error with known variance  $\sigma_{iw}^2$ . Assume  $\text{Corr}(e_{it}, w_{it}) = \text{Corr}(e_{it}, \varepsilon_{it}) = \text{Corr}(w_{it}, \varepsilon_{it}) = 0$ . In practice, the process adjustment  $\Delta g_{it}$  may not be directly made. Instead, it is assumed to be done by adjusting a controllable factor,  $x_{jt}$ , via the following model:

$$\Delta g_{it} = \beta_{ij}\Delta x_{jt}, \quad \beta_{ij} \neq 0, \quad (5)$$

where  $x_{jt}$  corresponds to the  $j$ th controllable factor and  $\beta_{ij}$  is called the process gain. For a SISO system, we can set  $j \equiv$

$i$ . If  $\Delta x_{j0} = 0$ , then  $\Delta g_{i0} = 0$ . Without loss of generality, the target  $\tau$  is assumed to be 0 in the rest of this article. For a positive time index  $T$ , we hope to determine the optimal  $\Delta x_{jt}$ ,  $t = 1, 2, \dots, T-1$  that satisfy

$$\min_{\{\Delta x_{jt}, t=1,2,\dots,T-1\}} \mathbb{E} \left( \sum_{t=1}^T y_{it}^2 \right). \quad (6)$$

The objective function  $\mathbb{E}(\sum_{t=1}^T y_{it}^2)$  is the process Mean Square Error (MSE), a measure of off-target cost for the process output  $y_{it}$  ( $t = 1, 2, \dots, T$ ). Here we assume the adjustment cost is neglectable in comparison with the off-target cost.

Brockwell and Davis (1991) gave two state-space representations for a general ARMA( $p, q$ ) process. Let  $r = \max(p, q + 1)$ ,  $\phi_{ik} = 0$  for  $k > p$ ,  $\theta_{ik} = 0$  for  $k > q$  and  $\theta_{i0} = 1$ . Following their first representation, we can write  $n_{it}$  as follows:

$$\begin{aligned} n_{it} &= \boldsymbol{\theta}_i^\tau \mathbf{v}_{it}, \\ \mathbf{v}_{i(t+1)} &= \mathbf{G}_i \mathbf{v}_{it} + \mathbf{c} \varepsilon_{i(t+1)}, \quad t = 0, \pm 1, \dots, \end{aligned}$$

where  $\boldsymbol{\theta}_i = (\theta_{i(r-1)} \theta_{i(r-2)} \dots \theta_{i0})^\tau$ ,  $\mathbf{v}_{it} = (v_{i(t-r+1)} v_{i(t-r+2)} \dots v_{it})^\tau$ ,  $\mathbf{c} = (0 \dots 0 1)^\tau$ ,  $\mathbf{c}^\# = r$ , and

$$\mathbf{G}_i = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ \phi_{ir} & \phi_{i(r-1)} & \phi_{i(r-2)} & \dots & \phi_{i1} \end{pmatrix}_{r \times r}.$$

From Equations (3) to (5) and the state-space representation of  $n_{it}$ , it is easy to get  $y_{it}$ 's state-space representation as follows:

$$y_{it} = \mathbf{h}_i^\tau \mathbf{l}_{it} + e_{it}, \quad e_{it} \sim N(0, \sigma_{ie}^2), \quad (7)$$

$$\begin{aligned} \mathbf{l}_{i(t+1)} &= \mathbf{A}_i \mathbf{l}_{it} + \mathbf{j} \beta_{ij} \Delta x_{jt} + \mathbf{z}_{it}, \quad \mathbf{l}_{i0} \sim N(\hat{\mathbf{l}}_{i0}, P_{i0}) \\ \mathbf{z}_{it} &\sim N(\mathbf{0}, \Sigma_{iz}), \end{aligned} \quad (8)$$

where  $\mathbf{h}_i = (1 \ \boldsymbol{\theta}_i^\tau)^\tau$ ,  $\mathbf{l}_{it} = (g_{it} \ \mathbf{v}_{it}^\tau)^\tau$ ,  $\mathbf{j} = (1 \ \mathbf{0})^\tau$ ,  $\mathbf{z}_{it} = (w_{it} \ \mathbf{0} \ \varepsilon_{it})^\tau$ ,  $\hat{\mathbf{l}}_{i0} = (\mu_{id}, \mathbf{0})^\tau$ ,  $P_{i0} = \text{diag}(\sigma_{id}^2, \mathbf{0})_{(r+1) \times (r+1)}$ ,  $\mathbf{A}_i = \text{diag}(1, \mathbf{G}_i)_{(r+1) \times (r+1)}$ ,  $\Sigma_{iz} = \text{diag}(\sigma_{iw}^2, 0, \dots, 0, \sigma_{i\varepsilon}^2)_{(r+1) \times (r+1)}$ , and  $\mathbf{h}_i^\# = \mathbf{l}_{it}^\# = \hat{\mathbf{l}}_{i0}^\# = \mathbf{j}^\# = \mathbf{z}_{it}^\# = r + 1$ .

The family of a general ARMA( $p, q$ ) process is a broad class and includes many common disturbances. For example, as Table 1 shows, when  $\phi_{i1}$  and  $\theta_{i1}$  of the ARMA(1,1)-type disturbance takes certain values, it reduces to some special time series. It is worth noting that although the conventional ARMA(1,1) model only approaches the non-stationary IMA(1,1) model approximately when  $\phi_{i1} \rightarrow 1$ , all of the derivations and results presented in the rest of this work still hold even if  $\phi_{i1} = 1$  exactly. Therefore, the method proposed in this work also solves the control of processes with an IMA(1,1) disturbance series exactly rather than in an approximate way.

**Table 1.** Types of disturbance included in ARMA(1,1)

	$\phi_{i1} = 0$	$ \phi_{i1}  < 1$	$\phi_{i1} = 1$
$\theta_{i1} = 0$	WN	AR(1)	Random walk
$ \theta_{i1}  < 1$	MA(1)	ARMA(1,1)	IMA(1,1)

## 2.2. Model setup for a MIMO system

Now we consider a complex linear system with  $s$  outputs  $y_{it}$  ( $i = 1, 2, \dots, s$ ) and  $m$  controllable factors  $x_{jt}$  ( $j = 1, 2, \dots, m$ ) and quickly extend the state-space framework of the SISO system to the  $(m \times s)$  MIMO system, where  $m \geq s$ . Again, we let the  $s$  outputs modeled as deviations from target and suppose that the MIMO system can be described as follows:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{g}_t + \mathbf{n}_t + \mathbf{e}_t, \quad \mathbf{e}_t \sim N(\mathbf{0}, \Sigma_e), \\ \mathbf{g}_{t+1} &= \mathbf{g}_t + B\Delta\mathbf{x}_t + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(\mathbf{0}, \Sigma_w), \end{aligned}$$

where  $\mathbf{y}_t = (y_{1t} \ y_{2t} \ \dots \ y_{st})^\tau$ ;  $\mathbf{g}_t = (g_{1t} \ g_{2t} \ \dots \ g_{st})^\tau$ ;  $\mathbf{n}_t = (n_{1t} \ n_{2t} \ \dots \ n_{st})^\tau$ ;  $\mathbf{e}_t = (e_{1t} \ e_{2t} \ \dots \ e_{st})^\tau$ ;  $\mathbf{w}_t = (w_{1t}w_{2t} \ \dots \ w_{st})^\tau$ ;  $B = (\beta_{ij})_{s \times m}$ , and  $\Delta\mathbf{x}_t = (\Delta x_{1t} \ \Delta x_{2t} \ \dots \ \Delta x_{mt})^\tau$ . Each component of vector  $\mathbf{n}_t$  can be modeled by a general ARMA( $p, q$ ) process. The orders  $p$  and  $q$  are determined by the maximum orders of the ARMA disturbances for all of the  $s$  components. Similar to a SISO system, we can get  $\mathbf{n}_t$ 's state-space representation as follows:

$$\begin{aligned} \mathbf{n}_t &= \Theta^\tau \mathbf{v}_t, \\ \mathbf{v}_{t+1} &= G\mathbf{v}_t + C\mathbf{e}_{t+1}, \end{aligned}$$

where  $\mathbf{e}_t = (\varepsilon_{1t} \ \varepsilon_{2t} \ \dots \ \varepsilon_{st})^\tau$ ,  $\Theta = \text{diag}(\theta_1, \theta_2, \dots, \theta_s)_{rs \times s}$ ,  $G = \text{diag}(G_1, G_2, \dots, G_s)_{rs \times s}$ ,  $C = \text{diag}(\mathbf{c}, \mathbf{c}, \dots, \mathbf{c})_{rs \times s}$ ,  $\mathbf{v}_t = (\mathbf{v}_{1t}^\tau \ \mathbf{v}_{2t}^\tau \ \dots \ \mathbf{v}_{st}^\tau)^\tau$ , and  $\mathbf{v}_t^\# = rs$ . Then we can derive  $\mathbf{y}_t$ 's state-space representation as

$$\mathbf{y}_t = H^\tau \mathbf{l}_t + \mathbf{e}_t, \quad \mathbf{e}_t \sim N(\mathbf{0}, \Sigma_e), \quad (9)$$

$$\begin{aligned} \mathbf{l}_{t+1} &= A\mathbf{l}_t + JB\Delta\mathbf{x}_t + \mathbf{z}_t, \quad \mathbf{l}_0 \sim N(\hat{\mathbf{l}}_0, P_0) \\ \mathbf{z}_t &\sim N(\mathbf{0}, \Sigma_z), \end{aligned} \quad (10)$$

where  $A = \text{diag}(A_1, A_2, \dots, A_s)_{(r+1)s \times (r+1)s}$ ,  $H = \text{diag}(\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_s)_{(r+1)s \times s}$ ,  $J = \text{diag}(\mathbf{j}, \mathbf{j}, \dots, \mathbf{j})_{(r+1)s \times s}$ ,  $\mathbf{l}_t = (\mathbf{l}_{1t}^\tau \ \mathbf{l}_{2t}^\tau \ \dots \ \mathbf{l}_{st}^\tau)^\tau$ ,  $\mathbf{z}_t = (\mathbf{z}_{1t}^\tau \ \mathbf{z}_{2t}^\tau \ \dots \ \mathbf{z}_{st}^\tau)^\tau$ ,  $\hat{\mathbf{l}}_0 = (\hat{\mathbf{l}}_{10}^\tau \ \hat{\mathbf{l}}_{20}^\tau \ \dots \ \hat{\mathbf{l}}_{s0}^\tau)^\tau$ ,  $P_0 = \text{diag}(P_{10}, P_{20}, \dots, P_{s0})_{(r+1)s \times (r+1)s}$ ,  $\Sigma_e = \text{diag}(\sigma_{1e}^2, \sigma_{2e}^2, \dots, \sigma_{se}^2)_{s \times s}$ ,  $\Sigma_z = \text{diag}(\Sigma_{1z}, \Sigma_{2z}, \dots, \Sigma_{sz})_{(r+1)s \times (r+1)s}$ , and  $\mathbf{l}_t^\# = \hat{\mathbf{l}}_t^\# = \mathbf{z}_t^\# = (r+1)s$ . Then the optimization problem for the MIMO system is for a positive time index  $T$ , we hope to determine the optimal  $\Delta\mathbf{x}_t$ ,  $t = 1, 2, \dots, T-1$  that satisfy

$$\min_{\{\Delta\mathbf{x}_t, t=1,2,\dots,T-1\}} E \left( \sum_{t=1}^T \mathbf{y}_t^\tau \mathbf{y}_t \right). \quad (11)$$

## 3. The ARMA controller

Let us define the posterior mean of  $\mathbf{l}_{t-1}$  as  $\hat{\mathbf{l}}_{t-1} = E(\mathbf{l}_{t-1} | \mathbf{y}_1^{t-1})$  and the posterior variance of  $\mathbf{l}_{t-1}$  as  $P_{t-1} =$

$\text{Var}(\mathbf{l}_{t-1} | \mathbf{y}_1^{t-1})$ . Then we have the posterior distribution of  $\mathbf{l}_{t-1}$  given the observations  $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{t-1}\}$  as

$$\mathbf{l}_{t-1} | \mathbf{y}_1^{t-1} \sim N(\hat{\mathbf{l}}_{t-1}, P_{t-1}).$$

Then based on Equation (10), we can get that the prior distribution of  $\mathbf{l}_t$  at time  $t-1$  is

$$\mathbf{l}_t | \mathbf{y}_1^{t-1} \sim N(\hat{A}\hat{\mathbf{l}}_{t-1} + JB\Delta\mathbf{x}_{t-1}, AP_{t-1}A^\tau + \Sigma_z). \quad (12)$$

With the setup above, solution to the optimization problems (9)–(11) can be derived via the classic LQG formulation. In Appendix A, we give a complete solution to a more general problem in Lemma A1. The following Theorem 1 only presents the optimal controller we are interested in this article.

**Theorem 1.** (ARMA controller) *If  $B$  is an invertible square matrix, then the optimal control algorithm that solves Equation (11) for models (9)–(10) is*

$$\Delta\mathbf{x}_t = -B^{-1}H^\tau \hat{\mathbf{l}}_t, \quad t = 1, 2, \dots \quad (13)$$

$$\text{where } \hat{\mathbf{l}}_t = (A - JH^\tau A)\hat{\mathbf{l}}_{t-1} + K_t\mathbf{y}_t, \quad (14)$$

$$K_t = (AP_{t-1}A^\tau + \Sigma_z)H[H^\tau(AP_{t-1}A^\tau + \Sigma_z)H + \Sigma_e]^{-1}, \quad (15)$$

$$P_t = (I - K_tH^\tau)(AP_{t-1}A^\tau + \Sigma_z), \quad (16)$$

$$\begin{aligned} \hat{\mathbf{l}}_0 &= (\hat{\mathbf{l}}_{10}^\tau \ \hat{\mathbf{l}}_{20}^\tau \ \dots \ \hat{\mathbf{l}}_{s0}^\tau)^\tau \text{ and } P_0 \\ &= \text{diag}(P_{10}, P_{20}, \dots, P_{s0})_{(r+1)s \times (r+1)s}. \end{aligned} \quad (17)$$

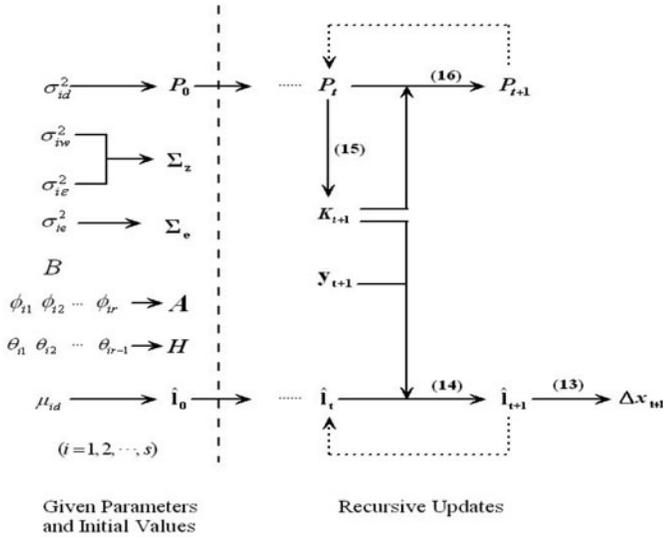
It is worth noting that Equations (13) to (17) are all unrelated to  $T$ , so the ARMA controller also solves

$$\min_{\{\Delta\mathbf{x}_t, t=1,2,\dots\}} E \left( \sum_{t=1}^{\infty} \mathbf{y}_t^\tau \mathbf{y}_t \right). \quad (18)$$

That is to say, the ARMA controller is an optimal controller to both the short-run (finite-horizon) problem and the long-run (infinite-horizon) problem. Note that Theorem 1 is naturally applicable to a SISO system ( $m = s = 1$ ) where  $B = \beta \neq 0$ . The proof of Theorem 1 is relegated to Appendix A. For the purpose of reference, we present in Corollary 1 the ARMA controller's version for a SISO system.

**Corollary 1.** *The optimal control algorithm that solves Equation (6) or*

$$\min_{\{\Delta x_{jt}, t=1,2,\dots\}} E \left( \sum_{t=1}^{\infty} y_{it}^2 \right), \quad (19)$$



**Fig. 1.** The implementation procedure of the ARMA controller. The equation used for each step is indicated by the equation number.

for models (7)–(8) is

$$\Delta x_{jt} = -\frac{1}{\beta_{ij}} \mathbf{h}_i^T A_i \widehat{\mathbf{I}}_{it}, \quad t = 1, 2, \dots, \quad (20)$$

$$\text{where } \widehat{\mathbf{I}}_{it} = (A_i - \mathbf{j} \mathbf{h}_i^T A_i) \widehat{\mathbf{I}}_{i(t-1)} + K_{it} y_{it}, \quad (21)$$

$$K_{it} = (A_i P_{i(t-1)} A_i^T + \Sigma_{iz}) \mathbf{h}_i [\mathbf{h}_i^T (A_i P_{i(t-1)} A_i^T + \Sigma_{iz}) \mathbf{h}_i + \sigma_{ie}^2]^{-1}, \quad (22)$$

$$P_{it} = (I - K_{it} \mathbf{h}_i^T) (A_i P_{i(t-1)} A_i^T + \Sigma_{iz}), \quad (23)$$

$$\widehat{\mathbf{I}}_{i0} = (\mu_{id}, \mathbf{0})^T \quad \text{and} \quad (24)$$

$$P_{i0} = \text{diag}(\sigma_{id}^2, \mathbf{0})_{(r+1) \times (r+1)}.$$

The implementation of the ARMA controller is illustrated in Fig. 1. Given parameters and initial values in the offline procedure, all of the values recursively update online according to the algorithm. At the startup,  $K_1$  can be derived from Equation (15) based on  $P_0$ . Upon the observation of  $\mathbf{y}_1$ ,  $\widehat{\mathbf{I}}_1$  can be computed from Equation (14) based on  $K_1$  and  $\widehat{\mathbf{I}}_0$ . Then the first adjustment  $\Delta \mathbf{x}_1$  can be obtained by applying Equation (13). From Equation (16), we can compute  $P_1$  using  $K_1$  and  $P_0$ . Then  $K_2$  is derived from Equation (15) in the way  $K_1$  is obtained, and the control process continues.

Equation (20) is the optimal control rule for a SISO system with a general ARMA( $p, q$ ) disturbance. Note that  $n_{it}$  will become a WN when  $\phi_{ik} = \theta_{ik} = 0$  ( $k = 1, 2, \dots, r$ ). At this time, if we set  $\beta_{ij} = 1$ , Equation (20) will become  $\Delta x_{jt} = -\widehat{g}_{it}$ , which is exactly the adjustment rule derived by Del Castillo *et al.* (2003). In Appendix B, we shown that  $E(\sum_{t=1}^T y_i^T \mathbf{y}_t)$  must be greater than all the sums of  $\sigma_{iw}^2$ ,  $\sigma_{ie}^2$ , and  $\sigma_{ie}^2$  ( $i = 1, 2, \dots, s$ ), which makes sense because control schemes cannot correct the adjustment error, the measurement error, or the intrinsic error within the process

disturbance in spite of their capabilities of eliminating the “effects” of the process disturbance.

#### 4. Stability conditions

In order to implement the ARMA controller, some parameters need to be estimated first. Specifically, the invertible matrix  $B$ , which is the process gain, is usually unknown. Therefore, with an offline estimate  $\widehat{B}$ , the adjustment rule for  $\mathbf{x}_t$  becomes

$$\Delta \mathbf{x}_t = -\widehat{B}^{-1} H^T \widehat{\mathbf{A}}_t. \quad (25)$$

Based on Equations (9) and (12),  $E(\mathbf{y}_t)$  can be written as

$$E(\mathbf{y}_t) = E[E(\mathbf{y}_t | \mathbf{y}_1^{t-1})] = H^T [AE(\widehat{\mathbf{I}}_{t-1}) + JBE(\Delta \mathbf{x}_{t-1})]. \quad (26)$$

Combining Equation (26) with Equation (25), we get

$$E(\mathbf{y}_t) = (I - B\widehat{B}^{-1})H^T AE(\widehat{\mathbf{I}}_{t-1}), \quad (27)$$

and from Equation (A11) in Appendix A and Equation (25), we get

$$\widehat{\mathbf{I}}_t = (A - JH^T A)\widehat{\mathbf{I}}_{t-1} + K_t y_t + (J - K_t)(I - B\widehat{B}^{-1})H^T \widehat{\mathbf{A}}_{t-1}. \quad (28)$$

Note that when  $\widehat{B} = B$ ,  $E(\mathbf{y}_t) = 0$  for any  $t \geq 1$  and Equation (28) will become Equation (14). The stability conditions for the ARMA controller are given in Theorem 2.

**Theorem 2.** Let  $D = A - JB\widehat{B}^{-1}H^T A$ , then the asymptotical stability conditions of the control rule (25) for model (9)–(10) with  $B$  and  $\widehat{B}$  invertible matrices are:

$$(i) \quad 0 < \lambda_1(D^T D) < 1; \quad (29)$$

(ii) There exist  $x^* < \infty$  and  $y^* > 0$ , such that

$$\lambda_1(H^T A P_{t-1} A^T H + H^T D E(\widehat{\mathbf{I}}_{t-1} \widehat{\mathbf{I}}_{t-1}^T) D^T H) \leq x^* \text{ for all } t > y^*, \quad (30)$$

where  $\lambda_1(X)$  denotes the largest eigenvalue of matrix  $X$ ;  $P_t$  and  $\widehat{\mathbf{I}}_t$  are described in Equations (16) and (28) respectively;  $\widehat{\mathbf{I}}_0$  and  $P_0$  are given by Equation (17).

The proof of Theorem 2 can be found in Appendix C. Note that for a SISO system, the invertible matrices  $B$  and  $\widehat{B}$  are equivalent to  $\beta_{ij} (\neq 0)$  and  $\widehat{\beta}_{ij} (\neq 0)$ . It is easy to verify that the stability condition (29) is equivalent to  $0 < \beta_{ij}/\widehat{\beta}_{ij} < 2$  and the condition (30) is equivalent to the fact that there exist  $x^* < \infty$  and  $y^* > 0$  such that  $\lambda_1(P_{t-1} + (1 - \beta_{ij}/\widehat{\beta}_{ij})^2 E(\widehat{\mathbf{I}}_{t-1} \widehat{\mathbf{I}}_{t-1}^T)) \leq x^*$  for all  $t > y^*$ .

#### 5. ARMA controller versus EWMA controller

An interesting question is how the ARMA controller relates to the widely used EWMA controller. Without loss of generality, we only focus on a SISO system. The following

**Table 2.** Influence of the measurement error

$\sigma_e$	$n_t$ is ARMA(1,1) $\phi_1 = 0.859$ $\theta_1 = -0.164$				$n_t$ is IMA(1,1) $\theta_1 = -0.8$			
	ARMA controller ( $\hat{\phi}_1 = 0.859,$ $\hat{\theta}_1 = -0.164$ )		PID controller ( $k_P = 0.24, k_I = 0.58,$ $k_D = -0.08$ )		ARMA controller ( $\hat{\phi}_1 = 1,$ $\hat{\theta}_1 = -0.8$ )		EWMA controller ( $\omega = 1+$ $\hat{\theta}_1 = 0.2$ )	
	AMSE	(SEAMSE)	AMSE	(SEAMSE)	AMSE	(SEAMSE)	AMSE	(SEAMSE)
0	0.994	(0.004)	1.046	(0.005)	0.994	(0.004)	0.994	(0.004)
0.2	1.061	(0.005)	1.116	(0.005)	1.047	(0.005)	1.047	(0.005)
0.4	1.231	(0.006)	1.303	(0.006)	1.178	(0.006)	1.179	(0.006)
0.6	1.497	(0.007)	1.609	(0.007)	1.389	(0.006)	1.393	(0.006)
0.8	1.860	(0.008)	2.049	(0.009)	1.695	(0.008)	1.705	(0.008)
1.0	2.313	(0.010)	2.617	(0.012)	2.089	(0.009)	2.112	(0.010)
1.2	2.829	(0.013)	3.290	(0.016)	2.550	(0.012)	2.589	(0.012)
1.4	3.422	(0.014)	4.100	(0.020)	3.087	(0.013)	3.155	(0.014)
1.6	4.099	(0.019)	5.048	(0.025)	3.731	(0.017)	3.830	(0.018)
1.8	4.862	(0.022)	6.093	(0.031)	4.431	(0.020)	4.570	(0.021)
2.0	5.654	(0.025)	7.237	(0.035)	5.219	(0.023)	5.406	(0.024)

theorem shows that the EWMA controller is a special case of the ARMA controller.

**Theorem 3.** *The following control problem defined by Equations (31) to (33):*

$$y_{it} = \alpha_i + \beta_{ij}x_{jt(t-1)} + n_{it}, \tag{31}$$

$$a_{it} = \omega(y_{it} - \hat{\beta}_{ij}x_{jt(t-1)}) + (1 - \omega)a_{i(t-1)}, \quad a_{i0} = d, \tag{32}$$

$$x_{jt} = -\frac{a_{it}}{\hat{\beta}_{ij}}, \tag{33}$$

is equivalent to the control problem defined by Equations (34) to (38).

$$y_{it} = g_{it} + n_{it} + e_{it}, \tag{34}$$

$$g_{i(t+1)} = g_{it} + \beta_{ij}\Delta x_{jt} + w_{it}, \tag{35}$$

$$e_{it} = w_{it} = 0, \tag{36}$$

$$\Delta x_{jt} = -\frac{1}{\hat{\beta}_{ij}} \omega y_{it}, \tag{37}$$

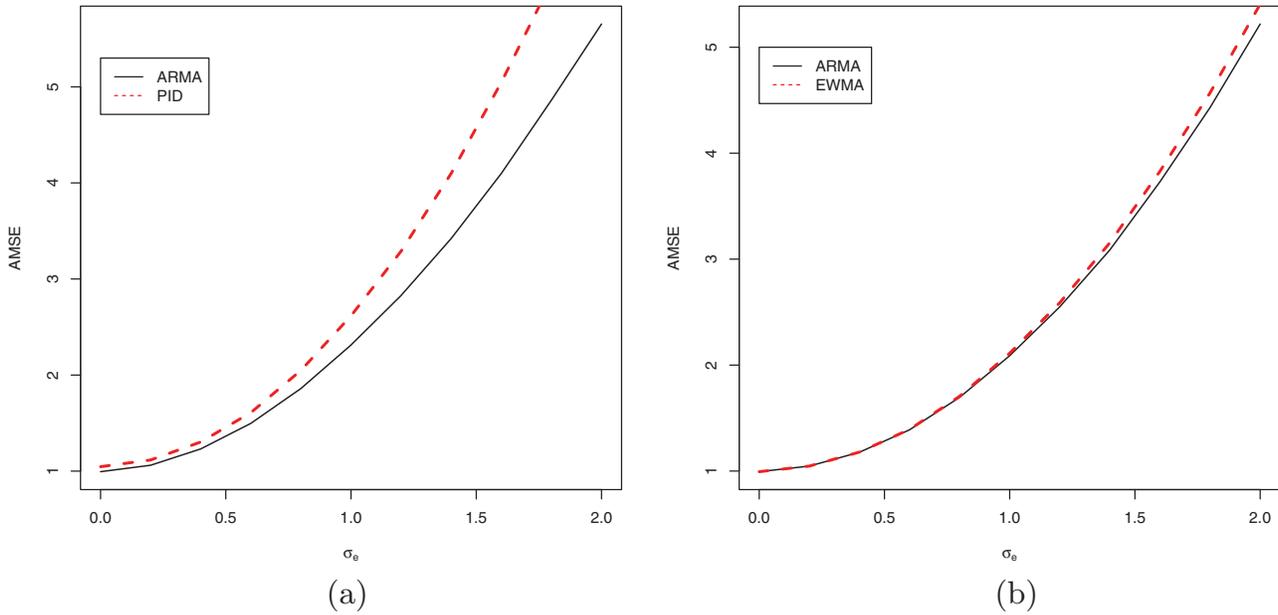
$$g_{i0} = d. \tag{38}$$

The proof of Theorem 3 can be found in Appendix D. From Theorem 3 we know that, in comparison with the ARMA controller, the traditional EWMA controller considers neither measurement error nor adjustment error (i.e.,  $e_{it} = w_{it} = 0$ , then  $\sigma_{ie} = \sigma_{iw} = 0$ ). Second, the initial bias  $d$  of  $y_{it}$  from the target is assumed to be unknown but a fixed constant in the traditional EWMA controller (i.e.,  $\mu_{id} = d$  and  $\sigma_{id} = 0$ ), whereas it is assumed to be a random variable with a prior distribution having mean and variance be  $\mu_{id}$  and  $\sigma_{id}^2$ , respectively, in the ARMA controller. Last but not least, the parameter  $\omega$  in the traditional EWMA controller is set haphazardly by practitioners. As a rule of thumb,  $\omega$  is usually between 0.1 and 0.3, regardless of the disturbance process  $n_{it}$  (Hunter, 1986). Thus, in

general  $\Delta x_{jt}$  defined by Equation (37) is not equal to that by Equation (20) even if  $\hat{\beta}_{ij} = \beta_{ij}$ , which means that the traditional EWMA controller is not optimal for a general ARMA( $p, q$ ) disturbance. Only when the following four conditions are satisfied does Equation (37) equal Equation (20): (i)  $n_{it}$  is an IMA(1,1) process with parameter  $\theta_{i1}$ ; (ii)  $\omega = 1 + \theta_{i1}$ ; (iii)  $d = 0$ ; and (iv)  $\hat{\beta}_{ij} = \beta_{ij}$ . This fact can be easily verified. Hence, the traditional EWMA controller is a special case of the ARMA controller and the EWMA controller could reach its optimality in a very rare case.

## 6. Performance analysis

For simplicity, in this section, we study the performance of the newly proposed ARMA controller under multiple scenarios through Monte Carlo simulations and only focus on the SISO system of Equations (3) to (5) and ignore the model parameters suffix  $i$  or  $j$  in the rest of this article; i.e.,  $n_t$  replaces  $n_{it}$ ,  $\beta$  replaces  $\beta_{ij}$ , etc. The proposed ARMA controller in this work can be applied to any general ARMA disturbance series. In the following, we first conduct a performance study assuming that the disturbance  $n_t$  follows an ARMA(1,1) process and an IMA(1,1) process, respectively, and then apply it to an ARMA(2,1) disturbance model. When  $n_t$  is an ARMA(1,1) process, the performance of the PID controller introduced in Tsung and Shi (1999) will also be evaluated and compared with that of the ARMA controller. As in that paper's example, we set  $\phi_1 = 0.895$  and  $\theta_1 = -0.164$  for the ARMA(1, 1) disturbance. The PID parameters  $k_P, k_I$ , and  $k_D$  are obtained by checking the PID design maps for ARMA(1,1) disturbance in Tsung and Shi (1999). Note that the parameter  $\theta_1$  of ARMA(1,1) process in this article corresponds to  $-\theta_1$  in Tsung and Shi (1999). When  $n_t$  is an IMA(1,1) process, the performance of the EWMA controller will be evaluated and compared with



**Fig. 2.** AMSE with different  $\sigma_e$  for the ARMA controller and other controllers when the disturbance is (a) ARMA(1,1) and (b) IMA(1,1), respectively. The black solid line is for the ARMA controller, whereas the red dashed line is for the alternative.

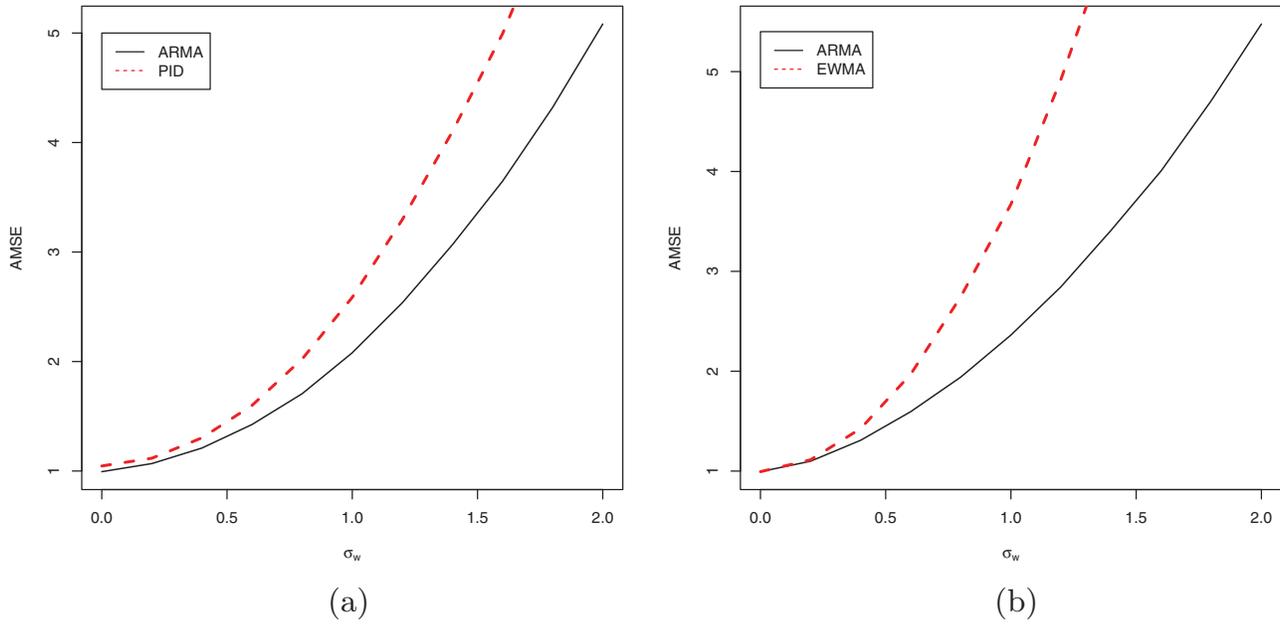
that of the ARMA controller. As we mentioned before, the EWMA controller may be an MMSE controller for an IMA(1,1) disturbance. We arbitrarily set  $\theta_1 = -0.8$  for the IMA(1, 1) disturbance and the EWMA parameter  $\omega$  is set to be  $1 + \theta_1 = 0.2$ . Measurement errors and adjustment errors are allowed in the simulations.

Seven scenarios are examined. In Scenario 1, the effect of the measurement error  $e_t$  on the ARMA, PID, and EWMA controllers is investigated; in Scenario 2, the effect of the adjustment error  $w_t$  is explored; in Scenario 3, we investigate the joint effect of both types of errors on the three

controllers; in Scenario 4, we look into the effect of the process initial bias  $d$ ; in Scenario 5, we study how the estimate of the process gain  $\hat{\beta}$  affects the controllers; in Scenario 6, we investigate the controllers' performance when the parameters in the disturbance  $n_t$  are unknown and cannot be estimated accurately; i.e.,  $\hat{\phi}_1 \neq \phi_1$  and  $\hat{\theta}_1 \neq \theta_1$ ; and in Scenario 7, we randomly draw one simulation and show the three controllers' performance on an ARMA(2,1) disturbance. Except for Scenario 6, we set  $\hat{\phi}_1 = \phi_1$  and  $\hat{\theta}_1 = \theta_1$  for simplicity. The standard deviation for the intrinsic error  $\sigma_\varepsilon$  is set to one in all simulations except for in Scenario 7.

**Table 3.** Influence of the adjustment error

$\sigma_w$	$n_t$ is ARMA(1, 1) $\phi_1 = 0.859$ $\theta_1 = -0.164$				$n_t$ is IMA(1, 1) $\theta_1 = -0.8$			
	ARMA controller ( $\hat{\phi}_1 = 0.859, \hat{\theta}_1 = -0.164$ )		PID controller ( $k_P = 0.24, k_I = 0.58, k_D = -0.08$ )		ARMA controller ( $\hat{\phi}_1 = 1, \hat{\theta}_1 = -0.8$ )		EWMA controller ( $\omega = 1 + \hat{\theta}_1 = 0.2$ )	
	AMSE	(SEAMSE)	AMSE	(SEAMSE)	AMSE	(SEAMSE)	AMSE	(SEAMSE)
0	0.994	(0.004)	1.046	(0.005)	0.994	(0.004)	0.994	(0.004)
0.2	1.068	(0.005)	1.117	(0.005)	1.099	(0.005)	1.112	(0.005)
0.4	1.210	(0.006)	1.303	(0.006)	1.308	(0.006)	1.426	(0.008)
0.6	1.425	(0.006)	1.600	(0.008)	1.595	(0.007)	1.970	(0.012)
0.8	1.706	(0.007)	2.023	(0.009)	1.941	(0.009)	2.747	(0.020)
1.0	2.080	(0.009)	2.586	(0.012)	2.362	(0.011)	3.672	(0.027)
1.2	2.537	(0.012)	3.297	(0.016)	2.846	(0.013)	4.925	(0.040)
1.4	3.067	(0.014)	4.101	(0.020)	3.412	(0.016)	6.363	(0.053)
1.6	3.647	(0.016)	4.994	(0.024)	4.007	(0.017)	7.961	(0.067)
1.8	4.321	(0.020)	6.092	(0.031)	4.710	(0.022)	9.655	(0.087)
2.0	5.082	(0.023)	7.247	(0.036)	5.476	(0.024)	11.836	(0.102)



**Fig. 3.** AMSE with different  $\sigma_w$  for the ARMA controller and other controllers when the disturbance is (a) ARMA(1,1) and (b) IMA(1,1), respectively. The black solid line is for the ARMA controller, whereas the red dashed line is for the alternative.

We performed 1000 replications for each scenario and ran 100 steps for  $y_t$  in each replication. The first 100-run MSE of  $y_t$  was computed. The Average MSE (AMSE) of the 1000 replications is reported in Tables 2 to 8. Also reported is the standard error in the AMSE (SEAMSE), which is

$$SEAMSE = \frac{SDMSE}{\sqrt{\text{Number of replicates}}}$$

where SDMSE is the standard deviation of MSEs, and the number of replicates used in the simulation was 1000. AMSE measures the performance of the controllers, and SEAMSE reflects the variability of the AMSE.

### 6.1. Effect of measurement error

In order to focus on the effect of the measurement error  $e_t$  on the controllers, we set both the adjustment error  $w_t$  and the initial process bias  $d$  to zero, and set the estimate of the process gain  $\hat{\beta}$  to the true value  $\beta$ .

Table 2 and Fig. 2 show that the AMSE of  $y_t$  when the standard deviation of the measurement error  $\sigma_e$  varies. We first compare the performance of the ARMA controller with the PID controller under the ARMA(1,1) disturbance and then the performance of the ARMA controller with the EWMA controller under the IMA(1,1) disturbance. We can observe that the AMSE of  $y_t$  increases for all of the controllers when  $\sigma_e$  increases, no matter whether  $n_t$  is an

**Table 4.** AMSE when both types of error exist

$(\sigma_e, \sigma_w)$	$n_t$ is ARMA(1, 1) ( $\phi_1 = 0.859$ $\theta_1 = -0.164$ )		$n_t$ is IMA(1, 1) ( $\theta_1 = -0.8$ )	
	ARMA controller $\hat{\phi}_1 = 0.859$ $\hat{\theta}_1 = -0.164$	PID controller $k_P = 0.24$ $k_I = 0.58$ $k_D = -0.08$	ARMA controller $\hat{\phi}_1 = 1.0$ $\hat{\theta}_1 = -0.8$	EWMA controller $\hat{\theta}_1 = -0.8$ $\omega = 1 + \hat{\theta}_1 = 0.2$
(0.0, 0.0)	0.994 (0.004)	1.046 (0.005)	0.994 (0.004)	0.994 (0.004)
(0.0, 0.4)	1.210 (0.006)	1.303 (0.006)	1.308 (0.006)	1.426 (0.008)
(0.4, 0.0)	1.231 (0.006)	1.303 (0.006)	1.178 (0.006)	1.179 (0.006)
(0.4, 0.4)	1.449 (0.006)	1.541 (0.007)	1.501 (0.007)	1.601 (0.008)
(0.4, 1.0)	2.355 (0.010)	2.836 (0.013)	2.583 (0.012)	3.833 (0.029)
(1.0, 0.4)	2.604 (0.012)	2.850 (0.014)	2.501 (0.007)	2.541 (0.011)
(1.0, 1.0)	3.658 (0.016)	4.148 (0.020)	3.769 (0.017)	4.788 (0.029)
(1.0, 2.0)	6.845 (0.030)	8.764 (0.043)	7.149 (0.032)	12.804 (0.102)
(2.0, 1.0)	7.678 (0.035)	8.836 (0.044)	7.535 (0.034)	8.116 (0.040)
(2.0, 2.0)	11.469 (0.051)	13.461 (0.067)	11.544 (0.052)	16.247 (0.111)

The corresponding SEAMSE values are enclosed in parentheses.

**Table 5.** AMSE when  $n_t$  is ARMA(1,1) with  $\phi_1 = 0.859$  and  $\theta_1 = -0.164$ 

$\mu_d$	ARMA controller $\hat{\phi}_1 = 0.859$ $\hat{\theta}_1 = -0.164$				PID controller $k_P = 0.24$ $k_I = 0.58$ $k_D = -0.08$					
	$\sigma_d = 0$	1	2	3	4	$\sigma_d = 0$	1	2	3	4
-4	1.164 (0.005)	1.180 (0.006)	1.217 (0.008)	1.258 (0.010)	1.339 (0.014)	1.304 (0.006)	1.320 (0.007)	1.366 (0.011)	1.434 (0.015)	1.553 (0.021)
-3	1.088 (0.005)	1.107 (0.005)	1.142 (0.007)	1.190 (0.009)	1.278 (0.013)	1.189 (0.005)	1.208 (0.006)	1.256 (0.009)	1.329 (0.012)	1.467 (0.019)
-2	1.037 (0.005)	1.055 (0.005)	1.104 (0.006)	1.137 (0.007)	1.201 (0.010)	1.108 (0.005)	1.125 (0.005)	1.191 (0.007)	1.248 (0.010)	1.347 (0.014)
-1	1.019 (0.005)	1.033 (0.005)	1.055 (0.005)	1.115 (0.007)	1.187 (0.009)	1.077 (0.005)	1.088 (0.005)	1.124 (0.006)	1.213 (0.009)	1.320 (0.013)
0	0.994 (0.004)	1.022 (0.005)	1.057 (0.005)	1.103 (0.006)	1.170 (0.009)	1.046 (0.005)	1.071 (0.005)	1.120 (0.006)	1.196 (0.008)	1.296 (0.013)
1	1.010 (0.004)	1.031 (0.004)	1.057 (0.005)	1.107 (0.006)	1.188 (0.009)	1.066 (0.005)	1.084 (0.005)	1.125 (0.006)	1.200 (0.009)	1.328 (0.013)
2	1.041 (0.005)	1.053 (0.005)	1.089 (0.006)	1.144 (0.008)	1.226 (0.010)	1.115 (0.005)	1.122 (0.006)	1.171 (0.007)	1.257 (0.011)	1.382 (0.015)
3	1.092 (0.005)	1.110 (0.005)	1.137 (0.006)	1.204 (0.009)	1.262 (0.013)	1.193 (0.006)	1.206 (0.006)	1.247 (0.008)	1.346 (0.013)	1.436 (0.019)
4	1.152 (0.005)	1.195 (0.006)	1.207 (0.007)	1.265 (0.010)	1.335 (0.014)	1.294 (0.006)	1.332 (0.007)	1.355 (0.010)	1.442 (0.015)	1.560 (0.021)

The corresponding SE/AMSE values are enclosed in parentheses. Neither measurement errors nor adjustment errors are considered in the simulations.

**Table 6.** AMSE when  $n_t$  is IMA(1,1) with  $\theta_1 = -0.8$ 

$\mu_d$	ARMA controller $\hat{\phi}_1 = 1$ $\hat{\theta}_1 = -0.8$				EWMA controller $\omega = 1 + \hat{\theta}_1 = 0.2$					
	$\sigma_d = 0$	1	2	3	4	$\sigma_d = 0$	1	2	3	4
-4	1.164 (0.005)	1.177 (0.006)	1.215 (0.008)	1.256 (0.010)	1.338 (0.014)	1.449 (0.006)	1.472 (0.010)	1.566 (0.017)	1.673 (0.024)	1.893 (0.035)
-3	1.088 (0.005)	1.105 (0.005)	1.141 (0.007)	1.188 (0.009)	1.277 (0.013)	1.252 (0.005)	1.267 (0.008)	1.366 (0.014)	1.495 (0.020)	1.730 (0.032)
-2	1.037 (0.005)	1.053 (0.005)	1.102 (0.006)	1.135 (0.007)	1.201 (0.010)	1.109 (0.005)	1.137 (0.006)	1.232 (0.010)	1.360 (0.016)	1.531 (0.024)
-1	1.019 (0.005)	1.031 (0.005)	1.052 (0.005)	1.114 (0.007)	1.186 (0.009)	1.036 (0.005)	1.062 (0.005)	1.131 (0.008)	1.282 (0.013)	1.475 (0.021)
0	0.994 (0.004)	1.020 (0.005)	1.054 (0.005)	1.100 (0.006)	1.168 (0.009)	0.994 (0.004)	1.030 (0.005)	1.118 (0.007)	1.254 (0.012)	1.447 (0.022)
1	1.010 (0.004)	1.029 (0.004)	1.054 (0.005)	1.104 (0.006)	1.187 (0.009)	1.027 (0.005)	1.059 (0.005)	1.133 (0.008)	1.270 (0.013)	1.484 (0.022)
2	1.041 (0.005)	1.050 (0.005)	1.088 (0.006)	1.142 (0.008)	1.224 (0.010)	1.113 (0.005)	1.129 (0.007)	1.220 (0.010)	1.365 (0.017)	1.589 (0.026)
3	1.092 (0.005)	1.108 (0.005)	1.134 (0.006)	1.202 (0.009)	1.259 (0.013)	1.253 (0.006)	1.272 (0.008)	1.338 (0.012)	1.521 (0.020)	1.684 (0.031)
4	1.152 (0.005)	1.192 (0.006)	1.205 (0.007)	1.263 (0.010)	1.333 (0.014)	1.434 (0.006)	1.490 (0.010)	1.545 (0.016)	1.701 (0.024)	1.903 (0.036)

The corresponding SE/AMSE values are included in parentheses. Neither measurement errors nor adjustment errors are considered in the simulations.

ARMA(1,1) or IMA(1,1) disturbance. However, the AMSE of  $y_t$  using the ARMA controller is always smaller than that using the other two controllers, at a significant level for almost all values of  $\sigma_e$ . Note that when  $\sigma_e = 0$  and  $n_t$  is ARMA(1,1) or IMA(1,1) disturbance, the AMSE of  $y_t$  using the ARMA controller arrives at its minimum value of one, which comes from the intrinsic error  $\varepsilon_t$ , whereas only when  $n_t$  does the IMA(1,1) disturbance does the AMSE of  $y_t$  using the EWMA controller achieve optimality when  $\sigma_e = 0$ .

### 6.2. Effect of adjustment error

For all three controllers, an upward trend is also seen for the AMSE of  $y_t$  as the standard deviation of the adjustment error  $\sigma_w$  increases when both the measurement error  $e_t$  and the initial process bias  $d$  are 0, and the estimate of the process gain  $\hat{\beta}$  equals the true value  $\beta$ . Table 3 and Fig. 3 show that the AMSE of  $y_t$  using the ARMA controller is consistently smaller than that using the PID controller or the EWMA controller for almost all the  $\sigma_w$  values except when  $\sigma_w = 0$  and  $n_t$  is an IMA(1,1) disturbance, in which case the ARMA controller and the EWMA controller are both the MMSE controller.

### 6.3. Joint effect of both types of error

Table 4 presents the performance of the three controllers when both measurement errors and adjustment errors exist. We arbitrarily set the values of pairs of  $(\sigma_e, \sigma_w)$  and repeated the simulations. As before, the estimate of the process gain  $\hat{\beta}$  was set to equal the true value  $\beta$ . The results show that the ARMA controller's performance dominates the other two controllers for all of the cases we analyzed.

### 6.4. Effect of initial bias

For varying sizes of  $\mu_d$  and  $\sigma_d$  in the prior distribution of the process initial bias, Table 5 shows the results of the AMSE of  $y_t$  using the ARMA controller and the PID controller for ARMA(1,1) disturbances; Table 6 shows the results of the AMSE of  $y_t$  using the ARMA controller and the EWMA controller for IMA(1,1) disturbances. For simplicity, we set  $e_t$  and  $w_t$  to 0 and the estimate of the process gain  $\hat{\beta}$  to the true value  $\beta$ . It can be seen that for all three controllers, the AMSE increases with  $\sigma_d$  when  $\mu_d$  is fixed and the AMSE increases with  $|\mu_d|$  when  $\sigma_d$  is fixed. For the same pair of values  $(\mu_d, \sigma_d)$ , the ARMA controller has a better performance than the other two controllers. When the PID or EWMA controller is implemented, the AMSE value increases more rapidly than when the ARMA controller is used as  $\mu_d$  deviates away from 0 and  $\sigma_d$  increases. The ARMA controller is the MMSE controller for both ARMA(1,1) and IMA(1,1) disturbances when the initial bias is 0 ( $\mu_d = \sigma_d = 0$ ), whereas the EWMA controller is

the MMSE controller only for IMA(1,1) disturbance in this situation.

### 6.5. Effect of estimation uncertainties in the process gain

In order to focus on the effect of the estimate of the process gain on the controller's performance, we assumed that measurement errors and adjustment errors were absent and the initial bias  $d$  was 0; i.e.,  $\mu_d = \sigma_d = 0$ . At this time, if  $n_t$  is an IMA(1,1) disturbance, the ARMA and EWMA controllers are exactly the same according to Theorem 3, so we only analyze the case that  $n_t$  is an ARMA(1,1) disturbance and compared the ARMA controller with the PID controller. Table 7 presents the sensitivity of the AMSE of  $y_t$  to  $\hat{\beta}$  for both the ARMA and PID controllers. The process gain  $\beta$  was set to 2.5. It can be seen from Table 7 that the ARMA controller outperforms the PID controller for all of the  $\hat{\beta}$  values we examined. Additionally, an underestimated  $\beta$  (i.e.,  $\hat{\beta}/\beta < 1$ ) will hurt the ARMA controller's performance more than an overestimated  $\beta$  (i.e.,  $\hat{\beta}/\beta > 1$ ). When  $\hat{\beta} = \beta$ , the MSE of the process output  $y_t$  reaches the minimum value of one when the ARMA controller is used.

### 6.6. Effect of estimation uncertainties in the disturbance

Now let us assume that there are estimation uncertainties in the disturbances; i.e.,  $\hat{\phi}_i \neq \phi_i$  and/or  $\hat{\theta}_i \neq \theta_i$  ( $i = 1, 2, \dots$ ). Again, measurement errors and adjustment errors were not considered and the initial bias  $d$  was set to 0. In this case the ARMA controller is exactly the same as the EWMA controller when  $n_t$  is an IMA(1,1) process from Theorem 3, so we only compared the ARMA controller and the PID controller when  $n_t$  is an ARMA(1,1) disturbance. We chose nine values of  $\phi_1$  and five values of  $\theta_1$  as the true parameters for the ARMA(1,1) disturbance, so a total of 45 disturbances were investigated. The disturbances were chosen representatively since from Table 1 it is evident that not only ARMA(1,1) disturbances but also WN, AR(1), Random Walk, MA(1), and IMA(1,1) disturbances are also included. Table 8 shows that the ARMA controller's performance is better than the PID controller's in most of the cases. Only when  $\phi_1$  is close or equal to one is the PID controller's performance better than the ARMA controller's performance. The results imply that if an unstationary disturbance is misidentified to be a stationary disturbance, the ARMA controller's performance will be affected more than the PID controller's performance. However, to control a weakly stationary disturbance, the ARMA controller is a better choice even if the disturbance parameters are misidentified.

We briefly explain why the ARMA controller has such good performance. From the results in Appendix B, we can derive that  $\text{AMSE} = \sigma_\varepsilon^2 + [(\hat{\theta}_1 + \hat{\phi}_1) - (\theta_1 + \phi_1)]^2 \hat{v}_1^2$ , where  $\hat{v}_1$  is the last component in  $\mathbf{I}_t$  and it represents the state of the ARMA(1,1) disturbance. First, note that the expression of AMSE is not related to  $\hat{g}_t$ , which means at every

**Table 7.** Performance with different estimates of the process gain

$n_t$ is $ARMA(1, 1)$ $\phi = 0.859$ $\theta = -0.164$ $\beta = 2.5$				
$\hat{\beta}$	<i>ARMA controller</i> ( $\hat{\phi} = 0.859$ $\hat{\theta} = -0.164$ )		<i>PID controller</i> ( $k_P = 0.24$ $k_I = 0.58$ $k_D = -0.08$ )	
	<i>AMSE</i>	<i>(SEAMSE)</i>	<i>AMSE</i>	<i>(SEAMSE)</i>
1.5	1.244	(0.007)	1.353	(0.007)
1.7	1.111	(0.005)	1.189	(0.006)
1.9	1.038	(0.005)	1.099	(0.005)
2.1	1.020	(0.005)	1.075	(0.005)
2.3	1.010	(0.004)	1.062	(0.005)
2.5	0.994	(0.004)	1.046	(0.005)
2.7	1.006	(0.005)	1.057	(0.005)
2.9	1.013	(0.005)	1.066	(0.005)
3.1	1.014	(0.005)	1.067	(0.005)
3.3	1.027	(0.005)	1.081	(0.005)
3.5	1.043	(0.005)	1.100	(0.005)

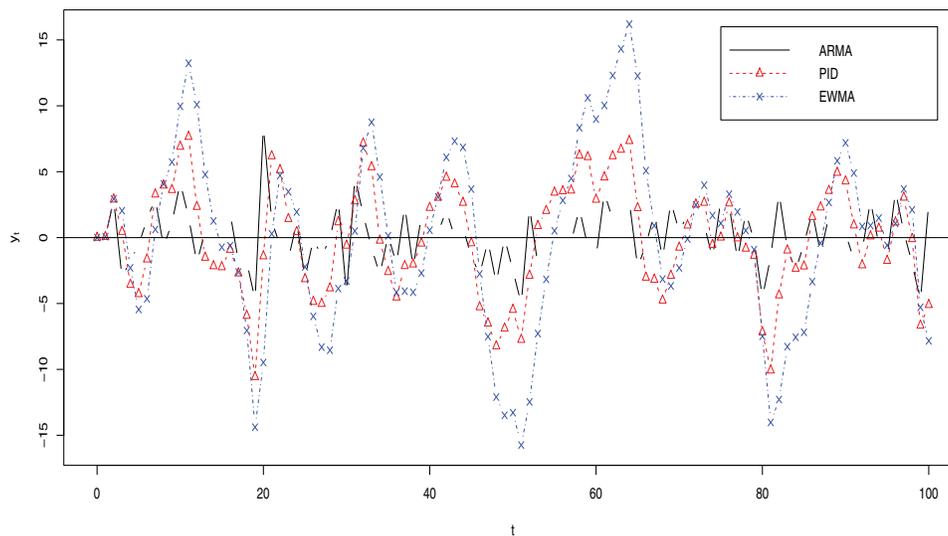
time  $t$  the process's bias can be removed by the control strategy, even if the parameters  $\theta_1$  and  $\phi_1$  are misidentified. This property explains why the ARMA controller can remove the initial bias very quickly as shown in Table 5 and Table 6. Second the expression for the AMSE implies that even if  $\hat{\theta}_1 \neq \theta_1$  and  $\hat{\phi}_1 \neq \phi_1$  but only if  $\hat{\theta}_1 + \hat{\phi}_1$  is close to  $\theta_1 + \phi_1$  can the controller's performance be close to optimal. Third, if the disturbance is weakly stationary,  $\hat{v}_t^2$  will not increase very much even if the disturbance parameters are misidentified. However, if a non-stationary disturbance—e.g., IMA(1,1) or Random Walk—is misidentified to be an ARMA(1,1) process,  $\hat{v}_t^2$  will rapidly increase. That is why the AMSE increases quickly when  $\phi_1 = 1$  in Table 8 for the ARMA controller.

### 6.7. Performance under an ARMA(2,1) disturbance model

The above studies were carried out based on an ARMA(1,1) or IMA(1,1) disturbance model. As previously mentioned, the proposed ARMA controller can be applied to any general ARMA( $p, q$ ) disturbance. For illustration purpose, we assume a process follows:

$$y_t = x_{t-1} + n_t$$

and  $n_t - 1.4385n_{t-1} + 0.6000n_{t-2} = \varepsilon_t + 0.5193\varepsilon_{t-1}$ ,  $\varepsilon_t$  is a white noise series having  $\sigma_\varepsilon^2 = 2.2120^2 = 4.8929$ ;  $y_t$  is the deviation from target. The disturbance model is the same as the ARMA(2,1) model introduced by Jiang *et al.* (2000). For simplicity, we further set  $\sigma_e = \sigma_w = 0$  and  $y_0 = 0$ . That is, we ignore measurement errors, adjustment errors, and



**Fig. 4.** One simulated path by using the ARMA, PID, and EWMA controllers for the ARMA(2,1) disturbance with  $n_t - 1.4385n_{t-1} + 0.6000n_{t-2} = \varepsilon_t + 0.5193\varepsilon_{t-1}$  with  $\sigma_\varepsilon = 2.2120$ .



developing optimal control algorithms for ARIMA  $(p, d, q)$  disturbances. Since ARIMA  $(p, d, q)$  could be widely used to model non-stationary or periodic processes, such an extension deserves more further research efforts.

### Acknowledgements

We are grateful to the Department Editor and the anonymous referees for the valuable comments, which have helped us significantly improve this work.

### Funding

Dr. Fangyi He's work was supported by the National Natural Science Foundation of China (grant. 71102145) and the SRF for ROCS, SEM. Dr. Kaibo Wang's work was supported by the National Natural Science Foundation of China (grant. 71072012) and the Tsinghua University Initiative Scientific Research Program.

### References

- Box, G.E.P., Jenkins, G.M., and Reinsel, G.C. (1994) *Time Series Analysis: Forecasting and Control*, 3 edition, Prentice Hall, Upper Saddle River, NJ.
- Brockwell, P.J. and Davis, R.A. (1991) *Time Series: Theory and Methods*, Springer-Verlag, New York, NY.
- Butler, S.W. and Stefani, J.A. (1994) Supervisory run-to-run control of polysilicon gate etch using in situ ellipsometry. *IEEE Transactions on Semiconductor Manufacturing*, **7**(2), 193–201.
- Del Castillo, (1999) Long run and transient analysis of a double EWMA feedback controller. *IIE Transactions*, **31**(12), 1157–1169.
- Del Castillo, E. (2006) Statistical process adjustment: a brief retrospective, current status, and some opportunities for further work. *Statistica Neerlandica*, **60**(3), 309–326.
- Del Castillo, E. and Hurwitz, A.M. (1997) Run-to-run process control: literature review and extensions. *Journal of Quality Technology*, **29**(2), 184–196.
- Del Castillo, E., Pan, R. and Colosimo, B.M. (2003) A unifying view of some process adjustment methods. *Journal of Quality Technology*, **35**(3), 286–293.
- Grubbs, F.E. (1954) An optimum procedure for setting machines or adjusting processes. *Industrial Quality Control*. Reprinted in *Journal of Quality Technology*, **15**(14), 186–189, 1983.
- He, F., Wang, K. and Jiang, W. (2009) A general harmonic rule controller for run-to-run process control. *IEEE Transactions on Semiconductor Manufacturing*, **22**(2), 232–244.
- Hong, H.M. and MacGregor, J.F. (1975) Identification and direct digital stochastic control of a continuous stirred tank process. *The Canadian Journal of Chemical Engineering*, **53**(2), 211–216.
- Hunter, J.S. (1986) The exponentially weighted moving average. *Journal of Quality Technology*, **18**(4), 203–210.
- Ingolfsson, A. and Sachs, E. (1993) Stability and sensitivity of an EWMA controller. *Journal of Quality Technology*, **25**(4), 271–287.
- Jiang, W., Tsui, K.L. and Woodall, W.H. (2000) A new SPC monitoring method: the ARMA chart. *Technometrics*, **42**(4), 399–410.

- Jin, M. and Tsung, F. (2009) A chart allocation strategy for multistage processes. *IIE Transactions*, **41**(9), 790–803.
- Lewis, F.L. (1986) *Optimal Estimation with an Introduction to Stochastic Control Theory*, John Wiley & Sons, New York, NY.
- Lian, Z., Colosimo, B.M. and Del Castillo, E. (2006) Setup adjustment of multiple lots using a sequential Monte Carlo method. *Technometrics*, **48**(3), 373–385.
- Lian, Z. and Del Castillo, E. (2006) Setup adjustment under unknown process parameters and fixed adjustment cost. *Journal of Statistical Planning and Inference*, **136**, 1039–1060.
- Lian, Z. and Del Castillo, E. (2007) Adaptive deadband control of a drifting process with unknown parameters. *Statistics & Probability Letters*, **77**, 843–852.
- Lin, J. and Wang, K. (2011) Online parameter estimation and run-to-run process adjustment using categorical observations. *International Journal of Production Research*, **49**(13), 4103–4117.
- MacGregor, J.F. (1976) Optimal choice of the sampling interval for discrete process control. *Technometrics*, **18**(2), 151–160.
- Pandit, S.M. and Wu, S.M. (1977) Modeling and analysis of closed-loop systems from operating data. *Technometrics*, **19**(4), 477–485.
- Trietsch, D. (1998) The harmonic rule for process setup adjustment with quadratic loss. *Journal of Quality Technology*, **30**(1), 75–84.
- Tseng, S.T., Chou, R.J. and Lee, S.P. (2002) Statistical design of a double EWMA controller. *Applied Stochastic Models in Business and Industry*, **18**(3), 313–322.
- Tseng, S.T., Tsung, F. and Liu, P.Y. (2007) Variable EWMA run-to-run controller for drifted processes. *IIE Transactions*, **39**(3), 291–301.
- Tseng, S.T., Yeh, B., Tsung, F. and Chan, I.Y. (2003) A study of variable EWMA controller. *IEEE Transactions on Semiconductor Manufacturing*, **16**(4), 633–643.
- Tsung, F. and Shi, J. (1999) Integrated design of run-to-run PID controller and SPC monitoring for process disturbance rejection. *IIE Transactions*, **31**, 517–527.
- Vander Wiel, S.A. (1996) Monitoring processes that wander using integrated moving average models. *Technometrics*, **38**(2), 139–151.
- Vander Wiel, S.A., Tucker, W.T., Faltin, F.W. and Doganaksoy, N. (1992) Algorithmic statistical process control: concepts and an application. *Technometrics*, **34**(3), 286–297.
- Wang, K. and Tsung, F. (2008) An adaptive  $T^2$  chart for monitoring dynamic systems. *Journal of Quality Technology*, **40**(1), 109–123.
- Wold, H.O.A. (1948) On prediction in stationary time series. *The Annals of Mathematical Statistics*, **19**(4), 558–567.

### Appendices

#### Appendix A

In order to prove Theorem 1, we need to use the following Lemma first. Lemma A1 can be proved easily by the results given in Lewis (1986, p. 315).

**Lemma A1.** Suppose that there are  $m$  controllable factors and  $s$  outputs modeled as deviations from the target value. Assume that the process is described by the linear equations:

$$\mathbf{y}_t = V^T \mathbf{I}_t + \mathbf{e}_t, \quad \mathbf{e}_t \sim N(\mathbf{0}, \Sigma_e),$$

$$\mathbf{I}_{t+1} = \Phi \mathbf{I}_t + \Gamma \Delta \mathbf{x}_t + \mathbf{z}_t, \quad \mathbf{I}_0 \sim N(\hat{\mathbf{I}}_0, P_0) \quad \mathbf{z}_t \sim N(\mathbf{0}, \Sigma_z),$$

and the objective function to minimize is quadratic and equals

$$\mathbb{E} \left[ \mathbf{I}_T^r \mathbf{Q} \mathbf{I}_T + \sum_{t=1}^{T-1} (\mathbf{I}_t^r \mathbf{Q} \mathbf{I}_t + \Delta \mathbf{x}_t^r R \Delta \mathbf{x}_t) \right]. \quad (\text{A1})$$

Then the optimal solution is as follows:

$$\Delta \mathbf{x}_t = -L_t \widehat{\mathbf{I}}_t, \quad (\text{A2})$$

$$\text{where} \quad (\text{A3})$$

$$L_t = (R + \Gamma^r S_{t+1} \Gamma)^{-1} \Gamma^r S_{t+1} \Phi, \quad t \leq T-1, \quad (\text{A4})$$

$$S_T = Q, \quad S_t = \Phi^r S_{t+1} \Phi + Q - L_t^r (R + \Gamma^r S_{t+1} \Gamma) L_t, \quad t \leq T-1, \quad (\text{A5})$$

$$\text{and } \widehat{\mathbf{I}}_t = \Phi \widehat{\mathbf{I}}_{t-1} + \Gamma \Delta \mathbf{x}_{t-1} + K_t [\mathbf{y}_t - V^r (\Phi \widehat{\mathbf{I}}_{t-1} + \Gamma \Delta \mathbf{x}_{t-1})], \quad (\text{A6})$$

$$K_t = (\Phi P_{t-1} \Phi^r + \Sigma_z) V [V^r (\Phi P_{t-1} \Phi^r + \Sigma_z) V + \Sigma_e]^{-1}, \quad (\text{A7})$$

$$P_t = (I - K_t V^r) (\Phi P_{t-1} \Phi^r + \Sigma_z). \quad (\text{A8})$$

Note that in Lemma A1,  $Q$  measures the off-target cost and  $R$  measures the adjustment cost. From Equation (9) we can get that

$$\mathbb{E} \left( \sum_{t=1}^T \mathbf{y}_t^r \mathbf{y}_t \right) = \mathbb{E} \left[ \mathbf{I}_T^r (HH^r) \mathbf{I}_T + \sum_{t=1}^{T-1} \mathbf{I}_t^r (HH^r) \mathbf{I}_t \right] + T \sum_{i=1}^s \sigma_{ie}^2.$$

That is, to minimize  $\mathbb{E}(\sum_{t=1}^T \mathbf{y}_t^r \mathbf{y}_t)$  is equivalent to minimizing

$$\mathbb{E} \left[ \mathbf{I}_T^r (HH^r) \mathbf{I}_T + \sum_{t=1}^{T-1} \mathbf{I}_t^r (HH^r) \mathbf{I}_t \right].$$

Comparing this objective function with that in Lemma A1, we get that  $Q = HH^r$  and  $R = 0$ , which means that the off-target cost is measured by  $HH^r$  and the adjustment cost is ignored. Using Lemma A1 directly by replacing  $V$ ,  $\Phi$ , and  $\Gamma$  with  $H$ ,  $A$ , and  $JB$ , respectively, Equations (A3)–(A8) become

$$L_t = (B^r J^r S_{t+1} JB)^{-1} B^r J^r S_{t+1} A, \quad t \leq T-1, \quad (\text{A9})$$

$$S_T = HH^r, \quad S_t = A^r S_{t+1} A + HH^r - L_t^r B^r J^r S_{t+1} J B L_t, \quad t \leq T-1, \quad (\text{A10})$$

$$\widehat{\mathbf{I}}_t = \widehat{\mathbf{A}}_{t-1} + JB \Delta \mathbf{x}_{t-1} + K_t [\mathbf{y}_t - H^r (\widehat{\mathbf{A}}_{t-1} + JB \Delta \mathbf{x}_{t-1})], \quad (\text{A11})$$

$$K_t = (AP_{t-1} A^r + \Sigma_z) H [H^r (AP_{t-1} A^r + \Sigma_z) H + \Sigma_e]^{-1}, \quad (\text{A12})$$

$$P_t = (I - K_t H^r) (AP_{t-1} A^r + \Sigma_z). \quad (\text{A13})$$

Using the fact that  $B$  is an invertible matrix and  $J^r H$  is an identity matrix, we can derive from Equations (A9)

and (A10) that  $S_t \equiv HH^r$  and  $L_t \equiv B^{-1} H^r A$  for any  $t = 1, 2, \dots, T-1$ ; i.e.,  $\Delta \mathbf{x}_t = -B^{-1} H^r \widehat{\mathbf{A}}_t$ . Substituting the expression of  $\Delta \mathbf{x}_t$  into Equation (A11), we get that

$$\begin{aligned} \widehat{\mathbf{I}}_t &= \widehat{\mathbf{A}}_{t-1} + JB(-B^{-1} H^r \widehat{\mathbf{A}}_{t-1}) + K_t \mathbf{y}_t \\ &\quad - K_t [H^r \widehat{\mathbf{A}}_{t-1} + H^r JB(-B^{-1} H^r \widehat{\mathbf{A}}_{t-1})] \\ &= (A - JH^r A) \widehat{\mathbf{I}}_{t-1} + K_t \mathbf{y}_t, \end{aligned}$$

which is Equation (14). Now, we have proved that for any  $T > 0$ , the algorithms (13) to (16) minimize the objective function  $\mathbb{E}(\sum_{t=1}^T \mathbf{y}_t^r \mathbf{y}_t)$ . This completes the proof.

### Appendix B

As

$$\mathbb{E} [\mathbf{I}_t^r (HH^r) \mathbf{I}_t] = \mathbb{E} \{ \mathbb{E} [\mathbf{I}_t^r (HH^r) \mathbf{I}_t | \mathbf{y}_1^{t-1}] \}, \quad (\text{A14})$$

we can get from Equations (12) and (13) that

$$\begin{aligned} &\mathbb{E} [\mathbf{I}_t^r (HH^r) \mathbf{I}_t | \mathbf{y}_1^{t-1}] \\ &= \mathbb{E} \left\{ [(\mathbf{I}_t | \mathbf{y}_1^{t-1}) - \mathbb{E}(\mathbf{I}_t | \mathbf{y}_1^{t-1})]^r HH^r [(\mathbf{I}_t | \mathbf{y}_1^{t-1}) - \mathbb{E}(\mathbf{I}_t | \mathbf{y}_1^{t-1})] \right\} + [\mathbb{E}(\mathbf{I}_t | \mathbf{y}_1^{t-1})]^r HH^r [\mathbb{E}(\mathbf{I}_t | \mathbf{y}_1^{t-1})] \\ &= \text{trace} \left\{ HH^r \mathbb{E} \left\{ [(\mathbf{I}_t | \mathbf{y}_1^{t-1}) - \mathbb{E}(\mathbf{I}_t | \mathbf{y}_1^{t-1})][(\mathbf{I}_t | \mathbf{y}_1^{t-1}) - \mathbb{E}(\mathbf{I}_t | \mathbf{y}_1^{t-1})]^r \right\} \right\} + [\mathbb{E}(\mathbf{I}_t | \mathbf{y}_1^{t-1})]^r HH^r [\mathbb{E}(\mathbf{I}_t | \mathbf{y}_1^{t-1})] \\ &= \text{trace} [HH^r (AP_{t-1} A^r + \Sigma_z)] + (\widehat{\mathbf{A}}_{t-1} + JB \Delta \mathbf{x}_{t-1})^r HH^r (\widehat{\mathbf{A}}_{t-1} + JB \Delta \mathbf{x}_{t-1}) \quad (\text{A15}) \\ &= \text{trace}(H^r AP_{t-1} A^r H) + \text{trace}(H^r \Sigma_z H) \\ &= \text{trace}(H^r AP_{t-1} A^r H) + \sum_{i=1}^s \sigma_{iw}^2 + \sum_{i=1}^s \sigma_{ie}^2; \end{aligned}$$

thus,

$$\begin{aligned} \mathbb{E} \left( \sum_{t=1}^T \mathbf{y}_t^r \mathbf{y}_t \right) &= \sum_{t=1}^T \mathbb{E} [\mathbf{I}_t^r (HH^r) \mathbf{I}_t] + T \sum_{i=1}^s \sigma_{ie}^2 \\ &= \sum_{t=1}^T \text{trace}(H^r AP_{t-1} A^r H) + T \sum_{i=1}^s \sigma_{iw}^2 \\ &\quad + T \sum_{i=1}^s \sigma_{ie}^2 + T \sum_{i=1}^s \sigma_{ie}^2. \quad (\text{A16}) \end{aligned}$$

Note that  $\text{trace}(H^r AP_{t-1} A^r H) > 0$  for all  $t = 1, 2, \dots$  since  $P_{t-1}$ 's are covariance matrices. That means that

$$\mathbb{E} \left( \sum_{t=1}^T \mathbf{y}_t^r \mathbf{y}_t \right) > T \sum_{i=1}^s (\sigma_{iw}^2 + \sigma_{ie}^2 + \sigma_{ie}^2).$$

### Appendix C

From Equations (27) and (28) we get that

$$\begin{aligned} E(\widehat{\mathbf{l}}_t) &= (A - JH^T A)E(\widehat{\mathbf{l}}_{t-1}) + K_t(I - B\widehat{B}^{-1})H^T AE(\widehat{\mathbf{l}}_{t-1}) \\ &\quad + (J - K_t)(I - B\widehat{B}^{-1})H^T AE(\widehat{\mathbf{l}}_{t-1}) \\ &= (A - JH^T A)E(\widehat{\mathbf{l}}_{t-1}) + J(I - B\widehat{B}^{-1})H^T AE(\widehat{\mathbf{l}}_{t-1}) \\ &= (A - JB\widehat{B}^{-1}H^T A)E(\widehat{\mathbf{l}}_{t-1}). \end{aligned} \quad (\text{A17})$$

Then combining (A17) and Equation (27), we get that

$$E(\mathbf{y}_t) = (I - B\widehat{B}^{-1})H^T A(A - JB\widehat{B}^{-1}H^T A)^{t-1}E(\widehat{\mathbf{l}}_0). \quad (\text{A18})$$

Let  $D = A - JB\widehat{B}^{-1}H^T A$ . Then, there exists an orthogonal matrix  $F$  such that

$$D^T D = F^T \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{(r+1)s}) F,$$

where  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{(r+1)s} \geq 0$ . Thus,

$$\begin{aligned} (D^{t-1})^T D^{t-1} &= (D^T D)^{t-1} \\ &= F^T \text{diag}(\lambda_1^{t-1}, \lambda_2^{t-1}, \dots, \lambda_{(r+1)s}^{t-1}) F. \end{aligned}$$

Using the Singular Value Decomposition theorem, we know that there exists another orthogonal matrix  $U$  such that

$$D^{t-1} = U^T \text{diag}\left(\sqrt{\lambda_1^{t-1}}, \sqrt{\lambda_2^{t-1}}, \dots, \sqrt{\lambda_{(r+1)s}^{t-1}}\right) F.$$

Under Condition (29), we get  $0 \leq \lambda_{(r+1)s} \leq \dots \leq \lambda_1 < 1$ , so  $D^{t-1} \rightarrow 0$  as  $t \rightarrow \infty$ . That is,  $\lim_{t \rightarrow \infty} E(\mathbf{y}_t) = 0$ . From Equations (9), (25), (A14), and (A15), we can get that

$$E(\mathbf{y}_t^T \mathbf{y}_t) = E[\mathbf{l}_t^T (HH^T) \mathbf{l}_t] + \sum_{i=1}^s \sigma_{ie}^2 \quad (\text{A19})$$

$$\begin{aligned} &= \text{trace}(H^T AP_{t-1} A^T H) + \sum_{i=1}^s (\sigma_{iw}^2 + \sigma_{ie}^2 + \sigma_{ie}^2) \\ &\quad + E[\widehat{\mathbf{l}}_{t-1}^T (A - JB\widehat{B}^{-1}H^T A)^T HH^T \\ &\quad \times (A - JB\widehat{B}^{-1}H^T A) \widehat{\mathbf{l}}_{t-1}] \end{aligned} \quad (\text{A20})$$

$$\begin{aligned} &= \text{trace}(H^T AP_{t-1} A^T H) + \sum_{i=1}^s (\sigma_{iw}^2 + \sigma_{ie}^2 + \sigma_{ie}^2) \\ &\quad + \text{trace}[H^T DE(\widehat{\mathbf{l}}_{t-1} \widehat{\mathbf{l}}_{t-1}^T) D^T H] \\ &= \text{trace}[H^T AP_{t-1} A^T H + H^T DE(\widehat{\mathbf{l}}_{t-1} \widehat{\mathbf{l}}_{t-1}^T) D^T H] \\ &\quad + \sum_{i=1}^s (\sigma_{iw}^2 + \sigma_{ie}^2 + \sigma_{ie}^2). \end{aligned} \quad (\text{A21})$$

Under Condition (30), we know  $\lim_{t \rightarrow \infty} E(\mathbf{y}_t^T \mathbf{y}_t) < \infty$ . As  $\text{Var}(\mathbf{y}_t) \leq E(\mathbf{y}_t^T \mathbf{y}_t)$ , we get  $\lim_{t \rightarrow \infty} \text{Var}(\mathbf{y}_t) < \infty$ . This completes the proof.  $\square$

### Appendix D

The EWMA controller does not consider measurement error and adjustment error, so  $e_{it} = w_{it} = 0$ . From Equations (32) and (33) it is easy to get that

$$a_{it} - a_{i(t-1)} = \omega y_{it}. \quad (\text{A22})$$

Also, from Equations (33) and (A22), we can get Equations (37). Let  $g_{it} = \alpha_i + \beta_{ij} x_{j(t-1)}$ , then Equation (31) becomes

$$y_{it} = g_{it} + n_{it}, \quad (\text{A23})$$

and

$$g_{i(t+1)} = g_{it} + \beta_{ij} \Delta x_{jt}. \quad (\text{A24})$$

In the EWMA controller, the initial bias  $d$  of  $y_t$  from

the target is assumed to be unknown but a fixed constant rather than a random variable. That means that it is equivalent to  $\mu_{id} = d$  and  $\sigma_{id} = 0$  in the ARMA controller's framework. This completes the proof of Theorem 3.

### Biographies

Fangyi He is an Associate Professor in the Department of Financial Engineering at the Southwestern University of Finance and Economics in China. He received B.S. and M.S. degrees in statistics from the University of Science and Technology of China (USTC) and a Ph.D. degree in Systems Engineering from Stevens Institute of Technology. His research interests include statistical process control, engineering process control, financial time series analysis, financial econometrics, and dynamic operation research. He has had articles published in *IEEE Transactions on Semiconductor Manufacturing*, *Probability in the Engineering and Information Sciences*, and *Quality Technology & Quantitative Management*.

Huilian Xie is an Assistant Professor in the Department of Public Health Sciences at the University of Miami in Miami, Florida. She received her B.S. in Finance from the University of Science and Technology of China and received an M.S. and a Ph.D. in statistics from the University of Iowa. Her research interests include variable selection in high-dimensional modeling, semi-parametric regression, latent variable modeling, and production control. Her research is sponsored by the NIH.

Kaibo Wang is an Associate Professor in the Department of Industrial Engineering, Tsinghua University, Beijing, China. He received his B.S. and M.S. degrees in Mechatronics from Xian Jiaotong University, Xian, China, and his Ph.D. in Industrial Engineering and Engineering Management from the Hong Kong University of Science and Technology, Hong Kong. He has published papers in journals such as *Journal of Quality Technology*, *IIE Transactions*, *Quality and Reliability Engineering International*, *International Journal of Production Research*, and others. His research is devoted to statistical quality control and data-driven complex system modeling, monitoring, diagnosis, and control, with a special emphasis on the integration of engineering knowledge and statistical theories for solving problems from industry. He is a member of IIE, ASQ, and INFORMS.