Cautious Control of Industrial Process Variability With Uncertain Input and Disturbance Model Parameters

Daniel W. Apley

Department of Industrial Engineering
and Management Sciences
Northwestern University
Evanston, IL 60208-3119
(*apley@northwestern.edu*)

Jeongbae Kim

Vendor Evaluation Division
Korea Telecom
Seoul, Korea
(*jeongbae@kt.co.kr*)

This article discusses a method for controlling variation in industrial processes when the model parameters are estimated from data and subject to uncertainty. A static input/output relationship with multiple input variables and an integrated moving average disturbance model are assumed. Most robust control methods use deterministic measures of uncertainty and a control objective that focuses on worst-case performance. This work uses a probabilistic measure of uncertainty and a control objective that relates more closely to minimizing variation, where parameter estimation errors are treated simply as an additional source of variability. We show that this approach results in a higher probability of closed-loop stability than the standard minimum variance control and can substantially lessen the adverse impact of parameter uncertainty on closed-loop variance. Guidelines for designing and evaluating the experiment used to estimate the model parameters are also discussed.

KEY WORDS: Cautious control; Engineering process control; Minimum variance control; Robust control; Statistical process control.

1. INTRODUCTION

The use of engineering process control (EPC) for reducing process variability in what has historically been the domain of statistical process control (SPC) is increasingly common in industry, due to a number of factors discussed by Box and Kramer (1992) and Box and Luceño (1997). In SPC, variability is reduced by detecting and eliminating assignable causes of variation. In EPC, variability is typically reduced by automatic, computer-controlled adjustment of some controllable process input variables. Although EPC strategies such as minimum variance (MV) control can be quite effective when an accurate process model is available, their performance may degrade substantially when using an empirical model that involves high levels of model uncertainty. This article discusses an EPC strategy for controlling variability under model uncertainty.

Suppose that a process to be controlled via EPC is described by the model

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

illustrated in Figure 1. The subscript t is a time index. Y_t is the measured (scalar) process output to be controlled. $\mathbf{X}_t = [X_{1,t}\ X_{2,t}\ \cdots\ X_{p,t}]^T$ is a p-dimensional vector of adjustable process inputs. d_t is a random disturbance that accounts for the variability in the process. β_0 and $\boldsymbol{\beta} = [\beta_1\ \beta_2\ \cdots\ \beta_p]^T$ are model parameters that describe the effects of the process inputs on the output. T is a target value for the process output. The disturbance is assumed to follow an integrated moving average (IMA) model of the form $d_t = d_{t-1} + a_t - \theta a_{t-1}$, where a_t is an iid, 0-mean sequence of random shocks with variance σ_a^2 and θ is the IMA parameter.

The model (1) is widely used to represent industrial processes that tend to wander if left uncontrolled (Box and Kramer 1992; Janakiram and Keats 1998; Montgomery, Keats, Runger, and Messina 1994; Tsung, Wu, and Nair 1998; Tsung and Apley 2002; Vander Wiel, Tucker, Faltin, and Doganaksoy 1992; Vander Wiel 1996; Del Castillo and Hurwitz 1997; Box and Luceño 1997; Luceño 1998; Chen and Guo 2001). For continuous-flow processes with short sampling intervals, we would typically use a model with input/output "dynamics," in which the output depends on both past and current inputs. For run-to-run or batch production processes, however, the static input/output relationship of (1) is typically assumed with $\mathbf{X}_t$ representing the output settings specified at the beginning of the tth batch. Most of the aforementioned work has considered only a single input variable ($p = 1$), although industrial processes often have multiple inputs available to manipulate. The multi-input model (1) is particularly common in semiconductor manufacturing (Hamby, Kabamba, and Khargonekar 1998; Ingolfsson and Sachs 1993; Sachs, Hu, and Ingolfsson 1995; Leang, Ma, Thompson, Bombay, and Spanos 1996; Gower-Hall, Boning, Rosenthal, and Wildhauer 2002). Although they did not necessarily assume a specific disturbance model, Hamby et al. (1998), Ingolfsson and Sachs (1993), Sachs et al. (1995), and many other authors focused on the popular exponentially weighted moving average (EWMA) controller, in which an EWMA is used to predict the disturbance, and the control inputs are then chosen so that the resulting output prediction equals the target value. If the EWMA parameter is set equal to $1 - \theta$, then the EWMA controller is equivalent to the MV controller for an assumed first-order IMA disturbance (Sachs et al. 1995).

TECHNOMETRICS, MAY 2004, VOL. 46, NO. 2
DOI 10.1198/004017004000000239

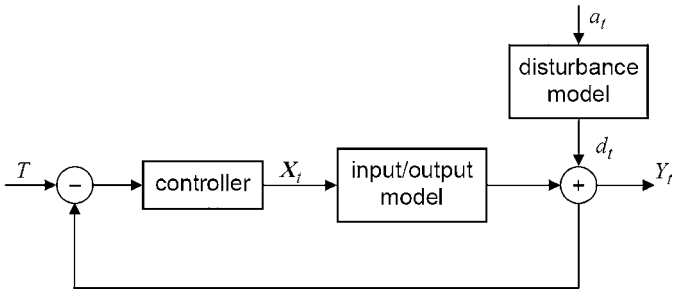

Figure 1. Block Diagram of the Assumed Process Model.

In practice, the model parameters $\boldsymbol{\beta}$ and θ must always be estimated from data, typically using experimental design, regression, and time series methods. If parameter uncertainty is large, then an MV controller (or any other controller) that is designed based on the estimates may result in output variance that is far from the minimum. The risk of increasing variability due to a poor understanding of process behavior (e.g., model uncertainty) was the essence of Deming's influential funnel experiment argument (Deming 1986; MacGregor 1990) discouraging the use of EPC, in lieu of SPC.

Despite the importance of model uncertainty, design of "robust" EPC strategies has received little attention from the SPC community. In contrast, robust control research has dominated the EPC literature over the past two decades. The field is now quite mature, and numerous comprehensive texts are available (e.g., Francis 1988; Green and Limebeer 1995; Morari and Zafiriou 1989; Zhou, Doyle, and Glover 1996). Robust control methods are not ideally suited for controlling variability in many types of manufacturing processes, however. To illustrate why, consider an estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$. Any well-defined notion of robust control must involve a well-defined measure of model uncertainty. Robust control methods use deterministic model uncertainty measures similar to that illustrated in Figure 2(a) for the case of two inputs. It is presumed that the true parameters lie with absolute certainty somewhere within a specified uncertainty region. The classic robust control design procedure (Morari and Zafiriou 1989) involves first designing a controller with "good" performance for the nominal model (i.e., when $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$) without consideration of uncertainty, then iteratively modifying the controller until certain robustness criteria are satisfied. Typical robustness criteria are stability and "acceptable" performance for all situations where the true parameters lie anywhere within the uncertainty region.

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. This is illustrated in Figure 2(b), which shows typical (multivariate normal) posterior probability contours for $\boldsymbol{\beta}$. Deterministic bounds that are not overly conservative to the point of being meaningless may be difficult to identify. Moreover, deterministic uncertainty measures lead to the fundamental principle behind robust control methods—that the robustness criteria be satisfied for *all* $\boldsymbol{\beta}$ lying within the uncertainty region. In this sense the design is based on the worst-case scenario, and points on the boundary of the uncertainty region in Figure 2(a) are implicitly assigned more weight than points near $\hat{\boldsymbol{\beta}}$. In actuality, it is much more likely that $\boldsymbol{\beta}$ lies in the neighborhood of $\hat{\boldsymbol{\beta}}$ than near the boundary of the uncertainty region. If probabilistic measures of parameter uncertainty are used, then all potential values for $\boldsymbol{\beta}$ could be assigned weights that are commensurate with their posterior probabilities.

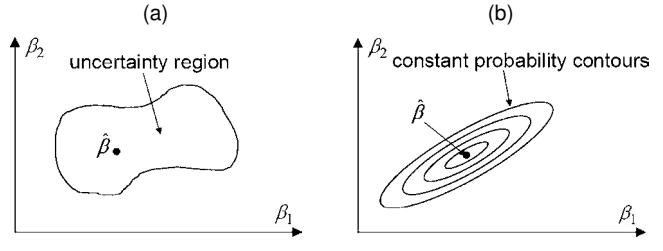

Figure 2. Comparison of Model Uncertainty Representations: (a) Deterministic Bounds Assumed in Robust Control; and (b) Probabilistic Measures Assumed in This Article.

This article discusses an EPC strategy for controlling process variability that considers probabilistic uncertainty in the model parameters. The objective is to minimize process variability with parameter uncertainty treated simply as an additional source of variability. A Bayesian approach is adopted in which the measure of uncertainty is the posterior covariance of $\boldsymbol{\beta}$ and θ, given the data from which their point estimates are obtained. Although the control strategy in this article is nonadaptive, a similar Bayesian approach is well known in the adaptive control literature and is referred to as cautious adaptive control (Åström and Wittenmark 1995). We thus refer to the approach as cautious minimum variance (CMV) control.

The format of the remainder of the article is as follows. In Section 2 we analyze the effects of parameter uncertainty in MV control. In Sections 3 and 4 we derive the CMV control law and provide a performance analysis and comparison with MV control. In Section 5 we use the results to develop guidelines for designing and evaluating the experiment used to estimate the parameters. Although the primary focus of this article is on parameter uncertainty, we briefly treat the issue of model structure uncertainty in Section 6. Throughout, we illustrate the concepts with an example from run-to-run control of the photolithography process considered by Leang et al. (1996).

2. THE EFFECTS OF MODEL UNCERTAINTY ON MINIMUM VARIANCE CONTROL

First suppose that the parameters $\boldsymbol{\beta}$ and θ are known. The MV control law is the fixed control law (an expression for $\mathbf{X}_t$ as a function of all past data) that minimizes the closed-loop output variance about the target T. It is well known (Box, Jenkins, and Reinsel 1994; Ingolfsson and Sachs 1993; Sachs et al. 1995) that for first-order IMA disturbances, the MV control law takes the form of the so-called EWMA controller,

$$\mathbf{X}_t = \mathbf{X}_{t-1} - \frac{\boldsymbol{\beta}}{\boldsymbol{\beta}^T\boldsymbol{\beta}}(1-\theta)(Y_{t-1} - T), \tag{2}$$

in which the EWMA parameter is $1-\theta$. Although with multiple inputs the control law that minimizes output variance is not unique, (2) is attractive in that it is the solution that minimizes the norm of the input adjustment vector $\mathbf{X}_t - \mathbf{X}_{t-1}$ (Ingolfsson

and Sachs 1993). We note that Ingolfsson and Sachs (1993) and Sachs et al. (1995) focused on a different, but equivalent form of the control law that uses an EWMA to predict the disturbance and then sets the resulting prediction of the output equal to the target value.

Now suppose that we are using estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\theta}$ of the parameters to design the MV controller, where the estimates are obtained from some suitable off-line experiment. We do not need an estimate of β_0, because the offset parameter is largely redundant with a nonstationary IMA disturbance. (If d_t follows an IMA model, then $\beta_0 + d_t$ follows the same model.) Let B denote the time series backward-shift operator (Box et al. 1994), and define $\nabla \mathbf{X}_t = (1 - B)\mathbf{X}_t = \mathbf{X}_t - \mathbf{X}_{t-1}$ to be the amount the control inputs are adjusted between time steps $t - 1$ and t. The MV control law becomes

$$\nabla \mathbf{X}_t = -\mathbf{g}_{\mathrm{MV}}(Y_{t-1} - T), \tag{3}$$

where

$$\mathbf{g}_{\mathrm{MV}} = \frac{\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}^T \hat{\boldsymbol{\beta}}}(1 - \hat{\theta}) \tag{4}$$

is the MV controller gain vector.

Multiplying both sides of (1) by the difference operator $(1 - B)$ and substituting (3) for $\nabla \mathbf{X}_{t-1}$ gives $(1 - B)(Y_t - T) = -\boldsymbol{\beta}^T \mathbf{g}_{\mathrm{MV}}(Y_{t-1} - T) + (1 - \theta B)a_t$. We have used the relationship $(1 - B)c = 0$ for any constant c. Rearranging this gives

$$Y_t - T = \frac{1 - \theta B}{1 - \rho B} a_t, \tag{5}$$

where $\rho = 1 - \boldsymbol{\beta}^T \mathbf{g}_{\mathrm{MV}}$ is the root of the closed-loop characteristic equation. From this, we see that with parameter estimation errors, the deviation between the closed-loop output and the target behaves as a first-order autoregressive moving average (ARMA) process, the variance of which is (Box et al. 1994)

$$\sigma_y^2 = E[(Y_t - T)^2] = \sigma_a^2 \left[1 + \frac{(\rho - \theta)^2}{1 - \rho^2} \right],$$

which is valid for $|\rho| < 1$. For $|\rho| \geq 1$, the closed-loop system is unstable. The closed-loop variance is a function of the true and estimated parameters via the dependence of ρ on $\boldsymbol{\beta}, \hat{\boldsymbol{\beta}}$, and $\hat{\theta}$. Note that $\sigma_y^2 = \sigma_a^2$ when the parameters coincide with their estimates.

Throughout this article, we assume that standard Bayesian methods are used to obtain the posterior (given the off-line experimental data) distribution of $\{\boldsymbol{\beta}, \theta\}$ and that the asymptotic multivariate normal approximation to the posterior distribution is valid. Appendix A provides details on calculating the posterior distribution. We take $\hat{\boldsymbol{\beta}}$ and $\hat{\theta}$ to be the posterior mean, and let $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$, Σ_θ, and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}\theta}$ denote the posterior covariance of $\boldsymbol{\beta}$, the posterior variance of θ, and the posterior covariance between $\boldsymbol{\beta}$ and θ. To quantify the effects of parameter uncertainty, define the relative increase in closed-loop variance

$$\eta = \frac{\sigma_y^2 - \sigma_a^2}{\sigma_a^2} = \frac{(\rho - \theta)^2}{1 - \rho^2}, \tag{6}$$

and consider the distribution of η with respect to variations in $\{\boldsymbol{\beta}, \theta\}$ over their posterior distribution. In Appendix B we show that the probability density $f_\eta(\cdot)$ of η $(0 < \eta < \infty)$ can be calculated by numerically evaluating the integral

$$f_\eta(\eta) = \frac{1}{2} \int_{-1}^{1} \sqrt{\frac{1 - \rho^2}{\eta}} \times \left[f_{\theta|\rho}\left(\rho + \sqrt{\eta(1 - \rho^2)}\right) + f_{\theta|\rho}\left(\rho - \sqrt{\eta(1 - \rho^2)}\right) \right] f_\rho(\rho)\, d\rho, \tag{7}$$

where $f_\rho(\cdot)$ is the posterior density of ρ, and $f_{\theta|\rho}(\cdot)$ is the posterior density of θ given ρ. Both $f_\rho(\cdot)$ and $f_{\theta|\rho}(\cdot)$ are normal densities, and expressions for their means and variances are provided in Appendix B.

Example. To illustrate MV control and the use of (7) to evaluate the effects of parameter uncertainty, consider the photolithography process in semiconductor manufacturing described by Leang et al. (1996). The output variable Y (for the spin-coat and bake steps) to be controlled is the resist thickness. The three control variables are the spin speed (*SPS*), baking time (*BTI*), and baking temperature (*BTE*). Using experimental methods, Leang et al. (1996) fit the following model relating the process inputs to the output:

$$Y_t = 1{,}292 + 928{,}233 \frac{1}{\sqrt{SPS_t}} - 1.62 BTI_t - 19.5 BTE_t + d_t,$$

where Y, *SPS*, *BTI*, and *BTE* are in units of Angstroms, rpm, seconds, and degrees Celsius. The target is $T = 12{,}700$ Angstroms. As suggested by Leang et al. (1996) for scaling purposes, let $X_{1,t}, X_{2,t}$, and $X_{3,t}$ be coded values of $SPS_t^{-1/2}$, BTI_t, and BTE_t with $X_{1,t} = \pm 1$ corresponding to $SPS_t = \{4{,}800, 4{,}700\}$, $X_{2,t} = \pm 1$ corresponding to $BTI_t = \{55, 70\}$, and $X_{3,t} = \pm 1$ corresponding to $BTE_t = \{85, 95\}$. The estimated model becomes

$$Y_t = 12{,}904 + 70.9 X_{1,t} - 12.2 X_{2,t} - 97.5 X_{3,t} + d_t.$$

Although Leang et al. (1996) did not discuss the details of the experiment used to estimate the parameters, for illustrative purposes we assume that the experiment consisted of $n = 64$ runs over which the sinusoidal input profiles shown in Figure 3 were applied. As we discuss in Section 5, this was not a particularly

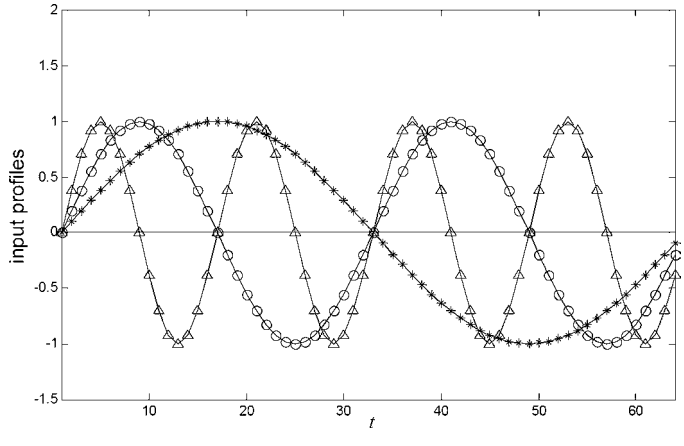

Figure 3. Input Profiles Over the Experiment Used to Estimate the Parameters (—✳*— $X_{1,t}$; —○— $X_{2,t}$; —△— $X_{3,t}$).*

effective experiment. It may appear to be a well-designed experiment on the surface, however, because the three input profiles are orthogonal and have equal average squared magnitude. Because no specific disturbance model was discussed by Leang et al. (1996) either, we also assume a first-order IMA disturbance model with $\hat{\theta} = .13$ and $\sigma_a = 60$ Angstroms. This model provided a reasonable fit to the data shown in figure 11a of Leang et al. (1996), which was generated with all inputs held constant (so that the output was the disturbance plus a constant).

From (3) and (4) with $\hat{\boldsymbol{\beta}} = [70.9 \ \ -12.2 \ \ -97.5]^T$, the controller gain and MV control law become $\mathbf{g}_{\text{MV}} = [.0042 \ \ -.0007 \ \ -.0058]^T$ and

$$\nabla \mathbf{X}_t = \begin{bmatrix} -.0042 \\ .0007 \\ .0058 \end{bmatrix} (Y_{t-1} - T). \tag{8}$$

Because the experiment was conducted in the open loop, we have [see App. A, (A.10)] $\boldsymbol{\Sigma}_{\boldsymbol{\beta}\theta} = \mathbf{0}$, $\Sigma_\theta = (1 - \hat{\theta}^2)/n = .0154$, and

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \sigma_a^2 \left[\sum_{t=1}^{n} \mathbf{U}_t \mathbf{U}_t^T \right]^{-1} = \begin{bmatrix} 9{,}256 & 200 & 99 \\ 200 & 2{,}330 & 49 \\ 99 & 49 & 599 \end{bmatrix}. \tag{9}$$

From Appendix B, the posterior density of θ given ρ is normal with mean $\mu_{\theta|\rho} = \hat{\theta} = .13$ and variance $\Sigma_{\theta|\rho} = \Sigma_\theta = .0154$, and the posterior density of ρ is normal with mean $\mu_\rho = 1 - \hat{\boldsymbol{\beta}}^T \mathbf{g}_{\text{MV}} = .13$ and variance $\Sigma_\rho = \mathbf{g}_{\text{MV}}^T \boldsymbol{\Sigma}_{\boldsymbol{\beta}} \mathbf{g}_{\text{MV}} = .179$. If we numerically integrate (7) with these densities substituted for $f_{\theta|\rho}(\cdot)$ and $f_\rho(\cdot)$, the resulting posterior probability density function (pdf) of η is shown in Figure 4. Also shown is the cumulative distribution function (cdf) of η, obtained by numerically integrating the pdf. The cdf of η is particularly revealing. There is a .19 probability that parameter uncertainty will inflate the closed-loop variance σ_y^2 by more than 50% ($\eta > .5$), relative to the minimum value σ_a^2 that would be achieved with no parameter uncertainty. There is a 0.1 probability that σ_y^2 is more than double σ_a^2 ($\eta > 1.0$) and a .05 probability that σ_y^2 is more than four times σ_a^2 ($\eta > 3$). The probability that the closed-loop system is unstable ($|\rho| > 1$) is .024, which follows from the N(.13, .179) distribution of ρ. In the following section we derive a control strategy that incorporates parameter covariance information and substantially reduces the risk of inflating the closed-loop variance.

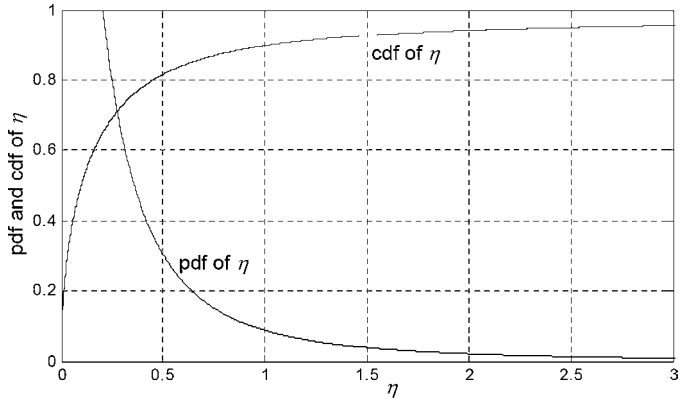

Figure 4. The pdf and cdf of the Relative Increase in Closed-Loop Variance for MV Control.

3. THE CAUTIOUS MINIMUM VARIANCE CONTROL LAW

Hamby et al. (1998) proposed one approach for taking into account probabilistic uncertainty in the input parameters $\boldsymbol{\beta}$. They considered an EWMA control law of the form $\nabla \mathbf{X}_t = -\mathbf{g}(Y_{t-1} - T)$ with controller gain vector $\mathbf{g} = \hat{\boldsymbol{\beta}}\omega$ for some scalar design parameter ω. Given a posterior distribution for $\boldsymbol{\beta}$, they selected ω to minimize the probability that the closed-loop system is unstable. In other words, ω was chosen to minimize the probability that the magnitude of the closed-loop root $\rho = 1 - \boldsymbol{\beta}^T \mathbf{g}$ is greater than 1.

In this article we consider a different strategy for treating probabilistic parameter uncertainty that relates more directly to minimizing closed-loop variance. Let $\mathbf{Y}^t = \{Y_t, Y_{t-1}, Y_{t-2}, \ldots, Y_1\}$ denote the set of all past "on-line" measurements up to current time t. We assume that $\mathbf{Y}^t$ is independent of the set of off-line experimental data used to estimate the model parameters. If there is no parameter uncertainty, the MV control strategy is equivalent to selecting the input adjustments $\nabla \mathbf{X}_t$ at each time t to minimize $E_{a_t}[(Y_t - T)^2 | \mathbf{Y}^{t-1}]$. The subscript a_t indicates the expectation is with respect to the distribution of the random shock at time t, which, given $\mathbf{Y}^{t-1}$, is the only random variable on which Y_t depends. In analogy with this for the present scenario in which the parameters are uncertain, we select $\nabla \mathbf{X}_t$ to minimize the CMV loss function

$$J_t \equiv E_{a_t, \gamma}[(Y_t - T)^2 | \mathbf{Y}^{t-1}], \tag{10}$$

where the expectation is with respect to the posterior distribution of the parameter vector $\boldsymbol{\gamma} = [\boldsymbol{\beta}^T, \theta]^T$, as well as the distribution of a_t.

Control strategies of this form are referred to in the adaptive control literature as *cautious adaptive control* (Åström and Wittenmark 1995). The *cautious* part of the term refers to the fact that when parameter uncertainty is large, (10) tends to result in more cautious (generally smaller) control input adjustments than if there were no uncertainty. The *adaptive* part refers to the fact that a strict use of (10) results in a control law that is adaptive in the following sense. Because the output is functionally dependent on $\boldsymbol{\gamma}$, the parameters are clearly not independent of the on-line data $\mathbf{Y}^{t-1}$. Conditioning the expectation in (10) on $\mathbf{Y}^{t-1}$ therefore has the effect of continuously adapting the posterior distribution of the parameters on-line, as each new observation is taken.

In addition to increased complexity of implementation, adaptive control laws have certain practical limitations, such as the bursting phenomenon (Anderson 1985; Bodson 1993; Radenkovic and Michel 1993), that we wish to avoid. Bursting means that periods of good control and estimation performance alternate with temporary bursts of divergent parameter estimates and poor control performance. The problem of bursting is greatly exacerbated when a cautious adaptive control law is used (Åström and Wittenmark 1995). The intuitive reason for this is that if parameter uncertainty begins to increase, then a cautious control law calls for smaller control adjustments. This in turn results in poorer excitation for the system and causes even larger parameter uncertainty. Eventually, the parameter uncertainty grows so large that the control law becomes completely ineffective or even temporarily unstable.

When this happens, there is a sudden burst of large output variability and large input adjustments. This, however, provides the excitation needed to improve the parameter estimates. When the parameter estimates improve, the control law becomes effective once again, and the cycle repeats indefinitely. This phenomenon has been discussed in more detail by Åström and Wittenmark (1995). To avoid an adaptive control law, we take the expectation in (10) with respect to the posterior distribution of $\boldsymbol{\gamma}$ given only the off-line data and not the on-line data $\mathbf{Y}^{t-1}$. In other words, we do not update the posterior distribution of $\boldsymbol{\gamma}$ on-line.

After subtracting T from both sides of (1) and multiplying by $(1-\theta B)^{-1}(1-B)$, we have $(1-\theta B)^{-1}(1-B)(Y_t-T) = \boldsymbol{\beta}^T(1-\theta B)^{-1}\nabla\mathbf{X}_t + a_t$. Using the identity $(1-\theta B)^{-1}(1-B) = 1-(1-\theta)(1-\theta B)^{-1}B$ gives

$$Y_t - T = (1-\theta)(1-\theta B)^{-1}(Y_{t-1}-T) + \boldsymbol{\beta}^T(1-\theta B)^{-1}\nabla\mathbf{X}_t + a_t. \quad (11)$$

The primary difficulty with substituting this into (10) and evaluating the result is that $Y_t - T$ depends on θ in a complex nonlinear manner. The proposed strategy is to use a first-order Taylor approximation of $Y_t - T$ about $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}$. In Appendix C we show that substituting the Taylor approximation into (10) yields the approximate CMV loss function

$$J_t \cong \left\{(1-\hat{\theta})(1-\hat{\theta}B)^{-1}(Y_{t-1}-T) + \hat{\boldsymbol{\beta}}^T\mathbf{U}_t\right\}^2 + \sigma_a^2 + \mathbf{U}_t^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\mathbf{U}_t + \Sigma_\theta w_{t-1}^2 - 2w_{t-1}\mathbf{U}_t^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}\theta}, \quad (12)$$

where $\mathbf{U}_t = (1-\hat{\theta}B)^{-1}\nabla\mathbf{X}_t$ and $w_t = (1-\hat{\theta}B)^{-1}\hat{a}_t$ are filtered versions of the control adjustments and model prediction errors (see App. A).

By inspecting (11), we can view the first term in (12) as the square of the one-step-ahead prediction of $Y_t - T$ using the estimated parameters and the second term as the loss due to the unpredictable component a_t of the disturbance. A similar expansion of the MV loss function $E_{a_t}[(Y_t-T)^2|\mathbf{Y}^{t-1}]$ with $\boldsymbol{\gamma}$ assumed equal to $\hat{\boldsymbol{\gamma}}$ would consist of only these two terms. The remaining three terms are functions of the parameter covariances and represent the loss due to parameter uncertainty. When the covariances are large, the loss due to parameter uncertainty may be large if the control adjustments are not chosen properly.

Setting the partial derivative (with respect to $\mathbf{U}_t$) of (12) equal to 0 and solving for $\mathbf{U}_t$ gives

$$\mathbf{U}_t = -\left[\boldsymbol{\Sigma}_{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^T\right]^{-1} \times \left\{(1-\hat{\theta})(1-\hat{\theta}B)^{-1}(Y_{t-1}-T)\hat{\boldsymbol{\beta}} - \boldsymbol{\Sigma}_{\boldsymbol{\beta}\theta}w_{t-1}\right\}$$

or

$$\nabla\mathbf{X}_t = (1-\hat{\theta}B)\mathbf{U}_t = -\left[\boldsymbol{\Sigma}_{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^T\right]^{-1}\left\{(1-\hat{\theta})(Y_{t-1}-T)\hat{\boldsymbol{\beta}} - \boldsymbol{\Sigma}_{\boldsymbol{\beta}\theta}\hat{a}_{t-1}\right\}.$$

Using the relationship $[\boldsymbol{\Sigma}_{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^T]^{-1}\hat{\boldsymbol{\beta}} = (1+\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\hat{\boldsymbol{\beta}})^{-1} \times \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\hat{\boldsymbol{\beta}}$, the CMV control law becomes

$$\nabla\mathbf{X}_t = -\mathbf{g}_{\mathrm{CMV}}(Y_{t-1}-T) + \left[\boldsymbol{\Sigma}_{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^T\right]^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\beta}\theta}\hat{a}_{t-1}, \quad (13)$$

where

$$\mathbf{g}_{\mathrm{CMV}} = \frac{\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\hat{\boldsymbol{\beta}}(1-\hat{\theta})}{1+\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\hat{\boldsymbol{\beta}}} \quad (14)$$

is the CMV controller gain vector.

For the common scenario where the parameters are estimated in the open loop, $\boldsymbol{\Sigma}_{\boldsymbol{\beta}\theta}$ is 0 (see App. A), and the CMV control law simplifies to

$$\nabla\mathbf{X}_t = -\mathbf{g}_{\mathrm{CMV}}(Y_{t-1}-T). \quad (15)$$

The CMV control law therefore takes the familiar form of an EWMA controller and is identical to the MV control law (3), except that the controller gain vector differs. The presence of $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}$ in $\mathbf{g}_{\mathrm{CMV}}$ accounts for parameter uncertainty by changing the magnitude and direction of the controller gain. Generally speaking, an increase in uncertainty reduces the magnitude of the controller gain and the input adjustments, resulting in a more "cautious" control law. In the next section we demonstrate that this always increases the probability of stability, relative to MV control, and appears to result in a more favorable distribution for the closed-loop variance.

The CMV control law (15) is pure integral control. Although the control law (13) for nonzero $\boldsymbol{\Sigma}_{\boldsymbol{\beta}\theta}$ is not, it does contain integral action. In general, integral action is a desirable property for a controller, because constant step disturbances (e.g., a step change in the process mean) are completely compensated asymptotically (Åström and Wittenmark 1990). Although we only explicitly consider the control of process variation, it is of some comfort to know that the controller would also be effective in compensating mean shifts.

4. PERFORMANCE ANALYSIS WHEN $\boldsymbol{\Sigma}_{\boldsymbol{\beta}\theta}$ IS $\mathbf{0}$

Consider the CMV control law (15) for the case where $\boldsymbol{\Sigma}_{\boldsymbol{\beta}\theta}$ is $\mathbf{0}$. Because this is of the same form as the MV control law (3) but with different controller gain, (6) is still a valid expression for the relative increase η in closed-loop variance. The only difference is that for CMV control, the closed-loop root is $\rho_{\mathrm{CMV}} = 1 - \boldsymbol{\beta}^T\mathbf{g}_{\mathrm{CMV}}$, whereas for MV control, the closed-loop root is $\rho_{\mathrm{MV}} = 1 - \boldsymbol{\beta}^T\mathbf{g}_{\mathrm{MV}}$. Substituting $\mathbf{g} = \mathbf{g}_{\mathrm{CMV}}$ from (14) into (B.1) and (B.2) in Appendix B, the posterior mean and variance of ρ_{CMV} are

$$\mu_{\rho,\mathrm{CMV}} = \frac{1+\hat{\theta}\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\hat{\boldsymbol{\beta}}}{1+\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\hat{\boldsymbol{\beta}}} \quad (16)$$

and

$$\Sigma_{\rho,\mathrm{CMV}} = (1-\hat{\theta})^2\frac{\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\hat{\boldsymbol{\beta}}}{(1+\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\hat{\boldsymbol{\beta}})^2}. \quad (17)$$

Similarly, with $\mathbf{g} = \mathbf{g}_{\mathrm{MV}}$ from (3), the posterior mean and variance of ρ_{MV} are

$$\mu_{\rho,\mathrm{MV}} = \hat{\theta} \quad (18)$$

and

$$\Sigma_{\rho,\mathrm{MV}} = (1-\hat{\theta})^2\frac{\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}}{(\hat{\boldsymbol{\beta}}^T\hat{\boldsymbol{\beta}})^2}. \quad (19)$$

Substituting the parameter estimates and covariances from the example of Section 2 into (16)–(19), ρ_{CMV} and ρ_{MV} follow N(.18, $.20^2$) and N(.13, $.42^2$) distributions. Figure 5 shows

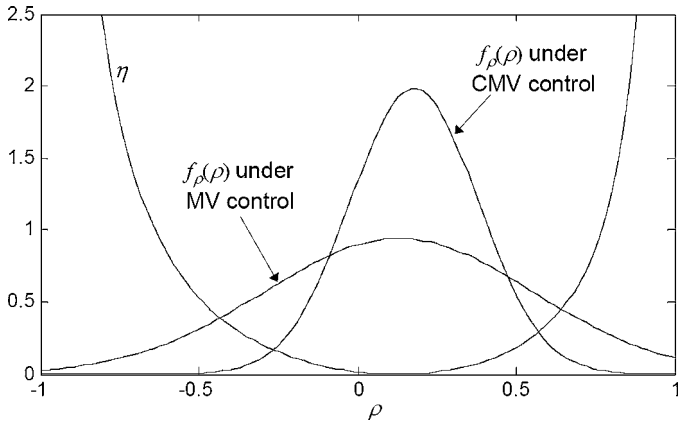

Figure 5. The Relative Increase in Closed-Loop Variance as a Function of ρ (for $\theta = \hat{\theta}$), Along With the Posterior Density of ρ for MV and CMV Control.

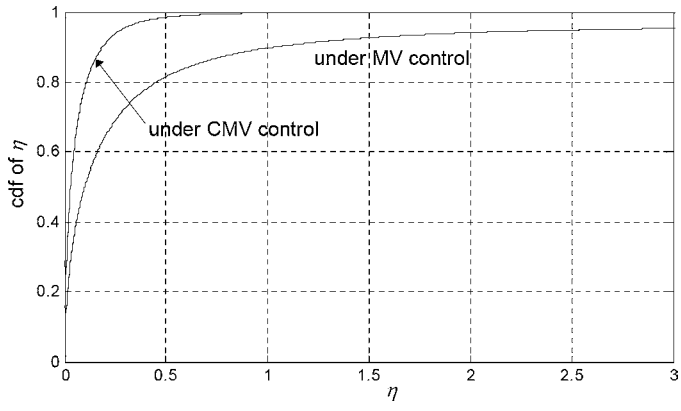

Figure 6. The cdf of the Relative Increase in Closed-Loop Variance for MV and CMV Control.

these posterior densities and also shows η from (6) as a function of ρ with θ fixed at $\hat{\theta} = .13$. Although the mean of ρ_{MV} is $\hat{\theta}$, which coincides with the value of ρ that minimizes η, the mean of ρ_{CMV} is slightly larger. This is more than compensated for by the fact that the standard deviation of ρ_{CMV} is roughly half that of ρ_{MV}. The end result is that there is a much higher probability that ρ_{MV} assumes a value for which η is large than does ρ_{CMV}. We show in Appendix D that for an arbitrary set of parameter estimates and covariances, the variance of ρ_{CMV} is always strictly less than the variance of ρ_{MV}.

Because the closed-loop system is stable if and only if $|\rho| < 1$, the probability of stability depends entirely on the distribution of ρ. We also show in Appendix D that the probability of stability is always higher for CMV control than for MV control, as long as the estimated disturbance model is invertible ($|\hat{\theta}| < 1$).

Equation (7) can also be used to calculate the posterior pdf and cdf of η for CMV control if we use (16) and (17) for the mean and variance of the (normal) density $f_\rho(\cdot)$. As we discuss in Appendix B, when $\boldsymbol{\Sigma}_{\boldsymbol{\beta}\theta}$ is $\mathbf{0}$ we have that $f_{\theta|\rho}(\cdot) = f_\theta(\cdot)$ is the normal density with mean $\hat{\theta}$ and variance Σ_θ. The example of Section 2 is continued next to illustrate the effects of incorporating parameter uncertainty information into the CMV control law.

Example (Continued). With $\hat{\boldsymbol{\beta}} = [70.9 \quad -12.2 \quad -97.5]^T$, $\hat{\theta} = .13$, and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ from (9), (14) results in the CMV controller gain $\mathbf{g}_{\text{CMV}} = [.0005 \quad -.0001 \quad -.0081]^T$. The CMV control law (15) then becomes

$$\nabla \mathbf{X}_t = \begin{bmatrix} -.0005 \\ .0001 \\ .0081 \end{bmatrix} (Y_{t-1} - T).$$

Because the third element of $\mathbf{g}_{\text{CMV}}$ is roughly 16 times larger than the first element, the CMV controller relies almost exclusively on the third input, $X_{3,t}$, to control output variability. This is intuitively reasonable, given that the effects of the third input are known with much higher certainty than the effects of the first input (i.e., the posterior variances of β_1 and β_3 are 9,256 and 599, resp.). In contrast, the first and third elements of $\mathbf{g}_{\text{MV}}$ are on the same order of magnitude, so that the MV controller of (8) attempts to use $X_{1,t}$ and $X_{3,t}$ equally in controlling output variability.

Figure 6 compares the cdf of η using MV control and CMV control. The cdf of η for MV control is the same as in Figure 4. For CMV control, we have substituted the N(.18, $.20^2$) density of ρ_{CMV} for $f_\rho(\cdot)$ in (7). Because $\boldsymbol{\Sigma}_{\boldsymbol{\beta}\theta}$ is $\mathbf{0}$, $f_{\theta|\rho}(\cdot) = f_\theta(\cdot)$ is the same for both MV and CMV control. From Figure 6, we see that the probability of parameter uncertainty severely inflating the closed-loop variance is much smaller for CMV control than for MV control. Whereas there was a .19 probability that $\eta > .5$ for MV control, this probability drops to .014 for CMV control. Likewise, the probability that $\eta > 1.0$ is .10 for MV and .0025 for CMV control.

The posterior cdf of η provides a means of comparing MV and CMV control from a Bayesian perspective in which the estimate $\hat{\boldsymbol{\gamma}}$ is fixed and variations in $\boldsymbol{\gamma}$ over its posterior distribution are considered. This is similar to what we used to derive the CMV control law. We also may be interested in a performance comparison from a frequentist perspective, in which a single fixed (but unknown) value for $\boldsymbol{\gamma}$ is assumed and variations in the estimate $\hat{\boldsymbol{\gamma}}$ are considered. Different values of $\hat{\boldsymbol{\gamma}}$ will result in different controller gain vectors and different closed-loop variances.

In a continuation of the previous example, we used the following Monte Carlo simulation to compare the performance of MV and CMV control from a frequentist perspective. For simulation purposes, we treated the estimates described previously as the true parameters $\boldsymbol{\beta}$ and θ. For each Monte Carlo replicate, we generated $n = 64$ random shocks from the N(0, σ_a^2) distribution with $\sigma_a = 60$. Using the input profiles shown in Figure 3, we then generated the 64 output observations representing the experimental data for that replicate via (1). We estimated the parameters $\boldsymbol{\beta}$, θ, and σ_a^2 and the parameter covariance $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ as described in Appendix A [with the MATLAB function PEM used to estimate the parameters, and $\hat{\sigma}_a^2$ used in place of σ_a^2 in (A.10) to estimate $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$]. We then calculated the controller gain vectors $\mathbf{g}_{\text{MV}}$ and $\mathbf{g}_{\text{CMV}}$ via (4) and (14). Finally, we substituted the closed-loop roots $\rho_{\text{MV}} = 1 - \boldsymbol{\beta}^T \mathbf{g}_{\text{MV}}$ and $\rho_{\text{CMV}} = 1 - \boldsymbol{\beta}^T \mathbf{g}_{\text{CMV}}$ into (6) to calculate η for MV control and CMV control. We repeated the entire procedure for a total of 10,000 Monte Carlo replicates; the empirical cdf's of η for MV and CMV control are shown in Figure 7. The results are quite similar to the Bayesian cdf's of η shown in Figure 6. Although derived from a Bayesian perspective, the performance improvement achieved with CMV control also extends to the frequentist scenario for this example.

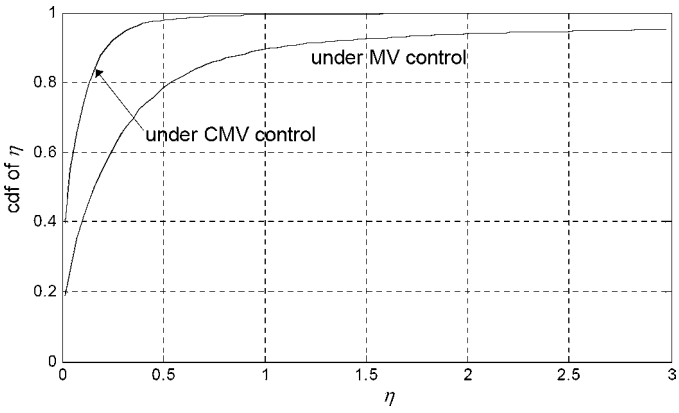

Figure 7. Empirical cdf of the Relative Increase in Closed-Loop Variance for MV and CMV Control in the Frequentist Scenario.

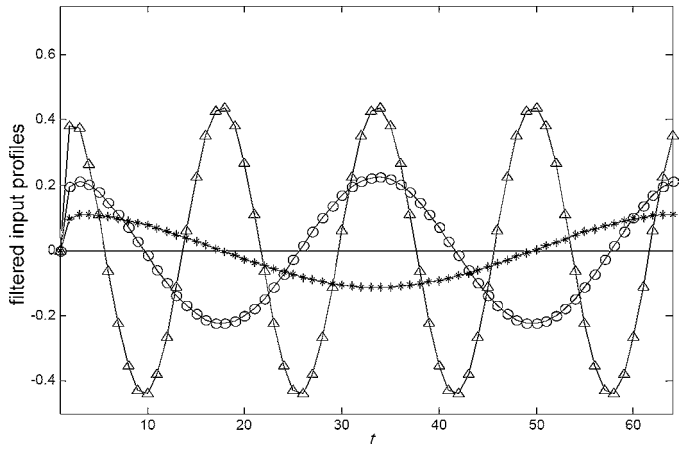

Figure 8. Filtered Input Profiles Corresponding to the Experimental Inputs in Figure 3 (— $U_{1,t}$; —o— $U_{2,t}$; —Δ— $U_{3,t}$).

5. EXPERIMENTAL DESIGN CONSIDERATIONS

Designing an effective experiment for estimating the model parameters is somewhat more involved than for the standard linear regression problem in which the disturbance is assumed to be iid. When a designed experiment is used to estimate the parameters in the open loop, the parameter covariances are (see App. A) $\Sigma_\theta = (1-\hat{\theta}^2)/n$ and $\mathbf{\Sigma}_{\boldsymbol{\beta}} = \sigma_a^2[\sum_{t=1}^n \mathbf{U}_t\mathbf{U}_t^T]^{-1}$, where $\mathbf{U}_t = (1-\hat{\theta}B)^{-1}(1-B)\mathbf{X}_t$ is a filtered version of the inputs over the experiment. Other than via sample size, the experimental design will not affect Σ_θ . In contrast, the design has a great effect on $\mathbf{\Sigma}_{\boldsymbol{\beta}}$. Ideally, we would design the experiment so that $\sum_{t=1}^n \mathbf{U}_t\mathbf{U}_t^T$ is a well-conditioned matrix. This cannot be accomplished without some prior knowledge of the parameters, however, because $\mathbf{U}_t$ depends on $\hat{\theta}$. It is not sufficient to select the input profiles so that $\sum_{t=1}^n \mathbf{X}_t\mathbf{X}_t^T$ is well conditioned, as is typically done in standard experimental designs.

To illustrate this, consider that the input profiles in Figure 3 are orthogonal with equal average squared magnitude. The matrix $\sum_{t=1}^n \mathbf{X}_t\mathbf{X}_t^T$ is therefore a scalar multiple of the identity matrix, and the experiment may appear to have been well designed. The corresponding profiles for $\mathbf{U}_t$ shown in Figure 8 indicate that the design is not ideal, however. After the inputs are filtered by $(1-\hat{\theta}B)^{-1}(1-B)$, $U_{1,t}$ has much smaller average squared magnitude than $U_{3,t}$. The result is that there is much larger uncertainty in β_1 than in β_3, as can be seen from (9).

In light of this, it may be desirable to conduct the experiment in two phases. In the first phase, hold $\mathbf{X}_t$ constant and collect sufficient data to obtain a preliminary estimate of θ. Based on the preliminary estimate, one may then evaluate whether a particular set of input profiles will result in a well-conditioned $\sum_{t=1}^n \mathbf{U}_t\mathbf{U}_t^T$. If not, then the input profiles can be modified accordingly. In the second phase, $\mathbf{X}_t$ would be varied according to the chosen input profiles. Data from both phases would then be combined to obtain the final estimates of all parameters together.

When designing the experiment, one must also consider the trade-off in using large variations in the inputs. Large input variations generally will result in smaller uncertainty in $\boldsymbol{\beta}$ and hence lower output variance after the control law is implemented. On the other hand, excessively large input variations may result in unacceptably high output variance during the experiment. Ideally, we would like to use the smallest input variations that still result in acceptable levels of parameter uncertainty. One strategy would be to start with an initial experiment that uses relatively small input variations. After estimating the parameters and their covariances, we could use the methods described earlier to calculate the posterior cdf of the relative increase η in output variance and construct a plot similar to that in Figure 6. If the plot indicates an unacceptably large risk of parameter uncertainty inflating the closed-loop variance, then a follow-up experiment can be designed to improve the parameter estimates and reduce uncertainty.

Calculating the posterior cdf of η requires the numerical integration in (7). A simpler alternative for approximating the cdf of η is to ignore uncertainty in θ by assuming that $\theta = \hat{\theta}$ and only consider uncertainty in ρ. For a fixed value of η (say $\eta = c$) and for $\theta = \hat{\theta}$, (6) has two solutions for ρ:

$$\rho = \frac{\hat{\theta} \pm \sqrt{c(1+c-\hat{\theta}^2)}}{1+c}.$$

The cdf of η becomes

$$\Pr[\eta \le c] = \Pr\left[\frac{\hat{\theta} - \sqrt{c(1+c-\hat{\theta}^2)}}{1+c} \le \rho \le \frac{\hat{\theta} + \sqrt{c(1+c-\hat{\theta}^2)}}{1+c}\right],$$

where ρ follows a normal distribution with mean and variance given by (16) and (17) for CMV control.

6. MODEL STRUCTURE UNCERTAINTY

We have considered only parametric model uncertainty in developing the CMV control law. In practice, model structure uncertainty is equally important. Forms of structural uncertainty include unmodeled quadratic and interaction effects in the input/output relationship, inputs that are included in the model but that have no effect on the output, and disturbances that do not follow a first-order IMA model. A potential approach to treating model structure uncertainty is to consider a set of models of various structure and use Bayesian modeling to identify the posterior probability that each of the model structures holds, as well as to identify the posterior distribution for the parameters. Box and Meyer (1993) and Chipman (1998) discussed

such Bayesian techniques for the pure linear regression model. If $\{\pi_i : i = 1, 2, \ldots, M\}$ denotes the posterior probabilities for each model in a set of M candidate models, then we may consider using a CMV loss function of the form $\sum_{i=1}^{M} \pi_i J_{i,t}$, where each $J_{i,t}$ is analogous to (10) for the ith model structure. If nonlinearities were included in some of the models, then we would no longer have a closed-form solution for the CMV control inputs. Rather, we would have to resort to numerical methods for minimizing the CMV loss function, and the resulting control law would be nonlinear. This approach is currently under investigation. In the remainder of this section we illustrate the effects of certain model structure discrepancies in a continuation of the example of Sections 2 and 4.

First, suppose that we assume an IMA disturbance model, but that the disturbance actually follows a first-order autoregressive (AR) model, $d_t = .8d_{t-1} + a_t$. Because of the complexity of a Bayesian analysis when the actual and assumed model structures differ, we used a frequentist analysis and Monte Carlo simulation to evaluate the effects of model structure uncertainty. As in the frequentist analysis of Section 4, we assumed that the true parameters are $\boldsymbol{\beta} = [70.9 \ -12.2 \ -97.5]^T$ and $\sigma_a = 60$. For each Monte Carlo replicate, we generated $n = 64$ random shocks from the $N(0, 60^2)$ distribution. Using the input profiles shown in Figure 3, we then generated the 64 output observations representing the experimental data for that replicate via (1) with a first-order AR disturbance. With an incorrectly assumed IMA disturbance, we estimated the parameters $\boldsymbol{\beta}, \theta$, and σ_a^2 and the parameter covariance $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ as described earlier. We then calculated the controller gain vectors $\mathbf{g}_{\text{MV}}$ and $\mathbf{g}_{\text{CMV}}$ via (4) and (14). Because (6) is no longer valid when the disturbance follows an AR model, we used a different method to calculate η. A repetition of the results leading to (5) would show that the closed-loop output now follows the ARMA(2, 1) model

$$Y_t - T = \frac{1 - B}{(1 - \rho B)(1 - \phi B)} a_t,$$

where ρ is as before (i.e., $1 - \boldsymbol{\beta}^T \mathbf{g}_{\text{MV}}$ for MV control and $1 - \boldsymbol{\beta}^T \mathbf{g}_{\text{CMV}}$ for CMV control). We used the impulse response method (Box et al. 1994) to calculate the variance of the ARMA(2, 1) process $Y_t - T$ and then the performance measure $\eta = (\sigma_y^2 - \sigma_a^2)/\sigma_a^2$. Figure 9 shows the empirical cdf's of η for MV control and CMV control from 10,000 Monte Carlo replicates.

Comparing Figures 9 and 7, we see that the performance of CMV control is not substantially affected by the mismatch in the disturbance model structure. Although the minimum value of η is slightly larger than σ_a^2, the upper percentiles of the cdf of η are quite similar to those shown in Figure 7. Somewhat surprisingly, the performance of MV control actually improves when the disturbance is AR but we assume an IMA model. An intuitive explanation is that for MV control with an assumed IMA disturbance, we are really using an EWMA to predict the output one-step-ahead. In general, EWMA predictors are quite robust, albeit slightly less than optimal. Box and Luceño (1997, pp. 140–142) discussed in detail the robustness of MV control to errors in the IMA parameter. Box and Luceño (1997, pp. 122–127) also demonstrated considerable robustness with respect to structural errors in the disturbance model. They showed that if the disturbance actually follows Barnard's model (i.e., white noise added to a signal that experiences random sized jumps at random times) but an IMA model is assumed, then the resulting one-step-ahead output prediction will be nearly optimal.

We have also investigated the effects of assuming that all inputs are active (i.e., they are included in the model) when in fact some have no effect on the output. We repeated the frequentist example of Section 4 two more times. The first time, we set $\boldsymbol{\beta} = [0 \ -12.2 \ -97.5]^T$; in other words, the first input no longer has an effect on the output, but it is still included in the model. The second time, we set $\boldsymbol{\beta} = [70.9 \ -12.2 \ 0]^T$, so that the third input has no effect on the output. The empirical cdf's of η for these two cases are shown in Figures 10 and 11. Comparing Figures 10 and 7, we see that when the first input has no effect, the performance of MV control degrades. In contrast, the performance of CMV control remains largely unchanged. This is not surprising, given that CMV control made very little use of the first input because of the large uncertainty in β_1. In a sense, CMV control has an inherent level of robustness to the type of model structure uncertainty that results from overmodeling the number of active inputs. If the true effect of an input (say $X_{i,t}$) is 0, then the covariance terms in $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ associated with β_i generally will be large relative to the estimate $\hat{\beta}_i$. In this case, by

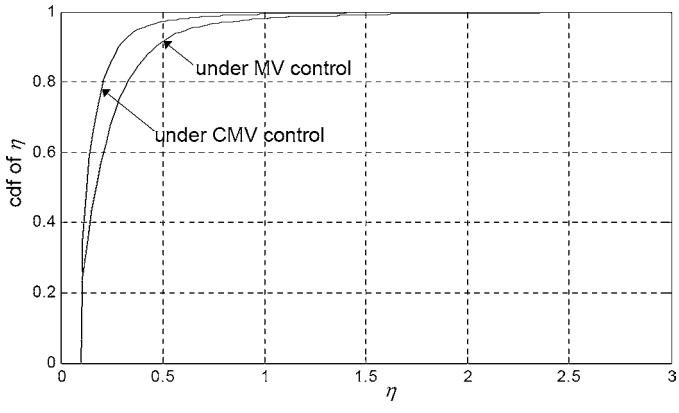

Figure 9. Empirical cdf of the Relative Increase in Closed-Loop Variance for MV and CMV Control With a Mismatch in the Disturbance Model Structure.

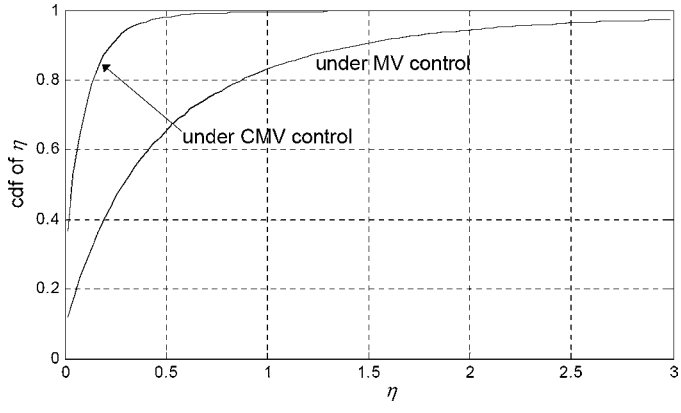

Figure 10. Empirical cdf of the Relative Increase in Closed-Loop Variance for MV and CMV Control When the First Input Has No Effect but Is Still Included in the Model.

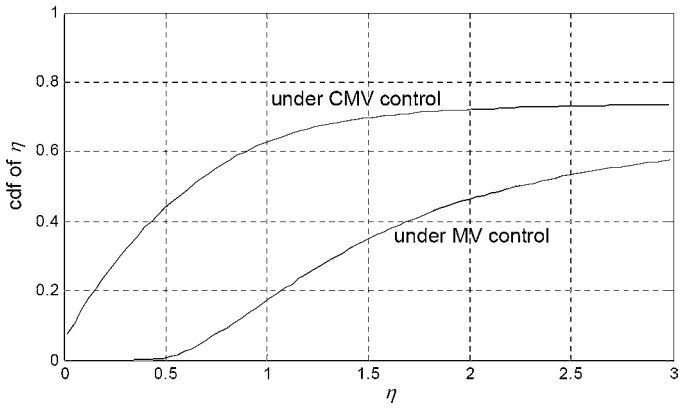

Figure 11. Empirical cdf of the Relative Increase in Closed-Loop Variance for MV and CMV Control When the Third Input Has No Effect but Is Still Included in the Model.

comparison of (4) and (14), CMV control would rely less heavily on $X_{i,t}$ to control the output than would MV control.

The situation is quite different for the case in which the third input has no effect on the output (Fig. 11). Although the performance of CMV control is still much better than MV control, neither is very good. The reason for this is that with $\boldsymbol{\beta} = [70.9 \ -12.2 \ 0]^T$, $X_{1,t}$ is the only input that has more than a negligible effect on the output. The uncertainty in its effect β_1 is so large, however, that $X_{1,t}$ cannot be used effectively to control the output. MV control performed worse than CMV control, because it did not take the uncertainty into account and still attempted to rely on $X_{1,t}$ to control the output. If after estimating the parameters, we had constructed a plot analogous to that in Figure 6 to assess the impact of uncertainty, then we most likely would have concluded that additional experimentation was necessary before implementing any control strategy.

Another form of structural uncertainty that may be of interest is when the effects $\boldsymbol{\beta}$ of the inputs vary randomly over time, as opposed to being constants. This situation may arise if we have additional unobservable random variables (say, $\mathbf{W}_t = [W_{1,t} \ W_{2,t} \ \cdots \ W_{m,t}]^T$) that interact with the inputs, as is often assumed in robust design studies (Wu and Hamada 2000). For example, suppose that the output behaves according to the model $Y_t = \beta_0 + \boldsymbol{\alpha}^T\mathbf{X}_t + \mathbf{W}_t^T\boldsymbol{\Gamma}\mathbf{X}_t + d_t$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\Gamma}$ are a $p \times 1$ vector and an $m \times p$ matrix of constant parameters and $\mathbf{W}_t$ is random with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{w}}$. We may represent this as $Y_t = \beta_0 + \boldsymbol{\beta}_t^T\mathbf{X}_t + d_t$, where $\boldsymbol{\beta}_t = \boldsymbol{\alpha} + \boldsymbol{\Gamma}^T\mathbf{W}_t$ varies randomly over time with mean $\boldsymbol{\mu}_{\boldsymbol{\beta}} = \boldsymbol{\alpha}$ and covariance $\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \boldsymbol{\Gamma}^T\boldsymbol{\Sigma}_{\mathbf{w}}\boldsymbol{\Gamma}$. Although this form of model uncertainty differs from the Bayesian parameter uncertainty due to estimation errors considered throughout this article, we may treat it in much the same manner if we substitute the foregoing mean and covariance of $\boldsymbol{\beta}_t$ in place of $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ in (14).

7. CONCLUSIONS

In this article we have investigated a cautious minimum variance approach to controlling variability in industrial processes with a linear static input/output relationship and a first-order IMA disturbance. The form of the CMV control law (for open-loop experiments) is identical to the standard MV EWMA controller, except that the controller gain vector differs. The CMV controller is therefore relatively straightforward to implement and requires only an estimate of the model parameters and their posterior covariance matrix. We have provided guidelines for designing the experiment used to estimate the parameters and evaluating the impact of parameter uncertainty on closed-loop variance, to determine whether additional experimentation is necessary. We have shown that CMV control always results in a higher probability of stability than does MV control if the assumed model structure holds. Simulation results for controlling the resist thickness in a photolithography process indicate that in situations where parameter uncertainty is large, CMV control may be substantially more robust to parameter errors than MV control.

We have treated the issue of model structure uncertainty only briefly. It appears that CMV control has an inherent robustness to certain types of structure uncertainty, such as a mismatch in the disturbance model structure or overmodeling the number of active inputs. Treatment of other types of uncertainty, such as unmodeled quadratic or interaction effects, would be more complex and would require nonlinear control methods. This is an important issue that deserves further attention.

ACKNOWLEDGMENTS

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.

APPENDIX A: THE POSTERIOR MEAN AND COVARIANCE OF THE PARAMETERS

Multiplying both sides of (1) by $(1 - B)$, the model becomes

$$Y_t = Y_{t-1} + \boldsymbol{\beta}^T\nabla\mathbf{X}_t + a_t - \theta a_{t-1}. \tag{A.1}$$

Suppose that an off-line experiment is designed, the data from which are to be used to estimate the parameters. Denote the output observations over the experiment by $\mathbf{Y} = [Y_1 \ Y_2 \ \cdots \ Y_n]^T$ and the input adjustment settings by $\mathbf{X} = [\nabla\mathbf{X}_1 \ \nabla\mathbf{X}_2 \ \cdots \ \nabla\mathbf{X}_n]^T$, where n is the sample size.

We use a Bayesian approach, in which the parameter vector $\boldsymbol{\gamma} = [\boldsymbol{\beta}^T\theta]^T$ is assumed to be random with some specified prior distribution $\pi(\boldsymbol{\gamma})$. Let $\pi(\boldsymbol{\gamma}|\mathbf{Y})$ denote the posterior distribution of $\boldsymbol{\gamma}$ given $\mathbf{Y}$, with mean $\hat{\boldsymbol{\gamma}}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$. Likewise, let $\pi(\mathbf{Y}|\boldsymbol{\gamma})$ denote the distribution of $\mathbf{Y}$ given $\boldsymbol{\gamma}$. We use the result (Carlin and Louis 1998) that for large n, the posterior distribution $\pi(\boldsymbol{\gamma}|\mathbf{Y})$ is approximately Gaussian with mean $\hat{\boldsymbol{\gamma}}$ equal to the mode of the posterior distribution and covariance

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} &= \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\beta}} & \boldsymbol{\Sigma}_{\boldsymbol{\beta}\theta} \\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}\theta}^T & \Sigma_\theta \end{bmatrix} \\ &= -\left[\frac{\partial^2 \log(\pi(\mathbf{Y}|\boldsymbol{\gamma})\pi(\boldsymbol{\gamma}))}{\partial\boldsymbol{\gamma}^2}\bigg|_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}}\right]^{-1},\end{aligned} \tag{A.2}$$

where the term in brackets denotes the Hessian (second derivative matrix) of $\log(\pi(\mathbf{Y}|\boldsymbol{\gamma})\pi(\boldsymbol{\gamma}))$ evaluated at $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}$. Note that the mode $\hat{\boldsymbol{\gamma}}$ of the posterior distribution is the $\boldsymbol{\gamma}$ that maximizes $\log(\pi(\boldsymbol{\gamma}|\mathbf{Y})) = \text{constant} + \log(\pi(\mathbf{Y}|\boldsymbol{\gamma})\pi(\boldsymbol{\gamma})) = \text{constant} + \log(\pi(\mathbf{Y}|\boldsymbol{\gamma})) + \log(\pi(\boldsymbol{\gamma}))$.

For the time being, assume that $\pi(\boldsymbol{\gamma})$ is noninformative (i.e., flat). In this case $\hat{\boldsymbol{\gamma}}$ maximizes $\log(\pi(\mathbf{Y}|\boldsymbol{\gamma}))$ and coincides with the maximum likelihood estimate (MLE) of $\boldsymbol{\gamma}$. Assuming that the a_t's are Gaussian, the approximate MLE of $\boldsymbol{\gamma}$ minimizes

$$-\log\big(\pi(\mathbf{Y}|\boldsymbol{\gamma})\big) = \text{constant} + (2\sigma_a^2)^{-1}\sum_{t=1}^{n}\hat{a}_t^2(\boldsymbol{\gamma},\mathbf{Y}), \quad \text{(A.3)}$$

where $\hat{a}_t(\boldsymbol{\gamma},\mathbf{Y})$ is the prediction error for the tth observation using a specified $\boldsymbol{\gamma}$. From (A.1), the prediction error can be obtained recursively via

$$\hat{a}_t(\boldsymbol{\gamma},\mathbf{Y}) = (1-\theta B)^{-1}[Y_t - Y_{t-1} - \boldsymbol{\beta}^T\nabla\mathbf{X}_t]. \quad \text{(A.4)}$$

Box et al. (1994) and Ljung (1987) discussed the approximate MLE of $\boldsymbol{\gamma}$ in detail and refered to it as nonlinear least squares and the prediction error method, respectively.

When $\pi(\boldsymbol{\gamma})$ is noninformative, the Hessian in (A.2) reduces to the Hessian of $\log(\pi(\mathbf{Y}|\boldsymbol{\gamma}))$, where $-\log(\pi(\mathbf{Y}|\boldsymbol{\gamma}))$ is given by (A.3). Following the arguments of Ljung (1987, chap. 9) for evaluating the Hessian and then substituting the result into (A.2), it can be shown that

$$\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = \frac{\sigma_a^2}{n}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varphi}}^{-1}, \quad \text{(A.5)}$$

where

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varphi}} = \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\varphi}_t\boldsymbol{\varphi}_t^T \quad \text{(A.6)}$$

is the sample covariance matrix of the $(p+1)$-length vector

$$\boldsymbol{\varphi}_t = \begin{bmatrix}\mathbf{U}_t\\ -w_{t-1}\end{bmatrix}, \quad \text{(A.7)}$$

with

$$\mathbf{U}_t = (1-\hat{\theta}B)^{-1}\nabla\mathbf{X}_t \quad \text{(A.8)}$$

and

$$w_t = (1-\hat{\theta}B)^{-1}\hat{a}_t. \quad \text{(A.9)}$$

Equations (A.8) and (A.9) represent filtered versions of the control adjustment vector and the prediction error,

$$\hat{a}_t = (1-\hat{\theta}B)^{-1}[Y_t - Y_{t-1} - \hat{\boldsymbol{\beta}}^T\nabla\mathbf{X}_t],$$

which is (A.4) with $\boldsymbol{\gamma}$ replaced by $\hat{\boldsymbol{\gamma}}$.

To summarize, when the prior distribution $\pi(\boldsymbol{\gamma})$ is noninformative, the posterior distribution $\pi(\boldsymbol{\gamma}|\mathbf{Y})$ is approximately multivariate normal with mean equal to the MLE $\hat{\boldsymbol{\gamma}}$ and covariance matrix given by (A.5). Suppose instead that a more informative prior distribution for $\boldsymbol{\gamma}$ is assumed. In particular, suppose that $\pi(\boldsymbol{\gamma})$ is multivariate normal with some specified prior mean $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$. In this case the posterior mean/mode $\hat{\boldsymbol{\gamma}}$ minimizes

$$\begin{aligned}-\log\big(\pi(\mathbf{Y}|\boldsymbol{\gamma})\pi(\boldsymbol{\gamma})\big) &= -\log\big(\pi(\mathbf{Y}|\boldsymbol{\gamma})\big) - \log(\pi(\boldsymbol{\gamma}))\\ &= \text{constant} + (2\sigma_a^2)^{-1}\sum_{t=1}^{n}\hat{a}_t^2(\boldsymbol{\gamma},\mathbf{Y})\\ &\quad + 2^{-1}[\boldsymbol{\gamma}-\boldsymbol{\mu}_0]^T\boldsymbol{\Sigma}_0^{-1}[\boldsymbol{\gamma}-\boldsymbol{\mu}_0].\end{aligned}$$

Because the Hessian of $-\log(\pi(\mathbf{Y}|\boldsymbol{\gamma})\pi(\boldsymbol{\gamma}))$ is the Hessian of $-\log(\pi(\mathbf{Y}|\boldsymbol{\gamma}))$ plus $\boldsymbol{\Sigma}_0^{-1}$, similar to (A.5), it follows that the posterior covariance is

$$\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = [n\sigma_a^{-2}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varphi}} + \boldsymbol{\Sigma}_0^{-1}]^{-1},$$

with $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varphi}}$ as in (A.6).

One common special case of particular interest is when the off-line experiment for estimating $\boldsymbol{\gamma}$ is conducted in the open loop (i.e., without feedback control) with the input adjustments determined from some designed experiment. In this situation, the $\mathbf{U}_t$ and w_{t-1} components of $\boldsymbol{\varphi}_t$ in (A.7) will generally be uncorrelated. In this case, (A.6) becomes

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varphi}} &= \frac{1}{n}\sum_{t=1}^{n}\begin{bmatrix}\mathbf{U}_t\mathbf{U}_t^T & -\mathbf{U}_t w_{t-1}\\ -\mathbf{U}_t^T w_{t-1} & w_{t-1}^2\end{bmatrix}\\ &\cong \begin{bmatrix}\frac{1}{n}\sum_{t=1}^{n}\mathbf{U}_t\mathbf{U}_t^T & \mathbf{0}\\ \mathbf{0} & \sigma_a^2/(1-\hat{\theta}^2)\end{bmatrix}.\end{aligned}$$

Taking the inverse of this matrix, we see that with a noninformative prior, the parameter covariances become $\boldsymbol{\Sigma}_{\boldsymbol{\beta}\theta} = \mathbf{0}$, $\Sigma_\theta = (1-\hat{\theta}^2)/n$, and

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \sigma_a^2\left[\sum_{t=1}^{n}\mathbf{U}_t\mathbf{U}_t^T\right]^{-1}. \quad \text{(A.10)}$$

APPENDIX B: DERIVATION OF THE POSTERIOR DENSITY OF η

For a specified $0 < \eta < \infty$ and $|\rho| < 1$, (6) has two solutions for θ,

$$\theta = \rho \pm \sqrt{\eta(1-\rho^2)}.$$

Because the abslolute value of the derivitive of (6) with respect to θ is

$$\left|\frac{d\eta}{d\theta}\right| = \frac{2|\rho-\theta|}{1-\rho^2} = 2\sqrt{\frac{\eta}{1-\rho^2}},$$

the conditional density of η given ρ is

$$\begin{aligned}f_{\eta|\rho}(\eta) = \frac{1}{2}\sqrt{\frac{1-\rho^2}{\eta}}\big[&f_{\theta|\rho}\big(\rho+\sqrt{\eta(1-\rho^2)}\big)\\ &+ f_{\theta|\rho}\big(\rho-\sqrt{\eta(1-\rho^2)}\big)\big].\end{aligned}$$

Integrating with respect to the posterior density of ρ over the range $|\rho| < 1$ gives (7) as the posterior density of η.

The densities $f_\rho(\cdot)$ and $f_{\theta|\rho}(\cdot)$ in (7) can be obtained using the multivariate normal approximation to the posterior distribution of $\{\boldsymbol{\beta},\theta\}$ discussed in Appendix A. Recall that $\rho = 1-\boldsymbol{\beta}^T\mathbf{g}$, where $\mathbf{g}$ denotes the controller gain vector [either (4) for MV control or (14) for CMV control]. It follows that $f_\rho(\cdot)$ is the normal density with mean

$$\mu_\rho = 1 - \hat{\boldsymbol{\beta}}^T\mathbf{g} \quad \text{(B.1)}$$

and variance

$$\Sigma_\rho = \mathbf{g}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\mathbf{g}. \quad \text{(B.2)}$$

Moreover, the posterior distribution of θ is normal with mean $\mu_\theta = \hat{\theta}$ and variance Σ_θ, and the covariance between θ and ρ

is $\Sigma_{\rho\theta} = E[(\rho - \mu_\rho)(\theta - \mu_\theta)] = E[-\mathbf{g}^T(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}^T)(\theta - \mu_\theta)] = -\mathbf{g}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}\theta}$. It follows that $f_{\theta|\rho}(\cdot)$ is the normal density with mean

$$\begin{aligned}\mu_{\theta|\rho} &= \mu_\theta + \Sigma_\rho^{-1}\Sigma_{\rho\theta}(\rho - \mu_\rho) \\ &= \hat{\theta} - \frac{\mathbf{g}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}\theta}}{\mathbf{g}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\mathbf{g}}(\rho - 1 + \hat{\boldsymbol{\beta}}^T\mathbf{g}) \qquad \text{(B.3)}\end{aligned}$$

and variance

$$\Sigma_{\theta|\rho} = \Sigma_\theta - \Sigma_\rho^{-1}\Sigma_{\rho\theta}^2 = \Sigma_\theta - \frac{(\mathbf{g}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}\theta})^2}{\mathbf{g}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\mathbf{g}}. \qquad \text{(B.4)}$$

Equations (B.1)–(B.4) can be used in (7) to calculate the posterior density of η via numerical integration. When $\boldsymbol{\Sigma}_{\boldsymbol{\beta}\theta}$ is $\mathbf{0}$ (e.g., with an open-loop experiment), (B.3) and (B.4) simplify to $\mu_{\theta|\rho} = \hat{\theta}$ and $\Sigma_{\theta|\rho} = \Sigma_\theta$.

APPENDIX C: DERIVATION OF (12) FOR THE CMV LOSS FUNCTION

The partial derivatives of (11) with respect to $\boldsymbol{\beta}$ and θ are

$$\left.\frac{\partial(Y_t - T)}{\partial\boldsymbol{\beta}}\right|_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = (1 - \hat{\theta}B)^{-1}\nabla\mathbf{X}_t = \mathbf{U}_t$$

and

$$\begin{aligned}\left.\frac{\partial(Y_t - T)}{\partial\theta}\right|_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} &= -(1 - \hat{\theta}B)^{-1}(Y_{t-1} - T) \\ &\quad + (1 - \hat{\theta})(1 - \hat{\theta}B)^{-2}B(Y_{t-1} - T) \\ &\quad + \hat{\boldsymbol{\beta}}^T(1 - \hat{\theta}B)^{-2}B\nabla\mathbf{X}_t \\ &= -(1 - \hat{\theta}B)^{-2}(1 - B)(Y_{t-1} - T) \\ &\quad + \hat{\boldsymbol{\beta}}^T(1 - \hat{\theta}B)^{-2}\nabla\mathbf{X}_{t-1} \\ &= -(1 - \hat{\theta}B)^{-1} \\ &\quad \times \left\{(1 - \hat{\theta}B)^{-1}[Y_{t-1} - Y_{t-2} - \hat{\boldsymbol{\beta}}^T\nabla\mathbf{X}_{t-1}]\right\} \\ &= -(1 - \hat{\theta}B)^{-1}\hat{a}_{t-1} \\ &= -w_{t-1}.\end{aligned}$$

Thus the first-order Taylor approximation of $Y_t - T$ about $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}$ is

$$\begin{aligned}Y_t - T &\cong (1 - \hat{\theta})(1 - \hat{\theta}B)^{-1}(Y_{t-1} - T) \\ &\quad + \hat{\boldsymbol{\beta}}^T(1 - \hat{\theta}B)^{-1}\nabla\mathbf{X}_t + a_t \\ &\quad + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T\left.\frac{\partial(Y_t - T)}{\partial\boldsymbol{\beta}}\right|_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} + (\theta - \hat{\theta})\left.\frac{\partial(Y_t - T)}{\partial\theta}\right|_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} \\ &= (1 - \hat{\theta})(1 - \hat{\theta}B)^{-1}(Y_{t-1} - T) \\ &\quad + \hat{\boldsymbol{\beta}}^T\mathbf{U}_t + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T\mathbf{U}_t - (\theta - \hat{\theta})w_{t-1} + a_t.\end{aligned}$$

Substituting this into (10) yields (12).

APPENDIX D: PROOF THAT CMV CONTROL PROVIDES A HIGHER PROBABILITY OF STABILITY THAN MV CONTROL

We show that $\Pr[|\rho_{\text{CMV}}| \geq 1] < \Pr[|\rho_{\text{MV}}| \geq 1]$, where $\Pr[\cdot]$ denotes probability. Because both $\mu_{\rho,\text{CMV}}$ and $\mu_{\rho,\text{MV}}$ lie in the open interval $(-1, 1)$ if $|\hat{\theta}| < 1$, the proof follows if we show that the statistical distances $|\pm 1 - \mu_{\rho,\text{CMV}}|\Sigma_{\rho,\text{CMV}}^{-1/2}$ to the points $\rho = \pm 1$ for CMV control are larger than the corresponding statistical distances $|\pm 1 - \mu_{\rho,\text{MV}}|\Sigma_{\rho,\text{MV}}^{-1/2}$ for MV control. From (16)–(19), the ratio of the squared statistical distances to the point $\rho = 1$ is

$$\begin{aligned}\frac{(1 - \mu_{\rho,\text{CMV}})^2/\Sigma_{\rho,\text{CMV}}}{(1 - \mu_{\rho,\text{MV}})^2/\Sigma_{\rho,\text{MV}}} &= \frac{\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\hat{\boldsymbol{\beta}}}{(\hat{\boldsymbol{\beta}}^T\hat{\boldsymbol{\beta}})^2/(\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}})} \\ &= \left[\frac{\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}^T\hat{\boldsymbol{\beta}}}\right]\left[\frac{\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}^T\hat{\boldsymbol{\beta}}}\right]. \qquad \text{(D.1)}\end{aligned}$$

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

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

It follows that the ratio in (D.1) is greater than or equal to 1, and therefore $\Pr[\rho_{\text{CMV}} \geq 1] \leq \Pr[\rho_{\text{MV}} \geq 1]$.

Now, from (17) and (19),

$$\begin{aligned}\frac{\Sigma_{\rho,\text{MV}}}{\Sigma_{\rho,\text{CMV}}} &= \frac{\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}}{(\hat{\boldsymbol{\beta}}^T\hat{\boldsymbol{\beta}})^2} \Big/ \frac{\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\hat{\boldsymbol{\beta}}}{(1 + \hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\hat{\boldsymbol{\beta}})^2} \\ &> \frac{\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}}{(\hat{\boldsymbol{\beta}}^T\hat{\boldsymbol{\beta}})^2} \Big/ \frac{\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\hat{\boldsymbol{\beta}}}{(\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\hat{\boldsymbol{\beta}})^2} \\ &= \left[\frac{\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}^T\hat{\boldsymbol{\beta}}}\right]\left[\frac{\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}^T\hat{\boldsymbol{\beta}}}\right] \\ &\geq 1\end{aligned}$$

or $\Sigma_{\rho,\mathrm{CMV}} < \Sigma_{\rho,\mathrm{MV}}$. We also have from (16) and (18) that

$$\mu_{\rho,\mathrm{CMV}} - \mu_{\rho,\mathrm{MV}} = \frac{1-\hat{\theta}}{1+\hat{\boldsymbol{\beta}}^T \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} \hat{\boldsymbol{\beta}}} > 0$$

or $\mu_{\rho,\mathrm{CMV}} > \mu_{\rho,\mathrm{MV}}$. Because $\mu_{\rho,\mathrm{CMV}} > \mu_{\rho,\mathrm{MV}}$ and $\Sigma_{\rho,\mathrm{CMV}} < \Sigma_{\rho,\mathrm{MV}}$, the ratio of squared statistical distances to the point $\rho = -1$ is

$$\frac{(-1-\mu_{\rho,\mathrm{CMV}})^2/\Sigma_{\rho,\mathrm{CMV}}}{(-1-\mu_{\rho,\mathrm{MV}})^2/\Sigma_{\rho,\mathrm{MV}}} > 1.$$

It follows that $\Pr[\rho_{\mathrm{CMV}} \le -1] < \Pr[\rho_{\mathrm{MV}} \le -1]$, which completes the proof.

[Received March 2002. Revised May 2003.]

REFERENCES

Anderson, B. D. O. (1985), "Adaptive Systems, Lack of Persistency of Excitation and Bursting Phenomena," *Automatica*, 21, 247–258.

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

——— (1995), *Adaptive Control* (2nd ed.), New York: Addison-Wesley.

Bodson, M. (1993), "Pseudo-Burst Phenomenon in Ideal Adaptive Systems," *Automatica*, 29, 929–940.

Box, G. E. P., Jenkins, G., and Reinsel, G. (1994), *Time Series Analysis, Forecasting, and Control* (3rd ed.), Englewood Cliffs, NJ: Prentice-Hall.

Box, G. E. P., 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*, New York: Wiley.

Box, G. E. P., and Meyer, R. D. (1993), "Finding the Active Factors in Fractional Screening Experiments," *Journal of Quality Technology*, 25, 94–105.

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

Chen, A., and Guo, R. (2001), "Age-Based Double EWMA Controller and Its Application to CMP Processes," *IEEE Transactions on Semiconductor Manufacturing*, 14, 11–19.

Chipman, H. (1998), "Handling Uncertainty in Analysis of Robust Design Experiments," *Journal of Quality Technology*, 30, 11–17.

Del Castillo, E., and Hurwitz, A. M. (1997), "Run-to-Run Process Control: Literature Review and Extensions," *Journal of Quality Technology*, 29, 184–196.

Deming, W. E. (1986), *Out of the Crisis*, Cambridge, MA: MIT Press.

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

Gower-Hall, A., Boning, D., Rosenthal, P., and Waldhauer, A. (2002), "Model-Based Uniformity Control for Epitaxial Silicon Deposition," *IEEE Transactions on Semiconductor Manufacturing*, 15, 295–309.

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

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.

Leang, S., Ma, S., Thompson, J., Bombay, B., and Spanos, C. (1996), "A Control System for Photolithographic Sequences," *IEEE Transactions on Semiconductor Manufacturing*, 9, 191–207.

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

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. (1990), "A Different View of the Funnel Experiment," *Journal of Quality Technology*, 22, 255–259.

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*, Englewood Cliffs, NJ: Prentice-Hall.

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*, 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.

Wu, C. F. J., and Hamada, M. (2000), *Experiments: Planning, Analysis, and Parameter Design Optimization*, New York: Wiley.

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