Improved Design of Robust Exponentially Weighted Moving Average Control Charts for Autocorrelated Processes

Hyun Cheol Lee
Samsung Electronics, Semiconductor Business
Quality Assurance Dept.
Banwol-Dong, Hwasung-City
Gyeonggi-Do, Korea 445-701
Phone: 82-31-208-4717
Fax: 82-31-208-6699
Email: hc3.lee@samsung.com

Daniel W. Apley (Corresponding Author)
Department of Industrial Engineering and Management Sciences
Northwestern University
Evanston, IL 60208-3119
Phone: 847-491-2397
Fax: 847-491-8005
Email: apley@northwestern.edu

Abstract: Residual-based control charts for autocorrelated processes are known to be sensitive to time series modeling errors, which can seriously inflate the false alarm rate. This article presents a design approach for a residual-based exponentially weighted moving average (EWMA) chart that mitigates this problem by modifying the control limits based on the level of model uncertainty. Using a Bayesian analysis, we derive the approximate expected variance of the EWMA statistic, where the expectation is with respect to the posterior distribution of the unknown model parameters. The result is a relatively clean expression for the expected variance as a function of the estimated parameters and their covariance matrix. We use control limits proportional to the square root of the expected variance. We compare our approach to two other approaches for designing robust residual-based EWMA charts and argue that our approach generally results in a more appropriate widening of the control limits.

Key Words: Residual-based control charts; exponentially weighted moving average; time series; autoregressive moving average models; robust design; model uncertainty

1. INTRODUCTION

Exponentially weighted moving average (EWMA) control charts are widely used in statistical process control (SPC) to detect changes in a process mean. If $\{x_t: t = 1, 2, \ldots\}$ denotes observations of the process, the EWMA statistic z_t, introduced by Roberts[1], is calculated recursively via $z_t = (1-\lambda)z_{t-1} + \lambda x_t$, where $0 < \lambda \leq 1$ is the EWMA parameter. If the process observations x_t are statistically independent, then the control limits for the EWMA chart are typically set as

$$CL = \pm L\,\hat{\sigma}_z\,, \tag{1}$$

where $\hat{\sigma}_z = \hat{\sigma}_x\,\lambda^{1/2}(2-\lambda)^{-1/2}$ is an estimate (estimates are indicated by the "^" symbol) of the steady-state standard deviation of the EWMA statistic, and L is a constant that provides a specific desired in-control ARL. Tables in Lucas and Saccucci[2] or Lu and Reynolds[3], for example, can be used to select L. Here we assume the in-control process mean has been subtracted from the observations, so that the resulting x_t has an in-control mean of zero.

When the process data are autocorrelated, however, applying the standard control chart with standard control limits results in far too frequent false alarms (see, e.g., Johnson and Bagshaw[4]; Harris and Ross[5]; Alwan[6]). Montgomery and Woodall[7], Woodall and Montgomery[8], and Stoumbos *et al.*[9] contain excellent discussions on the increasing prevalence of autocorrelated data in SPC applications due, in part, to measurement automation that results in steady streams of data. To represent the autocorrelation, one typically uses an autoregressive moving average (ARMA) model of the form (Box *et al.*[10])

$$x_t = \frac{\Theta(B)}{\Phi(B)}a_t\,, \tag{2}$$

where t is a time index, B is a backward shift operator defined such that $Bx_t = x_{t-1}$, $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 \cdots - \phi_p B^p$ is the AR polynomial of order p, $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 \cdots - \theta_q B^q$ is the

MA polynomial of order q, and a_t is assumed to be an identically independently distributed (i.i.d.) random process with mean zero and variance σ_a^2.

Control charting approaches for autocorrelated data typically involve calculating the residuals $e_t = \hat{\Theta}^{-1}(B)\hat{\Phi}(B)x_t$, of the estimated ARMA model fitted via time series modeling of a prior sample of size N observations of x_t. With no modeling errors, the residuals are uncorrelated, and traditional control charts can be applied with well understood in-control run length properties. Berthouex *et al.*[11], Alwan and Roberts[12], Montgomery and Mastrangelo[13], Superville and Adams[14], Wardell *et al.*[15], Runger *et al.*[16], Lin and Adams[17], Vander Weil[18], Apley and Shi[19], Lu and Reynolds[3], English *et al.*[20], and many others have investigated residual-based control charts. Perhaps the most common chart is a residual-based EWMA (e.g., Lu and Reynolds[3]) of the form $z_t = (1-\lambda)z_{t-1} + \lambda e_t$. One typically neglects ARMA modeling errors and uses the control limits (1) with $\hat{\sigma}_z = \hat{\sigma}_a \lambda^{1/2}(2-\lambda)^{-1/2}$.

In this paper we focus on the effects of ARMA modeling errors. Many authors (e.g., Kramer and Schmid[21]; Adams and Tseng[22]; Apley and Shi[19]; Lu and Reynolds[3]; Kramer and Schmid[23]; Apley and Lee[24]) have investigated the adverse effects of ARMA modeling errors on residual based charts, using either simulation or analytical methods. One serious adverse effect is a substantial increase in the false alarm rate if the modeling errors are such that the autocorrelation is underestimated, similar to the increased false alarm rate that results from ignoring autocorrelation altogether. Jensen *et al.*[25] provides a comprehensive discussion on the effects of parameter estimation errors on control chart performance, in general. Their focus is on independent data and errors in estimating the mean and variance, but they include a brief discussion of control charts for autocorrelated data.

Apley[26], Apley and Lee[27], and Testik[28] proposed methods for widening the control limits of a residual-based EWMA in order to avoid the excessive false alarms caused by ARMA parameter estimation errors. Let $\boldsymbol{\gamma} = [\phi_1\ \phi_2\ \cdots\ \phi_p\ \theta_1\ \theta_2\ \cdots\ \theta_q]^T$ and $\hat{\boldsymbol{\gamma}} = [\hat{\phi}_1\ \hat{\phi}_2 \cdots \hat{\phi}_p\ \hat{\theta}_1\ \hat{\theta}_2 \cdots \hat{\theta}_q]^T$ denote the vectors of ARMA parameters and their estimates, respectively. Apley[26] used the control limits

$$CL = \pm L\sqrt{\hat{E}[\sigma_z^2 \mid \boldsymbol{\gamma}]}\,, \tag{3}$$

where L is chosen as if there were no modeling errors, and $\hat{E}[\sigma_z^2 \mid \boldsymbol{\gamma}]$ denotes an estimate of the expected EWMA variance, where the expectation is with respect to the random parameter estimates $\hat{\boldsymbol{\gamma}}$, conditioned on the unknown true parameters $\boldsymbol{\gamma}$. They approximate $\hat{E}[\sigma_z^2 \mid \boldsymbol{\gamma}]$ using a first-order Taylor approximation of certain underlying quantities. Noting that the approach of Apley[26] may not widen the control limits enough, Apley and Lee[27] used the control limits

$$CL = \pm L\,\sigma_{z,\alpha}\,, \tag{4}$$

where $\sigma_{z,\alpha}$ denotes the upper boundary of an approximate upper one-sided $1-\alpha$ confidence interval on the EWMA standard deviation σ_z for some appropriate choice of α. They referred to these as "worst-case" control limits, because they provide protection against excessive false alarms for the worst (largest) value of σ_z within the confidence interval. Testik[28] used a similar approach for the control limits for an AR(1) process but with a different "worst-case" value for σ_z based on assuming a truncated normal distribution for $\hat{\phi}_1$.

As we will illustrate in later examples, the motivation for this work is that the approach of Apley[26] generally does not widen the control limits enough, and the worst-case approach of Apley and Lee[27] generally widens the control limits by more than is needed. To avoid being overly conservative in this regard, Apley and Lee[27] recommended the relatively large value of α = 0.2. We demonstrate later that this large choice for α tends to widen the control limits too much for large N and perhaps not enough for small N. Our approach is to use the control limits

$$CL = \pm L\sqrt{E[\sigma_z^2 \mid \hat{\boldsymbol{\gamma}}]}\,, \tag{5}$$

which is akin to Apley[26] but with two differences: 1) We use a better Taylor approximation of σ_z^2, and 2) we use a Bayesian analysis in which the expectation in (5) is with respect to the posterior distribution of the unknown true parameters $\boldsymbol{\gamma}$, conditioned on the estimate $\hat{\boldsymbol{\gamma}}$.

In Section 2 we derive the expected EWMA variance $E[\sigma_z^2 \mid \hat{\boldsymbol{\gamma}}]$ for use in (5), which turns out to be quite tractable and only requires an estimate of $\boldsymbol{\gamma}$ and its error covariance matrix, both of which are typically produced by commercial ARMA modeling software. In Section 3 we provide further simplified expressions for $E[\sigma_z^2 \mid \hat{\boldsymbol{\gamma}}]$ for the special cases of first-order ARMA processes. We provide design guidelines in Section 4. In Section 5 we present comparisons with the approaches of Apley[26] and Apley and Lee[27] and argue that the proposed approach usually results in more appropriate widening of the control limits. We also discuss sample size recommendations and other discussion points. Section 6 concludes the paper.

2. EXPECTED EWMA VARIANCE

For notational convenience, denote $\nu = 1-\lambda$. Combining Eq. (2), the equation $e_t = \hat{\Theta}^{-1}(B)\hat{\Phi}(B)x_t$ for generating the residuals, and the EWMA equation $z_t = (1-\lambda)z_{t-1} + \lambda e_t = \nu z_{t-1} + (1-\nu)e_t$ gives the following model (see Apley and Lee[27] for further details)

$$z_t = (1-\nu B)^{-1}(1-\nu)\hat{\Theta}(B)^{-1}\hat{\Phi}(B)\Phi(B)^{-1}\Theta(B)a_t = G(B)a_t = \sum_{j=0}^{\infty} g_j\, a_{t-j}, \tag{6}$$

where $G(B) = (1-\nu B)^{-1}(1-\nu)\hat{\Theta}(B)^{-1}\hat{\Phi}(B)\Phi(B)^{-1}\Theta(B) = \sum_{j=0}^{\infty} g_j B^j$, and g_j (j = 0, 1, 2, . . .) denote the impulse response coefficients of ARMA($p+q+1$,$p+q$) transfer function $G(B)$. Hence, using the impulse response method (see Box *et al.*[10]), the EWMA variance is

$$\sigma_z^2 = \sigma_a^2 \sum_{j=0}^{\infty} g_j^2. \tag{7}$$

We approximate the EWMA variance using a second-order Taylor approximation

$$\sigma_z^2 \cong \sigma_z^2\big|_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} + \left[\frac{\partial \sigma_z^2}{\partial \boldsymbol{\gamma}}\Big|_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}}\right]^T (\boldsymbol{\gamma}-\hat{\boldsymbol{\gamma}}) + \frac{1}{2}(\boldsymbol{\gamma}-\hat{\boldsymbol{\gamma}})^T\left[\frac{\partial^2 \sigma_z^2}{\partial \boldsymbol{\gamma}^2}\Big|_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}}\right](\boldsymbol{\gamma}-\hat{\boldsymbol{\gamma}})$$

about $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}$. Taking the expected value of this with respect to the posterior distribution of $\boldsymbol{\gamma}$, given the data from which $\hat{\boldsymbol{\gamma}}$ is calculated, gives the approximate expected EWMA variance

$$E[\sigma_z^2 \mid \hat{\boldsymbol{\gamma}}] \cong \hat{\sigma}_z^2 + \frac{1}{2} tr\left\{ \frac{\partial^2 \sigma_z^2}{\partial \boldsymbol{\gamma}^2}\Big|_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} \right\} \tag{8}$$

where $\hat{\sigma}_z^2 = \hat{\sigma}_a^2 (1-\nu)(1+\nu)^{-1}$ is the EWMA variance if there were no modeling errors (i.e., if $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}$, and $\sigma_a = \hat{\sigma}_a$), *tr* denotes the matrix trace operator, and $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ denotes the posterior covariance matrix of $\boldsymbol{\gamma}$.

In deriving (8), we assume a suitable approximate maximum likelihood estimator $\hat{\boldsymbol{\gamma}}$ and a noninformative prior for $\boldsymbol{\gamma}$. In this case, invoking the standard large sample results for maximum likelihood estimation (see Carlin and Louis[29]) implies that the posterior distribution of $\boldsymbol{\gamma} \mid \hat{\boldsymbol{\gamma}}$ is approximately multivariate normal with mean $\hat{\boldsymbol{\gamma}}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\gamma}}}$, where $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\gamma}}}$ denotes the standard (non-Bayesian) large sample estimate of the covariance matrix of $\hat{\boldsymbol{\gamma}}$. Most commercial time series modeling software will produce the estimates $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\gamma}}}$ based on approximate likelihood methods. See Box *et al.*[10] for further details on calculating the estimates $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\gamma}}}$ and Appendix B of Apley and Lee[27] for a straightforward numerical procedure for calculating $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\gamma}}}$ that can be implemented in spreadsheet software.

In the remainder of this section, we show that Eq. (8) reduces to a relatively simple function of $\hat{\boldsymbol{\gamma}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$. Towards this end, differentiate Eq. (7) twice with respect to $\boldsymbol{\gamma}$ to give

$$\frac{\partial^2 \sigma_z^2}{\partial \boldsymbol{\gamma}^2} = 2\sigma_a^2 \sum_{j=0}^{\infty} \left[g_j \frac{\partial^2 g_j}{\partial \boldsymbol{\gamma}^2} + \left[\frac{\partial g_j}{\partial \boldsymbol{\gamma}}\right]\left[\frac{\partial g_j}{\partial \boldsymbol{\gamma}}\right]^T \right].$$

Combining this with Eq. (8) gives

$$E[\sigma_z^2 \mid \hat{\boldsymbol{\gamma}}] \cong \hat{\sigma}_a^2 \left(\frac{1-\nu}{1+\nu}\right) + \hat{\sigma}_a^2 tr\left\{ \sum_{j=0}^{\infty} \hat{g}_j \hat{\mathbf{D}}_j \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} + \sum_{j=0}^{\infty} \hat{\mathbf{d}}_j \hat{\mathbf{d}}_j^T \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} \right\}, \tag{9}$$

where we denote $\hat{\mathbf{D}}_j = \partial^2 g_j / \partial \boldsymbol{\gamma}^2 \big|_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}}$, and $\hat{\mathbf{d}}_j = \partial g_j / \partial \boldsymbol{\gamma} \big|_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}}$.

We show in the Appendix that, after much tedious algebra, Eq. (9) simplifies to

$$E[\sigma_z^2 \mid \hat{\boldsymbol{\gamma}}] \cong \hat{\sigma}_a^2 \left(\frac{1-\nu}{1+\nu}\right)\left\{1+\frac{1}{N}\left[\frac{2\mathbf{V}_p^T\overline{\boldsymbol{\Sigma}}_{\Phi}\mathbf{V}_p}{\hat{\Phi}^2(\nu)}-\frac{2\mathbf{V}_p^T\overline{\boldsymbol{\Sigma}}_{\Phi\Theta}\mathbf{V}_q}{\hat{\Phi}(\nu)\hat{\Theta}(\nu)}+p+q\right.\right.$$

$$\left.\left.+\frac{2\left[\hat{\phi}_1, 2\hat{\phi}_2, 3\hat{\phi}_3, \cdots, p\hat{\phi}_p\right]\mathbf{V}_p}{\hat{\Phi}(\nu)}+\frac{2\left[\hat{\theta}_1, 2\hat{\theta}_2, 3\hat{\theta}_3, \cdots, q\hat{\theta}_q\right]\mathbf{V}_q}{\hat{\Theta}(\nu)}\right]\right\}, \qquad (10)$$

where $\hat{\Phi}(\nu)=1-\hat{\phi}_1\nu-\hat{\phi}_2\nu^2-\ldots-\hat{\phi}_p\nu^p$, $\hat{\Theta}(\nu)=1-\hat{\theta}_1\nu-\hat{\theta}_2\nu^2-\ldots-\hat{\theta}_q\nu^q$, $\mathbf{V}_p=[\nu \;\; \nu^2 \;\ldots\; \nu^p]^T$, $\mathbf{V}_q=[\nu \;\; \nu^2 \;\ldots\; \nu^q]^T$, and we have partitioned the $(p+q)\times(p+q)$ parameter covariance matrix (which is inversely proportional to sample size N) as

$$\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}=\begin{bmatrix}\boldsymbol{\Sigma}_{\Phi} & \boldsymbol{\Sigma}_{\Phi\Theta}\\ \boldsymbol{\Sigma}_{\Phi\Theta}^T & \boldsymbol{\Sigma}_{\Theta}\end{bmatrix}=\frac{1}{N}\begin{bmatrix}\overline{\boldsymbol{\Sigma}}_{\Phi} & \overline{\boldsymbol{\Sigma}}_{\Phi\Theta}\\ \overline{\boldsymbol{\Sigma}}_{\Phi\Theta}^T & \overline{\boldsymbol{\Sigma}}_{\Theta}\end{bmatrix}=\frac{1}{N}\overline{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}}.$$

Given the estimates $\hat{\boldsymbol{\gamma}}$, $\hat{\sigma}_a$, and $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ from time series modeling software, the size N of the sample of observations from which the estimates were obtained, and the EWMA parameter λ (= $1-\nu$), one calculates the expected EWMA variance using Eq. (10) and then substitutes this into the control limits (5). In Section 5 we provide examples illustrating the extent to which this widens the control limits, thereby protecting against excessive false alarms that can result from ARMA modeling errors.

Remark 1: The covariance matrix $N\overline{\boldsymbol{\Sigma}}_{\Theta}$ of the MA parameters does not appear directly in Eq. (10). This is not because the expected EWMA variance does not depend on $\overline{\boldsymbol{\Sigma}}_{\Theta}$. Indeed it does; but in the derivations in the Appendix, we have already substituted a standard expression for $\overline{\boldsymbol{\Sigma}}_{\Theta}$.

Remark 2: Apley[26] used a related approach to calculate $\hat{E}[\sigma_z^2 \mid \boldsymbol{\gamma}]$ in a non-Bayesian scenario. They used a first order Taylor approximation of z_t (with respect to $\hat{\boldsymbol{\gamma}}$, about $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}$) by differentiating Eq. (6) and then substituted the impulse response coefficients of the Taylor approximation into Eq. (7). In spite of these differences, their approach is equivalent to using Eq. (9) but excluding the terms involving the second derivative matrices $\hat{\mathbf{D}}_j$. We demonstrate in later examples that including the second derivative terms results in more reasonable widening of the control limits, especially for small sample sizes.

3. RESULTS FOR LOW-ORDER ARMA PROCESSES

For certain low-order ARMA processes, for which we have simple closed-form expressions for $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ as a function of the estimated ARMA parameters, the posterior variance of Eq. (10) further simplifies. We refer readers to Box *et al.*[10] for details and derivations of the expressions for $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ that we use in this section. For ARMA(1,1) processes, the parameter covariance is

$$\frac{1}{N}\overline{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}} = \frac{1}{N}\frac{\left(1-\hat{\phi}_1\hat{\theta}_1\right)}{\left(\hat{\phi}_1-\hat{\theta}_1\right)^2}\begin{bmatrix}\left(1-\hat{\phi}_1^2\right)\left(1-\hat{\phi}_1\hat{\theta}_1\right) & \left(1-\hat{\phi}_1^2\right)\left(1-\hat{\theta}_1^2\right)\\ \left(1-\hat{\phi}_1^2\right)\left(1-\hat{\theta}_1^2\right) & \left(1-\hat{\theta}_1^2\right)\left(1-\hat{\phi}_1\hat{\theta}_1\right)\end{bmatrix}.$$

Substituting this into Eq. (10) gives

$$E[\sigma_z^2 \mid \hat{\boldsymbol{\gamma}}] \cong \hat{\sigma}_a^2\left(\frac{1-\nu}{1+\nu}\right)\left\{1+\frac{1}{N}\left[\frac{2\nu^2\left(1-\hat{\phi}_1\hat{\theta}_1\right)^2\left(1-\hat{\phi}_1^2\right)}{\left(1-\hat{\phi}_1\nu\right)^2\left(\hat{\phi}_1-\hat{\theta}_1\right)^2}-\frac{2\nu^2\left(1-\hat{\phi}_1\hat{\theta}_1\right)\left(1-\hat{\phi}_1^2\right)\left(1-\hat{\theta}_1^2\right)}{\left(1-\hat{\phi}_1\nu\right)\left(1-\hat{\theta}_1\nu\right)\left(\hat{\phi}_1-\hat{\theta}_1\right)^2}\right.\right.$$

$$\left.\left.+1+1+\frac{2\hat{\phi}_1\nu}{1-\hat{\phi}_1\nu}+\frac{2\hat{\theta}_1\nu}{1-\hat{\theta}_1\nu}\right]\right\}$$

$$= \hat{\sigma}_a^2\left(\frac{1-\nu}{1+\nu}\right)\left[1+\frac{1}{N}\left[\frac{2\nu^2\left(1-\hat{\phi}_1\hat{\theta}_1\right)\left(1-\hat{\phi}_1^2\right)\left(\nu-\hat{\theta}_1\right)+2\left(\hat{\phi}_1-\hat{\theta}_1\right)\left(1-\hat{\phi}_1\nu\right)\left(1-\hat{\phi}_1\hat{\theta}_1\nu^2\right)}{\left(\hat{\phi}_1-\hat{\theta}_1\right)\left(1-\hat{\phi}_1\nu\right)^2\left(1-\hat{\theta}_1\nu\right)}\right]\right] \quad (11)$$

as the expected EWMA variance for ARMA(1,1) processes. The last equality follows after a great deal of algebra.

For AR(2) processes, the parameter covariance matrix is

$$\frac{1}{N}\overline{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}} = \frac{1}{N}\begin{bmatrix}1-\hat{\phi}_2^2 & -\hat{\phi}_1\left(1+\hat{\phi}_2\right)\\ -\hat{\phi}_1\left(1+\hat{\phi}_2\right) & 1-\hat{\phi}_2^2\end{bmatrix}.$$

When substituted into Eq. (10), this gives

$$E[\sigma_z^2 \mid \hat{\boldsymbol{\gamma}}] \cong \hat{\sigma}_a^2\left(\frac{1-\nu}{1+\nu}\right)\left[1+\frac{1}{N}\left[\frac{2-2\hat{\phi}_1\nu-2\hat{\phi}_1\nu^3-6\hat{\phi}_1\hat{\phi}_2\nu^3-\hat{\phi}_2^2\nu^2+\hat{\phi}_2^2\nu^4+\nu^2+\nu^4}{\left(1-\hat{\phi}_1\nu-\hat{\phi}_2\nu^2\right)^2}\right]\right]$$

as the expected EWMA variance for AR(2) processes.

For an AR(1) process, we have only a single parameter, and its variance is $N^{-1}\overline{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}} = N^{-1}\left(1-\hat{\phi}_1^2\right)$. Substituting this into Eq. (10) gives

$$E[\sigma_z^2 \mid \hat{\boldsymbol{\gamma}}] \cong \hat{\sigma}_a^2\left(\frac{1-\nu}{1+\nu}\right)\left[1+\frac{1}{N}\left[\frac{1-3\hat{\phi}_1^2\nu^2+2\nu^2}{\left(1-\hat{\phi}_1\nu\right)^2}\right]\right]$$

as the expected EWMA variance for AR(1) processes.

Similarly, using expressions for the parameter covariance matrices given in Box, et al. (1994), one can show that the expected EWMA variances for MA (2) and MA (1) processes are

$$E[\sigma_z^2 \mid \hat{\boldsymbol{\gamma}}] \cong \hat{\sigma}_a^2\left(\frac{1-\nu}{1+\nu}\right)\left[1+\frac{1}{N}\left[\frac{2+2\hat{\theta}_2\nu^2}{1-\hat{\theta}_1\nu-\hat{\theta}_2\nu^2}\right]\right] \text{ and}$$

$$E[\sigma_z^2 \mid \hat{\boldsymbol{\gamma}}] \cong \hat{\sigma}_a^2\left(\frac{1-\nu}{1+\nu}\right)\left[1+\frac{1}{N}\left[\frac{1+\hat{\theta}_1\nu}{1-\hat{\theta}_1\nu}\right]\right],$$

respectively.

The preceding result for MA(1) processes is exactly the same as the result for $\hat{E}[\sigma_z^2 \mid \boldsymbol{\gamma}]$ from Apley[26]. The reason is that g_j is a linear function of θ_1, and so the first-order Taylor approximation used by Apley[26] is the same as the second-order Taylor approximation that we use. The preceding results for other low-order ARMA processes are quite different than those of Apley[26]. We further discuss the differences in Section 5.2.

4. EWMA DESIGN PROCEDURE USING THE EXPECTED VARIANCE

As for any residual-based chart, the first step is to use appropriate time series modeling software to fit an ARMA model to a set of observations $\{x_1, x_2, \ldots, x_N\}$. If the model is one of the five special cases covered in Section 3, only the estimates $\hat{\boldsymbol{\gamma}}$ and $\hat{\sigma}_a$ are needed. One can then substitute these directly into one of the expressions for $E[\sigma_z^2 \mid \hat{\boldsymbol{\gamma}}]$ in Section 3. This, in turn, is substituted into Eq. (5) to give the EWMA control limits, suitably widened to protect against excessive false alarms due to ARMA modeling errors. If the model is not one of the special cases discussed in Section 3, the procedure is the same, except that one also needs $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ (in addition to $\hat{\boldsymbol{\gamma}}$

and $\hat{\sigma}_a$), which is typically produced by most commercial software packages for time series modeling. These are then substituted into Eq. (10) and the result of this, into Eq. (5). If the time series modeling software does not produce an estimate of $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$, one can implement the numerical procedure in Appendix B of Apley and Lee[27] using spreadsheet software to calculate $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$. All that are required for this are the parameter estimates $\hat{\boldsymbol{\gamma}}$ and the sample size N.

One must also select L and λ, prior to calculating the control limits. For this, we recommend the same approach that Apley and Lee[27] recommended (see Apley and Lee[27] for justification): Choose L and λ exactly as one would if parameter uncertainty were neglected. That is, one can choose λ to be sensitive to a certain size mean shift and then choose the value of L that would result in a desired in-control ARL if there were no modeling errors and the control limits $\pm L\hat{\sigma}_z$ were used. Lu and Reynolds[3] provide guidelines for this. For a specified λ, the tables of Lucas and Saccucci[2] can be used to choose L to give a desired in-control ARL.

To illustrate the design procedure, we use the Series A data from Box *et al.*[10], which are 197 concentration readings from a chemical production process. An ARMA(1,1) model was fitted to the data with estimated parameters $\hat{\phi}$ = 0.87, $\hat{\theta}$ = 0.48, and $\hat{\sigma}_a^2$ = 0.098. If we select λ = 0.1 and a desired in-control ARL of 500, then we would choose L = 2.814 from the table of Lucas and Saccucci[2]. If there were no parameter uncertainty, the EWMA variance would be $\hat{\sigma}_z = \hat{\sigma}_a \lambda^{1/2}(2-\lambda)^{-1/2}$ = 0.0718. Hence, the standard EWMA control limits are $\pm L\hat{\sigma}_z$ = ±0.202. Considering model uncertainty, the expected EWMA variance for this ARMA(1,1) model is calculated by substituting the estimated parameters, N = 197, and ν = 0.9 into Eq. (11). This gives $E[\sigma_z^2 \mid \hat{\boldsymbol{\gamma}}]$ = 0.00568, and $\sqrt{E[\sigma_z^2 \mid \hat{\boldsymbol{\gamma}}]}$ = 0.0754, which is 4.9% larger than $\hat{\sigma}_z$. Substituting this into Eq. (5) gives control limits of ±0.212, based on the expected variance, which are 4.9% wider than the standard limits. In comparison, one can show that the control limits of Apley[26] and Apley and Lee[27] are ±0.208 [from Eq. (3)] and ±0.227 [from Eq. (4) with α = 0.2], respectively. These methods widen the control limits by 3.0% and 12.3%, respectively. The proposed approach, which widens the control limits by 4.9%, falls somewhere between the other two approaches in terms of how conservatively it widens the control limits.

Figure 1 shows an EWMA control chart applied to 1000 simulated observations from the ARMA(1,1) model of the chemical process when the true parameters coincide with their estimates. All four sets of control limits [standard; Apley[26]; the proposed; and Apley and Lee[27]] are displayed in the figure. Figure 2 is analogous to Figure 1, except that when simulating the process, we introduced an error in the parameters: We let θ_1 and σ_a coincide with their estimates, but used $\phi_1 = 0.9$. Figure 2 is intended to illustrate the increased false alarms caused by parameter estimation errors and the mitigating effects of widening the control limits (later in the paper, we use Monte Carlo simulation to investigate the effects of widening the control limits on the in-control and out-of-control ARLs). Since there was no mean shift, all of the out-of-control signals in Figures 1 and 2 are false alarms. Over the 1,000 observations in Figure 1, there are three false alarms using the standard control limits, two using the control limits of Apley[26], one using the proposed control limits, and none using the control limits of Apley and Lee[27]. In Figure 2, the numbers of false alarms for the four sets of control limits increase to five, four, two, and one, respectively.

5. DISCUSSION

In this section we discuss several points of interest regarding the proposed method: We compare our approach to two other approaches, in terms of the extent to which the different methods widen the control limits to account for parameter uncertainty. We contrast the Bayesian paradigm we have adopted with a corresponding non-Bayesian paradigm. We also discuss guidelines for choosing the sample size large enough so that the effects of parameter uncertainty are not detrimental to the performance of the chart in detecting shifts.

5.1 Comparison with Other Methods for Widening the Control Limits

We compare our method with the two other design procedures discussed in the introduction for widening the control limits of a residual-based EWMA to account for parameter uncertainty: The methods of Apley[26] (hereafter A) and Apley and Lee[27] (hereafter A&L). The methods of A

and A&L use the control limits (3) and (4), respectively, and our method uses the control limits (5). To have a common basis for comparison, we use the same values of L and λ for all three methods.

Tables 1—3 compare the control limits for the three methods for various sample sizes (N = 50, 100, 200, and 500) and for various fitted ARMA(1,1) models. Table 1 is for EWMA parameter λ = 0.05, and Tables 2 and 3 are for λ = 0.10 and λ = 0.20, respectively. For each of the three values of λ, L was taken from the tables of Lucas and Saccucci[2] to give a desired in-control ARL of 500 for the situation in which there are no modeling errors. The "RI" columns show the relative increase in control limit width, i.e., the percentage increase relative to the standard control limits $\pm L\hat{\sigma}_a(1-\nu)^{1/2}(1+\nu)^{-1/2}$ that would be used if one neglected modeling errors. For simplicity, we have neglected errors in σ_a by assuming $\sigma_a = \hat{\sigma}_a = 1.0$. The A&L method involves selection of an additional parameter α, the confidence level for the worst-case control limits. In Tables 1—3, we have used the midrange value α = 0.2 recommended by A&L. Table 5, discussed below, shows analogous results for larger and smaller choices of α.

We can draw some general conclusions from Tables 1—3: For the examples considered, the proposed method always widened the control limits by a greater amount than method A. For small λ (λ = 0.05, Table 1) the RI was roughly three times larger for the proposed method than for method A, whereas for large λ (λ = 0.20, Table 3), the RI for the two methods were much more comparable. We believe this is desirable. Modeling errors cause autocorrelation in the residuals, and the effect of residual autocorrelation is much greater when one uses a small value of λ (see Apley and Lee[27], for a discussion of the reasons). Hence, a much larger widening of the control limits is in order. Method A, based on a cruder first-order Taylor approximation, does not sufficiently widen the control limits in this situation.

Another conclusion from the tables is that the RIs for the proposed method and for the A&L method are often comparable for small N, but for large N the RI is much larger for the A&L method. We believe this is a desirable characteristic of the proposed method and that the A&L method can be overly conservative for large N. Consider the situation in Table 1 for $\hat{\phi}_1$ = 0.9 and

$\hat{\theta}_1$ = 0.6. For the small sample size of N = 50, the proposed method and the A&L method both widen the control limits by roughly 30% (RI = 31.8% and 30.2%). This seems reasonable, because with such a small sample size the modeling errors may be quite large. In contrast, for the large sample size of N = 500, the proposed method only widens the control limits by 3.6%, whereas the A&L method still widens the control limits by 10.4%. For such a large sample size, a 10.4% RI seems unnecessarily conservative, the drawback being larger out-of-control ARLs, which we discuss shortly.

Other conclusions are more obvious: As N increases, the RI for all methods decreases, because the ARMA parameters are estimated with greater precision. Moreover, as λ increases, the RI decreases for all methods. The reason, as discussed earlier, is that autocorrelation in the residuals has a stronger effect on the false alarm rate when λ is small.

To give an idea of the effect of widening the control limits on the ARLs, Table 4 shows the ARLs for various size mean shifts for the Box *et al.*[10] ARMA(1,1) chemical data example considered earlier. Recall that the sample size was N = 197, $\hat{\phi}$ = 0.87, $\hat{\theta}$ = 0.48, and $\hat{\sigma}_a^2$ = 0.098. We will consider λ = 0.1 (for which L = 2.814 for an in-control ARL of 500). This is the same example considered in Apley and Lee[27], for which they chose α = 0.1. Hence, we will use α = 0.1 also. We used Monte Carlo simulation to calculate all ARL values, all of which are for the case that the parameters coincide with their estimates. From Table 4, the performance in detecting shifts is clearly adversely affected by widening the control limits. For the A method, the proposed method, and the A&L method, the control limits are widened by 3.0% (= .208/.202−1), 5.0% (= .212/.202−1), and 17.3% (= .237/.202−1), respectively. Consequently, as evident from Table 4, the worst-case A&L control limits result in substantially worse detection performance than the other EWMA charts, which are widened by much lesser amounts. For the smaller size shifts in Table 4, the worst-case limits of A&L result in almost double the ARL of the other charts. This illustrates the consequences of using overly-conservative (i.e., overly-widened) control limits.

The A&L method involves choosing an additional design parameter α, and this has a large effect on the extent to which the A&L method widens the control limits. Table 5 compares the RI values for the A&L method with α = 0.1, α = 0.2, and α = 0.3 for the case λ = 0.05. A&L recommended these relatively large (relative to what one typically chooses for confidence intervals in other contexts) α values to avoid overly overly-widened control limits and the resulting decrease in detection performance seen in Table 4. In the setting of Table 4, if we had used α = 0.2 in the A&L method, the control limits would have been ±0.226 or 11.6% wider than the standard control limits (compared to the ±0.237 control limits that were 17.3% wider for α = 0.1). Using these narrower control limits would have resulted in out-of-control performance somewhere between the proposed method and the A&L method with α = 0.1. Choosing an appropriate α to balance between widening the control limits to mitigate excessive false alarms versus keeping the control limits narrow enough to retain reasonable out-of-control detection power is somewhat subjective. We believe the approach of this paper constitutes a more reasonable and less subjective way of accomplishing this.

5.2 Bayesian Versus Non-Bayesian Approaches

In the approach of this paper, we used the Taylor expansion

$$\sigma_z^2 \cong \sigma_z^2\big|_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} + \left[\frac{\partial\sigma_z^2}{\partial\boldsymbol{\gamma}}\Big|_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}}\right]^T(\boldsymbol{\gamma}-\hat{\boldsymbol{\gamma}}) + \frac{1}{2}(\boldsymbol{\gamma}-\hat{\boldsymbol{\gamma}})^T\left[\frac{\partial^2\sigma_z^2}{\partial\boldsymbol{\gamma}^2}\Big|_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}}\right](\boldsymbol{\gamma}-\hat{\boldsymbol{\gamma}})$$

about $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}$ and, in a Bayesian paradigm, took the expected value of this with respect to the posterior distribution of $\boldsymbol{\gamma} \mid \hat{\boldsymbol{\gamma}}$ to give an approximate expression for $E[\sigma_z^2 \mid \hat{\boldsymbol{\gamma}}]$. An alternative approach would be to use the Taylor expansion

$$\sigma_z^2 \cong \sigma_z^2\big|_{\hat{\boldsymbol{\gamma}}=\boldsymbol{\gamma}} + \left[\frac{\partial\sigma_z^2}{\partial\hat{\boldsymbol{\gamma}}}\Big|_{\hat{\boldsymbol{\gamma}}=\boldsymbol{\gamma}}\right]^T(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}) + \frac{1}{2}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})^T\left[\frac{\partial^2\sigma_z^2}{\partial\hat{\boldsymbol{\gamma}}^2}\Big|_{\hat{\boldsymbol{\gamma}}=\boldsymbol{\gamma}}\right](\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})$$

about $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}$ and, in a non-Bayesian paradigm, take the expected value of this with respect to the distribution of $\hat{\boldsymbol{\gamma}}$. The result would be a function of the unknown true parameters $\boldsymbol{\gamma}$, but we could substitute $\hat{\boldsymbol{\gamma}}$ to give an estimate $\hat{E}[\sigma_z^2 \mid \boldsymbol{\gamma}]$.

Even though we are assuming a noninformative prior for $\boldsymbol{\gamma}$, in which case $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\gamma}}}$, these two approach would not yield $E[\sigma_z^2 \mid \hat{\boldsymbol{\gamma}}] = \hat{E}[\sigma_z^2 \mid \boldsymbol{\gamma}]$. This is because of the asymmetry of $G(B) = (1 - vB)^{-1}(1 - v)\hat{\Theta}(B)^{-1}\hat{\Phi}(B)\Phi(B)^{-1}\Theta(B)$ with respect to $\hat{\boldsymbol{\gamma}}$ and $\boldsymbol{\gamma}$. In the non-Bayesian approach, the second derivative of the impulse response coefficients $\{g_j: j = 1, 2, \ldots\}$ with respect to the estimated AR parameters are zero. In contrast, in the Bayesian approach, the second derivative with respect the true MA parameters are zero. It is straightforward to show that the non-Bayesian $\hat{E}[\sigma_z^2 \mid \boldsymbol{\gamma}]$ is of exactly the same form as the Bayesian $E[\sigma_z^2 \mid \hat{\boldsymbol{\gamma}}]$, except that the roles of the MA and AR parameters are reversed. We believe the Bayesian expression is more intuitively appealing, because it places more emphasis on the AR parameters than does the non-Bayesian expression.

5.3 Sample Size Requirements

From Eq. (10), as sample size $N \to \infty$, the expected EWMA variance approaches the standard EWMA variance when there are no modeling errors, in which case the control limits are not widened at all. Consequently, in order to mitigate the drawbacks (namely, decreased detection performance) of widening the control limits, one may prefer to collect a sufficiently large data sample when estimating the time series model. We recommend the same strategy recommended by Apley and Lee[27]: Choose a small value δ that represents the maximum acceptable percentage by which the control limits may be widened. The sample size N is then chosen to ensure that

$$\frac{\sqrt{E[\sigma_z^2 \mid \hat{\boldsymbol{\gamma}}]}}{\hat{\sigma}_z} \leq 1 + \delta .$$

Given preliminary guesses for the parameters (perhaps from a small pilot sample), one could substitute $E[\sigma_z^2 \mid \hat{\boldsymbol{\gamma}}]$ from Eq. (10) into the preceding equation and solve for the required N. One would then collect a larger sample of size N and refit the model.

Using Eq. (11), for ARMA(1,1) processes the required sample size is

$$N \geq \frac{2\nu^2\left(1-\hat{\phi}_1\hat{\theta}_1\right)\left(1-\hat{\phi}_1^2\right)\left(\nu-\hat{\theta}_1\right)+2\left(\hat{\phi}_1-\hat{\theta}_1\right)\left(1-\hat{\phi}_1\nu\right)\left(1-\hat{\phi}_1\hat{\theta}_1\nu^2\right)}{\left(\delta^2+2\delta\right)\left(\hat{\phi}_1-\hat{\theta}_1\right)\left(1-\hat{\phi}_1\nu\right)^2\left(1-\hat{\theta}_1\nu\right)}.$$

For example, for the ARMA(1,1) chemical process data example with preliminary estimates $\hat{\phi}_1 =$ 0.87 and $\hat{\theta}_1 = 0.48$, we would need $N \geq 310$ to ensure that the control limits are no more than 5% wider ($\delta = 0.05$) than the standard control limits when $\lambda = 0.05$. For $\delta = 0.01$ with the same λ, the required sample size increases to $N \geq 1600$.

Similarly, for AR (1) processes, the required sample size is

$$N \geq \frac{1-3\hat{\phi}_1^2\nu^2+2\nu^2}{\left(\delta^2+2\delta\right)\left(1-\hat{\phi}_1\nu\right)^2}.$$

6. CONCLUSIONS

We have presented an approach for widening the control limits of a residual-based EWMA to take into account uncertainty in the estimated ARMA parameters. Like the approach of Apley[26], we set the control limits proportional to the square root of the expected EWMA variance. However, we use a more accurate second-order Taylor approximation to the EWMA variance and a Bayesian analysis. For a number of scenarios, we compared the extent to which the control limits are widened using our approach, the approach of Apley[26], and the worst-case approach of Apley and Lee[27]. We argued that our approach generally results in a more reasonable and intuitively appealing widening of the control limits than do the other approaches. It usually widens the control limits more than the approach of Apley[26] but less than the overly-conservative worst-case approach of Apley and Lee[27]. The exception is when sample size is very small, in which case our approach widens the control limits by roughly the same amount as does Apley

and Lee[27] with midrange choice of α, which we believe is desirable. Another advantage of our approach is that it is less subjective than Apley and Lee[27], which requires choosing the additional design parameter α.

We have only considered parametric model uncertainty and have ignored any uncertainty in the assumed model structure. For example, in practice the model order (p,q) must also be estimated and is therefore subject to uncertainty. Treatment of model structure uncertainty would be substantially more complicated. Because considering model structure uncertainty would generally increase the overall level of uncertainty and further widen the control limits, the approach of this paper results in what should be viewed as the minimum amount by which the control limits should be widened to prevent excessive false alarms.

ACKNOWLEDGMENTS

This work was partially supported by the National Science Foundation under Grant CMMI-0758557.

REFERENCES

1. Roberts SW. Control Chart Tests Based on Geometric Moving Averages. *Technometrics* 1959; 1: 239-250.

2. Lucas JM, Saccucci MS. Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. *Technometrics* 1990; 32: 1-12.

3. Lu CW, Reynolds MR. EWMA Control Charts for Monitoring the Mean of Autocorrelated Processes. *Journal of Quality Technology* 1999; 31: 166-188.

4. Johnson RA, Bagshaw M. The Effect of Serial Correlation on the Performance of CUSUM tests. *Technometrics* 1974; 16: 103-112.

5. Harris TJ, Ross WH. Statistical Process Control Procedures for Autocorrelated Observations. *Canandian Journal of Chemical Engineering* 1991; 69: 48-57.

6. Alwan LC. Effects of Autocorrelation on Control Chart Performance. *Communications in Statistics : Theory and Methods* 1992; 21: 1025-1049.

7. Montgomery DC, Woodall WH. A Discussion on Statistically-Based Process Monitoring and Control. *Journal of Quality Technology* 1997; 29: 121-162.

8. Woodall WH, Montgomery DC. Research Issues and Ideas in Statistical Process Control. *Journal of Quality Technology* 1999; 31: 376-386.

9. Stoumbos ZG, Reynolds MR Jr., Ryan TP, Woodall WH. The State of Statistical Process Control as We Proceed into the 21st Century. *Journal of the American Statistical Association* 2000; 95: 992-998.

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

11. Berthouex PM, Hunter WG, Pallesen L. Monitoring Sewage Treatment Plants: Some Quality Control Aspects. *Journal of Quality Technology* 1978; 10: 139-149.

12. Alwan LC, Roberts HV. Time-Series Modeling for Statistical Process Control. *Journal of Business and Economic Statistics* 1988; 6: 87-95.

13. Montgomery DC, Mastrangelo CM. Some Statistical Process Control Methods for Autocorrelated Data. *Journal of Quality Technology* 1991; 23: 179-193.

14. Superville CR, Adams BM. An Evaluation of Forecast-Based Quality Control Schemes. *Communications in Statistics: Simulation and Computation* 1994; 23: 645-661.

15. Wardell DG, Moskowitz H, Plante RD. Run-Length Distributions of Special-Cause Control Charts for Correlated Processes. *Technometrics* 1994; 36: 3-17.

16. Runger GC, Willemain TR, Prabhu S. Average Run Lengths for Cusum Control Charts Applied to Residuals. *Communications in Statistics: Theory and Methods* 1995; 24: 273-282.

17. Lin WS, Adams BM. Combined Control Charts for Forecast-Based Monitoring Schemes. *Journal of Quality Technology* 1996; 28: 289-301.

18. Vander Wiel SA. Monitoring Processes That Wander Using Integrated Moving Average Models. *Technometrics* 1996; 38: 139-151.

19. Apley DW, Shi J. The GLRT for Statistical Process Control of Autocorrelated Processes. *IIE Transactions* 1999; 31: 1123-1134.

20. English JR, Lee S, Martin TW, Tilmon C. Detecting changes in autoregressive processes with $\overline{X}$ and EWMA charts. *IIE Transactions* 2000; 32: 1103-1113.

21. Kramer H, Schmid W. Control Charts for Time Series. *Nonlinear Analysis, Theory, Methods & Applications* 1997; 30: 4007-4016.

22. Adams BM, Tseng IT. Robustness of Forecast-Based Monitoring Schemes. *Journal of Quality Technology* 1998; 30: 328-339.

23. Kramer H, Schmid W. The Influence of Parameter Estimation on the ARL of Shewhart Type Charts for Time Series. *Statistical Papers*, 2000; 41: 173-196.

24. Apley DW, Lee HC. Robustness Comparison of Exponentially Weighted Moving-Average Charts on Autocorrelated Data and on Residuals. *Journal of Quality Technology* 2008; 40: 428-447.

25. Jensen WA, Jones-Farmer LA, Champ CW, Woodall WH. Effects of parameter estimation on control chart properties: A literature review. *Journal of Quality Technology* 2006; 38: 349—364.

26. Apley DW. Time Series Control Charts in the Presence of Model Uncertainty. *Journal of Manufacturing Science and Engineering* 2002; 124: 891-898.

27. Apley DW, Lee HC. Design of Exponentially Weighted Moving Average Control Charts for Autocorrelated Processes With Model Uncertainty. *Technometrics* 2003; 45: 187-198.

28. Testik MC. Model Inadequacy and Residuals Control Charts for Autocorrelated Processes. *Quality and Reliability Engineering International* 2005; 21: 115–130.

29. Carlin BP, Louis TA. *Bayes and Empirical Bayes Methods for Data Analysis*, (*2nd edn*), Chapman & Hall/CRC: Boca Raton, FL, 2000.

30. Apley DW, Lee HC. The Effects of Model Parameter Deviations on the Variance of a Linearly Filtered Time Series. *Naval Research Logistics* 2010; in press. DOI: 10.1002/nav.20414.

APPENDIX: DERIVATION OF THE EXPECTED EWMA VARIANCE OF EQ. (10)

To show that Eq. (9) simplifies to Eq. (10), we need simplified expressions for $\hat{\mathbf{D}}_j = \partial^2 g_j / \partial \boldsymbol{\gamma}^2 \big|_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}}$ and $\hat{\mathbf{d}}_j = \partial g_j / \partial \boldsymbol{\gamma} \big|_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}}$. Define $\mathbf{D}_j = \partial^2 g_j / \partial \boldsymbol{\gamma}^2$, and $\mathbf{d}_j = \partial g_j / \partial \boldsymbol{\gamma}$. Apley and Lee[30] derived expressions for $\mathbf{d}_j$, and we extend their approach to find expressions for $\mathbf{D}_j$ as well. We also show that when these are combined with the standard expression for $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$, the right-most summation in Eq. (9) simplifies considerably.

Denote the elements of $\mathbf{D}_j$ by $D_j^{\phi_i,\phi_l}$, $D_j^{\phi_i,\theta_l}$, and $D_j^{\theta_i,\theta_l}$ ($1 \le i \le p$, $1 \le l \le p$) and the elements of $\mathbf{d}_j$ by $d_j^{\phi_i}$ and $d_j^{\theta_l}$ ($1 \le i \le p$, $1 \le l \le p$). For example,

$$D_j^{\phi_i,\theta_l} = \frac{\partial^2 g_j}{\partial \phi_i \partial \phi_l} \text{ and } d_j^{\phi_i} = \frac{\partial g_j}{\partial \phi_i}.$$

As in Apley and Lee[30], we write Eq. (6) as $z_t = H(B)x_t = G(B)a_t$, where $H(B) = (1-vB)^{-1}(1-v)\hat{\Theta}(B)^{-1}\hat{\Phi}(B) = \sum_{j=0}^{\infty} h_j B^j$ with impulse response coefficients denoted by $\{h_j: j =$

0, 1, 2, . . . }, and $G(B) = \Phi^{-1}(B)\Theta(B)H(B)$. From the latter, we have $g_j - \phi_1 g_{j-1} - \phi_2 g_{j-2} - \cdots - \phi_p g_{j-p} = h_j - \theta_1 h_{j-1} - \theta_2 h_{j-2} - \cdots - \theta_q h_{j-q}$, from which, upon differentiating both sides with respect to ϕ_i and θ_i, we get

$$d_j^{\phi_i} - \phi_1 d_{j-1}^{\phi_i} - \cdots - \phi_p d_{j-p}^{\phi_i} - g_{j-i} = 0, \text{ and} \tag{A1}$$

$$d_j^{\theta_i} - \phi_1 d_{j-1}^{\theta_i} - \cdots - \phi_p d_{j-p}^{\theta_i} = -h_{j-i}. \tag{A2}$$

Here it is understood that $g_j = h_j = d_j^{\phi_i} = d_j^{\theta_i} = 0$ for $j < 0$. Viewing $d_j^{\phi_i}$ and $d_j^{\theta_i}$ as sequences in the index j (so that $B\, d_j^{\phi_i} \equiv d_{j-1}^{\phi_i}$), we can write (A1) as

$$\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right) d_j^{\phi_i} = g_{j-i}$$

or, equivalently, as

$$d_j^{\phi_i} = \Phi^{-1}(B)\, g_{j-i} = \Phi^{-1}(B) G(B)\, \delta_{j-i}, \tag{A3}$$

where δ_k denotes the Kronecker delta function (i.e., the impulse function defined as $\delta_k = 1$ for $k = 0$, and $\delta_k = 0$ for $k \neq 0$). In other words, $d_j^{\phi_i}$ is the impulse response of the filter $\Phi^{-1}(B)G(B)$, delayed by i timesteps. Proceeding similarly from (A2) gives

$$d_j^{\theta_i} = -\Phi^{-1}(B)\, h_{j-i} = -\Phi^{-1}(B)\, \Theta^{-1}(B)\Phi(B)\, g_{j-i} = -\Theta^{-1}(B)\, g_{j-i} = -\Theta^{-1}(B) G(B)\, \delta_{j-i}. \tag{A4}$$

From (A3) and (A4), we have

$$\hat{d}_j^{\phi_i} = \hat{\Phi}^{-1}(B)\hat{G}(B)\, \delta_{j-i} = \hat{\Phi}^{-1}(B) \frac{(1-v)\hat{\Phi}(B)\hat{\Theta}(B)}{(1-vB)\hat{\Theta}(B)\hat{\Phi}(B)} \delta_{j-i}$$

$$= (1-v)\hat{\Phi}^{-1}(B)(1-vB)^{-1} \delta_{j-i}, \text{ and} \tag{A5}$$

$$\hat{d}_j^{\theta_i} = -\hat{\Theta}^{-1}(B)\hat{G}(B)\, \delta_{j-i} = -\hat{\Theta}^{-1}(B) \frac{(1-v)\hat{\Phi}(B)\hat{\Theta}(B)}{(1-vB)\hat{\Theta}(B)\hat{\Phi}(B)} \delta_{j-i}$$

$$= -(1-v)\hat{\Theta}^{-1}(B)(1-vB)^{-1} \delta_{j-i}. \tag{A6}$$

The elements of $\mathbf{D}_j$ are derived similarly, as follows. Differentiating both sides of (A1) with respect to ϕ_l gives

$$D_j^{\phi_i,\phi_l} - \phi_1 D_{j-1}^{\phi_i,\phi_l} - \cdots - \phi_p D_{j-p}^{\phi_i,\phi_l} - d_{j-i}^{\phi_l} - d_{j-l}^{\phi_i} = 0,$$

which we write as

$$D_j^{\phi_i,\phi_l} = \Phi^{-1}(B)\left(d_{j-i}^{\phi_l} + d_{j-l}^{\phi_i}\right) = 2\Phi^{-2}(B)G(B)\delta_{j-i-l}, \tag{A7}$$

where in the right-most equality we have used Eq. (A3) for $d_{j-i}^{\phi_l}$. Likewise, differentiating both sides of (A1) with respect to θ_l gives

$$D_j^{\phi_i,\theta_l} - \phi_1 D_{j-1}^{\phi_i,\theta_l} - \cdots - \phi_p D_{j-p}^{\phi_i,\theta_l} - d_{j-i}^{\theta_l} = 0,$$

or

$$D_j^{\phi_i,\theta_l} = \Phi^{-1}(B)\, d_{j-i}^{\theta_l} = -\Phi^{-1}(B)\Theta^{-1}(B)G(B)\delta_{j-i-l}, \tag{A8}$$

where in the right-most equality we have used Eq. (A4) for $d_{j-i}^{\theta_l}$. Furthermore, it is straightforward to show that for all $1 \le i,l \le q$,

$$D_j^{\theta_i,\theta_l} = 0. \tag{A9}$$

Consequently, from (A7)—(A9), we have

$$\hat{D}_j^{\phi_i,\phi_l} = 2\hat{\Phi}^{-2}(B)\hat{G}(B)\delta_{j-i-l} = 2(1-v)\hat{\Phi}^{-2}(B)(1-vB)^{-1}\delta_{j-i-l}, \tag{A10}$$

$$\hat{D}_j^{\phi_i,\theta_l} = -\hat{\Phi}^{-1}(B)\hat{\Theta}^{-1}(B)\hat{G}(B)\delta_{j-i-l} = -(1-v)\hat{\Phi}^{-1}(B)\hat{\Theta}^{-1}(B)(1-vB)^{-1}\delta_{j-i-l}, \text{ and} \tag{A11}$$

$$\hat{D}_j^{\theta_i,\theta_l} = 0. \tag{A12}$$

Eqs. (A5), (A6), (A10), (A11), and (A12) show that all of the elements of $\hat{\mathbf{D}}_j$ and $\hat{\mathbf{d}}_j$ (j = 1, 2, . . .) are the impulse response coefficients of various AR transfer functions. For example, from Eq. (A11), $\hat{D}_j^{\phi_i,\theta_l}$ is the (j–i–l)th impulse response coefficient of the (p+q+1)th-order AR transfer function $-(1-v)\hat{\Phi}^{-1}(B)\hat{\Theta}^{-1}(B)(1-vB)^{-1}$. Hence, one could simulate the response of these transfer functions to a single impulse at time 0 to calculate the elements of $\hat{\mathbf{D}}_j$ and $\hat{\mathbf{d}}_j$, and then substitute these into Eq. (9) to calculate $E[\sigma_z^2 \mid \hat{\boldsymbol{\gamma}}]$. However, this is unnecessary. In the

remainder of this Appendix, we show that the special structures of $\hat{\mathbf{D}}_j$ and $\hat{\mathbf{d}}_j$ lead to the closed-form expression in Eq. (10).

To evaluate the term $\sum_{j=0}^{\infty}\hat{g}_j\hat{\mathbf{D}}_j$ in Eq. (9), notice that the impulse response coefficients of $\hat{G}(B)=(1-\nu)(1-\nu B)^{-1}$ are $\hat{g}_j = (1-\nu)\nu^j$, and introduce the notation $\{\eta\}_j$ to denote the jth impulse response coefficient of any ARMA transfer function $\eta(B) = \sum_{j=0}^{\infty}\{\eta\}_j B^j$. Thus, from Eq. (A10) we have

$$\sum_{j=0}^{\infty}\hat{g}_j\hat{D}_j^{\phi_i,\phi_l} = 2(1-\nu)\sum_{j=0}^{\infty}\nu^j\left\{\hat{\Phi}^{-2}\hat{G}\right\}_{j-i-l} = 2(1-\nu)\nu^{i+l}\sum_{k=0}^{\infty}\nu^k\left\{\hat{\Phi}^{-2}\hat{G}\right\}_k$$
$$= 2(1-\nu)\nu^{i+l}\hat{\Phi}^{-2}(\nu)\hat{G}(\nu) = \frac{2(1-\nu)\nu^{i+l}(1-\nu)}{\hat{\Phi}^2(\nu)(1-\nu^2)} = \frac{2(1-\nu)\nu^{i+l}}{\hat{\Phi}^2(\nu)(1+\nu)}.$$

The third equality follows by noting that for any ARMA transfer function $\eta(B) = \sum_{j=0}^{\infty}\{\eta\}_j B^j$, we define $\eta(\nu) = \sum_{j=0}^{\infty}\{\eta\}_j \nu^j$.

Similarly, from Eq. (A11) we have

$$\sum_{j=0}^{\infty}\hat{g}_j\hat{D}_j^{\phi_i,\theta_l} = -(1-\nu)\sum_{j=0}^{\infty}\nu^j\left\{\hat{\Phi}^{-1}\hat{\Theta}^{-1}\hat{G}\right\}_{j-i-l} = -(1-\nu)\nu^{i+l}\sum_{k=0}^{\infty}\nu^k\left\{\hat{\Phi}^{-1}\hat{\Theta}^{-1}\hat{G}\right\}_k$$
$$= -(1-\nu)\nu^{i+l}\hat{\Phi}^{-1}(\nu)\hat{\Theta}^{-1}(\nu)\hat{G}(\nu) = \frac{-(1-\nu)\nu^{i+l}(1-\nu)}{\hat{\Phi}(\nu)\hat{\Theta}(\nu)(1-\nu^2)} = \frac{-(1-\nu)\nu^{i+l}}{\hat{\Phi}(\nu)\hat{\Theta}(\nu)(1+\nu)}.$$

Inserting these into $tr\{\sum_{j=0}^{\infty}\hat{g}_j\hat{\mathbf{D}}_j\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\}$ in Eq. (9) gives

$$tr\{\sum_{j=0}^{\infty}\hat{g}_j\hat{\mathbf{D}}_j\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\} = \frac{1-\nu}{1+\nu}tr\left\{\left[\begin{array}{c|c}\dfrac{2\mathbf{V}_p\mathbf{V}_p^T}{\hat{\Phi}^2(\nu)} & \dfrac{-\mathbf{V}_p\mathbf{V}_q^T}{\hat{\Phi}(\nu)\hat{\Theta}(\nu)} \\ \hline \dfrac{-\mathbf{V}_q\mathbf{V}_p^T}{\hat{\Phi}(\nu)\hat{\Theta}(\nu)} & \mathbf{0}\end{array}\right]\left[\begin{array}{c|c}\boldsymbol{\Sigma}_{\Phi} & \boldsymbol{\Sigma}_{\Phi\Theta} \\ \hline \boldsymbol{\Sigma}_{\Phi\Theta}^T & \boldsymbol{\Sigma}_{\Theta}\end{array}\right]\right\}$$
$$= \frac{2(1-\nu)}{1+\nu}\left(\frac{\mathbf{V}_p^T\boldsymbol{\Sigma}_{\Phi}\mathbf{V}_p}{\hat{\Phi}^2(\nu)} - \frac{\mathbf{V}_p^T\boldsymbol{\Sigma}_{\Phi\Theta}\mathbf{V}_q}{\hat{\Phi}(\nu)\hat{\Theta}(\nu)}\right) \qquad \text{(A13)}$$

where $\mathbf{0}$ denotes a $q\times q$ matrix of zeros.

All that remains in deriving Eq. (10) is to simplify the expression $tr\{\sum_{j=0}^{\infty}\hat{\mathbf{d}}_j\hat{\mathbf{d}}_j^T\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\}$. For this, we use the following asymptotic expression for $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ (recall that $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\gamma}}}$) from Box *et al.*[10]:

$$\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = \frac{\sigma_a^2}{N}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} \tag{A14}$$

where $\boldsymbol{\Sigma}_{\mathbf{w}}$ is the steady-state covariance matrix of the random vector $\mathbf{w}_t$, defined as $\mathbf{w}_t = [u_t \;\; u_{t-1} \;\ldots\; u_{t-p+1} \;\; v_t \;\; v_{t-1} \;\ldots\; v_{t-q+1}]^T$, where the random processes u_t and v_t are defined as $u_t = \hat{\Phi}^{-1}(B)a_t$ and $v_t = -\hat{\Theta}^{-1}(B)a_t$.

Define $\mathbf{y}_t = (1-\nu)(1-\nu B)^{-1}\mathbf{w}_t = (1-\nu)\sum_{k=0}^{\infty}\nu^k\mathbf{w}_{t-k}$, and note that the elements of $\mathbf{y}_t$ are time-delayed versions of $(1-\nu)(1-\nu B)^{-1}\hat{\Phi}^{-1}(B)a_t$ and $-(1-\nu)(1-\nu B)^{-1}\hat{\Theta}^{-1}(B)a_t$. Because Eqs. (A5) and (A6) imply that $\hat{d}_j^{\phi_i}$ and $\hat{d}_j^{\theta_l}$ are the delayed impulse responses of the filters $(1-\nu)(1-\nu B)^{-1}\hat{\Phi}^{-1}(B)$ and $-(1-\nu)(1-\nu B)^{-1}\hat{\Theta}^{-1}(B)$, it follows that

$$\boldsymbol{\Sigma}_{\mathbf{y}} = \sigma_a^2\sum_{j=0}^{\infty}\hat{\mathbf{d}}_j\hat{\mathbf{d}}_j^T, \text{ or } \sum_{j=0}^{\infty}\hat{\mathbf{d}}_j\hat{\mathbf{d}}_j^T = \frac{1}{\sigma_a^2}\boldsymbol{\Sigma}_{\mathbf{y}}. \tag{A15}$$

But from $\mathbf{y}_t = (1-\nu)\sum_{k=0}^{\infty}\nu^k\mathbf{w}_{t-k}$, we also have

$$\boldsymbol{\Sigma}_{\mathbf{y}} = E[\mathbf{y}_t\mathbf{y}_t^T] = (1-\nu)^2\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\nu^j\nu^k\,E[\mathbf{w}_{t-j}\mathbf{w}_{t-k}^T].$$

Combining this with (A14) and (A15) gives

$$\begin{aligned} tr\{\textstyle\sum_{j=0}^{\infty}\hat{\mathbf{d}}_j\hat{\mathbf{d}}_j^T\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\} &= tr\{\frac{1}{\sigma_a^2}\boldsymbol{\Sigma}_{\mathbf{y}}\frac{\sigma_a^2}{N}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\} \\ &= \frac{(1-\nu)^2}{N}tr\{\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\nu^j\nu^k\,E[\mathbf{w}_{t-j}\mathbf{w}_{t-k}^T]\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\} \\ &= \frac{(1-\nu)^2}{N}tr\{\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\nu^j\nu^k\,E[\mathbf{w}_t\mathbf{w}_{t-|k-j|}^T]\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\} \end{aligned} \tag{A16}$$

To evaluate this, we write $\mathbf{w}_t$ as

$\mathbf{w}_t = \mathbf{A}\mathbf{w}_{t-1} + \mathbf{b}a_t$,

where $\mathbf{A} = \left[\begin{array}{ccccc:ccccc} \hat{\phi}_1 & \hat{\phi}_2 & \cdots & \cdots & \hat{\phi}_p & 0 & 0 & \cdots & \cdots & 0 \\ 1 & 0 & & & 0 & 0 & 0 & & & \vdots \\ 0 & 1 & 0 & & \vdots & \vdots & & \ddots & & 0 \\ \vdots & & \ddots & \ddots & 0 & \vdots & & & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \hdashline 0 & 0 & \cdots & \cdots & 0 & \hat{\theta}_1 & \hat{\theta}_2 & \cdots & \cdots & \hat{\theta}_q \\ 0 & 0 & & & \vdots & 1 & 0 & & & 0 \\ \vdots & & \ddots & & 0 & 0 & 1 & 0 & & \\ \vdots & & & \ddots & \vdots & \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & 0 \end{array}\right]$, and $\mathbf{b} = \left[\begin{array}{c} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \\ \hdashline -1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{array}\right]$

Then for any j and k, we have $\mathbf{w}_t = \mathbf{A}^{|k-j|}\mathbf{w}_{t-|k-j|}$ plus a function of $\{a_t, a_{t-1}, \ldots, a_{t-|k-j|+1}\}$ that is independent of $\mathbf{w}_{t-|k-j|}$. Consequently,

$$E[\mathbf{w}_t \mathbf{w}_{t-|k-j|}^T] = E[\mathbf{A}^{|k-j|}\mathbf{w}_{t-|k-j|}\, \mathbf{w}_{t-|k-j|}^T] = \mathbf{A}^{|k-j|}\boldsymbol{\Sigma}_{\mathbf{w}}. \tag{A17}$$

Combining (A16) and (A17) gives

$$\begin{aligned} tr\{\textstyle\sum_{j=0}^{\infty}\hat{\mathbf{d}}_j\hat{\mathbf{d}}_j^T\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\} &= \frac{(1-\nu)^2}{N}\, tr\{\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\nu^j\nu^k\, \mathbf{A}^{|k-j|}\} \\ &= \frac{(1-\nu)^2}{N}\, tr\{2\sum_{j=0}^{\infty}\sum_{k=j}^{\infty}\nu^j\nu^k\, \mathbf{A}^{|k-j|} - \sum_{j=0}^{\infty}\nu^{2j}\, \mathbf{I}\} \\ &= \frac{(1-\nu)^2}{N}\, [2tr\{\sum_{j=0}^{\infty}\nu^{2j}\sum_{l=0}^{\infty}(\nu\mathbf{A})^l\} - (p+q)\sum_{j=0}^{\infty}\nu^{2j}] \\ &= \frac{(1-\nu)}{(1+\nu)N}\, [2tr\{[\mathbf{I}-\nu\mathbf{A}]^{-1}\} - (p+q)] \end{aligned} \tag{A18}$$

where $\mathbf{I}$ denotes the $(p+q)\times(p+q)$ identity matrix.

Because of the trace operation, we only need the diagonal elements of the matrix $[\mathbf{I}-\nu\mathbf{A}]^{-1}$ in (A18), which we derive as follows. Partition

$$\mathbf{A} = \left[\begin{array}{c:c} \mathbf{A}_{\Phi} & 0 \\ \hdashline 0 & \mathbf{A}_{\Theta} \end{array}\right],$$

and define $\mathbf{M} = [\mathbf{I} - \nu\mathbf{A}_{\boldsymbol{\Phi}}]^{-1}$, which is the upper-left block of $[\mathbf{I} - \nu\mathbf{A}]^{-1}$. Let e_i = [0 0 …0 1 0 …0]^T denote a column vector of zeros with a one as its ith element. From the structure of $\mathbf{A}_{\boldsymbol{\Phi}}$, we have

$$\mathrm{e}_1^T \mathbf{A}_{\boldsymbol{\Phi}} = \sum_{j=1}^{p} \hat{\phi}_j e_j^T \,, \tag{A19}$$

$$e_i^T \mathbf{A}_{\boldsymbol{\Phi}} = e_{i-1}^T, \qquad \text{for } 2 \le i \le p, \tag{A20}$$

and

$$\mathrm{e}_i^T \mathbf{I}\, \mathrm{e}_j = \begin{cases} 1: & i = j \\ 0: & i \ne j \end{cases} \tag{A21}$$

where $\mathbf{I}$ denotes the $p \times p$ identity matrix.

Denote by $\mathbf{M}_{ii}$ the ith diagonal element of $\mathbf{M}$. Using (A19)–(A21) and repeatedly substituting $\mathbf{M} = \mathbf{I} + \nu\mathbf{A}_{\boldsymbol{\Phi}}\mathbf{M}$ gives

$$\begin{aligned} \mathbf{M}_{ii} &= \mathrm{e}_i^T \mathbf{M}\, \mathrm{e}_i = \mathrm{e}_i^T \left[\mathbf{I} + \nu\mathbf{A}_{\boldsymbol{\Phi}}\mathbf{M}\right]\mathrm{e}_i = 1 + \nu \mathrm{e}_{i-1}^T \mathbf{M}\, \mathrm{e}_i = 1 + \nu \mathrm{e}_{i-1}^T \left[\mathbf{I} + \nu\mathbf{A}_{\boldsymbol{\Phi}}\mathbf{M}\right]\mathrm{e}_i \\ &= 1 + \nu^2 \mathrm{e}_{i-2}^T \mathbf{M}\, \mathrm{e}_i = \cdots = 1 + \nu^i \left(\sum_{j=1}^{p} \hat{\phi}_j e_j^T \right) \mathbf{M}\, e_i \,, \end{aligned} \tag{A22}$$

which holds for $1 \le i \le p$. In a similar manner, repeatedly substituting $\mathbf{M} = \mathbf{I} + \nu\mathbf{A}_{\boldsymbol{\Phi}}\mathbf{M}$ gives, for $j < i$,

$$\begin{aligned} \mathrm{e}_j^T \mathbf{M}\, \mathrm{e}_i &= \mathrm{e}_j^T \left[\mathbf{I} + \nu\mathbf{A}_{\boldsymbol{\Phi}}\mathbf{M}\right]\mathrm{e}_i = \nu \mathrm{e}_{j-1}^T \mathbf{M}\, \mathrm{e}_i = \nu \mathrm{e}_{j-1}^T \left[\mathbf{I} + \nu\mathbf{A}_{\boldsymbol{\Phi}}\mathbf{M}\right]\mathrm{e}_i \\ &= \nu^2 \mathrm{e}_{j-2}^T \mathbf{M}\, \mathrm{e}_i = \cdots = \nu^j \left(\sum_{k=1}^{p} \hat{\phi}_k e_k^T \right) \mathbf{M}\, e_i = \nu^{j-i} \left(\mathbf{M}_{ii} - 1 \right), \end{aligned}$$

and for $j \ge i$,

$$\begin{aligned} \mathrm{e}_j^T \mathbf{M}\, \mathrm{e}_i &= \mathrm{e}_j^T \left[\mathbf{I} + \nu\mathbf{A}_{\boldsymbol{\Phi}}\mathbf{M}\right]\mathrm{e}_i = \nu \mathrm{e}_{j-1}^T \mathbf{M}\, \mathrm{e}_i = \nu \mathrm{e}_{j-1}^T \left[\mathbf{I} + \nu\mathbf{A}_{\boldsymbol{\Phi}}\mathbf{M}\right]\mathrm{e}_i \\ &= \nu^2 \mathrm{e}_{j-2}^T \mathbf{M}\, \mathrm{e}_i = \cdots = \nu^{j-i} \mathrm{e}_i^T \mathbf{M}\, \mathrm{e}_i = \nu^{j-i} \mathbf{M}_{ii} \,. \end{aligned}$$

Substituting these into (A22) gives

$$\begin{aligned} \mathbf{M}_{ii} &= 1 + \nu^i \sum_{j=1}^{i-1} \hat{\phi}_j \nu^{j-i} \left(\mathbf{M}_{ii} - 1 \right) + \nu^i \sum_{j=i}^{p} \hat{\phi}_j \nu^{j-i}\, \mathbf{M}_{ii} \\ &= 1 + \nu^i \sum_{j=1}^{p} \hat{\phi}_j \nu^{j-i}\, \mathbf{M}_{ii} - \nu^i \sum_{j=i}^{i-1} \hat{\phi}_j \nu^{j-i} \end{aligned}$$

$$= 1 + \sum_{j=1}^{p} \hat{\phi}_j \nu^j \, \mathbf{M}_{ii} - \sum_{j=i}^{i-1} \hat{\phi}_j \nu^j$$

Thus, the *i*th diagonal element of $[\mathbf{I} - \nu \mathbf{A}_{\Phi}]^{-1}$ is

$$\mathbf{M}_{ii} = \frac{1 - \sum_{j=1}^{i-1} \hat{\phi}_j \nu^j}{1 - \sum_{j=1}^{p} \hat{\phi}_j \nu^j} = \frac{1 - \sum_{j=1}^{i-1} \hat{\phi}_j \nu^j}{\hat{\Phi}(\nu)},$$

where it is understood that the summation in the numerator is zero for $i = 1$.

Notice that the submatrix $\mathbf{A}_{\Theta}$ is of identical structure to $\mathbf{A}_{\Phi}$, except that the AR coefficients are replaced by the MA coefficients. In light of this, a straightforward repetition of the preceding derivation gives

$$\frac{1 - \sum_{j=1}^{i-1} \hat{\theta}_j \nu^j}{\hat{\Theta}(\nu)}$$

as the *i*th diagonal element of $[\mathbf{I} - \nu \mathbf{A}_{\Theta}]^{-1}$, $i = 1, 2, \ldots, q$. Using these diagonal values gives

$$2tr\{[\mathbf{I} - \nu \mathbf{A}]^{-1}\} - (p+q)$$

$$= \frac{2\left\{1 + \left(1 - \hat{\phi}_1 \nu\right) + \left(1 - \hat{\phi}_1 \nu - \hat{\phi}_2 \nu^2\right) + \cdots + \left(1 - \hat{\phi}_1 \nu - \hat{\phi}_2 \nu^2 - \cdots - \hat{\phi}_{p-1} \nu^{p-1}\right)\right\}}{\hat{\Phi}(\nu)} - p$$

$$+ \frac{2\left\{1 + \left(1 - \hat{\theta}_1 \nu\right) + \left(1 - \hat{\theta}_1 \nu - \hat{\theta}_2 \nu^2\right) + \cdots + \left(1 - \hat{\theta}_1 \nu - \hat{\theta}_2 \nu^2 - \cdots - \hat{\theta}_{q-1} \nu^{q-1}\right)\right\}}{\hat{\Theta}(\nu)} - q$$

$$= p + q + \frac{2\left[\hat{\phi}_1, 2\hat{\phi}_2, 3\hat{\phi}_3, \cdots, p\hat{\phi}_p\right]\mathbf{V}_p}{\hat{\Phi}(\nu)} + \frac{2\left[\hat{\theta}_1, 2\hat{\theta}_2, 3\hat{\theta}_3, \cdots, q\hat{\theta}_q\right]\mathbf{V}_q}{\hat{\Theta}(\nu)}.$$

Substituting this into Eq. (A18) gives

$$tr\{\textstyle\sum_{j=0}^{\infty} \hat{\mathbf{d}}_j \hat{\mathbf{d}}_j^T \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\} =$$

$$\frac{(1-\nu)}{(1+\nu)N}\left[p + q + \frac{2\left[\hat{\phi}_1, 2\hat{\phi}_2, 3\hat{\phi}_3, \cdots, p\hat{\phi}_p\right]\mathbf{V}_p}{\hat{\Phi}(\nu)} + \frac{2\left[\hat{\theta}_1, 2\hat{\theta}_2, 3\hat{\theta}_3, \cdots, q\hat{\theta}_q\right]\mathbf{V}_q}{\hat{\Theta}(\nu)}\right]$$

Finally, substituting this and Eq. (A13) into Eq. (9) gives Eq. (10).

TABLES

$\hat{\phi}_1$	$\hat{\theta}_1$	N	Proposed		A		A&L, α=0.2	
			CL	RI	CL	RI	CL	RI
0.9	0.6	50	0.5517	31.8%	0.4827	15.3%	0.5252	30.2%
		100	0.4898	17.0%	0.4519	7.9%	0.5114	22.1%
		200	0.4556	8.8%	0.4356	4.0%	0.4861	16.1%
		500	0.4339	3.6%	0.4256	1.6%	0.4625	10.4%
0.9	0.4	50	0.5413	29.3%	0.4775	14.0%	0.5443	30.0%
		100	0.4839	15.6%	0.4491	7.2%	0.5107	22.0%
		200	0.4525	8.1%	0.4342	3.7%	0.4856	16.0%
		500	0.4326	3.3%	0.4250	1.5%	0.4621	10.4%
0.8	0.6	50	0.5455	30.3%	0.4624	10.4%	0.5335	27.4%
		100	0.4863	16.1%	0.4411	5.3%	0.5026	20.0%
		200	0.4538	8.4%	0.4301	2.7%	0.4796	14.5%
		500	0.4331	3.4%	0.4233	1.1%	0.4582	9.4%
0.8	0.4	50	0.5182	23.8%	0.4570	9.1%	0.5301	26.6%
		100	0.4711	12.5%	0.4383	4.7%	0.5001	19.4%
		200	0.4457	6.4%	0.4286	2.4%	0.4777	14.1%
		500	0.4297	2.6%	0.4227	1.0%	0.4569	9.1%

Table 1 Comparison of the extent to which the three methods widen the control limits for λ = 0.05 (L = 2.615). RI denotes the percentage increase in the control limits, relative to the standard limits $\pm L\,\hat{\sigma}_a\,(1-\nu)^{1/2}(1+\nu)^{-1/2} = \pm 0.4187$ that result if parameter uncertainty is neglected.

$\hat{\phi}_1$	$\hat{\theta}_1$	N	Proposed		A		A&L, $\alpha = 0.2$	
			CL	RI	CL	RI	CL	RI
0.9	0.6	50	0.7715	19.5%	0.7239	12.1%	0.7378	14.3%
		100	0.7113	10.2%	0.6859	6.2%	0.7121	10.3%
		200	0.6792	5.2%	0.6660	3.2%	0.6932	7.4%
		500	0.6592	2.1%	0.6538	1.3%	0.6761	4.7%
0.9	0.4	50	0.7648	18.5%	0.7169	11.0%	0.7378	14.3%
		100	0.7077	9.6%	0.6821	5.7%	0.7121	10.3%
		200	0.6774	4.9%	0.6641	2.9%	0.6932	7.4%
		500	0.6585	2.0%	0.6531	1.2%	0.6761	4.7%
0.8	0.6	50	0.7753	20.1%	0.7042	9.1%	0.7355	13.9%
		100	0.7134	10.5%	0.6755	4.6%	0.7103	10.0%
		200	0.6803	5.4%	0.6607	2.3%	0.6920	7.2%
		500	0.6597	2.2%	0.6517	0.9%	0.6753	4.6%
0.8	0.4	50	0.7537	16.7%	0.6969	8.0%	0.7344	13.8%
		100	0.7017	8.7%	0.6717	4.0%	0.7095	9.9%
		200	0.6742	4.4%	0.6588	2.0%	0.6914	7.1%
		500	0.6572	1.8%	0.6509	0.8%	0.6749	4.5%

Table 2 Comparison of the extent to which the three methods widen the control limits for $\lambda = 0.10$ ($L = 2.814$). RI denotes the percentage increase in the control limits, relative to the standard limits $\pm L\,\hat{\sigma}_a\,(1-\nu)^{1/2}(1+\nu)^{-1/2} = \pm 0.6456$ that result if parameter uncertainty is neglected.

$\hat{\phi}_1$	$\hat{\theta}_1$	N	Proposed		A		A&L, $\alpha = 0.2$	
			CL	RI	CL	RI	CL	RI
0.9	0.6	50	1.0889	10.3%	1.0724	8.6%	1.0797	9.4%
		100	1.0394	5.3%	1.0308	4.4%	1.0535	6.7%
		200	1.0137	2.7%	1.0093	2..2%	1.0346	4.8%
		500	0.9980	1.1%	0.9962	0.9%	1.0175	3.1%
0.9	0.4	50	1.0853	9.9%	1.0642	7.8%	1.0786	9.2%
		100	1.0375	5.1%	1.0265	4.0%	1.0527	6.6%
		200	1.0127	2.6%	1.0071	2.0%	1.0340	4.7%
		500	0.9976	1.0%	0.9953	0.8%	1.0171	3.0%
0.8	0.6	50	1.0902	10.4%	1.0579	7.1%	1.0806	9.4%
		100	1.0400	5.3%	1.0232	3.6%	1.0541	6.8%
		200	1.0140	2.7%	1.0054	1.8%	1.0350	4.8%
		500	0.9981	1.1%	0.9946	0.7%	1.0177	3.1%
0.8	0.4	50	1.0820	9.6%	1.0495	6.3%	1.0806	9.4%
		100	1.0358	4.9%	1.0189	3.2%	1.0541	6.8%
		200	1.0118	2.5%	1.0032	1.6%	1.0350	4.8%
		500	0.9972	1.0%	0.9937	0.6%	1.0177	3.1%

Table 3 Comparison of the extent to which the three methods widen the control limits for $\lambda = 0.20$ ($L = 2.962$). RI denotes the percentage increase in the control limits, relative to the standard limits $\pm L\,\hat{\sigma}_a\,(1-\nu)^{1/2}(1+\nu)^{-1/2} = \pm 0.9873$ that result if parameter uncertainty is neglected.

		mean shift magnitude (in units of σ_a)					
chart	control limits	0	1	2	3	4	5
EWMA (λ = 0.1)	0.202 (Standard)	500	101	23.8	8.11	3.54	2.22
EWMA (λ = 0.1)	0.212 (Proposed)	729	129	27.7	9.24	4.00	2.39
EWMA (λ = 0.1)	0.208 (A method)	612	115	25.5	8.58	3.79	2.30
EWMA (λ = 0.1)	0.237 (A&L method)	2020	247	43.3	13.3	5.29	2.89
Shewhart	0.967 (Standard)	500	366	168	49.1	7.83	1.38

Table 4 ARL values for various size mean shifts for the ARMA(1,1) example when the ARMA parameters coincide with their estimates.

$\hat{\phi}_1$	$\hat{\theta}_1$	N	A&L, α=0.1		A&L, α=0.2		A&L, α=0.3	
			CL	RI	CL	RI	CL	RI
0.9	0.6	50	0.6008	43.5%	0.5252	30.2%	0.5013	19.7%
		100	0.5537	32.2%	0.5114	22.1%	0.4786	14.3%
		200	0.5178	23.7%	0.4861	16.1%	0.4619	10.3%
		500	0.4838	15.5%	0.4625	10.4%	0.4465	6.6%
0.9	0.4	50	0.5995	43.2%	0.5443	30.0%	0.5007	19.6%
		100	0.5527	32.0%	0.5107	22.0%	0.4781	14.2%
		200	0.5171	23.5%	0.4856	16.0%	0.4615	10.2%
		500	0.4833	15.4%	0.4621	10.4%	0.4463	6.6%
0.8	0.6	50	0.5846	39.6%	0.5335	27.4%	0.4934	17.8%
		100	0.5413	29.3%	0.5026	20.0%	0.4728	12.9%
		200	0.5085	21.4%	0.4796	14.5%	0.4576	9.3%
		500	0.4775	14.0%	0.4582	9.4%	0.4437	6.0%
0.8	0.4	50	0.5799	38.5%	0.5301	26.6%	0.4911	17.3%
		100	0.5377	28.4%	0.5001	19.4%	0.4711	12.5%
		200	0.5058	20.8%	0.4777	14.1%	0.4564	9.0%
		500	0.4756	13.6%	0.4569	9.1%	0.4429	5.8%

Table 5. The effect of choice of α on the control limits of the A&L method for λ = 0.05.

FIGURES

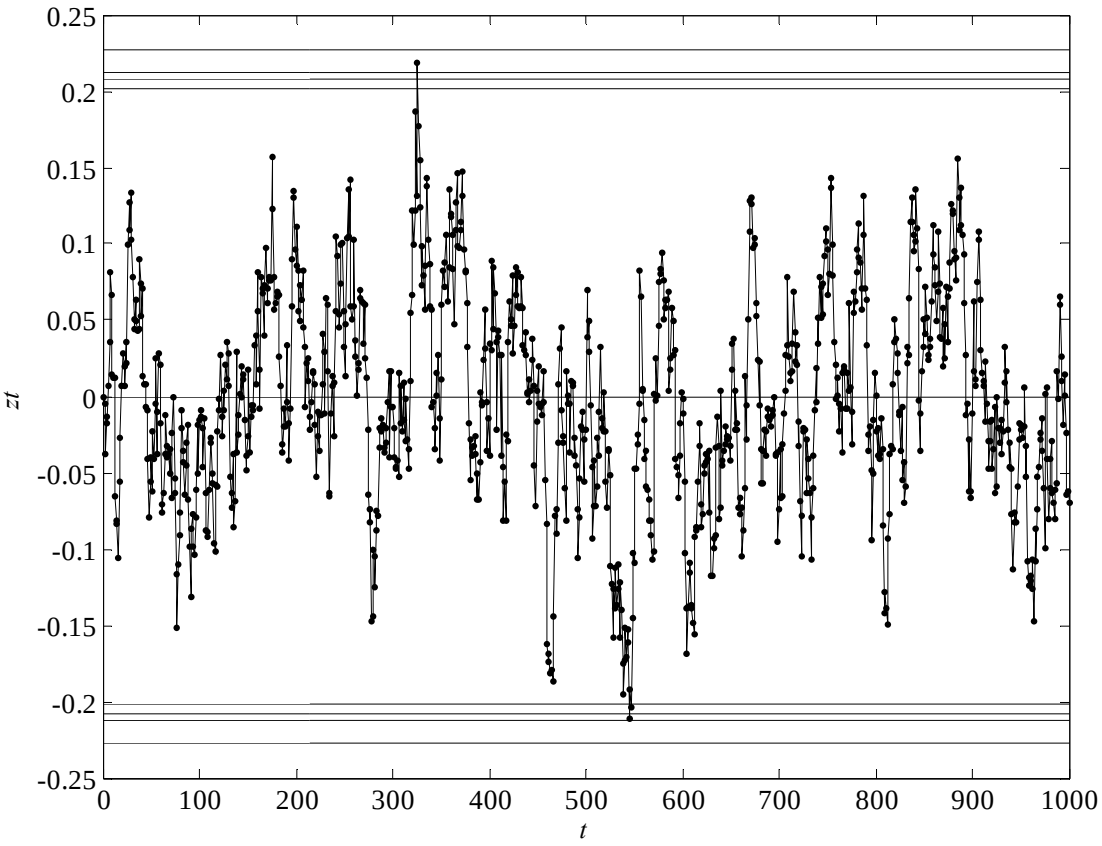

Figure 1. Simulated ARMA(1,1) chemical process example demonstrating false alarm occurrences when there are no parameter errors for an EWMA chart with four different sets of control limits: ±0.202 (standard), ±0.208 (Apley[26]), ±0.212 (proposed), and ±0.227 (Apley and Lee[27], with α=0.2).

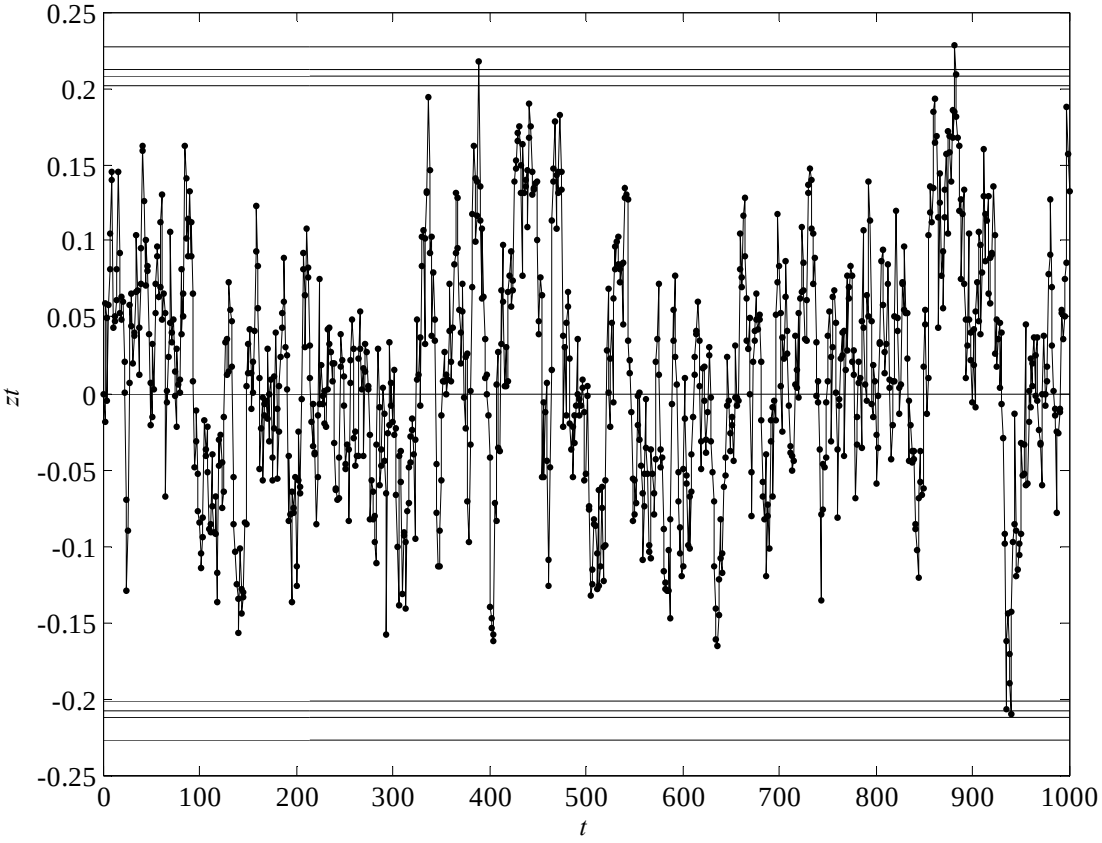

Figure 2. Simulated ARMA(1,1) chemical process example demonstrating false alarm occurrences when there are parameter errors for an EWMA chart with four different sets of control limits: ±0.202 (standard), ±0.208 (Apley[26]), ±0.212 (proposed), and ±0.227 (Apley and Lee[27], with α=0.2).