

Design of Exponentially Weighted Moving Average Control Charts for Autocorrelated Processes With Model Uncertainty

Daniel W. APLEY and Hyun Cheol LEE

Department of Industrial Engineering
Texas A&M University
College Station, TX 77843
(apley@tamu.edu)

Residual-based control charts are popular methods for statistical process control of autocorrelated processes. To implement these methods, a time series model of the process is required. The model must be estimated from data, in practice, and model estimation errors can cause the actual in-control average run length to differ substantially from the desired value. This article develops a method for designing residual-based exponentially weighted moving average (EWMA) charts under consideration of the uncertainty in the estimated model parameters. The resulting EWMA control limits are widened by an amount that depends on a number of factors, including the level of model uncertainty.

KEY WORDS: Autoregressive moving average model; Exponentially weighted moving average chart; Mean shift detection; Residual-based control chart

1. INTRODUCTION

Statistical process control (SPC) is widely used to monitor and improve quality in industrial processes. Traditional SPC techniques are based on the assumption that process data are independent. Significant advances in measurement and data collection technology—particularly in the area of in-process sensing—have created the potential for much more frequent inspection. As a result, autocorrelated data are now common (Montgomery and Woodall 1997). The run-length properties of traditional SPC methods like cumulative sum (CUSUM) and \bar{X} charts are strongly affected by data autocorrelation, and the in-control average run length (ARL) can be much shorter than intended if the autocorrelation is positive (Johnson and Bagshaw 1974; Vasilopoulos and Stamboulis 1978). Consequently, there has been considerable research in recent years on designing SPC procedures suitable for autocorrelated processes (see, e.g., Montgomery and Woodall 1997; Lu and Reynolds 1999; and the references therein).

The most widely investigated methods for SPC of autocorrelated processes are residual-based control charts (e.g., Alwan and Roberts 1988; Apley and Shi 1999; Berthouex, Hunter, and Pallesen 1978; English, Krishnamurthi, and Sastri 1991; Lin and Adams 1996; Lu and Reynolds 1999; Montgomery and Mastrangelo 1991; Runger, Willemain, and Prabhu 1995; Superville and Adams 1994; Vander Wiel 1996; Wardell, Moskowitz, and Plante 1994). One usually assumes that the process data x_t (t is a time index) follow an autoregressive moving average (ARMA) model with AR order p and MA order q , denoted by ARMA(p, q). Using standard time series notation (see Box, Jenkins, and Reinsel 1994) with the backward shift operator B defined such that $Bx_t = x_{t-1}$, an ARMA model can be written as

$$x_t = \frac{\Theta(B)}{\Phi(B)} a_t, \quad (1)$$

where $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$, $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$, and a_t is an independently identically distributed (iid), 0-mean sequence of random shocks with variance σ_a^2 . It is assumed that the in-control process mean has been subtracted, so that x_t is 0-mean until a shift occurs. For notational convenience, the results in this article are derived for ARMA processes, although a straightforward extension to autoregressive integrated moving average (ARIMA) processes is discussed in Section 2.

The basic idea behind residual-based charts is to directly monitor the residuals (the one-step-ahead prediction errors) generated via $e_t = \Theta^{-1}(B)\Phi(B)x_t$. From (1), e_t is exactly the iid sequence a_t after any initial transients have died out. Thus traditional Shewhart, CUSUM, and exponentially weighted moving average (EWMA) control charts can be applied to the uncorrelated residuals with well-understood in-control run length properties. In practice, however, the model parameters $\{\phi_i : i = 1, 2, \dots, p\}$, $\{\theta_i : i = 1, 2, \dots, q\}$, and σ_a^2 must always be estimated from process data. Using the “ $\hat{\cdot}$ ” symbol to denote estimates of the parameters and the resulting ARMA polynomials, the residuals generated via the estimated model behave as the ARMA($p + q, p + q$) process,

$$e_t = \frac{\hat{\Phi}(B)}{\hat{\Theta}(B)} x_t = \frac{\hat{\Phi}(B)\Theta(B)}{\hat{\Theta}(B)\Phi(B)} a_t, \quad (2)$$

and are no longer iid.

Suppose that an EWMA of the form $y_t = (1 - \lambda)y_{t-1} + \lambda e_t$, for some EWMA parameter $0 < \lambda \leq 1$, is applied to the residuals. Defining $\nu = 1 - \lambda$, the EWMA statistic, y_t , can be written as the ARMA($p + q + 1, p + q$) process

$$y_t = \frac{1 - \nu}{1 - \nu B} e_t = \frac{(1 - \nu)\hat{\Phi}(B)\Theta(B)}{(1 - \nu B)\hat{\Theta}(B)\Phi(B)} a_t. \quad (3)$$

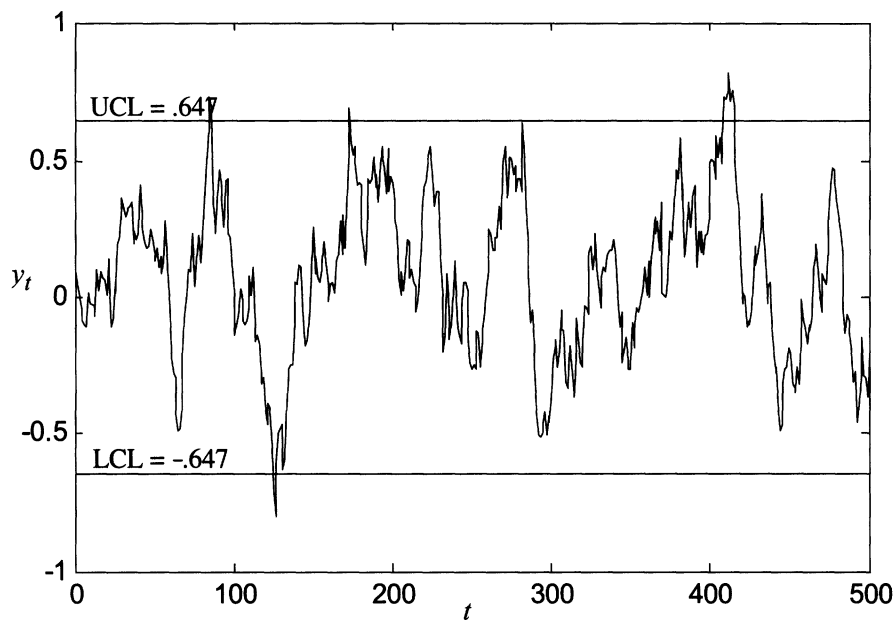


Figure 1. Sample EWMA Chart for an In-Control AR(1) Process With ϕ_1 Underestimated. The desired in-control ARL is 500, whereas the actual ARL is much lower due to frequent false alarms.

For a specified λ , the typical EWMA design procedure is to set the upper control limit (UCL) and lower control limit (LCL) on y_t at (Lu and Reynolds 1999)

$$\{\text{LCL}, \text{UCL}\} = \pm L \hat{\sigma}_y, \quad (4)$$

where $\hat{\sigma}_y = \hat{\sigma}_a(1 - \nu)^{1/2}(1 + \nu)^{-1/2}$ is the steady-state standard deviation of y_t assuming that the estimated model is perfect (Montgomery 2001), and the constant L is chosen to provide a desired in-control ARL. Lucas and Saccucci (1990) developed tables for monitoring iid data that give the values of L that result in several in-control ARL values for various choices of λ . To improve the sensitivity to mean shifts that occur when the control chart is first initiated, time-varying control limits that gradually widen to the steady-state limits (4) can also be used (Montgomery 2001). This article considers only constant steady-state control limits.

Let σ_y^2 denote the actual variance of the EWMA statistic (3), which is a function of the true parameters and their estimates. Because the EWMA is a weighted average of the past residuals, residual autocorrelation due to estimation errors can have a substantial effect on σ_y and the resulting in-control ARL (Adams and Tseng 1998; Apley and Shi 1999; Lu and Reynolds 1999). If the true and estimated parameters are such that the residual autocorrelation is positive, then σ_y generally will be larger than $\hat{\sigma}_y$, the in-control ARL will be shorter than intended, and the control chart may be plagued with frequent false alarms.

To illustrate the effects of modeling errors, suppose that x_t is an AR(1) process with $\phi_1 = .9$ and $\sigma_a^2 = 1.0$ and that the estimated parameters are $\hat{\phi}_1 = .85$ and $\hat{\sigma}_a^2 = 1.0$. Using an EWMA with $\lambda = .1$ and treating the estimates as perfect, the assumed EWMA variance is $\hat{\sigma}_y^2 = \hat{\sigma}_a^2(1 - \nu)(1 + \nu)^{-1} = .053$. For a desired in-control ARL of 500, $L = 2.814$ (Lucas and Saccucci 1990) and the control limits $\pm L \hat{\sigma}_y = \pm .647$ would be used. Using (3) and any of the methods for calculating the variance of an ARMA process discussed by Box et al. (1994), however, it

can be shown that the actual EWMA variance is $\sigma_y^2 = .084$ —roughly 60% larger than the assumed variance. If the control limits based on the assumed variance are used, then Monte Carlo simulation (refer to Sec. 4 for details) reveals that the actual in-control ARL is approximately 165, which is substantially shorter than intended. Figure 1, which shows the EWMA statistic for 500 simulated observations with the $\pm .647$ control limits, illustrates the frequent false alarms that result in this situation.

To account for uncertainty in the estimated parameters and guard against a situation in which the in-control ARL is substantially shorter than desired, a reasonable precaution is to use control limits that are wider than those used when the model is assumed to be perfect. This article presents a method for widening the EWMA control limits based on the following “worst-case” design approach. For a specified λ and a given set of ARMA parameter estimates, (3) implies that σ_y is a function of the true, unknown parameters. Considering the uncertainty in the true parameters, Section 2 derives an approximate upper one-sided $1 - \alpha$ confidence interval for σ_y for some user-selected $0 < \alpha < 1$. Let $\sigma_{y,\alpha}$ denote the upper boundary of this confidence interval, which can be viewed as a worst-case (maximum) value for the true EWMA standard deviation.

The proposed method is to monitor the EWMA statistic (3), but to use the worst-case control limits,

$$\{\text{LCL}, \text{UCL}\} = \pm L \sigma_{y,\alpha}, \quad (5)$$

instead of the standard control limits (4). Section 3 discusses guidelines for selecting the design parameters L , λ , and α . L can be chosen so that the worst-case ARL (roughly, the in-control ARL that would result if σ_y assumed its worst-case value) approximately equals some desired ARL value specified by the user. Widened control limits will inevitably increase the out-of-control ARL for any size mean shift and reduce the power of the chart. Section 4 discusses this drawback of the worst-case

design approach and illustrates this with examples. It also discusses sample size requirements and compares the EWMA with a Shewhart individual chart, which is less powerful for small to moderate mean shifts but more robust to modeling errors.

2. THE WORST-CASE EWMA VARIANCE

The EWMA statistic (3) can be rewritten as

$$y_t = (1 - \nu)G(B)a_t = (1 - \nu) \sum_{j=0}^{\infty} G_j a_{t-j},$$

where

$$G(B) = \frac{\hat{\Phi}(B)\Theta(B)}{(1 - \nu B)\hat{\Theta}(B)\Phi(B)} = \sum_{j=0}^{\infty} G_j B^j$$

and $\{G_j : j = 0, 1, 2, \dots\}$ are the impulse response coefficients of the ARMA($p + q + 1, p + q$) transfer function $G(B)$. For a fixed set of ARMA parameters and their estimates, the EWMA variance is (Box et al. 1994)

$$\sigma_y^2 = \sigma_a^2(1 - \nu)^2 \sum_{j=0}^{\infty} G_j^2. \tag{6}$$

Define the ARMA parameter vector $\boldsymbol{\gamma} = [\phi_1 \phi_2 \dots \phi_p \theta_1 \theta_2 \dots \theta_q \sigma_a^2]^T$, and let $\hat{\boldsymbol{\gamma}}$ denote a point estimate. To find an approximate confidence interval for σ_y , we use a first-order Taylor approximation of the ratio $\sigma_y^2/\hat{\sigma}_y^2$ about $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}$. If the parameter error vector is defined as $\tilde{\boldsymbol{\gamma}} = \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}$, then the first-order Taylor approximation is

$$\sigma_y^2/\hat{\sigma}_y^2 \cong 1 + \mathbf{V}^T \tilde{\boldsymbol{\gamma}}, \tag{7}$$

where

$$\mathbf{V} = \left[\begin{array}{ccccccc} -2\nu & -2\nu^2 & \dots & -2\nu^p & 2\nu & 2\nu^2 & \dots & 2\nu^q & -\sigma_a^{-2} \end{array} \right]^T,$$

with $\Phi(\nu) = \Phi(B)|_{B=\nu} = 1 - \phi_1\nu - \phi_2\nu^2 - \dots - \phi_p\nu^p$, and $\Theta(\nu) = \Theta(B)|_{B=\nu} = 1 - \theta_1\nu - \theta_2\nu^2 - \dots - \theta_q\nu^q$. The Taylor approximation (7) is derived in Appendix A for the special case of a first-order ARMA process and was given by Apley (2003) for the more general ARMA(p, q) case.

Let N denote the number of observations in the sample used to estimate the ARMA parameters. For most estimation methods, the distribution of $\tilde{\boldsymbol{\gamma}}$ for large N is approximately multivariate normal with mean 0 and some covariance matrix $\boldsymbol{\Sigma}_\boldsymbol{\gamma}$ that is inversely proportional to N (Box et al. 1994; Brockwell and Davis 1991). Commercial statistical software packages for ARMA modeling often provide an estimate $\hat{\boldsymbol{\Sigma}}_\boldsymbol{\gamma}$ of the covariance along with the parameter estimates. Alternatively, the method outlined in Appendix B may be used to calculate $\hat{\boldsymbol{\Sigma}}_\boldsymbol{\gamma}$ when only the parameter estimates are available. Closed-form expressions for $\hat{\boldsymbol{\Sigma}}_\boldsymbol{\gamma}$ are also provided in Appendix B for the special case of first-order ARMA processes.

Using the multivariate normal approximation to the distribution of $\tilde{\boldsymbol{\gamma}}$, the ratio $\sigma_y^2/\hat{\sigma}_y^2$ in (7) is approximately normally distributed with mean 1 and variance $\mathbf{V}^T \boldsymbol{\Sigma}_\boldsymbol{\gamma} \mathbf{V}$. Thus, for any probability $0 < \alpha < 1$,

$$1 - \alpha \cong \Pr[\sigma_y^2/\hat{\sigma}_y^2 \leq 1 + z_\alpha(\mathbf{V}^T \boldsymbol{\Sigma}_\boldsymbol{\gamma} \mathbf{V})^{1/2}] \\ = \Pr[\sigma_y \leq \hat{\sigma}_y\{1 + z_\alpha(\mathbf{V}^T \boldsymbol{\Sigma}_\boldsymbol{\gamma} \mathbf{V})^{1/2}\}^{1/2}],$$

where z_α denotes the upper α percentile of the standard normal distribution. Substituting $\hat{\boldsymbol{\Sigma}}_\boldsymbol{\gamma}$ and

$$\hat{\mathbf{V}} = \left[\begin{array}{ccccccc} -2\nu & -2\nu^2 & \dots & -2\nu^p & 2\nu & 2\nu^2 & \dots & 2\nu^q & -\hat{\sigma}_a^{-2} \end{array} \right]^T \tag{8}$$

for $\boldsymbol{\Sigma}_\boldsymbol{\gamma}$ and \mathbf{V} leads to the approximate $1 - \alpha$ confidence interval

$$\sigma_y \leq \sigma_{y,\alpha} = \hat{\sigma}_y\{1 + z_\alpha(\hat{\mathbf{V}}^T \hat{\boldsymbol{\Sigma}}_\boldsymbol{\gamma} \hat{\mathbf{V}})^{1/2}\}^{1/2} \tag{9}$$

for the EWMA standard deviation. After L is selected as described in the following section, $\sigma_{y,\alpha}$ can be used in the worst-case control limits (5).

The Taylor approximation (7) has an interesting interpretation when the process is ARMA(1, 1). In this case, (7) reduces to

$$\sigma_y^2 \cong \hat{\sigma}_y^2 \left\{ 1 - \frac{2\nu(\hat{\phi}_1 - \phi_1)}{1 - \phi_1\nu} + \frac{2\nu(\hat{\theta}_1 - \theta_1)}{1 - \theta_1\nu} - \frac{\hat{\sigma}_a^2 - \sigma_a^2}{\sigma_a^2} \right\}.$$

The EWMA variance increases (relative to the assumed value $\hat{\sigma}_y^2$) when ϕ_1 is underestimated ($\hat{\phi}_1 < \phi_1$) and/or θ_1 is overestimated ($\hat{\theta}_1 > \theta_1$). The reason is that the autocorrelation of x_t is underestimated in this situation, resulting in residuals with positive autocorrelation. When the residuals are positively autocorrelated, the variance of their EWMA is larger than if the residuals were iid. This was discussed in more detail in Adams and Tseng (1998). The foregoing equation also indicates that the effects of parameter estimation errors are larger for larger values of ν . In the limiting case with $\nu = 0$ (a Shewhart individual chart on the residuals), errors in estimating ϕ_1 and θ_1 have very little effect on the EWMA variance, which is further discussed in Section 4.3.

The confidence interval (9) and the expressions for $\hat{\boldsymbol{\Sigma}}_\boldsymbol{\gamma}$ in Appendix B are also valid for ARIMA($p, 1, q$) processes of the form $x_t = (1 - B)^{-1}\Phi^{-1}(B)\Theta(B)a_t$. The reason is that when estimating the parameters of an ARIMA model, one fits an ARMA model to the differenced data $(1 - B)x_t$. Because the residuals are still generated via (2) with x_t replaced by the differenced data, the EWMA statistic follows the same ARMA($p + q + 1, p + q$) model (3). The parameter errors thus have the exact same effect on the EWMA variance as in the ARMA case.

3. SELECTING THE DESIGN PARAMETERS L, λ , AND α

When designing an EWMA chart for iid data with no consideration of model uncertainty, the parameters λ and L are often jointly selected to minimize the out-of-control ARL for a specified mean shift, while ensuring the in-control ARL equals some desired value. Lucas and Saccucci (1990) provided tables for selecting values of λ and L that are optimal in this sense. For a residual-based EWMA with autocorrelated data, optimally selecting λ and L is complicated even when perfect models are assumed. The optimal λ and L depend on many factors, including the desired in-control ARL, the specified mean shift of interest, and the ARMA parameters. For first-order AR models, Lu and Reynolds (1999) provided tables for selecting the optimal λ and L for the specific cases of $\phi_1 = .4$ and $\phi_1 = .8$ with a desired in-control ARL of 370. When considering model uncertainty as in

this article, jointly selecting λ and L to satisfy some optimality criterion is prohibitively complex.

In light of this, it is recommended that one first select λ as if the estimated model were perfect. The rule of thumb $.05 < \lambda \leq .5$ (Lu and Reynolds 1999), where it is understood that smaller λ values result in better detection of small mean shifts, but slower detection of large shifts, may be used. (For more detailed guidelines, see the thorough discussions in Lucas and Saccucci 1990 and Lu and Reynolds 1999).

After specifying λ , suppose that the tables of Lucas and Saccucci (1990) are used to select L based on some desired in-control ARL (denoted ARL_d). If used in the standard EWMA control limits (4), this value of L would provide the desired ARL when there is no model uncertainty and the residuals are iid. With model uncertainty considered, using the same value of L in the worst-case EWMA control limits (5) is recommended. If the EWMA standard deviation σ_y happens to coincide with its worst-case value $\sigma_{y,\alpha}$, then the control limits (5) will provide an in-control ARL that approximately equals the desired value ARL_d . The examples in Section 4 indicate that this choice of L also results in an appealing Bayesian interpretation of the control chart: If an appropriate posterior distribution for the ARMA parameters is considered, then the posterior probability that the ARL is less than ARL_d is reasonably close to the α value specified in the confidence interval on σ_y .

Using a slightly smaller value of L in the control limits (5) also might have been considered for the following reason. When there are no modeling errors, and the standard control limits (4) are used, the value of L that provides a desired in-control ARL depends on λ . This is primarily because the autocorrelation of the EWMA statistic y_t depends on λ . As λ decreases, the autocorrelation of y_t increases, and the in-control ARL increases for any fixed L . Consequently, as λ decreases, smaller values of L will provide the same in-control ARL. When modeling errors are present, the errors also affect the autocorrelation of y_t . When the true parameters are such that σ_y coincides with $\sigma_{y,\alpha}$, the autocorrelation of the residuals generally will be positive, and the autocorrelation of y_t will be larger than when there are no modeling errors. Consequently, a slightly smaller value of L may provide the desired ARL when σ_y coincides with $\sigma_{y,\alpha}$. On the other hand, a first-order Taylor approximation of the EWMA variance was also used in developing the expression for $\sigma_{y,\alpha}$. This approximation tends to underestimate the EWMA variance, and the resulting $\sigma_{y,\alpha}$ is slightly smaller than what would result from a more exact confidence interval. Because the control limits (5) are the product of L and $\sigma_{y,\alpha}$, the effects of the Taylor approximation are partially compensated by taking L directly from the tables Lucas and Saccucci (1990) as recommended, as opposed to using a slightly smaller value.

Note that the ARL_d that one specifies in the design procedure should be viewed as a worst-case ARL that results when the EWMA variance equals its worst-case value (within the $1 - \alpha$ confidence interval). If the true ARMA parameters and the EWMA variance are close to their estimates, the ARL will generally be larger than ARL_d . To avoid overly conservative control limits, this should be kept in mind when selecting the remaining design parameter α . A small value such as $\alpha = .01$ may widen the control limits to an extent that makes it difficult

to detect most mean shifts of interest. This trade-off in using the worst-case control limits is discussed in more detail in Sections 4.1 and 4.2, with a recommended range of $.1 \leq \alpha \leq .3$.

The design procedure is illustrated with the Series A data from Box et al. (1994), which are $N = 197$ concentration measurements from a chemical process. Box et al. (1994) found that an ARMA(1, 1) model fit the data well, and the estimated parameters were (omitting their subscripts) $\hat{\phi} = .87$, $\hat{\theta} = .48$, and $\hat{\sigma}_a^2 = .098$. Using (B.4), the estimated parameter covariance is

$$\hat{\Sigma}_y = \begin{bmatrix} 2.75 & 3.64 & 0 \\ 3.64 & 8.71 & 0 \\ 0 & 0 & .098 \end{bmatrix} \times 10^{-3}.$$

If $\lambda = .1$ and $ARL_d = 500$ are selected, then the tables of Lucas and Saccucci (1990) indicate that $L = 2.814$ should be used. Because $\hat{\sigma}_y = \hat{\sigma}_a(1 - \nu)^{1/2}(1 + \nu)^{-1/2} = .0718$, the standard control limits (4) become $\pm L\hat{\sigma}_y = \pm .202$. If $\alpha = .1$ is also selected, then (8) and (9) result in $\hat{V} = [-8.29 \ 3.17 \ -10.20]^T$, and $\sigma_{y,\alpha} = .0849$. The worst-case control limits (5) are, therefore, $\pm L\sigma_{y,\alpha} = \pm .239$, which are 18% wider than the standard control limits.

Figure 2 shows an EWMA control chart applied to 500 simulated observations from the process when the true parameters assume the values $\phi = .917$, $\theta = .491$, and $\sigma_a^2 = .102$. These parameter values were chosen because the resulting Taylor approximation (7) of σ_y^2 (with V replaced by \hat{V}) equals the worst-case value $\sigma_{y,\alpha}^2$. One can also show that of all parameter combinations that result in a Taylor approximation equal to $\sigma_{y,\alpha}^2$, these values have the highest likelihood (minimize $\tilde{y}^T \hat{\Sigma}_y^{-1} \tilde{y}$). Both the standard and the worst-case control limits are shown in Figure 2. Because the mean of x_t was held at 0 throughout the simulation, all cases where the EWMA statistic fell outside the control limits were false alarms. The standard control limits resulted in false alarms around timesteps 50, 275, and 425, whereas the worst-case control limits eliminated the first two of these. In the following section, Monte Carlo simulation is used to provide a more comprehensive analysis of the control chart performance.

Figure 3, which is similar to Figure 2 except that the true ARMA parameters were chosen to coincide with their estimates, illustrates one drawback of using the worst-case control limits: If the true parameters happen to fall sufficiently close to their estimates, then the standard control limits provide the desired in-control ARL. The worst-case control limits are unnecessarily wide in this case, which inevitably decreases the power of the control chart. This is an inherent consequence of the worst-case design approach, which is intended to guard against the situation where the true parameters are not "sufficiently" close to their estimates. To mitigate this drawback, using both sets of control limits for the EWMA chart is recommended. An observation falling outside the worst-case control limits provides strong evidence that the process has changed. An observation falling within the worst-case control limits but outside the standard limits should be interpreted with more caution; it could mean that either the process has changed or that the ARMA parameters differ from their estimates. Section 4 provides a detailed discussion of the trade-offs involved in the worst-case design approach.

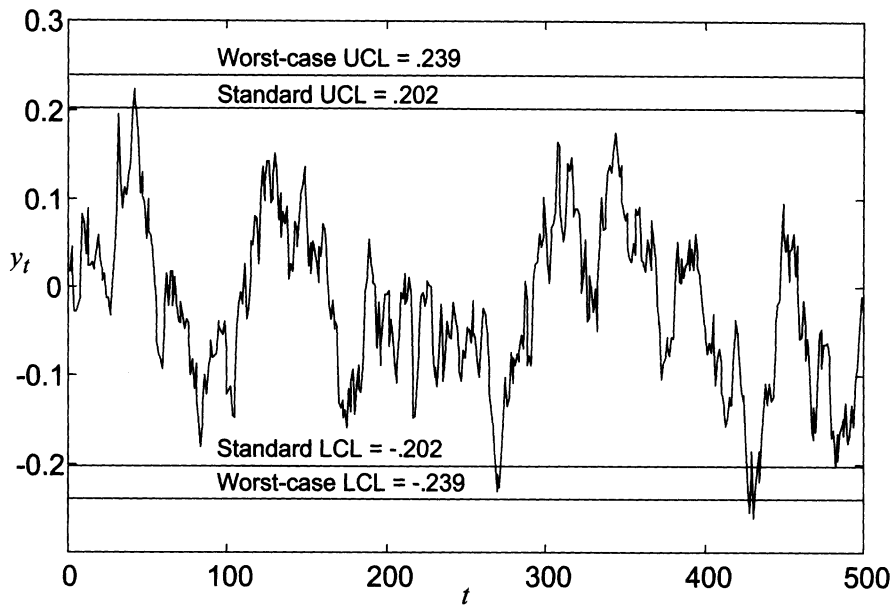


Figure 2. Example EWMA Chart With Standard and Worst-Case Control Limits, When σ_y Coincides With Its Worst-Case Value $\sigma_{y,\alpha}$.

4. DISCUSSION

Monte Carlo simulation is used throughout this section to investigate the ARL performance of EWMA charts with standard and worst-case control limits when the parameters differ from their estimates. For each simulation (with a specified combination of true and estimated parameters), 10,000 replicates were used. Thus the standard deviation of the ARL estimation error was approximately 1% of the true ARL. For each replicate, the a_t 's were generated as an independent sequence of random numbers from the normal distribution with mean 0 and variance σ_a^2 . The x_t 's were then generated from (1); the residuals, from (2). The first K (where K is some number that depends on ϕ , θ , $\hat{\phi}$, and $\hat{\theta}$) residuals were discarded, so it can be assumed that the remaining sequence of residuals has reached steady state. The EWMA for the remaining residuals was calculated via (3),

with y_0 initialized at 0. A signal occurred when y_t fell outside the control limits. The same procedure was followed for investigating an out-of-control ARL, except that a mean shift was added to x_t at timestep $K + 1$.

4.1 Bayesian Interpretations

Consider a Bayesian alternative to the worst-case design approach, where some posterior distribution for γ is assumed (given the data from which the parameters are estimated) and the control limits are selected to provide a desired average ARL with respect to the posterior distribution of γ . This section discusses why designing the control chart based on an average ARL would actually lead to control limits that are narrower than the standard limits. In addition, a Bayesian analysis is conducted to investigate the posterior probability that the ARL is

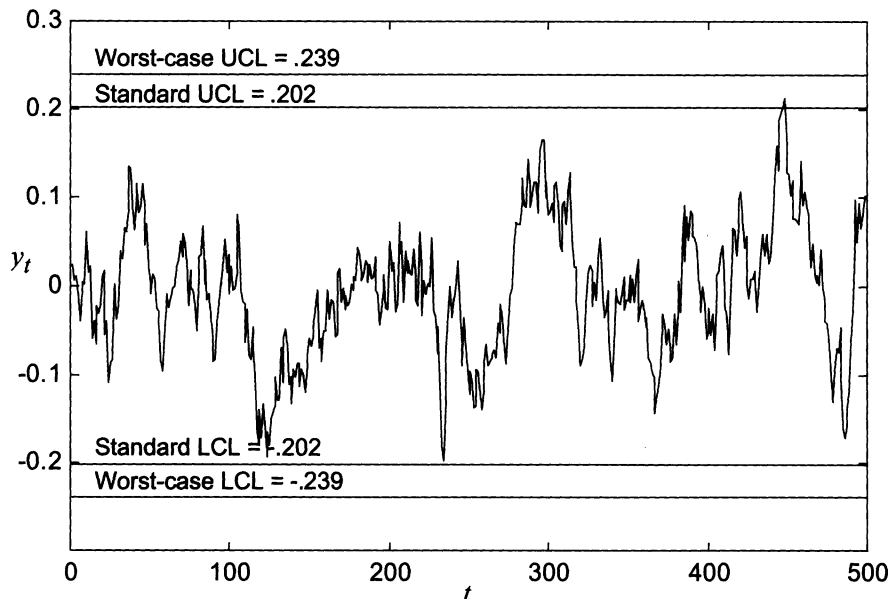


Figure 3. Example EWMA Chart With Standard and Worst-Case Control Limits, When σ_y Coincides With $\hat{\sigma}_y$.

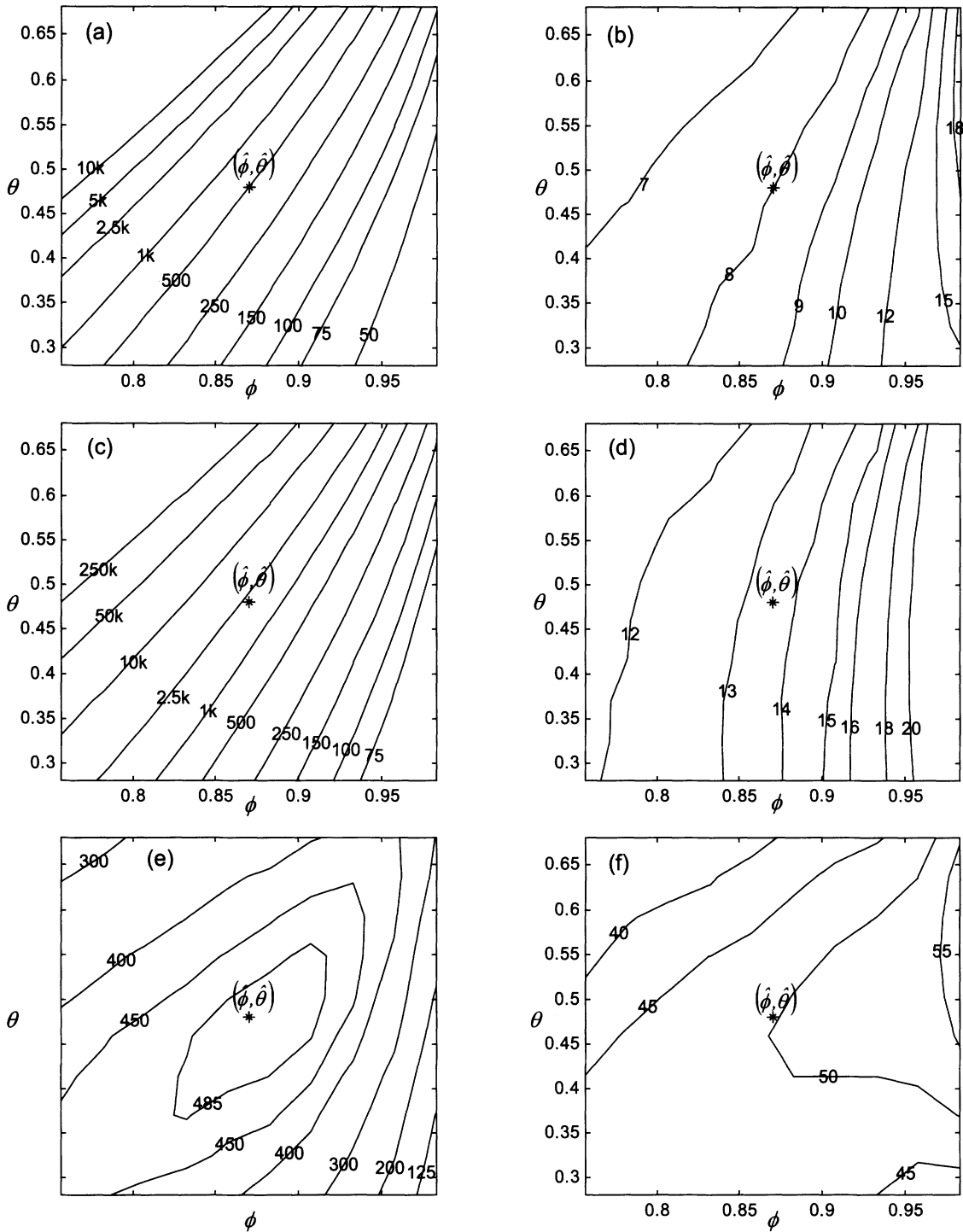


Figure 4. ARL Contours as a Function of ϕ and θ for the ARMA(1, 1) Example. (a), (c), and (e) In-control ARLs for the standard EWMA, worst-case EWMA, and Shewhart charts. (b), (d), and (f) Out-of-control ARLs for the three charts when the mean shift magnitude is $3\sigma_a$.

less than ARL_d when the worst-case control limits are used. For the examples considered here, this probability is reasonably close to the value of α specified in the confidence interval. For analysis purposes, it is assumed the posterior distribution of γ is approximately multivariate normal with mean $\hat{\gamma}$ and covariance $\hat{\Sigma}_\gamma$ (see App. B). This can be viewed as an asymptotic approximation when the prior distribution of γ is noninformative.

Reconsider the ARMA(1, 1) example introduced in Section 3, where the estimated parameters were $\hat{\phi} = .87$, $\hat{\theta} = .48$, and $\hat{\sigma}_a^2 = .098$. For simplicity, uncertainty in σ_a^2 is neglected

by modifying the earlier expression for $\hat{\Sigma}_\gamma$ so that its lower-right element (i.e., the variance of $\hat{\sigma}_a^2$) is 0. This results in $\sigma_{y,\alpha} = .0842$ and worst-case control limits $\pm L\sigma_{y,\alpha} = \pm .237$, which are only slightly narrower than when we also considered uncertainty in σ_a^2 . Figure 4 shows contour plots of the ARL as a function of ϕ and θ for $\sigma_a^2 = \hat{\sigma}_a^2$. Panel (a) is the in-control ARL contours for the standard EWMA with control limits of $\pm .202$. The parameter estimates are indicated by “*”. Because the EWMA was designed with $ARL_d = 500$, the $ARL = 500$ contour passes through the parameter estimates. Numerical in-

tegration of the ARL with respect to the assumed posterior density of γ gives a rough approximation of 730 for the average ARL of the EWMA chart with standard control limits. Somewhat surprisingly, this is larger than the desired ARL of 500 that results when the model is perfect. It may be concluded that an average ARL of 500 could be achieved with control limits even narrower than the standard control limits.

The reason the average ARL is larger than 500 is that the ARL is a highly skewed function of ϕ and θ , as can be seen in Figure 4(a). For $\phi < \hat{\phi}$ and $\theta > \hat{\theta}$, the ARL increases dramatically. The average ARL is misleading, however, because the ARL may decrease to unacceptably small values for $\phi > \hat{\phi}$ and $\theta < \hat{\theta}$. Numerical integration (of the posterior density over the $ARL < 250$ region) also reveals that there is a .24 probability that the ARL is less than 250, which is only half the desired ARL. Likewise, there is a .11 probability that the ARL is less than 150.

Figure 4(c) shows analogous in-control ARL contours for the EWMA chart with worst-case control limits of $\pm .237$. With the worst-case control limits, the probability that the ARL is less than the desired value 500 is approximately .13, reasonably close to the $\alpha = .1$ value selected when the chart was designed. Moreover, the probability that the ARL is less than 250 is only .05, compared with the .24 probability found with the standard control limits. The worst-case control limits clearly provide adequate protection against an unacceptably short in-control ARL. An additional benefit is that when the parameters coincide with their estimates, the in-control ARL will be even larger than the desired value. From Figure 4(c), the in-control ARL in this case is roughly 2,000, compared with an ARL of 500 with the standard control limits. The obvious disadvantage of widening the control limits, which is discussed in the following section, is the resulting decrease in the power of the chart for detecting mean shifts.

4.2 In-Control Versus Out-of-Control ARL Trade-off

For the same ARMA(1, 1) example introduced in Section 3 and continued in Section 4.1, Figures 4(b) and 4(d) show the out-of-control ARL contours for a mean shift with magnitude of $3\sigma_a$. Figure 4(b) is for the standard control limits of $\pm .202$, and Figure 4(d) is for the worst-case control limits of $\pm .237$. The worst-case control limits increase the out-of-control ARL by approximately 60% for most combinations of ϕ and θ . Note that even with the standard control limits, the ARL is approximately 8.0 when the ARMA parameters equal their estimates, which may seem large for a mean shift of $3\sigma_a$. After the initial occurrence of the mean shift, however, the mean of the residuals rapidly approaches a steady-state value of only $.75\sigma_a$. Superville and Adams (1994) and Apley and Shi (1999) discussed this “forecast recovery” phenomenon in detail. Table 1 presents

the out-of-control ARL values for other mean shifts for the specific case where the ARMA parameters coincide with their estimates. It also provides results for the Shewhart individual chart, discussed in Section 4.3. Widening the control limits clearly has a negative impact on the out-of-control ARL, particularly for small mean shifts. For a mean shift of size σ_a , which results in a steady-state residual mean of only $.25\sigma_a$, widening the control limits causes the out-of-control ARL to increase from 101 to 247. This is understandable, given that the in-control ARL (the ARL for a mean shift of size 0) increases from 500 to 2,020. The ARL increase is more moderate, but still substantial, for larger mean shifts.

As another example, with consideration of uncertainty in σ_a , suppose that the parameters of an AR(1) process are estimated using $N = 400$ observations and that the estimates are $\hat{\phi} = .5$ and $\hat{\sigma}_a^2 = 1.0$. If $\lambda = .1$ and a desired $ARL_d = 500$ are chosen, again $L = 2.814$. Because $\hat{\sigma}_y = \hat{\sigma}_a(1 - \nu)^{1/2}(1 + \nu)^{-1/2} = .2294$, the standard control limits (4) are $\pm L\hat{\sigma}_y = \pm .646$. Using (B.5), the parameter covariance is

$$\hat{\Sigma}_\gamma \cong \frac{1}{N} \begin{bmatrix} 1 - \hat{\phi}^2 & 0 \\ 0 & 2\hat{\sigma}_a^4 \end{bmatrix} = \begin{bmatrix} 1.88 & 0 \\ 0 & 5.0 \end{bmatrix} \times 10^{-3}.$$

If $\alpha = .1$ is selected, then (8) and (9) result in $\hat{V} = [-3.27 \ -1.00]^T$, and $\sigma_{y,\alpha} = .2516$. The worst-case control limits (5) are, therefore, $\pm L\sigma_{y,\alpha} = \pm .708$, which are roughly 10% wider than the standard control limits.

Figure 5 shows results for the AR(1) example that are analogous to Figure 4. Figures 5(a) and 5(c) show the in-control ARL contours as functions of ϕ and σ_a^2 for the standard and worst-case EWMA control limits. As in the Bayesian analysis of the previous section, suppose that the posterior distribution of γ is approximately multivariate normal with mean $\hat{\gamma}$ and covariance $\hat{\Sigma}_\gamma$. With the worst-case control limits, the probability that the ARL is less than 500 is roughly .105, almost identical to the selected value of α .

Figures 5(b) and 5(d) show the corresponding out-of-control ARL contours for a mean shift of magnitude $2\hat{\sigma}_a$. Table 2 gives the out-of-control ARL values for other mean shifts when the true parameters coincide with their estimates. Because the control limits are widened by a lesser extent than in the previous ARMA(1, 1) example, the worst-case design results in a much less severe increase in the out-of-control ARLs. For mean shifts with magnitude $2\hat{\sigma}_a$ or larger, the out-of-control ARLs increase by roughly 15%, whereas the in-control ARL doubles.

Given the decreased power of the chart that results from widening the control limits, to what extent (or even whether) they should be widened to account for model uncertainty would ideally depend on the costs associated with false alarms and the costs of failing to detect out-of-control conditions, as well as on the a priori probability of occurrence of out-of-control conditions. If the costs of false alarms are small, then it may not be

Table 1. ARL Values for Various Size Mean Shifts for the ARMA(1, 1) Example When the ARMA Parameters Coincide With Their Estimates

Chart	Control limits	Mean shift magnitude (in units of σ_a)					
		0	1	2	3	4	5
EWMA ($\lambda = .1$)	.202 (standard)	500	101	23.8	8.11	3.54	2.22
EWMA ($\lambda = .1$)	.237 (worst-case)	2,020	247	43.3	13.3	5.29	2.89
Shewhart	.967 (standard)	500	366	168	49.1	7.83	1.38

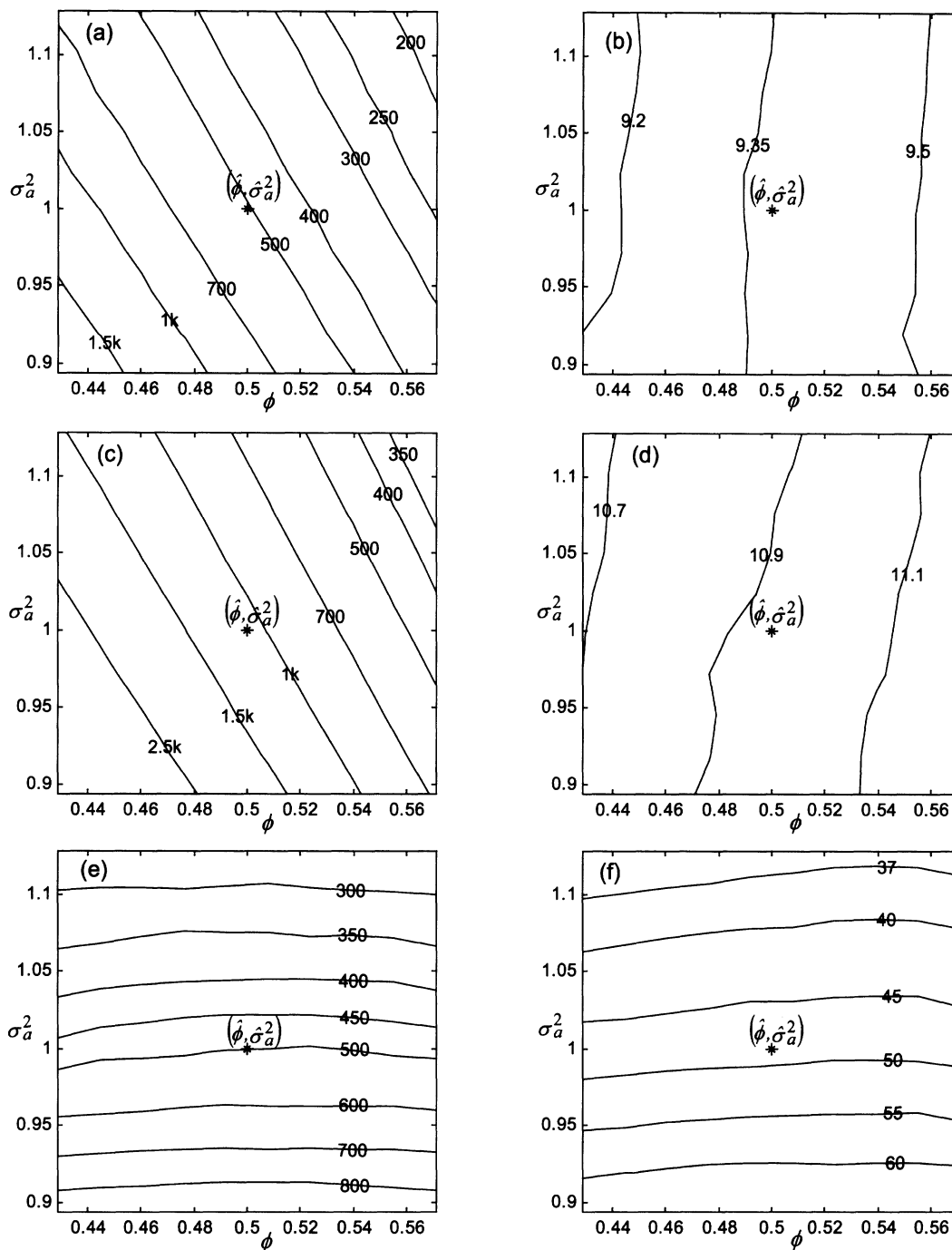


Figure 5. ARL Contours as a Function of ϕ and σ_a^2 for the AR(1) Example. (a), (c), and (e) In-control ARLs for the standard EWMA, worst-case EWMA, and Shewhart charts. (b), (d), and (f) Out-of-control ARLs for the three charts when the mean shift magnitude is $2\hat{\sigma}_a$.

desirable to widen the control limits. In the author's experience, however, the costs of frequent false alarms are often quite high when the hidden costs of unnecessary shutdowns, production delays, and operators who begin to ignore all alarms, including those that signal real out-of-control conditions, are consid-

ered. To lessen the severity of the trade-offs in using worst-case control limits, the best solution (when possible) would be to collect a larger sample of data to reduce the parameter uncertainty. Guidelines for sample size selection are discussed in Section 4.4.

Table 2. ARL Values for Various Size Mean Shifts for the AR(1) Example When the ARMA Parameters Coincide With Their Estimates

Chart	Control limits	Mean shift magnitude (in units of σ_a)					
		0	1	2	3	4	5
EWMA ($\lambda = .1$)	.646 (standard)	500	30.0	9.37	4.96	3.24	2.34
EWMA ($\lambda = .1$)	.708 (worst-case)	1,080	39.6	10.9	5.66	3.68	2.65
Shewhart	3.09 (standard)	500	199	48.1	10.6	2.32	1.10

4.3 Shewhart Individual Charts Versus EWMA Charts

Figures 4(e) and 4(f) show the in-control and out-of-control ARL contours for a Shewhart individual chart on the residuals in the previous ARMA(1, 1) example. Standard control limits of $\pm 3.09\hat{\sigma}_a = \pm .967$ were used, which provide an in-control ARL of 500 when there are no parameter errors. The mean shift magnitude for Figure 4(f) was $3\hat{\sigma}_a$, the same as for Figures 4(b) and 4(d). The ARL of the Shewhart chart is much less dependent on ϕ and θ than the ARL of an EWMA chart with small λ , because, unlike an EWMA, the Shewhart chart considers only individual residuals and does not take a weighted average of successive residuals. Consequently, residual autocorrelation has little effect on the Shewhart ARL if no supplementary run rules are used. Although an increase in the variance of the residuals will affect the Shewhart ARL, σ_a^2 was assumed to equal $\hat{\sigma}_a^2$ in this example, and small variations in ϕ and θ do not increase the residual variance substantially. Figures 5(e) and 5(f) show analogous results for a Shewhart chart applied to the residuals in the AR(1) example, where $\pm 3.09\hat{\sigma}_a$ control limits were again used. In this example, variations in σ_a^2 were also considered. Figure 5(e) shows that the in-control ARL depends predominantly on σ_a^2 and is nearly independent of ϕ over the range of values considered.

Given the relative insensitivity of the Shewhart individual chart with respect to parameter errors, an alternative to using an EWMA with worst-case control limits is to simply use a Shewhart chart with standard control limits. Because the out-of-control ARL for the EWMA is increased when its control limits are widened, one may speculate that the Shewhart chart with standard control limits could provide better detection of mean shifts. Tables 1 and 2 indicate that this is true only for large mean shifts in the examples considered. Even when the worst-case control limits are used, the EWMA still has substantially shorter out-of-control ARLs than the Shewhart chart for small to moderate mean shifts. Table 1 shows that for the ARMA(1, 1) example, the Shewhart chart does not surpass the worst-case EWMA in power until the mean shift is between $4\hat{\sigma}_a$ and $5\hat{\sigma}_a$. This is the same level of mean shift at which the Shewhart chart surpasses the EWMA with standard control limits. Table 2 demonstrates similar results for the AR(1) example. Moreover, comparing Figures 4(e) and 5(e) with Figures 4(c) and 5(c), the EWMA with worst-case control limits provides the additional benefit of substantially larger in-control ARLs for most parameter combinations.

4.4 Sample Size Requirements

In light of the decreased power that results from widening the EWMA control limits, one may wish to collect a sufficiently large sample of data to ensure that $\sigma_{y,\alpha}$ is close to $\hat{\sigma}_y$, in which case the worst-case control limits will be close to the standard control limits. It is difficult to provide general guidelines for sample size requirements without some knowledge of the ARMA parameters, because $\sigma_{y,\alpha}$ depends heavily on the parameter estimates. If initial estimates have been obtained from an initial set of data, however, this may be used to determine how much (or whether) additional data are needed. While waiting for the additional data to be collected, it may be desirable to use both the worst-case and the standard control limits

together (refer to Fig. 2) as temporary control limits until more accurate parameter estimates and new control limits can be calculated.

Suppose that initial parameter estimates have been obtained and λ and α have been selected. A reasonable strategy is to select the size N of the additional data sample sufficiently large so that the resulting percentage difference between $\sigma_{y,\alpha}$ and $\hat{\sigma}_y$ is less than some small value δ (e.g., $\delta = .05$). From (9), the requirement becomes

$$\frac{\sigma_{y,\alpha}}{\hat{\sigma}_y} = \left[1 + z_\alpha (\hat{\mathbf{V}}^T \hat{\Sigma}_y \hat{\mathbf{V}})^{1/2} \right]^{1/2} < 1 + \delta.$$

Define $\bar{\Sigma}_y = N \hat{\Sigma}_y$. As shown in Appendix B, $\bar{\Sigma}_y$ is a function of the parameter estimates but is otherwise independent of N . If this is substituted into the foregoing inequality, then the sample size requirement reduces to

$$N > \frac{z_\alpha^2 \hat{\mathbf{V}}^T \bar{\Sigma}_y \hat{\mathbf{V}}}{\delta^2 (2 + \delta)^2}. \tag{10}$$

To provide some insight into typical sample size requirements, Figure 6 shows contour plots of the required N from (10) as a function of $\hat{\phi}$ and $\hat{\theta}$ for an ARMA(1, 1) process with four different values of λ . The contour plots are for the specific case of $\delta = .05$ and $\alpha = .20$. Because neither $\bar{\Sigma}_y$ nor $\hat{\mathbf{V}}$ depends on δ and α , results for other δ and α are obtained by multiplying the values of N in Figure 6 by $z_\alpha^{-2} .05^2 2.05^2 z_\alpha^2 \delta^{-2} (2 + \delta)^{-2} = .0148 z_\alpha^2 \delta^{-2} (2 + \delta)^{-2}$. If, for example, a more conservative $\alpha = .1$ and the same δ are considered, then the required sample sizes are multiplied by 2.32. For a less conservative $\alpha = .3$ and the same δ , the same sizes are multiplied by .387. For small δ , (10) indicates that the required N is approximately inversely proportional to δ^2 .

In the ARMA(1, 1) example of Section 3 with $\hat{\phi} = .87$, $\hat{\theta} = .48$, and $N = 197$, the values $\lambda = .1$ and $\alpha = .1$ were selected. This resulted in worst-case control limits that were 18% wider than the standard control limits. Suppose that one wanted to collect a sample sufficiently large that the worst-case control limits were only 5% wider than the standard limits. From Figure 6, a sample size of approximately 1,270 would be required when $\alpha = .2$. For $\alpha = .1$, the required sample size is 2.32 times larger, or $N = 2,940$.

Note that the ridges in Figure 6 are at $\hat{\phi} = 1 - \lambda$ and $\hat{\theta} = 1 - \lambda$. For a specified λ , the EWMA chart is least robust when the parameter estimates coincide with $1 - \lambda$. This does not imply that one should avoid choosing a value of λ that coincides with $1 - \hat{\phi}$ or $1 - \hat{\theta}$, however. One may show that for any fixed positive values of $\hat{\phi}$ and $\hat{\theta}$, (10) increases monotonically as λ decreases.

Figure 7 shows contour plots of the required sample size as a function of $\hat{\phi}$ and λ for an AR(1) process with $\delta = .05$ and $\alpha = .20$. Results for other δ and α are again obtained by multiplying the values of N in Figure 7 by $.0148 z_\alpha^2 \delta^{-2} (2 + \delta)^{-2}$. Figure 7 also applies to first-order MA and IMA processes if $\hat{\phi}$ is replaced by $\hat{\theta}$, because of the symmetry of $\bar{\Sigma}_y$ and $\hat{\mathbf{V}}$ with respect to the AR and MA parameters. Figures 6 and 7 indicate that very large samples are often required to ensure that $\sigma_{y,\alpha}$ is no more than 5% larger than $\hat{\sigma}_y$. Even for an AR(1) process with $\alpha = .20$, sample sizes close to 1,000 are required for the typical values $\lambda \approx .1$ and $\hat{\phi} > .5$.

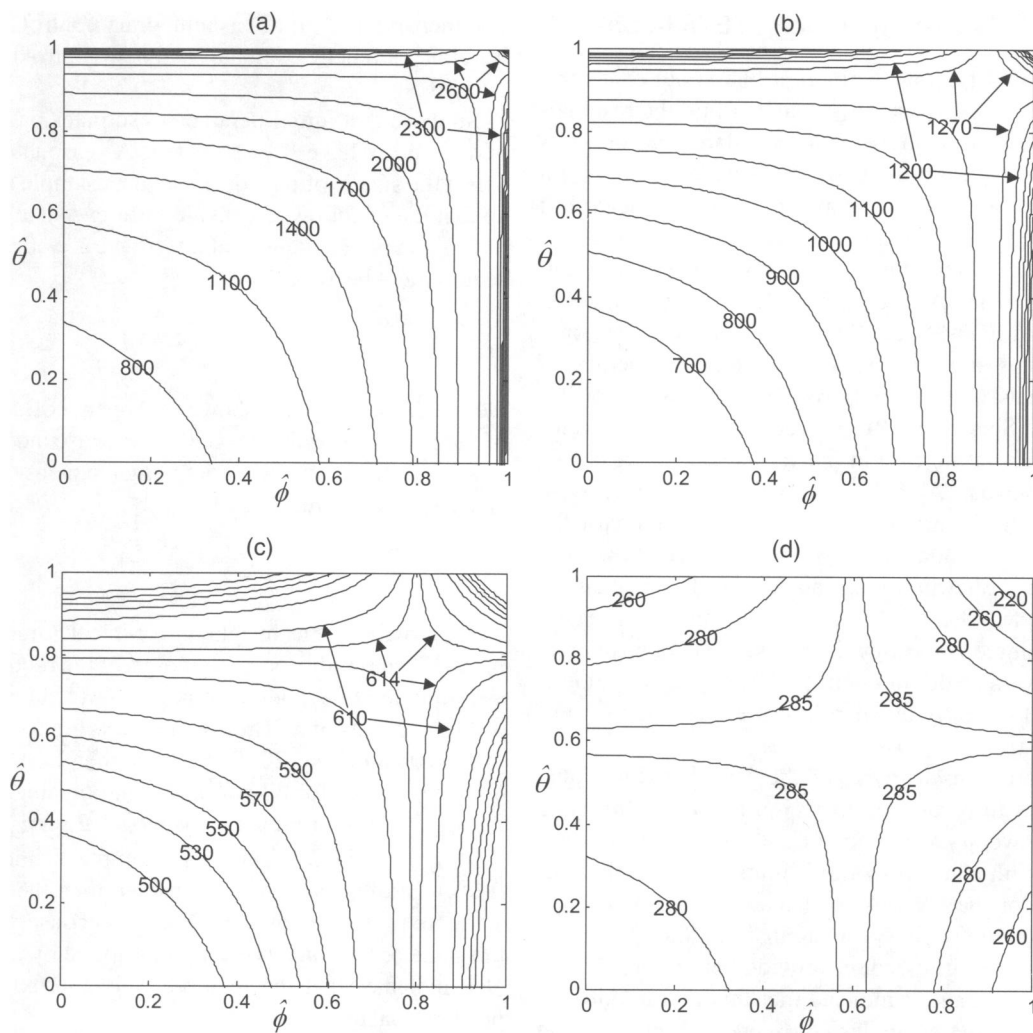


Figure 6. Contours of the Required Sample Size N with $\delta = .05$ and $\alpha = .20$ for an ARMA(1, 1) Process With $\lambda = .05$ (a), $.10$ (b), $.20$ (c) and $.40$ (d). For other values of δ and α , multiply the contours by $.0148z_{\alpha}^2\delta^{-2}(2 + \delta)^{-2}$.

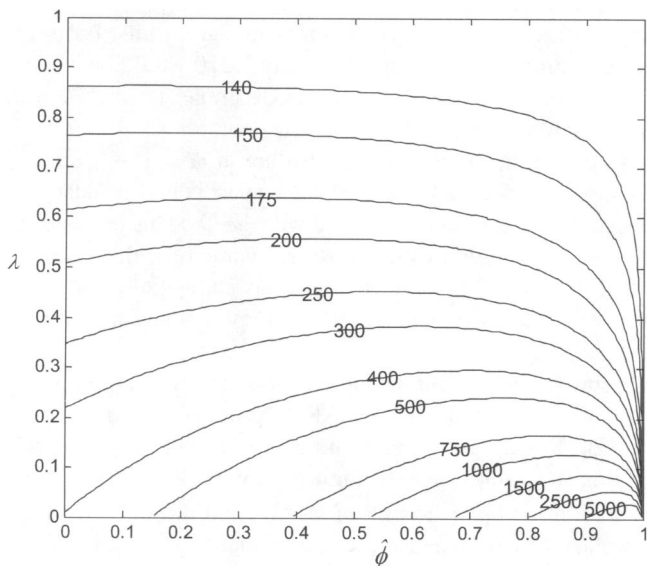


Figure 7. Contours of the Required Sample Size N with $\delta = .05$ and $\alpha = .20$ for an AR(1) process. For other values of δ and α , multiply the contours by $.0148z_{\alpha}^2\delta^{-2}(2 + \delta)^{-2}$. The results for first-order MA and IMA processes are identical if $\hat{\phi}$ is replaced by $\hat{\theta}$.

For iid processes, many authors have investigated the effects of estimating the process mean and variance on control chart performance. For example, Ghosh, Reynolds, and Van Hui (1981), Quesenberry (1993), and Del Castillo (1996), studied the effects of estimating the mean and/or variance on Shewhart-type charts. Bagshaw and Johnson (1975) studied the effects of estimating the variance on the ARL of a CUSUM chart. Jones, Champ, and Rigdon (2001) studied the effects of estimating the mean and variance on EWMA charts. A common conclusion is that sample sizes considerably larger than what one might expect are often needed when parameters are estimated. Jones et al. (2001) found that the required sample size depends strongly on the EWMA parameter λ , which is consistent with the foregoing results.

5. CONCLUSIONS

When designing a residual-based EWMA, a natural measure is to use wider control limits to account for uncertainty in the estimated parameters. The design approach of this article widens the control limits by an amount commensurate with the worst-case scenario, in which the ARMA parameters are such that

the EWMA variance equals the maximum value within an appropriate confidence interval. Assuming that an estimate of the parameter covariance matrix is available or can be calculated as described in Appendix B, the worst-case design approach involves little additional complexity relative to the standard design approach.

In light of the drawback of widening the control limits—decreased chart power—one may question whether residual-based control charts should be used. The impact of parameter uncertainty, however, is not unique to residual-based charts. Suppose that a CUSUM, \bar{X} , or EWMA chart is to be applied directly to an autocorrelated process x_t . The methods presented by Johnson and Bagshaw (1974), Vasilopoulos and Stamboulis (1978), and Zhang (1998) rely on an accurate ARMA process model (or, equivalently, the autocorrelation function of x_t) just as residual-based control charts do. The difference is that in residual-based control charts, the charted statistic depends on the model, whereas the control limits do not. In control charts applied directly to x_t , the control limits depend on the model, whereas the charted statistic does not. If the estimated model is inaccurate in either case, then the control limits will fail to provide the desired ARL.

Noting the lack of robustness of residual-based charts to parameter errors, Adams and Tseng (1998) recommended the alternative approach of removing (when possible) the autocorrelation by either removing the source or using feedback-control techniques. Effective removal of autocorrelation via feedback control also requires an accurate process model, however. With modeling errors, the feedback-controlled process output would have autocorrelation similar to the ARMA residuals. A control chart on the output would most likely be just as affected by modeling errors as a control chart on the residuals. An investigation of the relative robustness of different control charts (on the residuals, on x_t , and on the feedback-controlled output) would shed light on whether there are any significant differences in their sensitivity to modeling errors.

ACKNOWLEDGMENTS

The editor, the associate editor, and an anonymous referee have made numerous helpful comments that greatly improved the quality of this article. This work was supported by the State of Texas Advanced Technology Program grant 000512-0289-1999 and the National Science Foundation grant DMI-0093580.

APPENDIX A: DERIVATION OF THE FIRST-ORDER TAYLOR APPROXIMATION (7) FOR ARMA(1, 1) PROCESSES

The first-order Taylor approximation of the ratio

$$r = \frac{\sigma_y^2}{\hat{\sigma}_y^2} = \frac{\sigma_a^2(1 - \nu^2)}{\hat{\sigma}_a^2} \sum_{j=0}^{\infty} G_j^2$$

is sought, where (6) has been used for σ_y^2 . The approximation is

$$\begin{aligned} r &\cong r|_{\hat{\nu}=\nu} + \left. \frac{\partial r}{\partial \hat{\phi}} \right|_{\hat{\nu}=\nu} (\hat{\phi} - \phi) + \left. \frac{\partial r}{\partial \hat{\theta}} \right|_{\hat{\nu}=\nu} (\hat{\theta} - \theta) \\ &\quad + \left. \frac{\partial r}{\partial \hat{\sigma}_a^2} \right|_{\hat{\nu}=\nu} (\hat{\sigma}_a^2 - \sigma_a^2) \\ &= 1 + 2(1 - \nu^2) \sum_{j=0}^{\infty} \left(G_j \frac{\partial G_j}{\partial \hat{\phi}} \right) \Big|_{\hat{\nu}=\nu} \tilde{\phi} \\ &\quad + 2(1 - \nu^2) \sum_{j=0}^{\infty} \left(G_j \frac{\partial G_j}{\partial \hat{\theta}} \right) \Big|_{\hat{\nu}=\nu} \tilde{\theta} - \frac{(1 - \nu^2)}{\sigma_a^2} \sum_{j=0}^{\infty} G_j^2 |_{\hat{\nu}=\nu} \tilde{\sigma}_a^2. \end{aligned} \tag{A.1}$$

For ARMA(1, 1) processes, $G(B)$ can be expressed via its partial fraction expansion

$$\begin{aligned} G(B) &= \frac{(1 - \hat{\phi}B)(1 - \theta B)}{(1 - \nu B)(1 - \hat{\theta}B)(1 - \phi B)} \\ &= \frac{(\nu - \hat{\phi})(\nu - \theta)}{(\nu - \hat{\theta})(\nu - \phi)} \frac{1}{(1 - \nu B)} + \frac{(\phi - \hat{\phi})(\phi - \theta)}{(\phi - \nu)(\phi - \hat{\theta})} \frac{1}{(1 - \phi B)} \\ &\quad + \frac{(\hat{\theta} - \hat{\phi})(\hat{\theta} - \theta)}{(\hat{\theta} - \nu)(\hat{\theta} - \phi)} \frac{1}{(1 - \hat{\theta}B)}, \end{aligned}$$

which can be verified by straightforward but tedious algebra. For notational convenience, the subscripts on ϕ and θ have been omitted. Because $(1 - cB)^{-1} = \sum_{j=0}^{\infty} c^j B^j$ for any constant c with magnitude less than unity, the impulse response coefficients are

$$\begin{aligned} G_j &= \frac{(\nu - \hat{\phi})(\nu - \theta)}{(\nu - \hat{\theta})(\nu - \phi)} \nu^j + \frac{(\phi - \hat{\phi})(\phi - \theta)}{(\phi - \nu)(\phi - \hat{\theta})} \phi^j \\ &\quad + \frac{(\hat{\theta} - \hat{\phi})(\hat{\theta} - \theta)}{(\hat{\theta} - \nu)(\hat{\theta} - \phi)} \hat{\theta}^j. \end{aligned}$$

Hence

$$G_j |_{\hat{\nu}=\nu} = \nu^j,$$

$$\begin{aligned} \sum_{j=0}^{\infty} \left(G_j \frac{\partial G_j}{\partial \hat{\phi}} \right) \Big|_{\hat{\nu}=\nu} &= \sum_{j=0}^{\infty} \nu^j \left\{ \frac{-\nu^j}{\nu - \phi} + \frac{\phi^j}{\nu - \phi} \right\} \\ &= \frac{1}{\nu - \phi} \left\{ \frac{1}{1 - \phi\nu} - \frac{1}{1 - \nu^2} \right\} \\ &= \frac{-\nu}{(1 - \phi\nu)(1 - \nu^2)}, \end{aligned}$$

$$\begin{aligned} \sum_{j=0}^{\infty} \left(G_j \frac{\partial G_j}{\partial \hat{\theta}} \right) \Big|_{\hat{\nu}=\nu} &= \sum_{j=0}^{\infty} \nu^j \left\{ \frac{\nu^j}{\nu - \theta} - \frac{\theta^j}{\nu - \theta} \right\} \\ &= \frac{1}{\nu - \theta} \left\{ \frac{1}{1 - \nu^2} - \frac{1}{1 - \theta\nu} \right\} \\ &= \frac{\nu}{(1 - \theta\nu)(1 - \nu^2)}, \end{aligned}$$

and

$$\sum_{j=0}^{\infty} G_j^2 \Big|_{\hat{\nu}=\nu} = \sum_{j=0}^{\infty} \nu^{2j} = \frac{1}{1 - \nu^2}.$$

Substituting these into (A.1) gives (7).

APPENDIX B: CALCULATING THE PARAMETER COVARIANCE Σ_γ

Assume that the ARMA parameters are estimated using a method based on minimizing the sum of the squares of the model residuals, such as the nonlinear least squares or approximate maximum likelihood methods described by Box et al. (1994). For sample size N sufficiently large, the parameter covariance matrix is (Box et al. 1994)

$$\Sigma_\gamma \cong \begin{bmatrix} \Sigma_\eta & \mathbf{0} \\ \mathbf{0}^T & 2N^{-1}\sigma_a^4 \end{bmatrix} = \frac{1}{N} \begin{bmatrix} \sigma_a^2 \Sigma_w^{-1} & \mathbf{0} \\ \mathbf{0}^T & 2\sigma_a^4 \end{bmatrix}, \quad (B.1)$$

where $\mathbf{0}$ denotes a column vector of $p + q$ 0s, and Σ_η denotes the covariance of $\eta = [\phi_1 \ \phi_2 \ \dots \ \phi_p \ \theta_1 \ \theta_2 \ \dots \ \theta_q]^T$. The matrix Σ_w is defined as the covariance matrix of the random vector $w_t = [u_t \ u_{t-1} \ \dots \ u_{t-p+1} \ v_t \ v_{t-1} \ \dots \ v_{t-q+1}]^T$, where u_t and v_t are defined via $u_t = \Phi^{-1}(B)a_t$ and $v_t = -\Theta^{-1}(B)a_t$.

To calculate Σ_w , rewrite $u_t = \sum_{j=0}^{\infty} g_{\phi,j} a_{t-j}$ and $v_t = -\sum_{j=0}^{\infty} g_{\theta,j} a_{t-j}$, where the $g_{\phi,j}$'s and $g_{\theta,j}$'s are the impulse response coefficients of $\Phi^{-1}(B)$ and $\Theta^{-1}(B)$. Note that the impulse response coefficients can be calculated recursively for $j = 1, 2, \dots$, via

$$g_{\phi,j} = \phi_1 g_{\phi,j-1} + \phi_2 g_{\phi,j-2} + \dots + \phi_p g_{\phi,j-p}, \quad (B.2)$$

and

$$g_{\theta,j} = \theta_1 g_{\theta,j-1} + \theta_2 g_{\theta,j-2} + \dots + \theta_q g_{\theta,j-q} \quad (B.3)$$

with $g_{\phi,j} = g_{\theta,j} = 0$ for $j < 0$ and $g_{\phi,0} = g_{\theta,0} = 1$. If the matrix

$$\mathbf{H} = \begin{bmatrix} g_{\phi,0} & 0 & \dots & 0 & -g_{\theta,0} & 0 & \dots & 0 \\ g_{\phi,1} & g_{\phi,0} & \ddots & \vdots & -g_{\theta,1} & -g_{\theta,0} & \ddots & \vdots \\ g_{\phi,2} & g_{\phi,1} & \ddots & 0 & -g_{\theta,2} & -g_{\theta,1} & \ddots & 0 \\ \vdots & \vdots & & g_{\phi,0} & \vdots & \vdots & & -g_{\theta,0} \\ g_{\phi,p} & g_{\phi,p-1} & \dots & g_{\phi,1} & -g_{\theta,q} & -g_{\theta,q-1} & \dots & -g_{\theta,1} \\ g_{\phi,p+1} & g_{\phi,p} & \dots & g_{\phi,2} & -g_{\theta,q+1} & -g_{\theta,q} & \dots & -g_{\theta,2} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \end{bmatrix}$$

is constructed from the impulse response coefficients, then $\Sigma_w = \sigma_a^2 \mathbf{H}^T \mathbf{H}$ results, and $\sigma_a^2 \Sigma_w^{-1} = [\mathbf{H}^T \mathbf{H}]^{-1}$ can be substituted in (B.1). Because the impulse response coefficients decay exponentially for stable, invertible ARMA processes, the number of rows needed in \mathbf{H} will generally be reasonable.

Because the true ARMA parameters are unknown, their estimates must be substituted into (B.1)–(B.3) to calculate the estimate $\hat{\Sigma}_\gamma$ for use in the confidence interval (9). Box et al. (1994) showed that for first-order AR, MA, and ARMA processes, the estimated covariance of η reduces to the following:

$$\text{ARMA}(1, 1): \hat{\Sigma}_\eta = \frac{(1 - \hat{\phi}\hat{\theta})}{N(\hat{\phi} - \hat{\theta})^2} \times \begin{bmatrix} (1 - \hat{\phi}^2)(1 - \hat{\phi}\hat{\theta}) & (1 - \hat{\phi}^2)(1 - \hat{\theta}^2) \\ (1 - \hat{\phi}^2)(1 - \hat{\theta}^2) & (1 - \hat{\theta}^2)(1 - \hat{\phi}\hat{\theta}) \end{bmatrix}, \quad (B.4)$$

$$\text{AR}(1): \hat{\Sigma}_\eta = \frac{1 - \hat{\phi}^2}{N}, \quad (B.5)$$

and

$$\text{MA}(1): \hat{\Sigma}_\eta = \frac{1 - \hat{\theta}^2}{N}$$

[Received April 2000. Revised March 2003.]

REFERENCES

Adams, B. M., and Tseng, I. T. (1998), "Robustness of Forecast-Based Monitoring Schemes," *Journal of Quality Technology*, 30, 328–339.

Alwan, L. C., and Roberts, H. V. (1988), "Time-Series Modeling for Statistical Process Control," *Journal of Business and Economic Statistics*, 6, 87–95.

Apley, D. W. (2003), "The Sensitivity of EWMA Charts for Autocorrelated Processes," Technical Report 2003-001, Texas A&M University, Dept. of Industrial Engineering.

Apley, D. W., and Shi, J. (1999), "The GLRT for Statistical Process Control of Autocorrelated Processes," *IIE Transactions*, 31, 1123–1134.

Bagshaw, M., and Johnson, R. A. (1975), "The Influence of Reference Values and Estimated Variance on the ARL of CUSUM Tests," *Journal of the Royal Statistical Society, Ser. B*, 37, 413–420.

Berthouex, P. M., Hunter, W. G., and Pallesen, L. (1978), "Monitoring Sewage Treatment Plants: Some Quality Control Aspects," *Journal of Quality Technology*, 10, 139–149.

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

Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods* (2nd ed.), New York: Springer-Verlag.

Del Castillo, E. (1996), "Evaluation of Run Length Distribution for \bar{X} Charts With Unknown Variance," *Journal of Quality Technology*, 28, 116–122.

English, J. R., Krishnamurthi, M., and Sastri, T. (1991), "Quality Monitoring of Continuous Flow Processes," *Computers and Industrial Engineering*, 20, 251–260.

Ghosh, B. K., Reynolds, M. R., and Van Hui, Y. V. (1981), "Shewhart \bar{X} -Charts With Estimated Process Variance," *Communications in Statistics: Theory and Methods*, 18, 1797–1822.

Johnson, R. A., and Bagshaw, M. (1974), "The Effect of Serial Correlation on the Performance of CUSUM Tests," *Technometrics*, 16, 103–112.

Jones, L. A., Champ, C. W., and Rigdon, S. E. (2001), "The Performance of Exponentially Weighted Moving Average Charts With Estimated Parameters," *Technometrics*, 43, 156–167.

Lin, W. S., and Adams, B. M. (1996), "Combined Control Charts for Forecast-Based Monitoring Schemes," *Journal of Quality Technology*, 28, 289–301.

Lu, C. W., and Reynolds, M. R. (1999), "EWMA Control Charts for Monitoring the Mean of Autocorrelated Processes," *Journal of Quality Technology*, 31, 166–188.

Lucas, J. M., and Saccucci, M. S. (1990), "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements," *Technometrics*, 32, 1–12.

Montgomery, D. C. (2001), *Introduction to Statistical Quality Control* (4rd ed.), New York: Wiley.

Montgomery, D. C., and Mastrangelo, C. M. (1991), "Some Statistical Process Control Methods for Autocorrelated Data," *Journal of Quality Technology*, 23, 179–193.

Montgomery, D. C., and Woodall, W. H. (1997), "A Discussion on Statistically-Based Process Monitoring and Control," *Journal of Quality Technology*, 29, 121–162.

Quesenberry, C. P. (1993), "The Effect of Sample Size on Estimated Limits for \bar{X} and X Control Charts," *Journal of Quality Technology*, 25, 237–247.

Runger, G. C., Willemain, T. R., and Prabhu, S. (1995), "Average Run Lengths for Cusum Control Charts Applied to Residuals," *Communications in Statistics: Theory and Methods*, 24, 273–282.

Superville, C. R., and Adams, B. M. (1994), "An Evaluation of Forecast-Based Quality Control Schemes," *Communications in Statistics: Simulation and Computation*, 23, 645–661.

Vander Wiel, S. A. (1996), "Monitoring Processes That Wander Using Integrated Moving Average Models," *Technometrics*, 38, 139–151.

Vasilopoulos, A. V., and Stamboulis, A. P. (1978), "Modification of Control Chart Limits in the Presence of Data Correlation," *Journal of Quality Technology*, 10, 20–30.

Wardell, D. G., Moskowitz, H., and Plante, R. D. (1994), "Run-Length Distributions of Special-Cause Control Charts for Correlated Processes," *Technometrics*, 36, 3–17.

Zhang, N. F. (1998), "A Statistical Control Chart for Stationary Process Data," *Technometrics*, 40, 24–38.