IIE Transactions (2004) **36**, 417–432

ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/07408170490426134

A cautious minimum variance controller with ARIMA disturbances

DANIEL W. APLEY

Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208, USA
E-mail: apley@northwestern.edu

Received June 2001 and accepted May 2003

This article investigates a Cautious Minimum Variance (CMV) control approach for controlling industrial process variability when the model parameters are estimated from data and subject to uncertainty. CMV control has a number of advantages over traditional robust control methods. It incorporates probabilistic, as opposed to deterministic, measures of parameter uncertainty, which are more consistent with the statistical methods typically used to estimate industrial process models. CMV control is also more consistent with the objective of minimizing process variability, since parameter uncertainty is treated simply as an additional source of variation. CMV results have previously been derived for the case where the process disturbance follows a first-order integrated moving average model. This work extends the results to autoregressive moving average and autoregressive integrated moving average disturbances.

1. Introduction

Due to a number of factors, of which Box and Kramer (1992) provides an excellent discussion, Engineering Process Control (EPC) is an increasingly common tool for complementing statistical process control (SPC) in efforts to reduce industrial process variability. Minimum variance control, in which the EPC strategy is to select the input variable adjustments so that the resulting variation in the output variable is minimized, is of particular interest in an industrial SPC context (Åström and Wittenmark, 1990; Box and Luceño, 1997).

When an accurate process model is available, minimum variance control can be quite effective at reducing variation. Industrial process models always involve some level of uncertainty, however, because they must be estimated from process data, typically using experimental design, regression and time-series methods. When the uncertainty is large, the performance of minimum variance control deteriorates, possibly to the point that EPC actually increases process variability. The subject of this article is the design of an EPC strategy for minimizing variability when there is uncertainty in the process model parameters.

We assume that the process to be controlled is described by the model:

$$Y_t = \beta_0 + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_t + d_t, \tag{1}$$

which is illustrated in Fig. 1. The subscript t is a time index, and Y_t is a scalar process output. $\boldsymbol{X}_t = [X_{1,t}X_{2,t}\ldots X_{p,t}]^{\mathrm{T}}$ is a p-dimensional vector of adjustable process inputs. d_t is a random disturbance that accounts for the variability in the process. The model parameters β_0 and $\boldsymbol{\beta} = [\beta_1\beta_2\ldots\beta_p]^T$ describe the effects of the process inputs on the output. The control objective is to drive the output to a target value T with as low a variability about the target as possible. We assume that the disturbance follows an Autoregressive Integrated Moving Average (ARIMA) model with AR order n, MA order m, and integrator order q, denoted ARIMA(n,q,m). The integrator order q is assumed to be either one or zero. When $q = 0$, the ARIMA model becomes an ARMA model. Using standard time series notation (Box *et al.*, 1994) with the backward shift operator B defined such that $Bd_t = d_{t-1}$, the disturbance model can be written as;

$$d_t = \frac{\Theta(B)}{(1-B)^q\Phi(B)}a_t, \tag{2}$$

where $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 \ldots - \theta_m B^m$, $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 \ldots - \phi_n B^n$, and a_t is an independently, identically distributed (i.i.d.), zero-mean sequence of random shocks with variance σ_a^2. We assume that $\Theta(B)$ is invertible and $\Phi(B)$ is stable.

Since the model of Equation (1) involves no input/output "dynamics" or "memory" (i.e., the output is a function of only the current input values, and not past input values), it would not be suitable for representing continuous flow processes with short sampling intervals. The model of Equation (1) is widely used for discrete-parts processes and run-to-run batch production processes, however, although the focus is often on first-order ARMA or IMA disturbances and the single-input scenario. Vander Wiel *et al.* (1992) and Vander Wiel (1996) consider a single-input

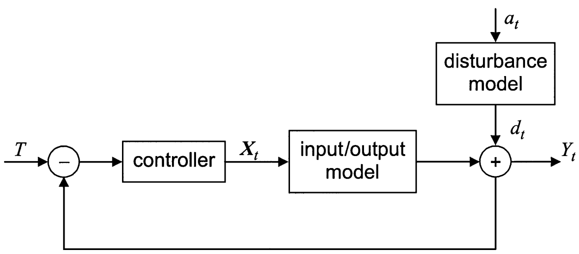

Fig. 1. Block diagram of assumed process model.

version of Equation (1) with first-order ARMA disturbance for representing batch polymerization and discrete-parts machining processes. Box and Kramer (1992), Box and Luceño (1997) and Luceño (1998) focus largely on the single-input version of Equation (1) with first-order IMA disturbance. MacGregor (1988) discusses the suitability of Equation (1) with ARMA or ARIMA disturbance for either continuous-flow processes with long sampling intervals or discrete-parts processes. Montgomery *et al.* (1994), Tsung *et al.* (1998) and Tsung and Apley (2002) all focus on the single-input version of Equation (1) with first-order AR, ARMA and ARIMA disturbances. Janakiram and Keats (1998) investigate the use of Equation (1) with first-order IMA disturbance in controlling a powder loading operation for an automobile air-bag initiator, a hybrid operation involving both continuous and discrete-parts production.

The multiple-input version of Equation (1) is particularly common in semiconductor manufacturing. Ingolfsson and Sachs (1993), Sachs *et al.* (1995), Hamby *et al.* (1998) and many of the references therein all design semiconductor manufacturing process controllers around the model of Equation (1) with multiple inputs. Del Castillo and Hurwitz (1997) consider the model of Equation (1) with only a single input. These works have all focused on the popular EWMA controller, which is of the form:

$$\boldsymbol{X}_t = \boldsymbol{X}_{t-1} - \frac{\boldsymbol{\beta}}{\boldsymbol{\beta}^T\boldsymbol{\beta}}\omega(Y_{t-1} - T), \tag{3}$$

where $0 \leq \omega \leq 1$ is a tuning parameter. Ingolfsson and Sachs (1993), Sachs *et al.* (1995) and Del Castillo and Hurwitz (1997) give an equivalent form of the EWMA controller that involves the EWMA prediction of the disturbance (plus the constant offset β_0), defined recursively via $\hat{d}_t = \omega(Y_t - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_t) + (1-\omega)\hat{d}_{t-1}$. The EWMA control strategy is to select the control inputs at time t such that $\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_t + \hat{d}_{t-1} = T$. It can be verified that the control law of Equation (3) accomplishes this, and for multiple inputs it is the solution that minimizes the norm of the input adjustment vector $\boldsymbol{X}_t - \boldsymbol{X}_{t-1}$ (Ingolfsson and Sachs, 1993). While these works did not assume a specific disturbance model except for simulation and/or analysis purposes, the EWMA controller can be derived as the minimum variance controller for the model of Equation (1) if the disturbance is a first-order IMA process with $\theta_1 = 1 - \omega$ (Sachs *et al.*, 1995).

This article focuses on minimizing variability about the target when there is uncertainty in the model parameter vector $\boldsymbol{\gamma} = [\boldsymbol{\beta}^{\mathrm{T}}, \beta_0, \boldsymbol{\theta}^{\mathrm{T}}, \boldsymbol{\phi}^{\mathrm{T}}]^{\mathrm{T}}$, where $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots \theta_m]^{\mathrm{T}}$ and $\boldsymbol{\phi} = [\phi_1, \phi_2, \ldots, \phi_n]^{\mathrm{T}}$. We assume the model structure of Equation (1) holds and an estimate of $\boldsymbol{\gamma}$, denoted $\hat{\boldsymbol{\gamma}}$, has been obtained from a set of off-line data. One may consider using one of the many traditional robust control methods described in (for example) Francis (1988), Morari and Zafiriou (1989), Green and Limebeer (1995) and Zhou *et al.* (1996) to account for the uncertainty in the parameters. As discussed in Hamby *et al.* (1998) and Apley and Kim (2004) traditional robust control methods are not ideally suited for controlling variability in many industrial processes for a number of reasons. One reason is that they incorporate deterministic uncertainty measures, in which the parameters are assumed to lie within some specified bounds with absolute certainty. *Deterministic* measures of parameter uncertainty are inconsistent with how manufacturing process models are typically estimated—from random process data collected over a period of production. In this situation, a *probabilistic* measure of parameter uncertainty is generally available in the form of a posterior (given the off-line data from which the parameter estimates are obtained) probability distribution for $\boldsymbol{\gamma}$.

Noting the shortcomings of the traditional robust control paradigm, Hamby *et al.* (1998) proposed using probabilistic uncertainty measures and considering the probability of closed-loop stability as the robustness criterion. Attention was restricted to the EWMA control law of Equation (3), in which case the robust control design task reduces to selecting the parameter ω that results in the greatest probability of closed-loop stability. Apley and Kim (2004) took a different approach and incorporated robustness considerations directly into the performance criterion. Specifically, they chose their control law to minimize output variability about the target, where parameter uncertainty was treated simply as an additional source of variability. The method requires a posterior mean (i.e., $\hat{\boldsymbol{\gamma}}$) and covariance matrix for $\boldsymbol{\gamma}$, which can be obtained using standard results from Bayesian estimation theory. The approach was referred to as cautious minimum variance (CMV) control, since it is a nonadaptive version of cautious adaptive control (Åström and Wittenmark, 1995).

The CMV control law derived in Apley and Kim (2003) applied to the specific case that the disturbance in Equation (1) follows a first-order IMA model. This article extends the results to ARIMA disturbances and also investigates the closed-loop performance and stability of the CMV control law. The format of the remainder of the article is as follows. We derive the CMV control laws for ARMA(*n*,*m*) and ARIMA(*n*,1,*m*) disturbances in Sections 2 and 3, respectively. Sections 4 and 5 present a simplified version of the control law and its stability/performance analysis for a common situation that occurs when the process parameters are estimated in the open-loop. Section 6 discusses issues such as experimental design considerations and the possibility of using adaptive control and illustrates with an example.

2. CMV control with ARMA (*n*,*m*) disturbances

Suppose we have designed and conducted an off-line experiment for the purpose of estimating γ. Denote the output observations of the experiment by $\mathbf{Y} = [Y_1\, Y_2 \ldots Y_N]^{\mathrm{T}}$, where N is the sample size. We adopt a Bayesian approach and assume that γ is random with some specified prior distribution $\pi(\gamma)$. $\pi(\gamma)$ may be informative or noninformative (i.e., flat). Let $\pi(\gamma \mid \mathbf{Y})$ denote the posterior distribution of γ, given the off-line data $\mathbf{Y}$, and let $\hat{\gamma}$ and $\boldsymbol{\Sigma}_\gamma$ denote the posterior mean and covariance matrix of γ. Appendix A describes an approach for calculating $\hat{\gamma}$ and $\boldsymbol{\Sigma}_\gamma$.

The objective is to design a control law that will be effective at reducing variability about the target, given the uncertainty (as measured by $\boldsymbol{\Sigma}_\gamma$) in γ. Similar to what was considered in Apley and Kim (2004), suppose the control inputs at time t are to be selected to minimize the CMV loss function:

$$J_t \equiv E_{a_t,\gamma}[(Y_t - T)^2 \mid \mathbf{Y}^{t-1}, \mathbf{Y}]. \tag{4}$$

The subscripts on the expectation operator E indicate the expectation is with respect to the distribution of the random shock a_t and the (posterior) distribution of γ. The expectation is conditioned on both the *off-line* data $\mathbf{Y}$ and the *on-line* data $\mathbf{Y}^{t-1} \equiv \{Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots\}$. The loss function of Equation (4) can be viewed as the conditional, one-step-ahead variance of Y_t about its target value, where the uncertainty in γ is treated simply as an additional source of variability.

If γ were a known, constant vector of parameters, the loss function of Equation (4) would reduce to $E_{a_t}[(Y_t - T)^2 \mid \mathbf{Y}^{t-1}]$. While this conditional variance is not equivalent to the unconditional variance $E[(Y_t - T)^2]$, the control laws that result from their minimizations are identical when the input/output model is static and $\Theta(B)$ is invertible (Åström and Wittenmark, 1990). This is referred to as Minimum Variance (MV) control, an approach that has been widely investigated for controlling industrial process variability (Åström and Wittenmark, 1990; Box and Kramer, 1992; Box and Luceño, 1997). Even when γ is unknown as in the context of this work, one could consider neglecting the uncertainty and assuming $\gamma = \hat{\gamma}$. In this case $E_{a_t}[(Y_t - T)^2 \mid \mathrm{Y}^{t-1}]$ would be minimized as if γ were known, and then γ would be replaced by $\hat{\gamma}$ in the resulting control law. The effectiveness of this approach, which is referred to as certainty equivalence MV control (Åström and Wittenmark, 1995; Mosca, 1995), depends largely on the level of parameter uncertainty.

A loss function such as that of Equation (4), in which the expectation is also with respect to the posterior distribution of the parameters, results in what is referred to as cautious control in the adaptive control literature (Åström and Wittenmark, 1995). Cautious control is usually considered in an adaptive control context, in which the posterior mean and covariance of the parameters are continuously updated on-line as each new observation is taken. A strict use of Equation (4) would, in fact, result in an adaptive control law. The reason is that γ is not independent of the on-line data $\mathbf{Y}^{t-1}$, since the output at any time depends on γ via the model of Equation (1). Consequently, the conditioning of the expectation in Equation (4) on $\mathbf{Y}^{t-1}$ would have the effect of adaptively updating the posterior distribution of γ as each new observation is obtained. However, one may wish to avoid a control law that is adaptive. Adaptive controllers are not only more complex to implement, they often behave erratically and are subject to covariance wind up and bursting (Anderson, 1985; Bodson, 1993; Radenkovic and Michel, 1993; Åström and Wittenmark, 1995). The potential pitfalls of adaptive control, as well as its potential advantages, will be illustrated in Section 6.3 with an adaptive version of CMV control. Aside from this, the remainder of this article focuses on nonadaptive CMV control.

In order to avoid an adaptive control law, we impose the assumption that the posterior distribution of γ given $\mathbf{Y}$ and $\mathbf{Y}^{t-1}$ is equivalent to the posterior distribution given only $\mathbf{Y}$. Strictly speaking, this assumption is not valid, for the reasons discussed above. It should be viewed as an artificially enforced assumption whose purpose is only to ensure that no on-line data will be used to update the posterior distribution of γ and result in an adaptive control law. This is discussed in more detail in Apley and Kim (2004).

In order to find the control adjustments that minimize Equation (4), rewrite Equations (1) and (2) with $q = 0$ as

$$\begin{aligned} Y_t &= \beta_0 + \boldsymbol{\beta}^T \boldsymbol{X}_t + \frac{\Theta(B)}{\Phi(B)} a_t \\ &= \beta_0 + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_t + \frac{H(B)}{\Phi(B)} a_{t-1} + a_t, \end{aligned} \tag{5}$$

where $H(B) \equiv B^{-1}[\Theta(B) - \Phi(B)] = (\phi_1 - \theta_1) + (\phi_2 - \theta_2)B + (\phi_3 - \theta_3)B^2 \ldots + (\phi_s - \theta_s)B^{s-1}$. Here, s denotes the larger of n and m, and the convention is to define $\phi_i \equiv \theta_j \equiv 0$ if $i > n$ or $j > m$. The purpose of writing the model as Equation (5) is to separate d_t into one component a_t and a second component $\Phi^{-1}(B)H(B)a_{t-1}$ that depends on only the past random shocks. Combining Equations (1) and (2), a_{t-1} can be expressed as a function of the unknown γ and the past data via:

$$a_{t-1}(\gamma) = \frac{\Phi(B)}{\Theta(B)}[Y_{t-1} - \beta_0 - \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{t-1}]. \tag{6}$$

Substituting this into Equation (5) gives:

$$Y_t = \beta_0 + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_t + \frac{H(B)}{\Theta(B)}[Y_{t-1} - \beta_0 - \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{t-1}] + a_t, \tag{7}$$

Define the parameter error vector $\tilde{\gamma} \equiv \gamma - \hat{\gamma}$ ($\tilde{\boldsymbol{\beta}}$, $\tilde{\theta}_i$, and $\tilde{\phi}_i$ are similarly defined). If we substitute Equation (7) into Equation (4) with γ replaced by $\hat{\gamma} + \tilde{\gamma}$, the result is an intractable function of the parameter errors $\tilde{\gamma}$, the conditional expectation of which would be impossible to evaluate. To overcome this difficulty, we propose to use a first-order Taylor approximation of Equation (7) about $\gamma = \hat{\gamma}$. In Appendix B we derive the Taylor approximation and show that when substituted into Equation (4) it yields:

$$\begin{aligned} J_t \cong & \left\{ \frac{\hat{\Phi}(B)}{\hat{\Theta}(B)}(\hat{\beta}_0 - T) + \hat{\boldsymbol{\beta}}^T \boldsymbol{u}_t + \frac{\hat{H}(B)}{\hat{\Theta}(B)}(Y_{t-1} - T) \right\}^2 \\ & + \boldsymbol{u}_t^T \boldsymbol{\Sigma}_\beta \boldsymbol{u}_t + \boldsymbol{v}_t^T \boldsymbol{\Sigma}_\alpha \boldsymbol{v}_t + \hat{g}^2 \boldsymbol{\Sigma}_{\beta_0} + 2\boldsymbol{u}_t^T \boldsymbol{\Sigma}_{\beta\alpha} \boldsymbol{v}_t \\ & + 2\boldsymbol{u}_t^T \boldsymbol{\Sigma}_{\beta\beta_0} \hat{g} + 2\hat{g} \boldsymbol{\Sigma}_{\beta_0\alpha} \boldsymbol{v}_t + \sigma_a^2, \end{aligned} \tag{8}$$

where the notation is as follows. Σ_γ has been partitioned as in Equation (19) below, where $\boldsymbol{\alpha} \equiv [\boldsymbol{\theta}^T \boldsymbol{\phi}^T]^T$ is the vector of ARMA parameters. The p-length vector:

$$\boldsymbol{u}_t \equiv \frac{\hat{\Phi}(B)}{\hat{\Theta}(B)} \boldsymbol{X}_t, \tag{9}$$

is a filtered version of the input vector. The scalar:

$$\hat{g} \equiv \left. \frac{\hat{\Phi}(B)}{\hat{\Theta}(B)} \right|_{B=1} = \frac{1 - \hat{\phi}_1 - \hat{\phi}_2 \ldots - \hat{\phi}_n}{1 - \hat{\theta}_1 - \hat{\theta}_2 \ldots - \hat{\theta}_m}.$$

The $(n + m)$-length vector:

$$\begin{aligned} \boldsymbol{v}_t \equiv [& -\hat{a}_{\Theta,t-1}, -\hat{a}_{\Theta,t-2}, \ldots, -\hat{a}_{\Theta,t-m}, \\ & \hat{a}_{\Phi,t-1}, \hat{a}_{\Phi,t-2}, \ldots, \hat{a}_{\Phi,t-n}]^T, \end{aligned} \tag{10}$$

where

$$\hat{a}_{\Phi,t} \equiv \frac{1}{\hat{\Phi}(B)} \hat{a}_t, \tag{11}$$

and

$$a_{\Theta,t} \equiv \frac{1}{\hat{\Theta}(B)} \hat{a}_t, \tag{12}$$

are filtered versions of the residual error:

$$\hat{a}_t \equiv a_t(\hat{\gamma}) = \frac{\hat{\Phi}(B)}{\hat{\Theta}(B)} [Y_t - \hat{\beta}_0 - \hat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{X}_t]. \tag{13}$$

The first term in Equation (8) is the square of the deviation between the target and the one-step ahead prediction of Y_t using the estimated parameters (see Appendix B). The last term is the loss due to the inherent randomness of the disturbance. The remaining terms are the loss due to parameter uncertainty. When there is no uncertainty ($\boldsymbol{\Sigma}_\gamma$ is zero), these terms are zero. When $\boldsymbol{\Sigma}_\gamma$ is large, however, the loss due to parameter uncertainty may be large if the control inputs $\boldsymbol{X}_t$ are not chosen properly. A similar expansion of the Certainty Equivalence Minimum Variance (CEMV) loss function (Equation (4) with γ assumed equal to $\hat{\gamma}$) would be identical, except that the terms associated with parameter uncertainty would be absent.

Appendix B also shows that the CMV control law (the input settings that minimize (8)) is:

$$\begin{aligned} \boldsymbol{X}_t = & -\frac{\boldsymbol{\Sigma}_\beta^{-1} \hat{\boldsymbol{\beta}}}{1 + \hat{\boldsymbol{\beta}}^T \boldsymbol{\Sigma}_\beta^{-1} \hat{\boldsymbol{\beta}}} \left[\frac{\hat{H}(B)}{\hat{\Phi}(B)}(Y_{t-1} - T) + (\hat{\beta}_0 - T) \right] \\ & - [\hat{\boldsymbol{\beta}} \hat{\boldsymbol{\beta}}^{\mathrm{T}} + \boldsymbol{\Sigma}_\beta]^{-1} \left[\boldsymbol{\Sigma}_{\beta\beta_0} + \boldsymbol{\Sigma}_{\beta\alpha} \frac{\hat{\Theta}(B)}{\hat{\Phi}(B)} \boldsymbol{v}_t \right]. \end{aligned} \tag{14}$$

The quantity $\hat{\Phi}^{-1}(B)\hat{\Theta}(B)\boldsymbol{v}_t$ can be calculated recursively using Equations (10) through (13), although it is not needed when $\boldsymbol{\Sigma}_{\beta\alpha}$ is zero, a common situation that will be discussed in Section 4.

CMV control directly incorporates parameter uncertainty information into the control law through the $\boldsymbol{\Sigma}_\beta$, $\boldsymbol{\Sigma}_{\beta\beta_0}$, and $\boldsymbol{\Sigma}_{\beta\alpha}$ terms in Equation (14). Since $\boldsymbol{\Sigma}_\alpha$ and $\boldsymbol{\Sigma}_{\beta_0}$ do not appear, uncertainty in β_0 and the disturbance parameters has only an indirect effect on the CMV control law, via the covariance between $\boldsymbol{\beta}$ and β_0 and $\boldsymbol{\alpha}$. An intuitive explanation is that $\boldsymbol{X}_t$ in Equation (1) interacts with the input parameters $\boldsymbol{\beta}$ but not with β_0 or the disturbance parameters. Consequently, the uncertainty in β_0 and the disturbance parameters cannot be mitigated by choice of $\boldsymbol{X}_t$, other than through their covariance with $\boldsymbol{\beta}$. Furthermore, Section 5 shows that as long as $\Phi(B)$ is a stable polynomial, the stability of the closed-loop system does not depend on the true disturbance parameters.

For comparison, we can obtain the CEMV control law by setting the partial derivative of only the first term in Equation (8) equal to zero. The solution is generally not unique with multiple inputs. An attractive approach in this case is to use the solution that minimizes the norm of $\boldsymbol{X}_t$, similar to what was done in Ingolfsson and Sachs (1993). This results in the CEMV control law:

$$\boldsymbol{X}_t = -\frac{\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}^T \hat{\boldsymbol{\beta}}} \left[\frac{\hat{H}(B)}{\hat{\Phi}(B)}(Y_{t-1} - T) + (\hat{\beta}_0 - T) \right]. \tag{15}$$

We will compare the closed-loop stability and performance of the CMV control law of Equation (14) and the CEMV control law of Equation (15) in Sections 5 and 6.

3. CMV control with ARIMA(*n*,*1*,*m*) disturbances

The results of the previous section can easily be extended to ARIMA(n,1,m) disturbances. With $q = 1$ in the disturbance model of Equation (2), Equation (1) becomes:

$$Y_t = Y_{t-1} + \boldsymbol{\beta}^{\mathrm{T}} \nabla \boldsymbol{X}_t + \frac{\Theta(B)}{\Phi(B)} a_t,$$

which is exactly Equation (5) with β_0 replaced by Y_{t-1}, and the control inputs $\boldsymbol{X}_t$ replaced by the input adjustments $\nabla \boldsymbol{X}_t \equiv (1 - B)\boldsymbol{X}_t = \boldsymbol{X}_t - \boldsymbol{X}_{t-1}$. Consequently, the results for ARMA disturbances in the previous section apply directly to ARIMA disturbances if we replace $\hat{\beta}_0$ by Y_{t-1} and replace $\boldsymbol{X}_t$ by $\nabla \boldsymbol{X}_t$ throughout. Since there is no uncertainty in Y_{t-1} (given $\boldsymbol{Y}^{t-1}$), all covariance terms associated with β_0 are taken to be zero. Specifically, the CMV control law for ARIMA disturbances becomes:

$$\nabla\boldsymbol{X}_t = -\frac{\boldsymbol{\Sigma}_\beta^{-1}\hat{\boldsymbol{\beta}}}{1+\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_\beta^{-1}\hat{\boldsymbol{\beta}}}\left[\frac{\hat{H}(B)}{\hat{\Phi}(B)}(Y_{t-1}-T)+(Y_{t-1}-T)\right] - [\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^T+\boldsymbol{\Sigma}_\beta]^{-1}\boldsymbol{\Sigma}_{\beta\alpha}\frac{\hat{\Theta}(B)}{\hat{\Phi}(B)}\boldsymbol{v}_t, \tag{16}$$

where $\hat{a}_{\Phi,t}$, $\hat{a}_{\Theta,t}$ and $\boldsymbol{v}_t$ are defined in Equations (10) through (12) with $\hat{a}_t \equiv \hat{\Theta}^{-1}(B)\hat{\Phi}(B)[Y_t - Y_{t-1} - \hat{\boldsymbol{\beta}}^T\nabla\boldsymbol{X}_t]$.

Equation (16) expresses the control law in terms of the input adjustments. Since $\nabla\boldsymbol{X}_t = (1-B)\boldsymbol{X}_t$, an equivalent form can be obtained by multiplying both sides of Equation (16) by the integral operator $(1-B)^{-1}$. The inputs $\boldsymbol{X}_t$ can then be expressed as the summation of the right-hand side of Equation (16) from time $t=1$ to the current time. The CMV control law for ARIMA disturbances therefore includes integral control action, which is one notable difference between it and the CMV control law for ARMA disturbances. An intuitive explanation is that since an ARIMA disturbance model contains an integrator, the controller must also contain an integrator in order to compensate for the effects of the disturbance. Integral control action is generally desirable, as it allows the asymptotic cancellation of constant mean shifts (Åström and Wittenmark, 1990). Note also that neither the estimate of β_0 nor its uncertainty appears in the CMV control law of Equation (16). This is reasonable since the constant offset β_0 in Equation (1) means little if the disturbance wanders in an unstable manner, as do ARIMA disturbances.

The corresponding CEMV control law for ARIMA disturbances is:

$$\nabla\boldsymbol{X}_t = -\frac{\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}^T\hat{\boldsymbol{\beta}}}\left[\frac{\hat{H}(B)}{\hat{\Phi}(B)}(Y_{t-1}-T)+(Y_{t-1}-T)\right], \tag{17}$$

which is obtained by replacing $\hat{\beta}_0$ by Y_{t-1} and replacing $\boldsymbol{X}_t$ by $\nabla\boldsymbol{X}_t$ in Equation (15). Note that this is the solution that minimizes the norm of $\nabla\boldsymbol{X}_t$. For the case of a first-order IMA disturbance, Equation (17) reduces to the EWMA controller of Equation (3) with $\omega = 1-\theta_1$.

A straightforward repetition of the derivations in Appendix A verifies that the expressions for the posterior mean and covariance of the parameters are also identical to the ARMA case with $\hat{\beta}_0$ and $\boldsymbol{X}_t$ replaced by Y_{t-1} and $\nabla\boldsymbol{X}_t$ and $\gamma \equiv [\boldsymbol{\beta}^T\ \boldsymbol{\alpha}^T]^T$. Thus, for a noninformative prior:

$$\boldsymbol{\Sigma}_\gamma = \begin{bmatrix} \boldsymbol{\Sigma}_\beta & \boldsymbol{\Sigma}_{\beta\alpha} \\ \boldsymbol{\Sigma}_{\beta\alpha}^T & \boldsymbol{\Sigma}_\alpha \end{bmatrix} = \frac{\sigma_a^2}{N}\boldsymbol{\Sigma}_z^{-1},$$

with $\boldsymbol{\Sigma}_z \equiv N^{-1}\sum_{t=1}^N \boldsymbol{z}_t\boldsymbol{z}_t^T$,

$$\boldsymbol{z}_t \equiv -[\partial a_t(\gamma)/\partial\gamma|_{\gamma=\hat{\gamma}}]^T = \left[\nabla\boldsymbol{u}_t^T\,\boldsymbol{v}_t^T\right]^T, \tag{18}$$

and $\nabla\boldsymbol{u}_t \equiv \boldsymbol{u}_t - \boldsymbol{u}_{t-1} = \hat{\Theta}^{-1}(B)\,\hat{\Phi}(B)\nabla\boldsymbol{X}_t$. The posterior mean $\hat{\gamma}$ minimizes Equation (A2) with $a_t(\gamma) \equiv \boldsymbol{\Theta}^{-1}(B)\boldsymbol{\Phi}(B)[Y_t - Y_{t-1} - \boldsymbol{\beta}^{\mathrm{T}}\nabla\boldsymbol{X}_t]$.

4. A simplified CMV control law for open-loop identification

This section discusses a more convenient form of the CMV control law in the common situation that $\boldsymbol{\Sigma}_{\beta\alpha}$ is zero. This could be the case if the off-line experimental data used to estimate the parameters includes a long period of production over which the inputs are held constant (e.g., a long period of normal production prior to the implementation of EPC). Although the constant-input data cannot be used to improve the accuracy of the input parameter estimates, it may allow the disturbance model parameters to be estimated with sufficient accuracy to assume $\boldsymbol{\Sigma}_{\beta\alpha}$ is zero.

Another situation in which $\boldsymbol{\Sigma}_{\beta\alpha}$ is zero is when the off-line experiment is conducted in the open-loop using predefined input profiles (as opposed to the closed-loop with the inputs allowed to depend on the observed outputs over the experiment). To see the reason why, consider the case of an ARMA disturbance. From Appendix A,

$$\boldsymbol{\Sigma}_\gamma = \begin{bmatrix} \boldsymbol{\Sigma}_\beta & \boldsymbol{\Sigma}_{\beta\beta_0} & \boldsymbol{\Sigma}_{\beta\alpha} \\ \boldsymbol{\Sigma}_{\beta\beta_0}^T & \boldsymbol{\Sigma}_{\beta_0} & \boldsymbol{\Sigma}_{\beta_0\alpha} \\ \boldsymbol{\Sigma}_{\beta\alpha}^T & \boldsymbol{\Sigma}_{\beta_0\alpha}^T & \boldsymbol{\Sigma}_\alpha \end{bmatrix} = \frac{\sigma_a^2}{N}\boldsymbol{\Sigma}_z^{-1} = \frac{\sigma_a^2}{N}\begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{u}} & \boldsymbol{\Sigma}_{\boldsymbol{u}\hat{g}} & \boldsymbol{\Sigma}_{\boldsymbol{uv}} \\ \boldsymbol{\Sigma}_{\boldsymbol{u}\hat{g}}^T & \Sigma_{\hat{g}} & \boldsymbol{\Sigma}_{\hat{g}\boldsymbol{v}} \\ \boldsymbol{\Sigma}_{\boldsymbol{uv}}^T & \boldsymbol{\Sigma}_{\hat{g}\boldsymbol{v}}^T & \boldsymbol{\Sigma}_{\boldsymbol{v}} \end{bmatrix}^{-1}, \tag{19}$$

where the partitioning of $\boldsymbol{\Sigma}_z$ is consistent with the partitioning of $\boldsymbol{z}_t = [\boldsymbol{u}_t^T, \hat{g}, \boldsymbol{v}_t^T]^T$ with $\boldsymbol{u}_t$ and $\boldsymbol{v}_t$ defined in Equations (9) and (10). Since $\hat{g}$ is a constant, $\boldsymbol{\Sigma}_{\boldsymbol{u}\hat{g}} = \bar{\boldsymbol{u}}\hat{g}$ and $\Sigma_{\hat{g}} = \hat{g}^2$, where $\bar{\boldsymbol{u}} = N^{-1}\sum_{t=1}^N \boldsymbol{u}_t$. If the experiment is conducted in the open-loop, $\boldsymbol{X}_t$, and thus $\boldsymbol{u}_t$, will be independent of $\boldsymbol{v}_t$. Since $\boldsymbol{v}_t$ is zero-mean, both $\boldsymbol{\Sigma}_{\boldsymbol{uv}}$ and $\boldsymbol{\Sigma}_{\hat{g}\boldsymbol{v}}$ will be zero. Using this in Equation (19), it can be verified by direct substitution that $\boldsymbol{\Sigma}_{\beta\alpha}$ and $\boldsymbol{\Sigma}_{\beta_0\alpha}$ are zero, $\boldsymbol{\Sigma}_\alpha = N^{-1}\sigma_a^2\Sigma_v^{-1}$, $\boldsymbol{\Sigma}_\beta = N^{-1}\sigma_a^2[\boldsymbol{\Sigma}_{\boldsymbol{u}} - \bar{\boldsymbol{u}}\bar{\boldsymbol{u}}^T]^{-1}$, and $\boldsymbol{\Sigma}_{\beta\beta_0} = -\boldsymbol{\Sigma}_\beta\bar{\boldsymbol{u}}\hat{g}^{-1}$. Note that when the process is identified in the closed-loop, it will not in general be true that $\boldsymbol{\Sigma}_{\boldsymbol{uv}}$ is zero. Hence, $\boldsymbol{\Sigma}_{\beta\alpha}$ will generally differ from zero for closed-loop identification.

Substituting $\boldsymbol{\Sigma}_{\beta\beta_0} = -\boldsymbol{\Sigma}_\beta\bar{\boldsymbol{u}}\hat{g}^{-1}$ into Equation (14) and using Equation (A9), the CMV control law for ARMA disturbances when $\boldsymbol{\Sigma}_{\beta\alpha}$ is zero reduces to:

$$\boldsymbol{X}_t = -\frac{\boldsymbol{\Sigma}_\beta^{-1}\hat{\boldsymbol{\beta}}}{1+\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_\beta^{-1}\hat{\boldsymbol{\beta}}}\left[\frac{\hat{H}(B)}{\hat{\Phi}(B)}(Y_{t-1}-T) + (\hat{\beta}_0 - T) + \frac{\hat{\boldsymbol{\beta}}^T\bar{\boldsymbol{u}}}{\hat{g}}\right] + \frac{\bar{\boldsymbol{u}}}{\hat{g}}. \tag{20}$$

For ARIMA disturbances, the CMV control law can be obtained directly from Equation (16) with $\boldsymbol{\Sigma}_{\beta\alpha}$ equal to zero, which gives:

$$\nabla \boldsymbol{X}_t = -\frac{\boldsymbol{\Sigma}_\beta^{-1}\hat{\boldsymbol{\beta}}}{1+\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_\beta^{-1}\hat{\boldsymbol{\beta}}} \times \left[\frac{\hat{H}(B)}{\hat{\Phi}(B)}(Y_{t-1}-T)+(Y_{t-1}-T)\right]. \quad (21)$$

Equation (21) is identical to the CEMV control law of Equation (17), except that the "controller gain" vector is modified from $(\hat{\boldsymbol{\beta}}^T\hat{\boldsymbol{\beta}})^{-1}\hat{\boldsymbol{\beta}}$ for CEMV control to $(1+\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_\beta^{-1}\hat{\boldsymbol{\beta}})^{-1}\boldsymbol{\Sigma}_\beta^{-1}\hat{\boldsymbol{\beta}}$ for CMV control.

The convenience of Equations (20) and (21) is that the control law no longer depends on $\boldsymbol{v}_t$, which is a rather complicated function of the past observations. As discussed in the following section, the analysis of closed-loop stability and performance is also simplified.

5. Closed-loop performance and stability analysis

When $\boldsymbol{\Sigma}_{\beta\alpha}$ is zero, it is relatively straightforward to analyze the closed-loop behavior of the output for a fixed set of parameters and their estimates. First consider the case of ARMA disturbances. Substituting Equation (20) into Equation (1) and rearranging gives:

$$\left[1+\frac{C_{\text{cmv}}\hat{H}(B)B}{\hat{\Phi}(B)}\right](Y_t-T) = (\beta_0 - C_{\text{cmv}}\hat{\beta}_0)+(\boldsymbol{\beta}-C_{\text{cmv}}\hat{\boldsymbol{\beta}})^T\frac{\bar{\boldsymbol{u}}}{\hat{g}}+(C_{\text{cmv}}-1)T+d_t,$$

where

$$C_{\text{cmv}} \equiv \frac{\boldsymbol{\beta}^T\boldsymbol{\Sigma}_\beta^{-1}\hat{\boldsymbol{\beta}}}{1+\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_\beta^{-1}\hat{\boldsymbol{\beta}}}.$$

Since $B\hat{H}(B)=\hat{\Theta}(B)-\hat{\Phi}(B)$, the steady-state output response for CMV control becomes:

$$Y_t - T = \delta_{\text{cmv}} + \frac{\hat{\Phi}(B)}{\hat{\Phi}(B)+C_{\text{cmv}}(\hat{\Theta}(B)-\hat{\Phi}(B))}d_t, \quad (22)$$

where $d_t = \Phi^{-1}(B)\Theta(B)a_t$, and

$$\begin{aligned}\delta_{\text{cmv}} &\equiv \left.\frac{\hat{\Phi}(B)}{\hat{\Phi}(B)+C_{\text{cmv}}(\hat{\Theta}(B)-\hat{\Phi}(B))}\right|_{B=1} \times \left[(\beta_0-C_{\text{cmv}}\hat{\beta}_0)+(\boldsymbol{\beta}-C_{\text{cmv}}\hat{\boldsymbol{\beta}})^T\frac{\bar{\boldsymbol{u}}}{\hat{g}}+(C_{\text{cmv}}-1)\,T\right],\\ &= \frac{\hat{g}}{\hat{g}+C_{\text{cmv}}(1-\hat{g})} \times \left[(\beta_0-C_{\text{cmv}}\hat{\beta}_0)+(\boldsymbol{\beta}-C_{\text{cmv}}\hat{\boldsymbol{\beta}})^T\frac{\bar{\boldsymbol{u}}}{\hat{g}}+(C_{\text{cmv}}-1)T\right],\end{aligned}$$

is the steady-state mean deviation between the output and the target.

For CEMV control with ARMA disturbances, if we define:

$$C_{\text{cemv}} \equiv \frac{\boldsymbol{\beta}^T\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}^T\hat{\boldsymbol{\beta}}},$$

we can similarly show that the steady-state output response is:

$$Y_t - T = \delta_{\text{cemv}} + \frac{\hat{\Phi}(B)}{\hat{\Phi}(B)+C_{\text{cemv}}(\hat{\Theta}(B)-\hat{\Phi}(B))}d_t, \quad (23)$$

where

$$\delta_{\text{cemv}} \equiv \frac{\hat{g}}{\hat{g}+C_{\text{cemv}}(1-\hat{g})}[(\beta_0 - C_{\text{cemv}}\hat{\beta}_0)+(C_{\text{cemv}}-1)T].$$

When considering a fixed set of parameters and their estimates, Equations (22) and (23) can be used to calculate the steady-state variance, mean deviation from target, and mean square error when using CMV or CEMV control with ARMA disturbances. If we let C denote either C_{cmv} or C_{cemv}, the steady-state variance is just $\sigma_a^2\sum_{j=0}^{\infty}g_j^2$, where:

$$G(B) \equiv \sum_{j=0}^{\infty} g_j B^j \equiv \frac{\hat{\Phi}(B)}{\hat{\Phi}(B)+C\left(\hat{\Theta}(B)-\hat{\Phi}(B)\right)}\frac{\Theta(B)}{\Phi(B)}, \quad (24)$$

is the impulse response of the transfer function relating a_t to $Y_t - T$. Box *et al.* (1994) discuss in detail this approach to calculating the variance of an ARMA process.

Equations (22) and (23) cannot be used to verify the performance of the controllers in practice, since we would not know the values of the true parameters. They do provide some insight into performance, however, and can be used to calculate the probability of stability when we consider variations in $\boldsymbol{\gamma}$ over its posterior distribution. From Equations (22) and (23), for a fixed $\hat{\boldsymbol{\gamma}}$, closed-loop stability depends entirely on C, which depends only on the input parameters $\boldsymbol{\beta}$. The closed-loop system is clearly stable for $C=1$, since the ARMA disturbance model and its estimate are assumed stable and invertible. If C deviates too far from unity, however, the closed-loop system will become unstable. The set of values of C that result in a stable system, denoted Ω_c, can be found by analyzing the closed-loop denominator polynomial $\hat{\Phi}(B)+C[\hat{\Theta}(B)-\hat{\Phi}(B)]$ in Equations (22) and (23). The roots of the characteristic equation are defined as the $s \equiv \max\{n, m\}$ values of λ such that:

$$\hat{\Phi}(\lambda^{-1})+C[\hat{\Theta}(\lambda^{-1})-\hat{\Phi}(\lambda^{-1})] \equiv 1-[\hat{\phi}_1 - C(\hat{\phi}_1-\hat{\theta}_1)]\lambda^{-1} - \cdots - [\hat{\phi}_s - C(\hat{\phi}_s-\hat{\theta}_s)]\lambda^{-s} = 0. \quad (25)$$

The closed-loop system is stable if and only if all roots have a magnitude strictly less than unity or, equivalently, all roots fall within the unit circle in the complex plane (Åström and Wittenmark, 1990). Hence, Ω_c can be determined by numerically calculating the roots for a range of values of C.

Root locus plots, which are plots of the loci of roots in the complex plane as C varies (Ogata, 1995), are useful in this regard.

The (posterior) probability that the closed-loop system is stable is $Pr[C \in \Omega_c]$. After determining Ω_c, we can calculate this by noting that since C_{cmv} and C_{cemv} are linear contrasts of $\boldsymbol{\beta}$, their posterior distributions are $C_{\text{cmv}} \sim \text{N}(\mu_{\text{cmv}}, \sigma^2_{\text{cmv}})$ and $C_{\text{cemv}} \sim \text{N}(\mu_{\text{cemv}}, \sigma^2_{\text{cemv}})$, with

$$\mu_{\text{cmv}} = \frac{\hat{\boldsymbol{\beta}}^T \boldsymbol{\Sigma}_{\beta}^{-1} \hat{\boldsymbol{\beta}}}{1 + \hat{\boldsymbol{\beta}}^T \boldsymbol{\Sigma}_{\beta}^{-1} \hat{\boldsymbol{\beta}}}, \tag{26}$$

$$\sigma_{\text{cmv}} = \frac{(\hat{\boldsymbol{\beta}}^T \boldsymbol{\Sigma}_{\beta}^{-1} \hat{\boldsymbol{\beta}})^{1/2}}{1 + \hat{\boldsymbol{\beta}}^T \boldsymbol{\Sigma}_{\beta}^{-1} \hat{\boldsymbol{\beta}}}, \tag{27}$$

$$\mu_{\text{cemv}} = 1, \tag{28}$$

and

$$\sigma_{\text{cemv}} = \frac{(\hat{\boldsymbol{\beta}}^T \boldsymbol{\Sigma}_{\beta} \hat{\boldsymbol{\beta}})^{1/2}}{\boldsymbol{\beta}^T \hat{\boldsymbol{\beta}}}. \tag{29}$$

To illustrate the use of root locus plots for determining Ω_c, suppose the estimated disturbance model is AR(3) with $\hat{\Phi}(B) = (1 - 0.9B)^3 = 1 - 2.7B + 2.43B^2 - 0.729B^3$. Disturbances of this type can be viewed as a series of cascaded AR(1) processes and are discussed in more detail in English and Case (1990). With $\hat{\Theta}(B) = 1$, the characteristic equation becomes $1 - (1 - C)2.7\lambda^{-1} + (1 - C)2.43\lambda^{-2} - (1 - C)0.729\lambda^{-3} = 0$. Figure 2 is a root locus plot showing the roots of the characteristic equation as C varies, and

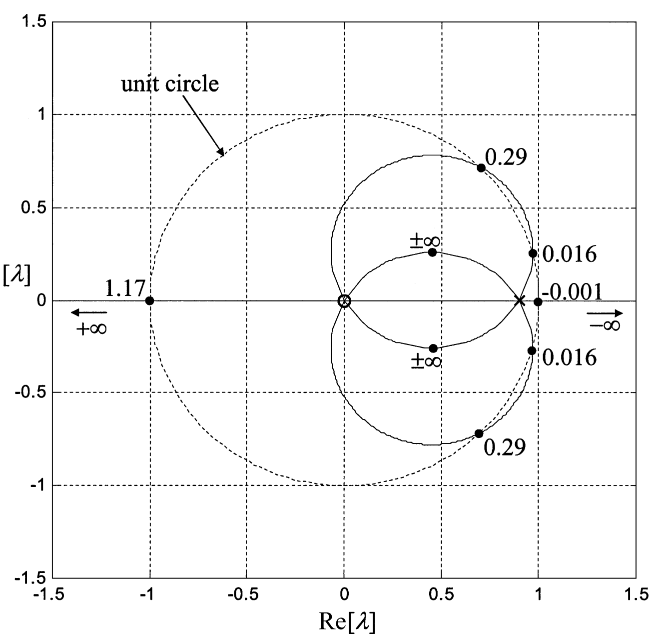

Fig. 2. Root locus plot for the AR(3) disturbance example. The numbers are the values of C corresponding to the indicated roots.

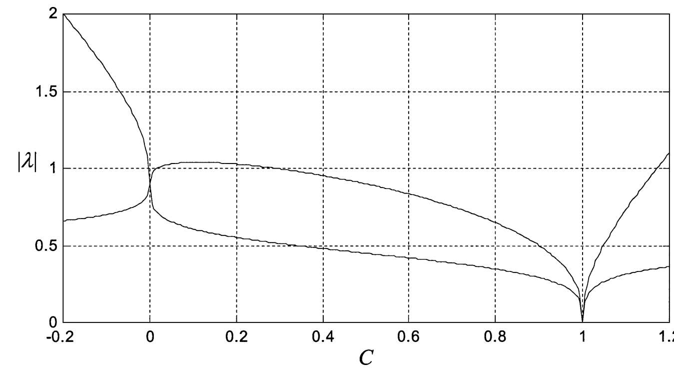

Fig. 3. Plot of the magnitude of the roots as a function of C for the AR(3) disturbance example.

Fig. 3 shows the magnitude of the roots as a function of C. The numbers in Fig. 2 are the values of C at which the roots leave the unit circle and the system becomes unstable. The $\times$ symbols indicate the roots for $C = 0$, which from Equation (25) are the roots of $\hat{\Phi}(\lambda^{-1}) = 0$ (a triple root at $\lambda = 0.9$). The O symbols indicate the roots for $C = 1$, which from Equation (25) are the roots of $\hat{\Theta}(\lambda^{-1}) = 0$ (a root at $\lambda = 0$). One root locus lies entirely on the real line and the two circular root loci correspond to the pair of complex-conjugate roots. Note that only two curves appear in Fig. 3, since the complex conjugate pair of roots have the same magnitude. From Figs. 2 and 3, the root locus that falls on the real line is inside the unit circle for $-0.001 < C < 1.17$, and the complex conjugate loci fall inside the unit circle for $C < 0.016$ and $C > 0.29$. Hence, $\Omega_c = (-0.001, 0.016) \cup (0.29, 1.17)$, although the interval $(-0.001, 0.016)$ is so small that it would contribute little to the probability of stability. An interval of stability around $C = 0$ is characteristic of any ARMA disturbance, since the roots of $\hat{\Phi}(\lambda^{-1}) = 0$ fall within the unit circle by assumption. Likewise, there will always be an interval of stability around $C = 1$. This may be observed in Fig. 3. Given $\hat{\beta}$, $\boldsymbol{\Sigma}_{\beta}$, and Ω_c, one could then use Equations (26) through (29) to compare the probability of stability of CMV and CEMV control. A numerical example is provided in Section 6.2.

As another example, suppose the disturbance is ARMA(1,1), in which case the analysis becomes simpler. The only root of the characteristic equation $1 - [\hat{\phi}_1 - C(\hat{\phi}_1 - \hat{\theta}_1)]\lambda^{-1} = 0$ is $\lambda = \hat{\phi}_1 - C(\hat{\phi}_1 - \hat{\theta}_1)$. For $\hat{\phi}_1 > \hat{\theta}_1$, which will be the case with positively autocorrelated disturbances, the stability region for C is the interval:

$$\Omega_c = \left(\frac{\hat{\phi}_1 - 1}{\hat{\phi}_1 - \hat{\theta}_1}, \frac{\hat{\phi}_1 + 1}{\hat{\phi}_1 - \hat{\theta}_1} \right). \tag{30}$$

Appendix C shows that CMV control always results in a higher probability of stability than CEMV control for ARMA(1,1) disturbances. It is not clear, however, whether this holds for higher-order ARMA disturbances. The probability of stability and closed-loop variance is further discussed in Section 6.2 with regards to determining if an experiment has resulted in a “sufficiently accurate” estimate of the model.

The closed-loop output responses of Equations (22) and (23) apply to ARMA disturbances. Using the CMV and CEMV control laws of Equations (21) and (17) for ARIMA(n,1,m) disturbances, a repetition of the above derivations reveals that the steady-state output response is:

$$Y_t - T = \frac{(1-B)\hat{\Phi}(B)}{(1-C)(1-B)\hat{\Phi}(B) + C\hat{\Theta}(B)} d_t, \qquad (31)$$

where $d_t = (1-B)^{-1}\Phi^{-1}(B)\Theta(B)a_t$. As before, C denotes either C_{cmv} or C_{cemv}, depending on whether CMV or CEMV control is used. Note that the steady-state mean of Y_t is equal to the target value, the reason for which is that Equations (21) and (17) contain integral action. For ARIMA disturbances, stability analyses can be conducted based on the roots of the denominator polynomial in Equation (31).

6. Example and discussion

This section provides an example illustrating some of the performance differences between CMV and CEMV control. We also use the example to illustrate how to evaluate whether additional experimentation is necessary and to explore issues involved in implementing an adaptive version of CMV.

6.1. *An example comparing CMV and CEMV control*

In the Bayesian scenario in which the CMV law was derived, γ is a random vector with some prior distribution $\pi(\gamma)$, the estimate $\hat{\gamma}$ becomes fixed given the off-line data, and variations in γ over its posterior distribution are considered. Although the *derivation* of the CMV law is purely from this Bayesian perspective, a *performance analysis* from a frequentist perspective may be of more interest. In a frequentist scenario, a single fixed (but unknown) value for γ is assumed, and variations in the estimate $\hat{\gamma}$ are considered. For a fixed γ, the distribution of $\hat{\gamma}$ depends on the distribution of the off-line data $\mathbf{Y}$. Each hypothetical value for $\hat{\gamma}$ will result in a different CMV control law and different closed-loop performance. The following example uses Monte Carlo simulation to compare the performance of CMV and CEMV control in this frequentist scenario.

Suppose we have a process with three inputs that follows the model:

$$Y_t = X_{1,t} + X_{2,t} + X_{3,t} + \frac{1-0.3B}{1-0.95B} a_t,$$

with $\sigma_a^2 = 1.0$ and a target value $T = 0$. In this case the model parameters are $\beta_0 = 0$, $\boldsymbol{\beta} = [1, 1, 1]^T$, $\theta_1 = 0.3$ and $\phi_1 = 0.95$. Also suppose that in the experiment used to estimate the parameters: (i) $\boldsymbol{X}_t$ is varied over 64 timesteps according to the sinusoidal profiles shown in Fig. 4; and (ii) an additional 200 observations are collected during which time all three inputs are held fixed at zero. The 200 additional observations are included to improve the estimates of β_0 and the ARMA parameters, although they will obviously not improve the estimate of $\boldsymbol{\beta}$. This is a reasonable strategy in practice, since additional observations with the inputs held constant may come at little additional cost. For example, there may be a long period of production prior to the implementation of feedback control, during which time the inputs are not varied.

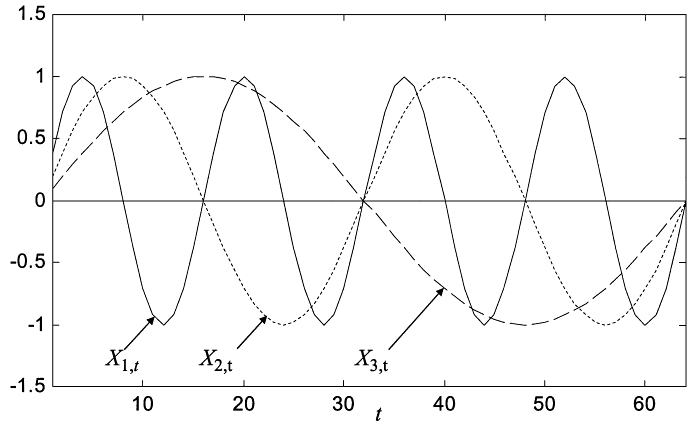

Fig. 4. Input profiles used to estimate γ.

As we will see in Section 6.2, this experimental design results in β_3 having much larger uncertainty than β_1. In this case we would intuitively prefer that a control law make heavier use of $X_{1,t}$ and lighter use of $X_{3,t}$. Since the premultiplier $\boldsymbol{\Sigma}_{\beta}^{-1}$ in the CMV control law of Equation (20) does exactly this, we might expect CMV to perform better than CEMV in this example. The following Monte Carlo simulation was used to compare their performances. For each Monte Carlo replicate, $N = 264$ a_ts were generated as random numbers drawn from the standard normal distribution, and d_t was calculated from Equation (2) with $q = 0$, $\Theta(B) = 1 - 0.3B$, and $\Phi(B) = 1 - 0.95B$. The Y_ts were generated from Equation (1) with $\boldsymbol{X}_t$ equal to zero for the first 200 observations and $\boldsymbol{X}_t$ given by the profiles in Figure 4 for the remaining 64 observations. $\hat{\gamma}$ and $\boldsymbol{\Sigma}_{\gamma}$ were then calculated as described in Appendix A and Section 4 (since the process was identified in the open-loop). Equations (22) and (23) were then used to calculate the steady-state mean-square error:

$$J \equiv E[(Y_t - T)^2 \mid \gamma, \hat{\gamma}],$$

for CMV and CEMV control. The entire procedure was repeated for 10,000 Monte Carlo replicates.

The empirical cumulative distribution function (cdf) of J for both CMV and CEMV control is shown in Fig. 5. As a point of reference, the mean square error J that would result if minimum variance control were used with perfect knowledge of the parameters is $\sigma_a^2 = 1.0$. While the 50th

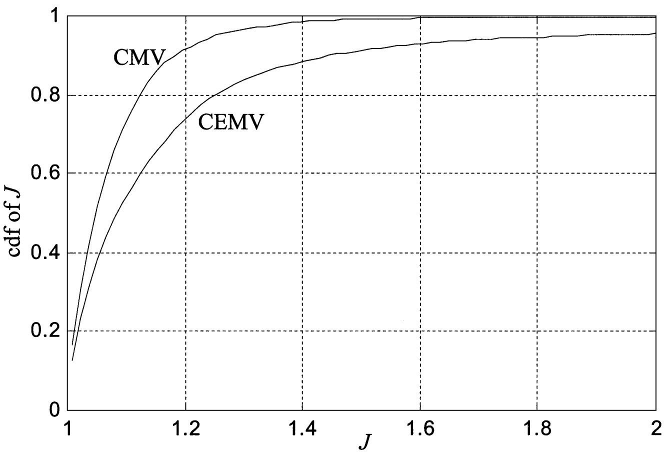

Fig. 5. Cdfs of the mean-square error J for CMV and CEMV control.

percentiles for CMV and CEMV are comparable and only slightly larger than σ_a^2, the higher percentiles are substantially larger for CEMV. For example, the 90th percentile using CEMV control is roughly 1.5. Hence, there is a 10% chance that CEMV control inflates the mean-square error (relative to the $J = 1.0$ value that results with perfect knowledge of the parameters) by more than 50%. In comparison, there is roughly a 1% chance that CMV control inflates the mean square error by more than 50%. Similarly, CEMV and CMV control will inflate the mean-square error by more than 20% with probability 0.26 and 0.08, respectively. Although not apparent from Fig. 5, CEMV and CMV control were unstable with probability 0.0134 and 0.0006, respectively.

Since the derivation of the CMV control law was based on the first-order Taylor approximation of Y_t in Equation (A8), we may be interested in how close the approximation is to the actual Y_t. It is difficult to generalize the accuracy of the Taylor approximation, since it will depend on the specific values for the true and estimated parameters, as well as the past and current values of the control inputs and output. The following provides some insight into the accuracy, however, for the specific example considered above. After estimating the parameters in each replicate of the preceding Monte Carlo simulation, the on-line system was simulated for 2000 timesteps in the open loop with the control inputs $\boldsymbol{X}_t$ generated as an i.i.d. sample of multivariate normal random variables with zero mean and covariance $\mathbf{I}$, the identity matrix. For each timestep, the actual Y_t was compared to the Taylor approximation from Equation (A8). Denoting their difference by e_t, the root mean square $s_e \equiv (2000^{-1} \sum_{t=1}^{2000} e_t^2)^{1/2}$ was calculated for each replicate of the Monte Carlo simulation. Over the 10,000 Monte Carlo replicates, s_e was approximately normally distributed with mean 0.33 and standard deviation 0.10, which should be considered relative to the standard deviation of Y_t. From Equation (1), the variance of Y_t is $\mathrm{Var}(Y_t) = \sigma_x^2 \boldsymbol{\beta}^T \boldsymbol{\beta} + \mathrm{Var}(d_t)$ when $\boldsymbol{X}_t$ is generated randomly in the open-loop with covariance matrix $\sigma_x^2 \mathbf{I}$. In this case, the standard deviation of Y_t was 2.89, so that the average root-mean-square of the Taylor approximation error was roughly 10% of the standard deviation of Y_t. Repeating the Monte Carlo simulation with $\sigma_x^2 = 9$, the average value of s_e was 0.41, whereas the standard deviation of Y_t was 5.69.

6.2. *Evaluating the experimental design*

Designing an effective experiment for estimating the parameters of Equation (1) is generally more complicated than for standard linear regression problems. Consider the case of open-loop identification with ARMA disturbances discussed in Section 4, and recall that $\boldsymbol{\Sigma}_\beta = N^{-1}\sigma_a^2[\boldsymbol{\Sigma}_{\boldsymbol{u}} - \bar{\boldsymbol{u}}\bar{\boldsymbol{u}}^T]^{-1}$. To minimize the uncertainty in $\boldsymbol{\beta}$, we would ideally design the experiment so that $\boldsymbol{\Sigma}_{\boldsymbol{u}} \equiv N^{-1} \sum_{t=1}^{N} \boldsymbol{u}_t \boldsymbol{u}_t^T$ is a well-conditioned matrix. This cannot be guaranteed without some *a priori* knowledge of the ARMA parameters, however, since $\boldsymbol{u}_t = \hat{\Theta}^{-1}(B)\hat{\Phi}(B)\boldsymbol{X}_t$ will depend on the ARMA parameter estimates.

To illustrate, consider the experiment in Section 6.1. Over the entire $N = 264$ observations, the three input profiles are orthogonal and their average squared magnitudes are equal. If the disturbance were i.i.d. and one were using pure linear regression to estimate $\boldsymbol{\beta}$, the resulting $\boldsymbol{\Sigma}_\beta$ would be diagonal with equal diagonal elements. Hence, the uncertainty in each of the input parameters would be identical. This is not the case when the disturbance is ARMA. If the disturbance parameter estimates happen to coincide with their true values (i.e., $\hat{\sigma}_a^2 = 1.0$, $\hat{\theta}_1 = 0.3$, and $\hat{\phi}_1 = 0.95$), for example, the input profiles in Fig. 4 result in:

$$\boldsymbol{\Sigma}_\beta = N^{-1}\sigma_a^2[\boldsymbol{\Sigma}_{\boldsymbol{u}} - \bar{\boldsymbol{u}}\bar{\boldsymbol{u}}^T]^{-1} = \begin{bmatrix} 0.116 & 0.005 & 0.008 \\ 0.005 & 0.411 & 0.015 \\ 0.008 & 0.015 & 1.348 \end{bmatrix}.$$

In this case, the posterior standard deviation of β_3 is 3.4 times larger than the standard deviation of β_1.

Apley and Kim (2004) discusses this and other design considerations in detail. They recommend a two-phase experimental procedure in which the ARMA parameters are estimated in the first phase while holding the inputs constant. Based on these initial ARMA parameter estimates, the second phase is designed to result in an acceptable $\boldsymbol{\Sigma}_\beta$. After conducting the experiment, Apley and Kim (2004) also recommends a procedure for determining whether the uncertainty is acceptable or whether additional experimentation is necessary to improve the parameter estimates. This can be accomplished using the closed-loop stability and performance analysis of Section 5.

To illustrate, suppose the experiment resulted in $\hat{\beta}_0 = 0$, $\hat{\boldsymbol{\beta}} = [1\ 1\ 1]^T$ and the above disturbance parameter estimates (coinciding with their true values). If we substitute $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma}_\beta$ from above into Equations (26) through (29), the posterior distributions of C_{cmv} and C_{cemv} are N(0.92, 0.27^2)

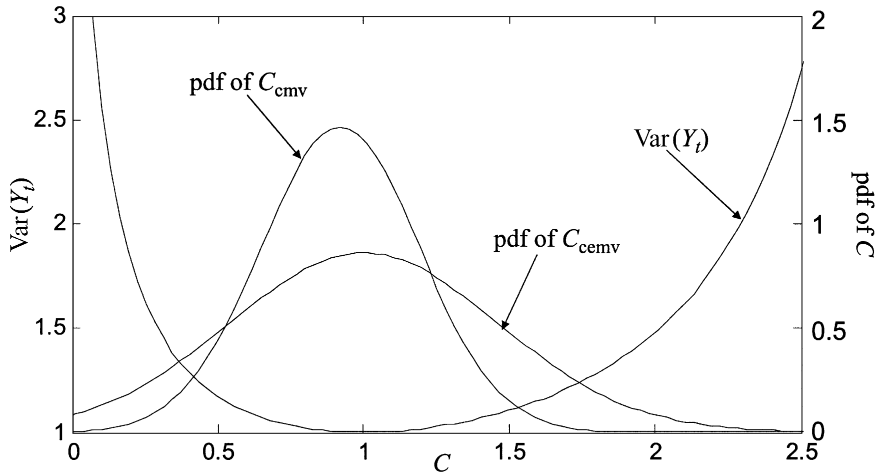

Fig. 6. The variance of Y_t as a function of C, with the pdfs of C_{cmv} and C_{cemv}.

and N(1, 0.47^2), respectively. For either CMV or CEMV control, Equation (30) implies the closed-loop system will be stable for $C \in \Omega_c = (-0.08, 3.00)$. Hence, the posterior probability of instability is only 1.2×10^{-4} for CMV control, which may indicate that further experimentation is unnecessary. Before concluding this, one may also consider the closed-loop performance as C varies over its posterior distribution. From Equation (24), closed-loop performance will clearly depend on the unknown $\Phi(B)$ and $\Theta(B)$, although closed-loop stability will not. Suppose we neglect uncertainty in the ARMA parameters and assume $\Phi(B) = \hat{\Phi}(B)$ and $\Theta(B) = \hat{\Theta}(B)$. In this case, Equation (24) becomes $G(B) = (1 - [\hat{\phi}_1 + C(\hat{\theta}_1 - \hat{\phi}_1)]B)^{-1}(1 - \hat{\theta}_1 B)$. For each potential value of C, one can calculate the impulse response coefficients (refer to Box *et al.* (1994) for details) and then the closed-loop $\mathrm{Var}(Y_t) = \sigma_a^2 \sum_{j=0}^{\infty} g_j^2$. Figure 6 shows the closed-loop variance as a function of C, along with the probability density functions (pdfs) of C_{cmv} and C_{cemv}. This type of figure could aid in determining whether the uncertainty in C is too large, in which case further experimentation should be used to reduce the uncertainty in the parameters. It could also be used to determine whether the uncertainty is small enough that CMV provides a negligible improvement over CEMV.

6.3. *Adaptive CMV control*

In the nonadaptive version of CMV, the on-line data are not used to update the parameter estimates or their covariance information. It is natural to consider using the on-line data in an attempt to improve the parameter estimates and reduce their uncertainty via an adaptive version of CMV. The adaptive version is identical to the nonadaptive version, except that $\hat{\gamma}$ and $\boldsymbol{\Sigma}_\gamma$ are continuously updated as each new on-line observation is obtained. An adaptive version of the approximate maximum likelihood method discussed in Appendix A, often referred to as the recursive prediction error method, is given by the recursions (Ljung, 1987):

$$\hat{\gamma}_t = \hat{\gamma}_{t-1} + \mathbf{R}_t^{-1}\mathbf{z}_t\hat{a}_t, \tag{32}$$

and

$$\mathbf{R}_t = \rho\mathbf{R}_{t-1} + \mathbf{z}_t\mathbf{z}_t^T, \tag{33}$$

where $\rho \in (0,1]$ is the “forgetting factor”, and $\mathbf{z}_t$ is defined in Equation (A4) for ARMA disturbances and in Equation (18) for ARIMA disturbances. The quantity $\hat{a}_t$ and the elements of $\mathbf{z}_t$ are updated at time t using the parameter estimate $\hat{\gamma}_{t-1}$ (refer to Ljung (1987) for details). An intuitive interpretation of the forgetting factor is that the recursive prediction error estimate $\hat{\gamma}_t$ attempts to minimize the exponentially weighted sum-of-the-squares of the residual errors $\sum_{j=0}^{t-1} \rho^j \hat{a}_{t-j}^2$. An approximate expression for the parameter covariance at time t is $\boldsymbol{\Sigma}_{\gamma,t} = \hat{\sigma}_a^2(1+\rho)^{-1}\mathbf{R}_t^{-1}$. The recursions can be initialized by setting $\hat{\gamma}_0$ equal to the estimate of γ from the off-line experiment and $\mathbf{R}_0$ equal to $\hat{\sigma}_a^2(1+\rho)^{-1}$ multiplied by the inverse of the parameter covariance matrix from the off-line experiment.

In a continuation of the example of Section 6.1, adaptive CMV control was applied with $\rho = 0.995$ and $\hat{\gamma}_0$ and $\mathbf{R}_0$ obtained from the off-line experiment described in Section 6.1. Figure 7 shows a typical closed-loop process output for 6000 timesteps of the simulation. The rather large number of timesteps was used primarily to illustrate a common problem in adaptive control, where periods of good control performance alternate with periods of poor performance. The sample mean-square error for the first 2000 timesteps was 1.01, very close to σ_a^2. The reason the control performance eventually deteriorated is related to Fig. 8, which shows wildly drifting estimates of $\boldsymbol{\beta}$ and β_0 over the first 1500 timesteps of the simulation. By timestep 2000 the parameter estimates had reached extremely large values (on the order of 10^5), and shortly thereafter the control performance deteriorated.

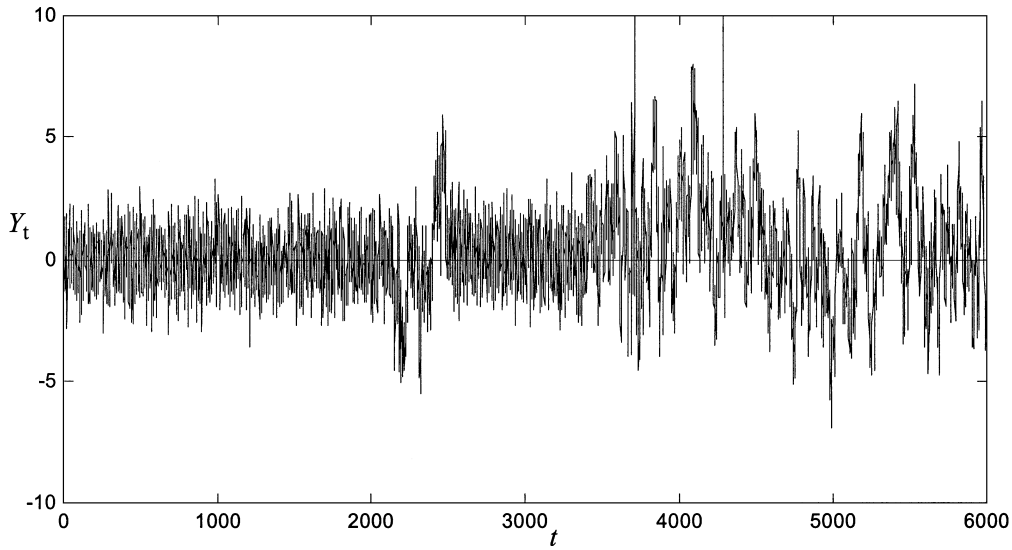

Fig. 7. Closed-loop output for adaptive CMV control with $\rho = 0.995$.

Considering Equations (32) and (33), there is a simple explanation for why the parameter estimates drift wildly. In the closed-loop, the control law of Equation (20) tends to confine the control inputs $\boldsymbol{X}_t$ to a one-dimensional subspace (along the direction of $\boldsymbol{\Sigma}_{\beta,t}^{-1}\hat{\boldsymbol{\beta}}_t$). The $\boldsymbol{u}_t$ component of $\boldsymbol{z}_t$ is also confined to this subspace, so that $\mathbf{R}_t$ in Equation (33) eventually becomes nearly singular. When this occurs, the scaled covariance matrix $\mathbf{R}_t^{-1}$ becomes very large, and the update equation, Equation (32), causes the parameter estimates to vary wildly. This is referred to as "covariance wind up" and the cause is referred to as "lack of persistent excitation" (Åström and Wittenmark, 1995). It is interesting to note that the control performance may be good even when $\hat{\boldsymbol{\beta}}_t$ differs substantially from $\boldsymbol{\beta}$, which can be observed from timesteps 1000 to 1500 in Figs. 7 and 8. From Equation (22), the closed-loop output variance depends on $\hat{\boldsymbol{\beta}}_t$ only via $C_{\text{cmv}} = (1 + \hat{\boldsymbol{\beta}}_t^T \boldsymbol{\Sigma}_{\beta,t}^{-1}\hat{\boldsymbol{\beta}}_t)^{-1}\hat{\boldsymbol{\beta}}_t^T \boldsymbol{\Sigma}_{\beta,t}^{-1}\hat{\boldsymbol{\beta}}_t$ which may remain close to unity even though $\hat{\boldsymbol{\beta}}_t$ differs from $\boldsymbol{\beta}$. Eventually, however, the parameter estimates drift in a manner that causes C_{cmv} to drift also, and control performance deteriorates. This may or may not provide enough excitation to improve the parameter estimates to the point that control performance again becomes acceptable.

One way to avoid covariance wind up is to use a forgetting factor of unity. A forgetting factor that is strictly less than unity (e.g., $\rho = 0.995$) is necessary if one wishes to allow the possibility of tracking time-varying parameters in adaptive control. If one believes that the parameters do not vary over time, as has been the assumption in this article, there is no reason to use $\rho < 1$. If $\rho = 1$, $\mathbf{R}_t$ is

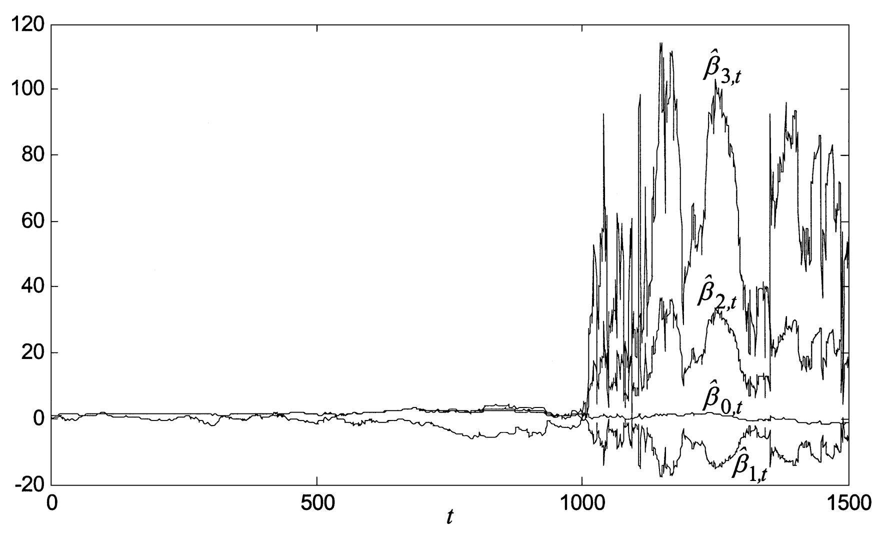

Fig. 8. Parameter estimates for adaptive CMV control corresponding to the first 1500 timesteps of Fig. 7.

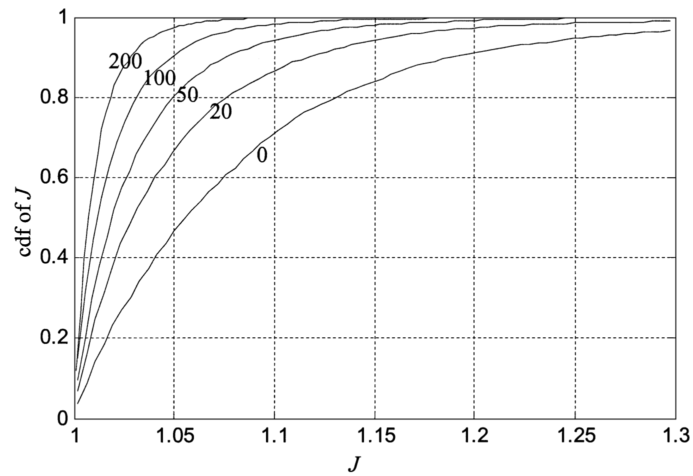

Fig. 9. Cdfs of J for adaptive CMV control with $\rho = 1$ using the parameter estimates after the first 0, 20, 50, 100 and 200 on-line observations.

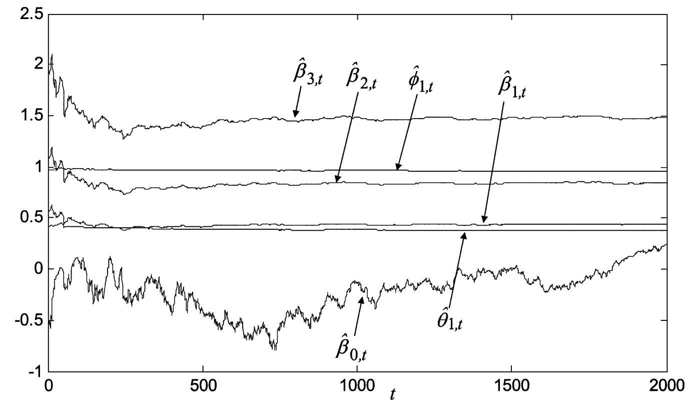

Fig. 10. Parameter estimates for adaptive CMV control with $\rho = 1$.

nondecreasing over time, its inverse remains bounded, and covariance wind up is avoided. 10 000 Monte Carlo replicates of the above adaptive CMV simulation with $\rho = 1$ were conducted. For each replicate, the parameter estimates were recorded after 20, 50, 100 and 200 on-line observations were obtained. The mean-square error for the steady-state response of Equation (22) using the parameter estimates at each of these points in time was then calculated. The empirical cdfs of the mean square errors are shown in Fig. 9. The cdf for zero additional on-line observations corresponds to using the parameter estimates from the off-line experiment with no on-line adaptation.

From Fig. 9, on-line adaptation with $\rho = 1$ clearly reduces the mean-square error. When deciding whether on-line adaptation outweighs the additional complexity of its implementation, the procedure described in Section 6.2 for evaluating the adequacy of the off-line experiment may be useful. If it appears that the off-line experiment has provided sufficiently accurate parameter estimates in terms of the probability of achieving an acceptable closed-loop variance (see, e.g., Fig. 6), one may wish to forego on-line adaptation. One should also consider that even though on-line adaptation with $\rho = 1$ improves the controller performance, the parameter estimates generally do not converge to the true parameters. This is illustrated in Fig. 10, which shows the adaptive parameter estimates over 2000 timesteps of a typical Monte Carlo replicate. Over the 10 000 Monte Carlo replicates, the average values of $\|\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}\|$ after $t = 0$, 20, 50, 100, 200 and 2000 on-line observations were 1.24, 1.11, 1.06, 1.02, 0.99 and 0.98, respectively. There was little further improvement after 2000 timesteps.

Åström and Wittenmark (1995) discusses a number of strategies for avoiding covariance wind up with $\rho < 1$. One of the most popular is the constant trace algorithm, in which $\mathbf{R}_t^{-1}$ is scaled at each timestep so that its trace is equal to some prespecified design parameter. Åström and Wittenmark (1995) also recommend adding a small scalar multiple of the identity matrix to $\mathbf{R}_t^{-1}$ at each timestep. Additional simulation results, which are omitted for brevity, revealed that the constant trace algorithm with appropriate choice of design parameters was able to largely eliminate the covariance wind up and parameter drift illustrated in Fig. 8. The parameters estimates still did not converge, however, and the mean-square error of the closed-loop output was often larger than for nonadaptive CMV. Much trial and error simulation was required to find a combination of design parameters that reduced the parameter drift and simultaneously improved the closed-loop mean-square error.

7. Conclusions

This article has extended CMV control to the case of ARMA and ARIMA disturbances. An analytical investigation of the closed-loop stability and performance of CMV has shown that for ARMA(1,1) disturbances, CMV always results in a higher probability of stability than does the corresponding CEMV controller. In addition, for the example considered in Section 6, CMV control clearly resulted in a better closed-loop performance than CEMV control. Since the form of the CMV control law is structurally similar to the CEMV controller, in particular when the parameters are estimated in the open-loop, its implementation involves only a modest increase in complexity. The primary difference is that the controller gain vector is altered in a manner that takes into account the uncertainty in the input parameters.

Only parametric uncertainty has been considered here. Apley and Kim (2004) investigated the effects of model structure uncertainty and found that both CMV and CEMV control were relatively robust to errors in the disturbance model structure if an ARIMA model was assumed, which is consistent with the findings in Box and

Luceño (1997). In addition, they found that CMV control was less adversely affected than CEMV control when the input/output model structure was over-modeled. Apley and Kim (2003) also discuss approaches for considering other forms of model structure uncertainty when designing the CMV control law.

The parametric model uncertainty considered in this article is due purely to statistical estimation error, as opposed to uncertainty due to time-varying parameters. In light of the covariance wind up problem that plagues adaptive control when $\rho < 1$, it would be difficult to tune an adaptive CMV controller to effectively track time-varying parameters. If one were able to measure the uncertainty due to time-varying parameters and represent this uncertainty in the form of a parameter covariance matrix, it is possible that nonadaptive CMV control would provide a control strategy that is reasonably robust to the parameter variations. Additional work on representing and quantitatively identifying other forms of structural and parametric model uncertainty is needed.

Acknowledgements

This work was supported by the State of Texas Advanced Technology Program under grant 000512-0289-1999 and the National Science Foundation under grant DMI-0093580. The authors would also like to thank three anonymous referees for many helpful comments that have improved this article.

References

Anderson, B.D.O. (1985) Adaptive systems, lack of persistency of excitation and bursting phenomena. *Automatica*, **21**, 247–258.

Apley, D.W. and Kim, J.B. (2004) Cautious control of industrial process variability with uncertain input and disturbance model parameters. *Technometrics*, (in press).

Åström, K.J. and Wittenmark, B. (1990) *Computer Controlled Systems: Theory and Design*, 2nd edn., Prentice Hall, Englewood Cliffs, NJ.

Åström, K.J. and Wittenmark, B. (1995) *Adaptive Control*, 2nd edn. Addison Wesley, New York, NY.

Bodson, M. (1993) Pseudo-burst phenomenon in ideal adaptive systems. *Automatica*, **29**, 929–940.

Box, G., Jenkins, G. and Reinsel, G. (1994) *Time Series Analysis, Forecasting, and Control*, 3rd edn., Prentice Hall, Englewood Cliffs, NJ.

Box, G. and Kramer, T. (1992) Statistical process monitoring and feedback adjustment—a discussion. *Technometrics*, **34**, 251–267.

Box, G.E.P. and Luceño, A. (1997) *Statistical Control by Monitoring and Feedback Adjustment*, Wiley, New York, NY.

Carlin, B.P. and Louis, T.A. (1998) *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman & Hall/CRC, New York, NY.

Del Castillo, E. and Hurwitz, A.M. (1997) Run-to-run process control: literature review and extensions. *Journal of Quality Technology*, **29**, 184–196.

English, J.R. and Case, K.E. (1990) Control charts applied as filtering devices within a feedback control loop. *IIE Transactions*, **22**, 255–269.

Francis, B.A. (1988) *A Course in H_∞ Control Theory*, Springer Verlag, New York, NY.

Green, M. and Limebeer, D. (1995) *Linear Robust Control*, Prentice Hall, Englewood Cliffs, NJ.

Hamby, E.S., Kabamba, P.T. and Khargonekar, P.P. (1998) A probabilistic approach to run-to-run control. *IEEE Transactions on Semiconductor Manufacturing*, **11**, 654–669.

Ingolfsson, A. and Sachs, E. (1993) Stability and sensitivity of an EWMA controller. *Journal of Quality Technology*, **25**, 271–287.

Janakiram, M. and Keats, J.B. (1998) Combining SPC and EPC in a hybrid industry. *Journal of Quality Technology*, **30**, 189–200.

Ljung, L. (1987) *System Identification: Theory for the User*, Prentice Hall, Englewood Cliffs, NJ.

Luceño, A. (1998) Performance of discrete feedback adjustment schemes with dead band, under stationary versus nonstationary stochastic disturbances. *Technometrics*, **40**, 223–233.

MacGregor, J.F. (1988) On-line statistical process control. *Chemical Engineering Progress*, **84**, 21–31.

Montgomery, D.C., Keats, J.B., Runger, G.C. and Messina, W.S. (1994) Integrating statistical process control and engineering process control. *Journal of Quality Technology*, **26**, 79–87.

Morari, M. and Zafiriou, E. (1989) *Robust Process Control*, Prentice Hall, Englewood Cliffs, NJ.

Mosca, E. (1995) *Optimal, Predictive, and Adaptive Control*, Prentice Hall, Englewood Cliffs, NJ.

Ogata, K. (1995) *Discrete-Time Control Systems*, 2nd edn., Prentice Hall, Englewood Cliffs, NJ.

Radenkovic, M.S. and Michel, A.N. (1993) Bursting phenomena in extended-least squares based self-tuning control, in *Proceedings of the American Control Conference*, Piscataway, NJ, pp. 296–300.

Sachs, E., Hu, S. and Infolfsson, A. (1995) Run by run process control: combining SPC and feedback control. *IEEE Transactions on Semiconductor Manufacturing*, **8**, 26–43.

Tsung, F. and Apley, D.W. (2002) The dynamic T^2 chart for monitoring feedback-controlled processes. *IIE Transactions*, **34**, 1043–1053.

Tsung, F., Wu, H. and Nair, V.N. (1998) On the efficiency and robustness of discrete proportional-integral control schemes. *Technometrics*, **40**, 214–222.

Vander Wiel, S.A. (1996) Monitoring processes that wander using integrated moving average models. *Technometrics*, **38**, 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**, 286–297.

Zhou, K., Doyle, J.C. and Glover, K. (1996) *Robust and Optimal Control*, Prentice Hall, Upper Saddle River, NJ.

Appendices

Appendix A: The posterior mean and covariance of γ for ARMA(n,m) disturbances

Let $\pi(\gamma)$, $\pi(\gamma \mid \mathbf{Y})$, and $\pi(\mathbf{Y} \mid \gamma)$ denote the assumed prior distribution of γ, the conditional distribution of γ given the off-line data $\mathbf{Y}$, and the conditional distribution of $\mathbf{Y}$ given γ, respectively. To obtain an approximate expression for $\mathbf{\Sigma}_\gamma$, we use the result (Carlin and Louis, 1998) that for large N the posterior distribution $\pi(\gamma \mid \mathbf{Y})$ is approximately Gaussian with mean $\hat{\gamma}$ (defined as the mode of the posterior distribution $\pi(\gamma \mid \mathbf{Y})$) and covariance matrix $\mathbf{\Sigma}_\gamma$ equal to the inverse of the "observed" Fisher information matrix:

$$\mathbf{\Sigma}_\gamma = -\left[\left. \frac{\partial^2 \log(\pi(\mathbf{Y} \mid \gamma)\pi(\gamma))}{\partial \gamma^2} \right|_{\gamma=\hat{\gamma}} \right]^{-1}. \tag{A1}$$

The term in brackets denotes the Hessian (second derivative matrix) of $\pi(\mathbf{Y}\mid\gamma)\,\pi(\gamma)$ evaluated at $\gamma=\hat{\gamma}$. The posterior mode $\hat{\gamma}$ is the value of γ that maximizes $\log(\pi(\gamma\mid\mathbf{Y}))$ or, equivalently, $\log(\pi(\mathbf{Y}\mid\gamma)\,\pi(\gamma))$.

Suppose, temporarily, that a noninformative prior distribution is assumed for γ. In this case, the Maximum Likelihood Estimate (MLE) of γ coincides with the mode of the posterior distribution $\pi(\gamma\mid\mathbf{Y})$. Assuming the a_ts are Gaussian, the MLE is approximately equivalent (Box *et al.*, 1994) to taking $\hat{\gamma}$ to be the γ that minimizes:

$$-\log(\pi(\mathbf{Y}\mid\gamma)) \cong \text{constant} + \left(2\sigma_a^2\right)^{-1}\Sigma_{t=1}^N a_t^2(\gamma), \tag{A2}$$

where $a_t(\gamma)$ is given by Equation (6). Ljung (1987) and Box *et al.* (1994) discuss this method of estimating γ in detail and refer to it as nonlinear least squares and the prediction error method, respectively. Numerous commercial statistical software packages are available for obtaining $\hat{\gamma}$.

To evaluate $\mathbf{\Sigma}_\gamma$ with a flat prior distribution, substituting Equation (A2) into Equation (A1) gives:

$$\begin{aligned}
&\left.\frac{\partial^2\log(\pi(\mathbf{Y}\mid\gamma)\pi(\gamma))}{\partial\gamma^2}\right|_{\gamma=\hat{\gamma}}\\
&\quad\cong -\sigma_a^{-2}\Sigma_{t=1}^N\left\{a_t(\hat{\gamma})\left.\frac{\partial^2 a_t(\gamma)}{\partial\gamma^2}\right|_{\gamma=\hat{\gamma}} + \boldsymbol{z}_t\boldsymbol{z}_t^T\right\}\\
&\quad\cong -\sigma_a^{-2}\Sigma_{t=1}^N\boldsymbol{z}_t\boldsymbol{z}_t^T,
\end{aligned}\tag{A3}$$

where $\boldsymbol{z}_t$ is defined as the transpose of $-\partial a_t(\gamma)/\partial\gamma$ evaluated at $\gamma=\hat{\gamma}$. The final approximate equality in Equation (A3) follows from the fact that the $a_t(\hat{\gamma})$s will be approximately zero-mean and independent of $\partial^2 a_t(\gamma)/\partial\gamma^2$ (refer to chapter 10 of Ljung (1987) for details). Following the same procedure as in Appendix B for differentiating with respect to γ, it can be shown that:

$$\boldsymbol{z}_t = \left[\boldsymbol{u}_t^T \quad \hat{g} \quad \boldsymbol{v}_t^T\right]^T, \tag{A4}$$

with $\boldsymbol{u}_t$, $\hat{g}$ and $\mathbf{v}_t$ defined in Section 2. Hence, Equation (A1) becomes:

$$\mathbf{\Sigma}_\gamma = \frac{\sigma_a^2}{N}\mathbf{\Sigma}_{\mathbf{z}}^{-1}, \tag{A5}$$

where

$$\mathbf{\Sigma}_z \equiv \frac{1}{N}\sum_{t=1}^N \boldsymbol{z}_t\boldsymbol{z}_t^T, \tag{A6}$$

is the sample "covariance matrix" of $\boldsymbol{z}_t$.

Suppose that instead of assuming $\pi(\gamma)$ is noninformative, it is assumed multivariate normal with some mean $\boldsymbol{\mu}_0$ and covariance matrix $\mathbf{\Sigma}_0$. The posterior mode $\hat{\gamma}$ then becomes the γ that minimizes:

$$\begin{aligned}-\log(\pi(\mathbf{Y}\mid\gamma)\pi(\gamma)) &= \text{constant} + \left(2\sigma_a^2\right)^{-1}\sum_{t=1}^N at^2(\gamma)\\
&\quad + 2^{-1}[\gamma-\boldsymbol{\mu}_0]^T\mathbf{\Sigma}_0^{-1}[\gamma-\boldsymbol{\mu}_0].\end{aligned}$$

Since the Hessian of $\log(\pi(\mathbf{Y}\mid\gamma)\pi(\gamma))$ is the Hessian of $\log(\pi(\mathbf{Y}\mid\gamma)$ minus $\mathbf{\Sigma}_0^{-1}$, similar to Equation (A5) it follows that the posterior covariance is:

$$\mathbf{\Sigma}_\gamma = \left[N\sigma_a^{-2}\mathbf{\Sigma}_z + \mathbf{\Sigma}_0^{-1}\right]^{-1},$$

with $\mathbf{\Sigma}_z$ as in Equation (A6).

Appendix B: Derivation of Equation (8) for J_t and the CMV control law of Equation (14)

To find the first-order Taylor approximation of Y_t about $\gamma=\hat{\gamma}$, first note that by the definition of $H(B)$:

$$1-\frac{BH(B)}{\Theta(B)} = \frac{\Phi(B)}{\Theta(B)}.$$

Differentiating Equation (7) therefore gives:

$$\begin{aligned}
\left.\frac{\partial Y_t}{\partial\beta_0}\right|_{\gamma=\hat{\gamma}} &= 1-\frac{\hat{H}(B)}{\hat{\Theta}(B)}1 = 1-\left.\frac{\hat{H}(B)}{\hat{\Theta}(B)}\right|_{B=1}\\
&= \left.\frac{\hat{\Phi}(B)}{\hat{\Theta}(B)}\right|_{B=1} = \hat{g},\\
\left.\frac{\partial Y_t}{\partial\boldsymbol{\beta}}\right|_{\gamma=\hat{\gamma}} &= \left[1-\frac{\hat{H}(B)B}{\hat{\Theta}(B)}\right]\boldsymbol{X}_t = \frac{\hat{\Phi}(B)}{\hat{\Theta}(B)}\boldsymbol{X}_t = \boldsymbol{u}_t,\\
\left.\frac{\partial Y_t}{\partial\boldsymbol{\phi}}\right|_{\gamma=\hat{\gamma}} &= \left.\frac{\partial H(B)}{\partial\boldsymbol{\phi}}\right|_{\gamma=\hat{\gamma}}\frac{1}{\hat{\Theta}(B)}[Y_{t-1}-\hat{\beta}_0-\hat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{X}_{t-1}]\\
&= [1, B, \ldots, B^{n-1}]\frac{1}{\hat{\Phi}(B)}\hat{a}_{t-1},\\
&= [\hat{a}_{\Phi,t-1},\;\; \hat{a}_{\Phi,t-2}, \ldots, \hat{a}_{\Phi,t-n}],
\end{aligned}\tag{A7}$$

and

$$\begin{aligned}
\left.\frac{\partial Y_t}{\partial\boldsymbol{\theta}}\right|_{\gamma=\hat{\gamma}} &= \left[-\frac{\hat{H}(B)}{\hat{\Theta}^2(B)}\left.\frac{\partial\Theta(B)}{\partial\boldsymbol{\theta}}\right|_{\gamma=\hat{\gamma}} + \frac{1}{\hat{\Theta}(B)}\left.\frac{\partial H(B)}{\partial\boldsymbol{\theta}}\right|_{\gamma=\hat{\gamma}}\right]\\
&\quad\times[Y_{t-1}-\hat{\beta}_0-\hat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{X}_{t-1}],\\
&= \left[\frac{\hat{H}(B)}{\hat{\Theta}(B)}[B, B^2, \ldots, B^m] - [1, B, \ldots, B^{m-1}]\right]\\
&\quad\times\frac{1}{\hat{\Theta}(B)}[Y_{t-1}-\hat{\beta}_0-\hat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{X}_{t-1}],\\
&= -\left[1-\frac{\hat{H}(B)B}{\hat{\Theta}(B)}\right][1, B, \ldots, B^{m-1}]\frac{1}{\hat{\Phi}(B)}\hat{a}_{t-1}\\
&= -\frac{\hat{\Phi}(B)}{\hat{\Theta}(B)}[1, B, \ldots, B^{m-1}]\frac{1}{\hat{\Phi}(B)}\hat{a}_{t-1},\\
&= [-\hat{a}_{\Theta,t-1}, -\hat{a}_{\Theta,t-2}, \ldots, -\hat{a}_{\Theta,t-m}].
\end{aligned}$$

The notational convention in Equation (A7) is that $\hat{\Theta}^{-1}(B)\hat{H}(B)1$ is the steady-state response of the filter $\hat{\Theta}^{-1}(B)\hat{H}(B)$ to a unit magnitude step, which by the final value theorem (Åström and Wittenmark, 1990) is $\hat{\Theta}^{-1}(B)\hat{H}(B)$ evaluated at $B=1$.

It also follows from Equation (7) that:

$$\begin{aligned}
Y_t|_{\gamma=\hat{\gamma}} &= \hat{\beta}_0 + \hat{\boldsymbol{\beta}}^T \boldsymbol{X}_t + \frac{\hat{H}(B)}{\hat{\Theta}(B)}[Y_{t-1} - \hat{\beta}_0 - \hat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{X}_{t-1}] + a_t, \\
&= \left[1 - \frac{\hat{H}(B)B}{\hat{\Theta}(B)}\right](\hat{\beta}_0 + \hat{\boldsymbol{\beta}}^T \boldsymbol{X}_t) + \frac{\hat{H}(B)}{\hat{\Theta}(B)} Y_{t-1} + a_t, \\
&= \frac{\hat{\Phi}(B)}{\hat{\Theta}(B)}\hat{\beta}_0 + \hat{\boldsymbol{\beta}}^T \boldsymbol{u}_t + \frac{\hat{H}(B)}{\hat{\Theta}(B)} Y_{t-1} + a_t.
\end{aligned}$$

Combining the above results, the first-order Taylor approximation of Y_t about $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}$ is:

$$\begin{aligned}
Y_t \cong\ & Y_t|_{\gamma=\hat{\gamma}} + \left.\frac{\partial Y_t}{\partial \boldsymbol{\gamma}}\right|_{\gamma=\hat{\gamma}} \tilde{\boldsymbol{\gamma}} = \frac{\hat{\Phi}(B)}{\hat{\Theta}(B)}\hat{\beta}_0 + \hat{\boldsymbol{\beta}}^T \boldsymbol{u}_t \\
&+ \frac{\hat{H}(B)}{\hat{\Theta}(B)} Y_{t-1} + \boldsymbol{u}_t^T \tilde{\boldsymbol{\beta}} + \tilde{\beta}_0 \hat{g} + \boldsymbol{v}_t^T \tilde{\boldsymbol{\alpha}} + a_t, \quad \text{(A8)}
\end{aligned}$$

so that

$$\begin{aligned}
Y_t - T \cong\ & \frac{\hat{\Phi}(B)}{\hat{\Theta}(B)}\left(\hat{\beta}_0 - T\right) + \hat{\boldsymbol{\beta}}^T \boldsymbol{u}_t + \frac{\hat{H}(B)}{\hat{\Theta}(B)}\left(Y_{t-1} - T\right) \\
&+ \boldsymbol{u}_t^T \tilde{\boldsymbol{\beta}} + \tilde{\beta}_0 \hat{g} + \boldsymbol{v}_t^T \tilde{\boldsymbol{\alpha}} + a_t.
\end{aligned}$$

Substituting this into Equation (4) gives Equation (8).

The CMV control law is determined by setting the partial derivative (with respect to $\boldsymbol{u}_t$) of Equation (8) equal to zero, which gives:

$$\begin{aligned}
\boldsymbol{u}_t = -[\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^T + \boldsymbol{\Sigma}_\beta]^{-1} \Bigg\{ \Bigg[& \frac{\hat{H}(B)}{\hat{\Theta}(B)}\left(Y_{t-1} - T\right) \\
&+ \frac{\hat{\Phi}(B)}{\hat{\Theta}(B)}(\hat{\beta}_0 - T)\Bigg]\hat{\beta} + \boldsymbol{\Sigma}_{\beta\beta_0}\hat{g} + \boldsymbol{\Sigma}_{\beta\alpha}\boldsymbol{v}_t \Bigg\},
\end{aligned}$$

or

$$\begin{aligned}
\mathbf{X}_t = \frac{\hat{\Theta}(B)}{\hat{\Phi}(B)}\boldsymbol{u}_t = -[\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^T + \boldsymbol{\Sigma}_\beta]^{-1} \Bigg\{ \Bigg[& \frac{\hat{H}(B)}{\hat{\Phi}(B)}(Y_{t-1} - T) \\
&+ (\hat{\beta}_0 - T)\Bigg]\hat{\boldsymbol{\beta}} + \boldsymbol{\Sigma}_{\beta\beta_0} + \boldsymbol{\Sigma}_{\beta\alpha}\frac{\hat{\Theta}(B)}{\hat{\Phi}(B)}\boldsymbol{v}_t \Bigg\}.
\end{aligned}$$

Equation (14) follows from the relationships

$$[\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^T + \boldsymbol{\Sigma}_\beta]^{-1} = \left[\mathbf{I} - \frac{\boldsymbol{\Sigma}_\beta^{-1}\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^T}{1 + \hat{\boldsymbol{\beta}}^T \boldsymbol{\Sigma}_\beta^{-1}\hat{\boldsymbol{\beta}}}\right]\boldsymbol{\Sigma}_\beta^{-1}, \quad \text{(A9)}$$

and

$$[\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^{\mathrm{T}} + \boldsymbol{\Sigma}_\beta]^{-1}\hat{\boldsymbol{\beta}} = \frac{\boldsymbol{\Sigma}_\beta^{-1}\hat{\boldsymbol{\beta}}}{1 + \hat{\boldsymbol{\beta}}^T \boldsymbol{\Sigma}_\beta^{-1}\hat{\boldsymbol{\beta}}}.$$

Appendix C: Proof that with ARMA(1,1) disturbances the probability of stability is higher for CMV control than for CEMV control

Consider the spectral decomposition $\boldsymbol{\Sigma}_\beta = \mathbf{E}\Lambda\mathbf{E}^T$, where $\mathbf{E} \equiv [\mathbf{e}_1\ \mathbf{e}_2 \ldots \mathbf{e}_p]$, $\Lambda \equiv \mathrm{diag}\{\lambda_i: i = 1, 2, \ldots, p\}$, and $\{\mathbf{e}_i, \lambda_i\}$ are the eigenvector/eigenvalue pairs of $\boldsymbol{\Sigma}_\beta$. Define $\mathbf{b} \equiv \mathbf{E}^T\hat{\boldsymbol{\beta}}/\|\hat{\boldsymbol{\beta}}\| = [b_1\ b_2 \ldots b_p]^T$, and note that $\mathbf{b}$ is unit norm. Using this and Equations (26) through (29):

$$\begin{aligned}
\frac{(\mu_{\mathrm{cmv}}/\sigma_{\mathrm{cmv}})^2}{(\mu_{\mathrm{cemv}}/\sigma_{\mathrm{cemv}})^2} &= \left(\frac{\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_\beta^{-1}\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}^T\hat{\boldsymbol{\beta}}}\right)\left(\frac{\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_\beta\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}^T\hat{\boldsymbol{\beta}}}\right) \\
&= (\mathbf{b}^T\Lambda^{-1}\mathbf{b})(\mathbf{b}^T\Lambda\mathbf{b}) \\
&= \left(\sum_{i=1}^p \frac{b_i^2}{\lambda_i}\right)\left(\sum_{j=1}^p \lambda_j b_j^2\right), \\
&= \sum_{i=1}^p b_i^4 + \sum_{i=1}^p \sum_{j=i+1}^p \left(\frac{\lambda_i}{\lambda_j} + \frac{\lambda_j}{\lambda_i}\right) b_i^2 b_j^2 \\
&\geq \sum_{i=1}^p b_i^4 + 2\sum_{i=1}^p \sum_{j=i+1}^p b_i^2 b_j^2 \\
&= \left(\sum_{i=1}^p b_i^2\right)\left(\sum_{i=1}^p b_j^2\right) = 1.
\end{aligned}$$

It is therefore always the case that $\mu_{\mathrm{cmv}}/\sigma_{\mathrm{cmv}} \geq \mu_{\mathrm{cemv}}/\sigma_{\mathrm{cemv}}$. Since C_{cmv} and C_{cemv} are normally distributed and the lower boundary of the stability region in Equation (30) is always less than zero, this implies that the probability that C_{cmv} falls below the lower boundary is less than or equal to the probability that C_{cemv} falls below the lower boundary. Moreover, $\mu_{\mathrm{cmv}}/\sigma_{\mathrm{cmv}} \geq \mu_{\mathrm{cemv}}/\sigma_{\mathrm{cemv}}$ and $\mu_{\mathrm{cemv}} \geq \mu_{\mathrm{cmv}}$ imply that $\sigma_{\mathrm{cemv}} \geq \sigma_{\mathrm{cmv}}$. Since the upper boundary of the stability region in Equation (30) is always greater than one, it follows that the probability that C_{cmv} falls above the upper boundary is also less than or equal to the probability that C_{cemv} falls above the upper boundary. Consequently, the probability of stability with CMV control is at least as large as the probability of stability with CEMV control. Note that although Equation (30) applies only to the case that $\hat{\phi}_1 > \hat{\theta}_1$, for $\hat{\phi}_1 < \hat{\theta}_1$ one can also show that the lower boundary of the stability region is less than zero and the upper boundary is greater than one.

Biography

Daniel W. Apley received B.S. and M.S. degrees in Mechanical Engineering, a M.S. degree in Electrical Engineering, and a Ph.D. degree in Mechanical Engineering in 1990, 1992, 1995 and 1997, respectively, all from the University of Michigan. From 1997 to 1998 he was a post-doctoral fellow with the Department of Industrial and Operations Engineering at the University of Michigan. Between 1998 and 2003 he was with Texas A&M University, where he was an Assistant Professor of Industrial Engineering. In 2003 he became an Associate Professor in the Department of Industrial Engineering and Management Sciences at Northwestern

University. His research area is manufacturing variation reduction via statistical process monitoring, diagnosis and automatic control and the utilization of large sets of in-process measurement data. His current work is sponsored by Ford, Solectron, Applied Materials, the National Science Foundation and the State of Texas Advanced Technology Program. He was an AT&T Bell Laboratories Ph.D. Fellow from 1993 to 1997 and received the NSF CAREER award in 2001. He is a member of IIE, IEEE, ASME, INFORMS and SME.

Contributed by the On-Line Quality Engineering Department