The Effects of Model Parameter Deviations on the Variance of a Linearly Filtered Time Series

Daniel W. Apley,[1] Hyun Cheol Lee[2]

[1] *Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208-3119*

[2] *Samsung Electronics, Semiconductor Business, Quality Assurance Department, Banwol-Dong, Hwasung-City, Gyeonggi-Do, Korea 445-701*

Received 7 May 2008; revised 4 March 2010; accepted 8 March 2010
DOI 10.1002/nav.20414
Published online 13 April 2010 in Wiley InterScience (www.interscience.wiley.com).

Abstract: We consider a general linear filtering operation on an autoregressive moving average (ARMA) time series. The variance of the filter output, which is an important quantity in many applications, is not known with certainty because it depends on the true ARMA parameters. We derive an expression for the sensitivity (i.e., the partial derivative) of the output variance with respect to deviations in the model parameters. The results provide insight into the robustness of many common statistical methods that are based on linear filtering and also yield approximate confidence intervals for the output variance. We discuss applications to time series forecasting, statistical process control, and automatic feedback control of industrial processes.
Naval Research Logistics 57: 460–471, 2010

Keywords: autoregressive moving average model; time series forecasting, parameter uncertainty; robustness; control chart

1. INTRODUCTION

Consider a time series $\{x_t : t = 1, 2, \ldots\}$ that follows an autoregressive moving-average (ARMA(p, q)) model of the form [12]

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} \cdots + \phi_p x_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \cdots - \theta_q a_{t-q},$$

where t is a time index, $\{\phi_1, \phi_2, \ldots \phi_p\}$ denote the AR parameters, $\{\theta_1, \theta_2, \ldots \theta_q\}$ denote the MA parameters, and $\{a_t : t = 1, 2, \ldots\}$ is an identically, independently distributed (iid) random sequence with mean zero and variance σ_a^2. The AR and MA orders are denoted by p and q, respectively. Let $\boldsymbol{\gamma} = [\phi_1\, \phi_2\, \ldots\, \phi_p\, \theta_1\, \theta_2\, \ldots\, \theta_q]^T$ denote the vector of ARMA parameters, and $\hat{\boldsymbol{\gamma}}$, an estimate obtained from a sample of size N observations.

Suppose one intends to filter x_t with some stable, time-invariant, but otherwise arbitrary, linear filter of the form

$$z_t = h_0 x_t + h_1 x_{t-1} + h_2 x_{t-2} + h_3 x_{t-3} + \ldots, \quad \text{for } t = 1, 2, 3, \ldots,$$

where z_t denotes the filter output, and $\{h_j\colon j = 0, 1, 2, \ldots\}$ denote the impulse response coefficients of the filter. Let σ_z^2 denote the asymptotic (as $t \to \infty$) variance of z_t.

In many such filtering operations in time series applications, one is interested in the variance of the filter output. For example, for each $k = 1, 2, \ldots$, the k-step-ahead forecasting error using ARMA forecasting methods is a linearly filtered version of x_t with $h_0 = 1$, $h_j = 0$ for $j = 1, 2, \ldots k - 1$, and $\{h_j\colon j = k, k + 1, \ldots\}$ determined by the form of the forecast equations. In this case, σ_z^2 represents the variance of the forecasting error, which is used when determining prediction intervals on the future observations.

Because in practice $\boldsymbol{\gamma}$ is unknown and must be estimated, the effect on σ_z^2 of errors in $\boldsymbol{\gamma}$ is also of general interest. In this article, we derive an expression for the sensitivity (i.e., the partial derivative) of σ_z^2 with respect to deviations in the model parameters and discuss how it can be used to analyze the effects of parameter errors. To facilitate the development, throughout the article we use standard time series notation [12] and let B denote the backward shift operator defined such that $Bx_t = x_{t-1}$. Defining $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 \ldots - \phi_p B^p$ and $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 \ldots - \theta_q B^q$, we can write the ARMA model as

$$x_t = \frac{\Theta(B)}{\Phi(B)} a_t. \tag{1}$$

Correspondence to: D.W. Apley (apley@northwestern.edu)

Likewise, we can write the filtering operation as

$$z_t = H(B)x_t, \quad (2)$$

where $H(B) = \sum_{j=0}^{\infty} h_j B^j$.

There has been a substantial amount of prior work that takes parameter uncertainty into account to modify the prediction intervals in time series forecasting. Bootstrapping methods have received quite a bit of attention (e.g., [37, 41, 44, 45]). In this article, we focus on analytical methods, which have also received considerable attention. References [11 (Appendix A7.3), 48, 49], used a first-order Taylor expansion (of the k-step-ahead forecast of x_{t+k}, about $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}$) to approximate the unconditional variance of the forecast errors with respect to the distributions of $\hat{\boldsymbol{\gamma}}$, of the past data $\{x_j\colon j = t, t-1, \ldots\}$, and of the future innovations, $\{a_j\colon j = t+1, t+2, \ldots\}$. References [3, 21, 34] also used a first-order Taylor approximation to calculate the conditional variance of the forecast errors with respect to the distributions of $\hat{\boldsymbol{\gamma}}$ and of the future innovations $\{a_j\colon j = t+1, t+2, \ldots\}$, conditioned on the past data $\{x_j\colon j = t, t-1, \ldots\}$. Reference 15 gives an excellent summary of this work and other issues involving forecasting and, in particular, prediction intervals in forecasting. Reference 27 compared the coverage probabilities of approximate prediction intervals based on conditional versus unconditional variances for AR(1) process.

Various later work extended these basic approaches to models other than ARMA. Reference 38 approximated the conditional forecast error variance for ARMA (p, q) processes with a linear trend and for ARIMA $(p, 1, q)$ processes with a constant (drift) term. From their asymptotic (with respect N) perspective, the uncertainty in the ARMA parameters was negligible. Hence, they considered parameter uncertainty in only the trend or drift parameters. Reference 24 approximated the conditional forecast error variance for a class of time series models that includes models in which the conditional mean of the k-step-ahead value can be written as linear function of past observations $\{x_t, x_{t-1}, \ldots, x_{t-d}\}$ for some finite d, which includes ARMA models if we allow $d \to \infty$. Various Bayesian prediction intervals have also been developed by calculating the posterior distribution of x_{t+k} given the past data $\{x_j\colon j = t, t-1, \ldots\}$, marginalized with respect to the posterior of $\boldsymbol{\gamma}$. See [15, Section 4.7] for a brief summary. It is conceptually straightforward to marginalize over the posterior of $\boldsymbol{\gamma}$, as discussed in [17] for example. However, simple closed-form expressions are lacking for all but the simplest cases.

In this article, we take a related but different approach and derive the partial derivative $\partial\sigma_z^2/\partial\boldsymbol{\gamma}$ of the unconditional variance of the filter output z_t, which can be used in a Taylor expansion to approximate σ_z^2. We introduce a relatively simple derivation, which results in quite tractable expressions for the partial derivatives. These provide insight into the robustness to parameter uncertainty and can be cleanly incorporated into subsequent analyses that require an approximate expression for σ_z^2, such as modifying prediction intervals to take into account parameter uncertainty.

Perhaps, the most significant distinction between this work and the prior time series forecasting work is that our results apply quite generally to the output of any linear filtering operation on an ARMA time series. The prior work applies only to forecasting, and then only to situations in which one is using a very specific form of predictor, namely, the predictor that would have minimum mean square prediction error (MSPE) if one were to assume the model structure is correct and $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}$. Reference 18 relaxed this slightly by assuming that one uses a predictor based on an AR model to predict a general ARMA time series, although they still required that the coefficients of the AR predictor are the minimum MSPE coefficients under the restricted AR predictor structure. There are reasons why one may wish to use a predictor other than the one that gives minimum MSPE for the assumed model, for example for implementation simplicity or if one doubts the assumed model and wishes to modify the predictor to be more robust to model uncertainty.

Moreover, the applicability of our results for a general linear filter structure has important implications in certain applications for which the objective is not purely to forecast the time series. One such application is statistical process control with autocorrelated data. Because of the increasing prevalence of autocorrelation in industrial process data, often due to sophisticated measurement and data collection technology that allows data to be automatically collected in steady streams, control charts for autocorrelated data have received a great deal of attention in the SPC literature ([30, 32, 40, 42, 47]). One class of methods involves fitting an ARMA model to the data and applying a traditional control chart (designed for temporally uncorrelated data) to the residuals. The ARMA residuals are calculated via the linear filtering operation $e_t = \hat{\Theta}^{-1}(B)\hat{\Phi}(B)x_t$, where the "^" symbol denotes an estimate of a quantity. The residuals are uncorrelated when there are no modeling errors. References 2, 7, 10, 16, 22, 23, 28, 31, 35, 43, 46 and many others have investigated residual-based control charts.

Perhaps, the most popular control chart for autocorrelated data is a residual-based exponentially weighted moving average (EWMA), in which one recursively calculates and charts the EWMA statistic $z_t = (1-\lambda)z_{t-1} + \lambda e_t$, where λ is a design parameter. At each time t, an alarm is sounded if z_t falls outside a set of lower/upper control limits $\{\text{LCL}, \text{UCL}\} = \pm L\sigma_z$, where L is a constant chosen by the user [28]. If there are no ARMA modeling errors, then $\sigma_z = \sigma_a \lambda^{1/2}(2-\lambda)^{-1/2}$, and L can be chosen to provide a desired false alarm rate or in-control average run length (ARL) [29, 30]. On the other hand, if there are modeling errors, then σ_z may be much different than believed. If σ_z is much larger than believed, the control chart will be plagued with far too many false alarms. This lack of robustness to modeling errors is a well-know drawback of residual-based control charts that has been observed empirically by many authors (e.g., [1, 7, 28]) and that will be demonstrated in Section 5. Because we can write the residual-based EWMA as Eq. (2) with $H(B) = \lambda[1-(1-\lambda)B]^{-1}\hat{\Theta}^{-1}(B)\hat{\Phi}(B)$, the results that we will derive in this article are applicable, providing analytical corroboration of the (lack of) robustness and a means of quantifying it, as well as insight into the mechanisms behind it.

We note that there are other methods of control charting autocorrelated data that also amount to filtering operations of the same form as Eq. (2). For example, one can apply an EWMA directly to the ARMA process x_t, instead of to the residuals, as proposed by [39, 40, 50]. In Section 3, we include an example that illustrates how our results can be used to assess the robustness of this control charting approach. Reference 6 provides a detailed performance and robustness comparison of a residual-based EWMA and an EWMA applied to x_t. Reference 26 discusses a broad variety of issues related to parameter estimation in control charting, with emphasis on iid processes.

The format of the remainder of this article is as follows. In Section 2, we derive expressions for the sensitivities of σ_z^2 with respect to deviations in the ARMA parameters $\{\phi_i: i = 1, 2, \ldots, p\}$ and $\{\theta_i: i = 1, 2, \ldots, q\}$. Here, we define the sensitivities as $\partial ln\sigma_z^2/\partial\boldsymbol{\gamma} = (\partial\sigma_z^2/\partial\boldsymbol{\gamma})/\sigma_z^2$. They provide measures of the extent to which parameter errors increase or decrease σ_z^2. The sensitivity measures do not directly incorporate any measures of uncertainty in $\boldsymbol{\gamma}$ (e.g., a parameter covariance matrix), and in this article, we do not develop any new approach for estimating the uncertainty in $\boldsymbol{\gamma}$. However, in Section 3, we combine the sensitivity results with standard results on the distribution of $\hat{\boldsymbol{\gamma}}$ to derive an approximate confidence interval for σ_z^2 that takes into account the uncertainty in the ARMA parameter estimates. In Sections 4—6, we provide three examples to illustrate how the results can be used to understand the impact of ARMA parameter estimation errors and to design filtering procedures that are more robust. Section 4 considers EWMA forecasting. Section 5 elaborates on the residual-based EWMA control charting application. Section 6 introduces an application to feedback control of run-to-run manufacturing processes with ARIMA disturbances. Section 7 concludes the paper.

2. THE SENSITIVITY OF σ_Z^2

Rewrite the filtering operation of Eq. (2) as

$$z_t = H(B)x_t = G(B)a_t = \sum_{j=0}^{\infty} g_j a_{t-j} \tag{3}$$

where $G(B) = \Phi^{-1}(B)\Theta(B)H(B)$ with impulse response coefficients denoted by $\{g_j: j = 0, 1, 2, \ldots\}$. Using the representation in Eq. (3), the (asymptotic) variance of z_t is given by [12]

$$\sigma_z^2 = \sigma_a^2 \sum_{j=0}^{\infty} g_j^2. \tag{4}$$

We assume that the filter $H(B)$ under consideration does not depend on $\boldsymbol{\gamma}$ (which is unknown in practice), although $H(B)$ will generally be designed based on the estimate $\hat{\boldsymbol{\gamma}}$.

Consider the following quantities, which we will use as measures of the sensitivity of σ_z^2 with respect to the ARMA parameters:

$$S_{\phi,i} = \frac{\frac{\partial\sigma_z^2}{\partial\phi_i}}{\sigma_z^2}: i = 1, 2, \ldots, p, \text{ and} \tag{5}$$

$$S_{\theta,i} = \frac{\frac{\partial\sigma_z^2}{\partial\theta_i}}{\sigma_z^2}: i = 1, 2, \ldots, q. \tag{6}$$

In the Appendix, we show that for the output $z_t = H(B)x_t$ of a stable, time-invariant linear filter applied to an ARMA (p, q) process x_t, the sensitivities are

$$S_{\phi,i} = 2\sum_{k=0}^{\infty} P_k\rho_{i+k}: i = 1, 2, \ldots, p, \text{ and} \tag{7}$$

$$S_{\theta,i} = -2\sum_{k=0}^{\infty} Q_k\rho_{i+k}: i = 1, 2, \ldots, q, \tag{8}$$

where ρ_k denotes the autocorrelation function of z_t at lag k, and $\{P_k: k = 0, 1, 2, \ldots\}$ and $\{Q_k: k = 0, 1, 2, \ldots\}$ denote the impulse response coefficients of $\Phi^{-1}(B) = \sum_{j=0}^{\infty} P_j B^j$ and $\Theta^{-1}(B) = \sum_{j=0}^{\infty} Q_j B^j$, respectively.

Equations (7) and (8) indicate that the sensitivities to the AR and MA parameters depend only on the nominal AR and MA polynomials, respectively, and the nominal autocorrelation function of the filter output z_t. By "nominal," we mean the quantities that result from using the ARMA parameter values at which the partial derivatives in Eqs. (5) and (6) are evaluated, such as a set of estimated ARMA parameters. More specifically, Eqs. (7) and (8) state that $S_{\phi,i}$ and $S_{\theta,i}$ are the weighted sums of the impulse response coefficients of $\Phi^{-1}(B)$ and $\Theta^{-1}(B)$, where the weights are given by the coefficients of the nominal autocorrelation function of z_t. The faster the autocorrelation function decays or, more generally, the smaller the autocorrelation function coefficients, the lower the sensitivities. In the extreme case that z_t is temporally uncorrelated, the sensitivities are zero. This is the case in many forecasting applications in which z_t represents one-step ahead forecasting errors. For example, ARMA-based forecasting typically results in uncorrelated one-step-ahead forecasting errors. The fact that the sensitivities are zero in this case is consistent with the fact that the parameters in ARMA-based forecast equations are chosen to minimize the variance of the one-step-ahead forecasting errors (i.e., the residuals). Many-step-ahead forecasting errors are generally autocorrelated, on the other hand, so their sensitivities will generally be nonzero. We demonstrate this in Section 4.

Because the sensitivities are partial derivatives, they can be used to assess the effects of parameter errors on σ_z^2. Notice that the sensitivities are the elements of $\sigma_z^{-2}\partial\sigma_z^2/\partial\boldsymbol{\gamma} = \partial ln(\sigma_z^2)/\partial\boldsymbol{\gamma}$. Hence, for a specified $\hat{\boldsymbol{\gamma}}$ and $\boldsymbol{\gamma}$, the sensitivities can be used in a first-order Taylor approximation of $ln(\sigma_z^2)$ to approximate the increase/decrease in σ_z^2 due to a specific set of parameter errors. We illustrate the calculations for this with an example in Section 5. If one characterizes the parameter uncertainty via a parameter covariance matrix (commonly produced by commercial time series modeling software) instead of by specifying hypothetical values for $\hat{\boldsymbol{\gamma}}$ and $\boldsymbol{\gamma}$, then the sensitivities can also be used in a Taylor approximation to derive a confidence interval for σ_z^2. This is the subject of Section 3.

One interesting characteristic of Eqs. (7) and (8) is the symmetry with respect to the AR and MA polynomials: $S_{\phi,i}$ depends on $\Phi(B)$ in exactly the same manner in which $S_{\theta,i}$ depends on $\Theta(B)$. Consider the special case that z_t is a first-order ARMA process, for which $\Phi^{-1}(B) = (1-\phi_1 B)^{-1} = 1 + \phi_1 B + \phi_1^2 B^2 + \ldots$, and $\Theta^{-1}(B) = (1-\theta_1 B)^{-1} = 1 + \theta_1 B + \theta_1^2 B^2 + \ldots$. In this case, $P_k = \phi_1^k$, and $Q_k = \theta_1^k$ so that Eqs. (7) and (8) reduce to

$$S_{\phi,1} = 2\sum_{k=0}^{\infty}\phi_1^k \rho_{k+1}, \text{ and}$$

$$S_{\theta,1} = -2\sum_{k=0}^{\infty}\theta_1^k \rho_{k+1}.$$

This implies that the sensitivities are generally smaller when the AR and MA parameters have smaller magnitudes. The insight into the robustness of the filtering operation that is provided by the sensitivity measures can be useful when designing and analyzing the filter, as will be illustrated with the examples in Sections 4–6.

3. A CONFIDENCE INTERVAL FOR σ_Z^2

Let $\hat{\sigma}_z^2$ denote the nominal variance of z_t that would result if $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}$. In other words, $\hat{\sigma}_z^2$ is the variance of the time series $\hat{G}(B)a_t$ with $\hat{G}(B) = \hat{\Phi}^{-1}(B)\hat{\Theta}(B)H(B)$. To derive a confidence interval for σ_z^2, one might consider deriving a first-order Taylor expansion of $\hat{\sigma}_z^2$ with respect to $\hat{\boldsymbol{\gamma}}$ (about $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}$) in a manner similar to how the sensitivities were derived in the preceding section. This approach would not be well formulated, however, because $H(B)$ will generally depend heavily on $\hat{\boldsymbol{\gamma}}$, either explicitly or implicitly. In Sections 4–6, we provide examples in which $H(B)$ depends on $\hat{\boldsymbol{\gamma}}$ explicitly.

Instead, we use the following approach. By definition of the sensitivities evaluated at $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}$, we have the first order Taylor approximation

$$\frac{\sigma_z^2}{\hat{\sigma}_z^2} \cong 1 + \mathbf{V}^T(\hat{\boldsymbol{\gamma}})[\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}}] \cong 1 + \mathbf{V}^T(\boldsymbol{\gamma})[\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}}], \tag{9}$$

where $\mathbf{V}(\boldsymbol{\gamma}) = [S_{\phi,1}\, S_{\phi,2}\, \ldots\, S_{\phi,p}\, S_{\theta,1}\, S_{\theta,2}\, \ldots\, S_{\theta,q}]^T$ evaluated at $\boldsymbol{\gamma}$, and $S_{\phi,i}$ and $S_{\theta,i}$ are given by Eqs. (7) and (8). The second equality in (9) follows by substituting a first-order Taylor approximation of $\mathbf{V}(\boldsymbol{\gamma})$ about $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}$ and ignoring second-order terms. If we invoke standard asymptotic approximations and assume that $\hat{\boldsymbol{\gamma}}$ follows a multivariate normal distribution with mean $\boldsymbol{\gamma}$ and covariance matrix denoted by $\Sigma_{\hat{\boldsymbol{\gamma}}}(\boldsymbol{\gamma})$, then the first-order approximation of $\sigma_z^2/\hat{\sigma}_z^2$ implies that it is approximately normally distributed with mean 1 and variance $\mathbf{V}^T(\boldsymbol{\gamma})\Sigma_{\hat{\boldsymbol{\gamma}}}(\boldsymbol{\gamma})\mathbf{V}(\boldsymbol{\gamma})$. Reference 12 discusses the asymptotic distribution of $\hat{\boldsymbol{\gamma}}$ and provides expressions for the asymptotic covariance [also see Eq. (19) below]. Note that we have written $\Sigma_{\hat{\boldsymbol{\gamma}}}(\boldsymbol{\gamma})$ as a function of $\boldsymbol{\gamma}$ to indicate that the asymptotic covariance is a function of the true parameters.

This yields the approximate two-sided $1-\alpha$ confidence interval

$$\hat{\sigma}_z^2\left(1 - z_{\alpha/2}\sqrt{\mathbf{V}^T(\hat{\boldsymbol{\gamma}})\Sigma_{\hat{\boldsymbol{\gamma}}}(\hat{\boldsymbol{\gamma}})\mathbf{V}(\hat{\boldsymbol{\gamma}})}\right) \le \sigma_z^2 \le \hat{\sigma}_z^2\left(1 + z_{\alpha/2}\sqrt{\mathbf{V}^T(\hat{\boldsymbol{\gamma}})\Sigma_{\hat{\boldsymbol{\gamma}}}(\hat{\boldsymbol{\gamma}})\mathbf{V}(\hat{\boldsymbol{\gamma}})}\right), \tag{10}$$

where $z_{\alpha/2}$ denotes the upper $\alpha/2$ quantile of the standard normal distribution, and we have further approximated by evaluating $\mathbf{V}$ and $\Sigma_{\hat{\boldsymbol{\gamma}}}$ at the estimates $\hat{\boldsymbol{\gamma}}$.

Depending on one's perspective, a Bayesian credible region [9] for σ_z^2 may be more meaningful than the confidence interval (10). In fact, it is straightforward to show that for a noninformative prior on $\boldsymbol{\gamma}$, a Bayesian credible region based on the approximate posterior distribution of σ_z^2 coincides with Eq. (10). This follows by noting that for the asymptotic multivariate normal distribution for $\hat{\boldsymbol{\gamma}}|\boldsymbol{\gamma}$ that we assumed in the development of Eq. (10), coupled with a noninformative prior on $\boldsymbol{\gamma}$, the posterior distribution of $\boldsymbol{\gamma}$ is also approximately multivariate normal with mean $\hat{\boldsymbol{\gamma}}$ and the same covariance matrix $\Sigma_{\boldsymbol{\gamma}}(\hat{\boldsymbol{\gamma}}) = \Sigma_{\hat{\boldsymbol{\gamma}}}(\hat{\boldsymbol{\gamma}})$. Substituting this into Eq. (9) gives (10) as a Bayesian credible region for σ_z^2.

In deriving the confidence interval (10), we have approximated the distribution of $\sigma_z^2/\hat{\sigma}_z^2$ as normal. Because in many related problems the log of the variance is often closer to normal than the variance, we might obtain a more accurate confidence interval by using a Taylor expansion of $ln(\sigma_z^2/\hat{\sigma}_z^2)$ about $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}$. Noticing that $\partial ln(\sigma_z^2)/\partial\boldsymbol{\gamma} = \sigma_z^{-2}\partial\sigma_z^2/\partial\boldsymbol{\gamma}$,

Table 1. Monte Carlo estimated coverage probabilities for the confidence intervals (10) and (12) for $1-\alpha=0.9$. The top number in each cell is for (12), and the bottom, for (10). The example is for an EWMA applied directly to an ARMA x_t.

	Sample size N		
Model	50	100	500
AR(1), $\phi=0.5$	0.897 0.853	0.893 0.882	0.900 0.897
AR(1), $\phi=0.7$	0.886 0.818	0.890 0.854	0.896 0.892
AR(1), $\phi=0.9$	0.848 0.747	0.878 0.797	0.892 0.874
ARMA(1), $\phi=0.9,\phi=0.5$	0.815 0.722	0.874 0.801	0.892 0.872

which are the sensitivity indices derived in Section 2, we have an alternative first order Taylor approximation

$$\ln\left(\frac{\sigma_z^2}{\hat{\sigma}_z^2}\right) \cong \mathbf{V}^T(\hat{\boldsymbol{\gamma}})[\boldsymbol{\gamma}-\hat{\boldsymbol{\gamma}}] \cong \mathbf{V}^T(\boldsymbol{\gamma})[\boldsymbol{\gamma}-\hat{\boldsymbol{\gamma}}]. \quad (11)$$

Proceeding as earlier with the asymptotic multivariate normal distribution assumed for $\hat{\boldsymbol{\gamma}}|\boldsymbol{\gamma}$ (equivalently, $\boldsymbol{\gamma}|\hat{\boldsymbol{\gamma}}$), we have the following alternative approximate two-sided $1-\alpha$ confidence interval (equivalently, Bayesian credible region) for σ_z^2:

$$\hat{\sigma}_z^2 \exp\left\{-z_{\alpha/2}\sqrt{\mathbf{V}^T(\hat{\boldsymbol{\gamma}})\Sigma_{\hat{\boldsymbol{\gamma}}}(\hat{\boldsymbol{\gamma}})\mathbf{V}(\hat{\boldsymbol{\gamma}})}\right\} \le \sigma_z^2 \le \hat{\sigma}_z^2 \exp\left\{z_{\alpha/2}\sqrt{\mathbf{V}^T(\hat{\boldsymbol{\gamma}})\Sigma_{\hat{\boldsymbol{\gamma}}}(\hat{\boldsymbol{\gamma}})\mathbf{V}(\hat{\boldsymbol{\gamma}})}\right\}. \quad (12)$$

We conducted a simulation to compare the coverage probabilities of the two confidence intervals (10) and (12). In the simulation examples, the filter output $z_t=(1-\lambda)z_{t-1}+\lambda x_t$ is an EWMA applied directly to an ARMA process x_t instead of to the residuals. References 39, 40, 50 investigated the use of this as a control chart statistic for autocorrelated data and discussed how to determine the control limits (without consideration of parameter uncertainty). Because the control limits are based on $\hat{\sigma}_z^2$, the effect of parameter uncertainty on σ_z^2 is of direct interest here. For this example, we have $H(B)=\lambda[1-(1-\lambda)B]^{-1}$ and $G(B)=\lambda[1-(1-\lambda)B]^{-1}\Phi^{-1}(B)\Theta(B)$. Table 1 compares the actual coverage probabilities for the two confidence intervals for a few different sample sizes and AR(1) or ARMA (1,1) models. In all cases, we used an EWMA parameter $\lambda=0.1$, a confidence level $1-\alpha=90\%$, and 10,000 Monte Carlo replicates. Hence, the standard errors for the estimated coverage probabilities are $[0.9(1-0.9)/10,000]^{1/2}=0.003$. Both methods appear to provide smaller coverage probability than the desired 0.9, although the confidence interval (12) based on the Taylor approximation of $ln(\sigma_z^2)$ is much more accurate. For all but the smallest sample size of $N=50$, it gives a coverage probability that is quite close to the desired 0.9. Hence, we recommend using (12) instead of (10).

When combined with standard expressions for estimating the parameter covariance matrix, Eq. (12) can also yield guidelines for selecting a sample size to provide ARMA parameter estimates that are sufficiently accurate, as determined by the context of the filtering application. The standard asymptotic expression for $\Sigma_{\hat{\boldsymbol{\gamma}}}(\hat{\boldsymbol{\gamma}})$ is inversely proportional to N [see, e.g., Eq. (19) below]. Consequently, with preliminary estimates or approximate guesses for the ARMA parameters, one could specify a desired width for the confidence interval of Eq. (10) and then solve for the required sample size. We illustrate this in Section 4 in the context of EWMA forecasting.

4. APPLICATION TO EWMA FORECASTING

When forecasting time series, the standard deviation of the forecasting error is an important quantity because confidence bands on the forecasts are directly proportional to it. In this section, we investigate the effect of parameter errors on the standard deviation of EWMA forecast errors. EWMA forecasting [36] is one of the most popular methods of forecasting nonseasonal time series. For $k=1,2,3,\ldots$ the k-step-ahead EWMA forecast of a time series x_t is

$$\hat{x}_{t+k|t}=\lambda\sum_{j=0}^{\infty}(1-\lambda)^j x_{t-j}=\frac{\lambda}{1-(1-\lambda)B}x_t,$$

where $0<\lambda\le 1$ is the EWMA parameter. It is well known that for a first-order integrated moving average (IMA) model of the form

$$x_t=\frac{1-\theta B}{1-B}a_t,$$

an EWMA forecast with $\lambda=1-\theta$ has minimum mean square error [12]. Hence, one popular approach for selecting λ is to fit a first-order IMA model to x_t and then set $\lambda=1-\hat{\theta}$.

Consider the k-step-ahead forecast error

$$z_t^{(k)}=x_t-\hat{x}_{t|t-k} = x_t-\frac{1-\hat{\theta}}{1-\hat{\theta}B}x_{t-k}=\left[1-\frac{(1-\hat{\theta})B^k}{1-\hat{\theta}B}\right]x_t,$$

which qualifies as a linear filtering operation of the same form as Eq. (3) with

$$G^{(k)}(B)=\left[1-\frac{(1-\hat{\theta})B^k}{1-\hat{\theta}B}\right]\frac{1-\theta B}{1-B}. \quad (13)$$

We have added a superscript (k) to $G(B)$ and to z_t to indicate the forecasting horizon. Likewise, a superscript (k) on other quantities will indicate that they are with respect to the k-step-ahead forecast. Because the forecasting error is a linearly filtered time series, we can apply the results of the previous sections to develop a confidence interval for the standard deviation of $z_t^{(k)}$ (denoted by $\sigma_z^{(k)}$), taking into account the error in estimating θ.

From (13), it follows that

$$\hat{G}^{(k)}(B) = \left[1 - \frac{(1-\hat{\theta})B^k}{1-\hat{\theta}B}\right]\frac{1-\hat{\theta}B}{1-B} = 1+(1-\hat{\theta})(B+B^2+\cdots B^{k-1}).$$

Hence, $z_t^{(k)}$ nominally follows a moving average model of order $(k-1)$, which has variance and autocorrelation function (see [12]):

$$\left(\hat{\sigma}_z^{(k)}\right)^2 = \hat{\sigma}_a^2\sum_{j=0}^{\infty}\hat{g}_j^2 = \hat{\sigma}_a^2[1+(k-1)(1-\hat{\theta})^2], \text{ and}$$

$$\hat{\rho}_j^{(k)} = \begin{cases} \dfrac{1-\hat{\theta}+(k-1-j)(1-\hat{\theta})^2}{1+(k-1)(1-\hat{\theta})^2} : & 1 \le j \le k-1 \\ 0 : & k \le j \end{cases} : \quad j = 1, 2, \ldots$$

Substituting the autocorrelation function into (8) with $i = 1$ and $\hat{Q}_j = \hat{\theta}^j$ gives the following sensitivity of $(\hat{\sigma}_z^{(k)})^2$ with respect to θ:

$$\begin{aligned} S_\theta^{(k)} &= -2\sum_{j=0}^{\infty}\hat{Q}_j\hat{\rho}_{1+j}^{(k)} \\ &= -2\sum_{j=0}^{k-2}\hat{\theta}^j\left[\frac{1-\hat{\theta}+(k-2-j)(1-\hat{\theta})^2}{1+(k-1)(1-\hat{\theta})^2}\right] \\ &= -2\left[\frac{1-\hat{\theta}+(k-2)(1-\hat{\theta})^2}{1+(k-1)(1-\hat{\theta})^2}\right]\sum_{j=0}^{k-2}\hat{\theta}^j \\ &\quad + 2\left[\frac{(1-\hat{\theta})^2}{1+(k-1)(1-\hat{\theta})^2}\right]\sum_{j=0}^{k-2}j\hat{\theta}^j \\ &= -2\left[\frac{1-\hat{\theta}+(k-2)(1-\hat{\theta})^2}{1+(k-1)(1-\hat{\theta})^2}\right]\left(\frac{1-\hat{\theta}^{k-1}}{1-\hat{\theta}}\right) \\ &\quad + 2\left[\frac{(1-\hat{\theta})^2}{1+(k-1)(1-\hat{\theta})^2}\right] \\ &\quad \times\left(\frac{\hat{\theta}-(k-1)\hat{\theta}^{k-1}+(k-2)\hat{\theta}^k}{(1-\hat{\theta})^2}\right) \\ &= -2\left[\frac{(1-\hat{\theta})(k-1)}{1+(k-1)(1-\hat{\theta})^2}\right]. \end{aligned} \tag{14}$$

The last equality follows from straightforward but tedious algebra.

Eq. (14) implies that the sensitivity $S_\theta^{(1)}$ of the variance of the one-step-ahead forecast error is zero. This is rather obvious if we notice that $S_\theta^{(1)} = 0$ means precisely that $\partial\sigma_z^2/\partial\theta = 0$, which must be true because the variance of the one-step-ahead prediction errors (i.e., the time series residuals) is minimized when $\theta = \hat{\theta}$. The more interesting implications of Eq. (14) are for $k > 1$. Equation (14) implies that the magnitude of the sensitivity monotonically increases in k, up to the limiting value of $2(1-\hat{\theta})^{-1}$.

A two-sided $1-\alpha$ confidence interval for the k-step-ahead forecast error variance is obtained by substituting Eq. (14) for $\mathbf{V}(\hat{\boldsymbol{\gamma}})$ into (12) with $\Sigma_{\hat{\gamma}}(\hat{\boldsymbol{\gamma}}) = N^{-1}(1-\hat{\theta}^2)$. The parameter covariance is an asymptotic expression that applies to most common methods of estimating θ based on nonlinear least squares [12]. The confidence interval for the standard deviation $\sigma_z^{(k)}$ becomes

$$\sigma_z^{(k)} \in \hat{\sigma}_z^{(k)}\exp\left\{\pm\frac{z_{\alpha/2}}{2N^{1/2}}\left[\frac{2|1-\hat{\theta}|(k-1)(1-\hat{\theta}^2)^{1/2}}{1+(k-1)(1-\hat{\theta})^2}\right]\right\}. \tag{15}$$

To illustrate the effect of parameter uncertainty on the standard deviation of the k-step-ahead forecast error, Fig. 1 plots $\hat{\sigma}_z^{(k)}$ and the upper and lower boundaries of the confidence interval (15) versus k for a specific case with $\hat{\theta} = 0.8$ (which corresponds to EWMA parameter $\lambda = 0.2$, a relatively common choice), sample size $N = 50$, $\sigma_a = \hat{\sigma}_a = 1.0$, and $\alpha = 0.05$. There is clearly substantial uncertainty in the standard deviation of the prediction error. For a prediction horizon of $k = 30$, the upper boundary of the confidence interval is 56% larger than the nominal standard deviation. The uncertainty in the prediction error standard deviation translates to

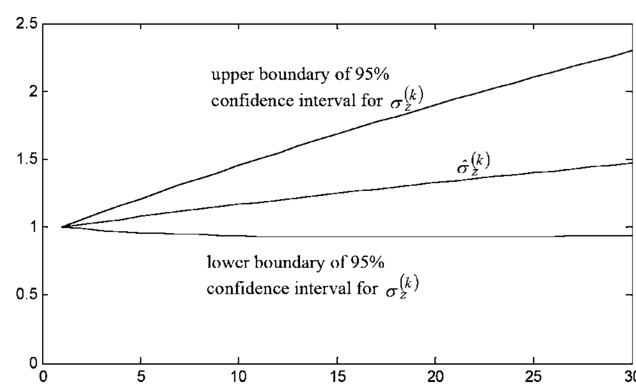

Figure 1. A plot of the nominal standard deviation of the k-step-ahead EWMA forecasting error versus k, together with the upper and lower boundaries of a 95% confidence interval for the true standard deviation from Eq. (15).

inaccurate confidence bands on the k-step-ahead forecasts, if parameter uncertainty is ignored.

If we have preliminary estimates of the parameters available, we can use the form of the confidence interval (15) to determine a suitable sample size that will yield sufficiently accurate ARMA parameter estimates. For example, suppose we want to insure that the upper boundary of a 95% confidence interval on $\sigma_z^{(30)}$ is no more than (say) 10% larger than the nominal value $\hat{\sigma}_z^{(30)}$. We can substitute $\sigma_z^{(30)}/\hat{\sigma}_z^{(30)} = 1.1$ into the upper boundary of (15) with the preliminary estimate $\hat{\theta} = 0.8$ and solve for the required N. This gives $N = 1098$ for this example, a relatively large sample size.

5. APPLICATION TO RESIDUAL-BASED EWMA CHARTS FOR AUTOCORRELATED DATA

The expressions (7) and (8) for the sensitivities simplify considerably when z_t is a residual-based EWMA, for which $H(B) = \lambda[1-(1-\lambda)B]^{-1}\hat{\Theta}^{-1}(B)\hat{\Phi}(B)$. At the nominal values $\Theta(B) = \hat{\Theta}(B)$ and $\Phi(B) = \hat{\Phi}(B)$, we have $G(B) = \hat{\Phi}^{-1}(B)\hat{\Theta}(B)H(B) = (1-\nu)[1-\nu B]^{-1}$, where $\nu = 1-\lambda$. Consequently, the nominal $z_t = (1-\nu)[1-\nu B]^{-1}a_t$ behaves as a first-order autoregressive process, and its autocorrelation function is $\rho_k = \nu^k$ [12]. Substituting this into Eqs. (7) and (8), the sensitivities for the residual-based EWMA become

$$S_{\phi,i} = 2\sum_{k=0}^{\infty} P_k \nu^{k+i} = 2\nu^i \sum_{k=0}^{\infty} P_k B^k|_{B=\nu} = 2\nu^i \Phi^{-1}(B)|_{B=\nu} = \frac{2\nu^i}{\Phi(\nu)} \quad (16)$$

and

$$S_{\theta,i} = -2\sum_{k=0}^{\infty} Q_k \nu^{k+i} = -2\nu^i \sum_{k=0}^{\infty} Q_k B^k|_{B=\nu} = -2\nu^i \Theta^{-1}(B)|_{B=\nu} = \frac{-2\nu^i}{\Theta(\nu)} \quad (17)$$

where $\Phi(\nu) = 1-\phi_1\nu-\phi_2\nu^2\ldots-\phi_p\nu^p$, and $\Theta(\nu) = 1-\theta_1\nu-\theta_2\nu^2\ldots-\theta_q\nu^q$.

The sensitivities have clearer interpretations if we consider the factorization $\Phi(B) = (1-\eta_1 B)(1-\eta_2 B)\ldots(1-\eta_p B)$ in terms of the roots $\{\eta_1,\eta_2,\ldots,\eta_p\}$ of $\Phi(B)$ and do similarly for $\Theta(B)$. The magnitude of the sensitivity in (16) becomes

$$|S_{\phi,i}| = 2\nu^i \prod_{j=1}^{p} \frac{1}{|1-\eta_j\nu|} \quad (18)$$

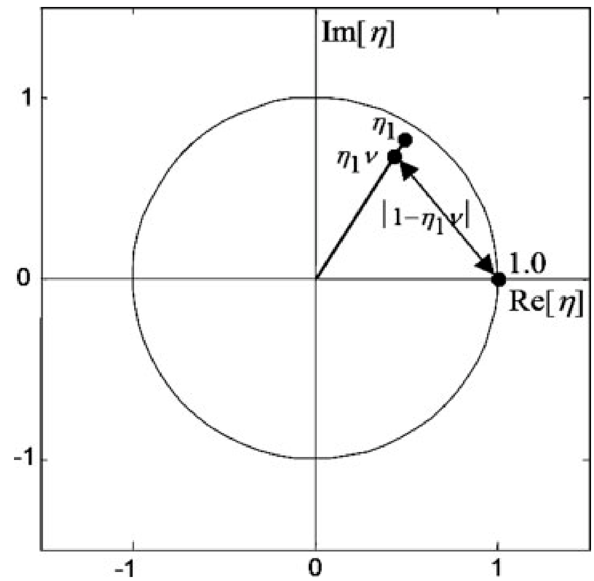

Figure 2. Illustration of the distance between the scaled root $\eta_1\nu$ and the point 1.0 in the complex plane. Smaller distances increase the sensitivity $S_{\phi,i}$.

Each term $|1-\eta_j\nu|$ in the denominator of (18) represents the distance between the scaled root $\eta_j\nu$ and the point 1.0 (i.e., the intersection of the unit circle and the positive real axis) in the complex plane. This is illustrated in Fig. 2 for a complex root and $\nu = 0.9$. The sensitivity will be large if any root is close to the point 1.0 and the EWMA parameter λ is small (ν close to 1). Complex conjugate roots near the stability boundary (the unit circle) do not necessarily result in large sensitivity. In contrast, roots on the positive real axis near the stability boundary always result in large sensitivity if λ is small. Analogous results hold for $S_{\theta,i}$ in terms of the roots of $\Theta(B)$.

This has important implications when designing a residual-based EWMA chart. If any roots of the AR or MA polynomials are close to 1.0, then one should be wary of choosing a small value for the EWMA parameter λ. A larger value for λ will provide a control chart whose false alarm rate is more robust to errors in the ARMA parameters. On the other hand, if none of the roots are close to 1.0, then one can use a smaller value of λ, if desired, without severe adverse consequences on the false alarm rate. We illustrate the robustness issues with the following example.

5.1. Example

Consider the Series A data from [12], which are 197 concentration measurements from a chemical process. Reference 12 found that an ARMA(1,1) model fit the data well, and the estimated parameters were $\hat{\phi}_1 = 0.87$, $\hat{\theta}_1 = 0.48$, and $\hat{\sigma}_a = 0.313$. The assumed standard deviation of the residual-based

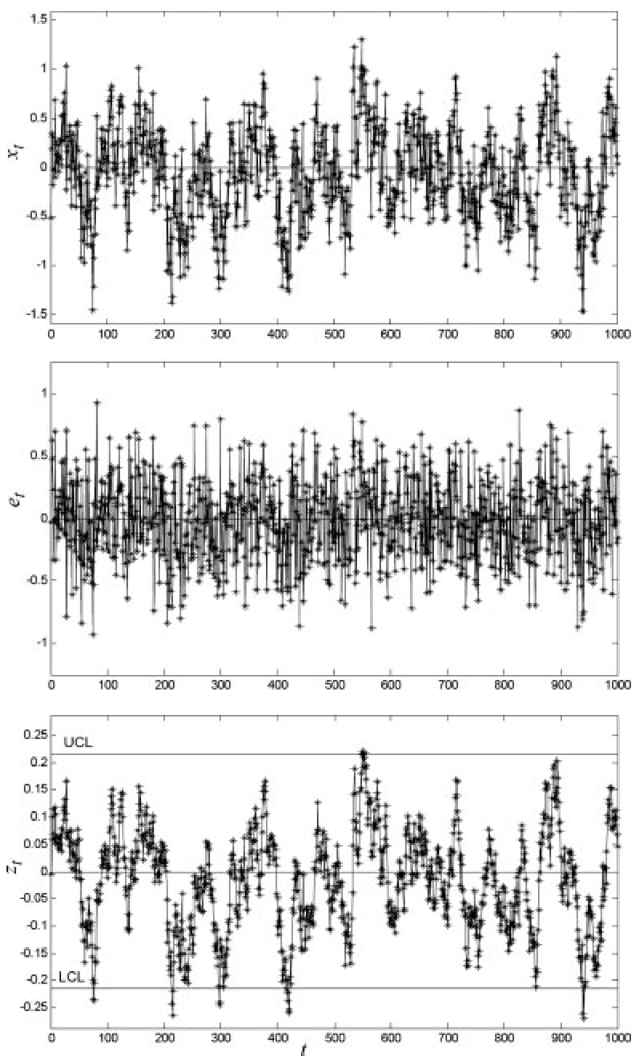

Figure 3. Simulated chemical concentration data example illustrating the increased false alarm rate for an EWMA chart with modeling errors. The three panels from top to bottom are the autocorrelated data x_t, the residuals e_t, and the residual-based EWMA z_t.

EWMA $z_t = (1-\lambda)z_{t-1} + \lambda e_t$ is $\hat{\sigma}_z = \hat{\sigma}_a \lambda^{1/2}(2-\lambda)^{-1/2}$ under the assumption that the true parameters coincide with the preceding estimates. If we choose an EWMA parameter $\lambda = 0.10$, the assumed standard deviation becomes $\hat{\sigma}_z = 0.0718$. Suppose we set the upper and lower control limits at $\{\text{LCL}, \text{UCL}\} = \pm 3\hat{\sigma}_z = \pm 0.216$.

To gage the effects of modeling errors on the EWMA, we can calculate the sensitivities from Eqs. (16) and (17), evaluated at $\phi_1 = \hat{\phi}_1$, and $\theta_1 = \hat{\theta}_1$, and $\nu = 0.9$. Equations (16) and (17) become $S_{\phi,1} = 8.29$ and $S_{\theta,1} = -3.17$. The sensitivities can be used in the Taylor approximation $\ln(\sigma_z^2/\hat{\sigma}_z^2) \cong \mathbf{V}^T(\hat{\boldsymbol{\gamma}})[\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}}]$ [see Eq. (11)] to approximate the increase in the EWMA standard deviation for specific values of the parameter errors. The Taylor approximation reduces to $\sigma_z/\hat{\sigma}_z \cong \exp\{\mathbf{V}^T(\hat{\boldsymbol{\gamma}})[\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}}]/2\} = \exp\{S_{\phi,1}(\phi - \hat{\phi}_1)/2\}$. For illustrative purposes, suppose that θ_1 and σ_a coincide with their estimates but that $\phi_1 = 0.90$. In this case, the approximate ratio is $\sigma_z/\hat{\sigma}_z \cong exp\{8.29(0.9-0.87)/2\} = 1.13$, which represents a 13% increase in the EWMA standard deviation.

A standard deviation that is 13% larger than believed will result in substantially more frequent false alarms. Consider the asymptotic false alarm probability, defined as $\alpha = \lim_{t\to\infty} Pr\{|z_t| > 3\hat{\sigma}_z\}$, under the premise that z_t is allowed to reach steady-state without resetting the control chart statistic after alarms are sounded. Assuming the a_t are normally distributed, $z_t \sim N(0, \sigma_z^2)$ in the limit. Hence, using the $\pm 3\hat{\sigma}_z$ control limits translates to an assumed asymptotic false alarm probability of $\alpha = 0.0027$. If σ_z is 13% larger than the assumed value $\hat{\sigma}_z = 0.0718$, we have $\sigma_z = 0.0813$. This results in an actual asymptotic false alarm probability of $\alpha = 0.0080$, which is three times larger than the assumed value.

The consequences of this are illustrated in Fig. 3. The top panel shows 1,000 simulated observations of x_t using $\theta_1 = 0.48$, $\sigma_a = 0.313$, and $\phi_1 = 0.90$ as the true parameters. The middle panel shows the residuals, calculated via $e_t = (1-\hat{\theta}_1 B)^{-1}(1-\hat{\phi}_1 B)x_t$ with $\hat{\phi}_1 = 0.87$ and $\hat{\theta}_1 = 0.48$. The bottom panel shows the EWMA on e_t with control limits ± 0.216, which were based on the assumed EWMA standard deviation $\hat{\sigma}_z = 0.0718$. All of the alarms in the bottom panel were false alarms because the mean was set at zero throughout the simulation. The excessive number of false alarms is a direct consequence of the inflation in σ_z due to the error in ϕ_1.

One can use the sensitivity results to take robustness into account when selecting an appropriate value for λ. Figure 4 plots the sensitivities, as functions of λ, for the chemical concentration example. The sensitivities monotonically converge to zero as λ increases. Figure 4 indicates that for values of λ

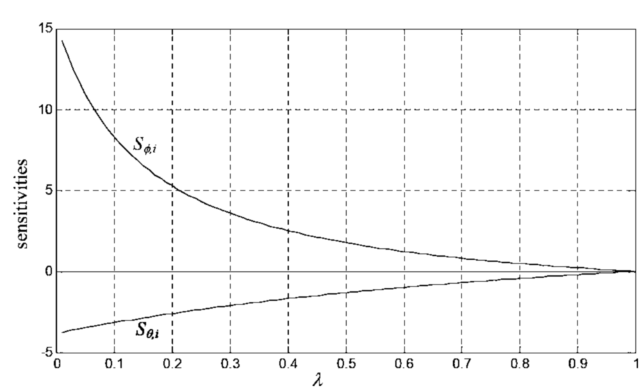

Figure 4. Plot of the AR and MA parameter sensitivities for the chemical concentration data example, as a function of the EWMA parameter λ.

smaller than 0.1, the sensitivity $S_{\phi,1}$ increases dramatically. Consequently, small values of λ should be used with caution, unless one is quite certain that there is little uncertainty in ϕ_1. Figure 4 also indicates that if one chooses a larger value of (say) $\lambda = 0.3$, then the AR sensitivity decreases from $S_{\phi,1} = 8.29$ to $S_{\phi,1} = 3.58$. Repeating the preceding calculations, this translates to a 5.5% increase in σ_z for $\phi_1 = 0.90$, which might be considered acceptable.

Instead of working directly with the parameter sensitivities and considering specific values of the parameter errors, one might prefer calculating the confidence interval for σ_z that is described in Section 3. An asymptotic expression for the ARMA(1,1) parameter covariance for this example is [12]

$$\Sigma_{\hat{\gamma}}(\hat{\boldsymbol{\gamma}}) = \frac{(1-\hat{\phi}\hat{\theta})}{N(\hat{\phi}-\hat{\theta})^2}\begin{bmatrix}(1-\hat{\phi}^2)(1-\hat{\phi}\hat{\theta}) & (1-\hat{\phi}^2)(1-\hat{\theta}^2)\\ (1-\hat{\phi}^2)(1-\hat{\theta}^2) & (1-\hat{\theta}^2)(1-\hat{\phi}\hat{\theta})\end{bmatrix} = \begin{bmatrix}2.75 & 3.64\\ 3.64 & 8.71\end{bmatrix}\times 10^{-3}, \quad (19)$$

and the vector of parameter sensitivities is $\mathbf{V}(\hat{\boldsymbol{\gamma}}) = [8.29\ {-3.17}]^{\mathrm{T}}$. Substituting these into Eq. (12) and taking the square root gives

$$0.751\hat{\sigma}_z \le \sigma_z \le 1.331\hat{\sigma}_z$$

as an approximate 95% confidence interval for σ_z. Consequently, the level of parameter uncertainty in this case could result in an EWMA standard deviation that is as much as 33% larger or 25% smaller than the assumed value $\hat{\sigma}_z = 0.0718$. A 33% increase in the EWMA standard deviation would cause a much greater increase in the number of false alarms than what is illustrated in Fig. 3, for which the EWMA standard deviation was only 13% larger than the assumed value. Similar to Fig. 4, one could plot the upper and lower bounds of a $1-\alpha$ confidence interval for σ_z versus λ, to aid in selecting a value of λ that provides an acceptable level of robustness.

If one desires a small value of λ for reasons other than robustness, and the sample size is not large enough to render parameter uncertainty negligible, then uncertainty in the EWMA standard deviation is unavoidable. In this case, the sensitivity results of Eqs. (16) and (17) also provide a basis for modifying the EWMA chart so as to avoid being plagued with false alarms. Reference 5 used the sensitivity results to widen the control limits by an amount commensurate with the largest EWMA standard deviation within a suitable confidence interval. By taking into account the fact that $\Sigma_{\hat{\gamma}}(\hat{\boldsymbol{\gamma}})$ is inversely proportional to the sample size, they also developed recommendations for minimum sample sizes when estimating the ARMA parameters for use in control charting.

To develop a confidence interval for σ_z for the residual-based EWMA, [5] used a different approach than what we have used to develop (10) and (12). For the residual-based EWMA, $H(B) = \lambda[1-(1-\lambda)B]^{-1}\hat{\Theta}^{-1}(B)\hat{\Phi}(B)$ is an explicit function of $\hat{\Theta}(B)$ and $\hat{\Phi}(B)$. Thus, σ_z^2 is an explicit function of $\hat{\Theta}(B)$ and $\hat{\Phi}(B)$ via Eq. (4). Noting this, [5] calculated the partial derivative of σ_z^2 with respect to $\hat{\boldsymbol{\gamma}}$ (as opposed to $\boldsymbol{\gamma}$) and used this to develop their confidence interval for σ_z. In spite of the different paradigms, the confidence interval of [5] was identical to (10) because of the symmetry (with respect to the true and estimated parameters) of $G(B) = \lambda[1-(1-\lambda)B]^{-1}\hat{\Theta}^{-1}(B)\hat{\Phi}(B)\Phi^{-1}(B)\Theta(B)$ for the residual-based EWMA.

6. APPLICATION TO FEEDBACK CONTROL OF RUN-TO-RUN PROCESSES WITH ARIMA DISTURBANCES

Run-to-run control of batch production processes is used extensively in the process industries such as semiconductor fabrication [19, 20]. A typical feedback control objective is disturbance rejection, in which one actively adjusts a controllable process input variable (denoted by u_t) so as to minimize the effects of a process disturbance variable (denoted by d_t) on some process output variable (denoted by y_t) that is of particular interest. Disturbances can be deterministic step or ramp functions, and/or random disturbances represented by ARIMA models. Run-to-run process are typically modeled as having a static input/output relationship of the form [4, 19]

$$y_t = \beta_0 + \beta_1 u_{t-1} + d_t. \quad (20)$$

Here, u_{t-1} represents the input variable setting that is determined at the end of batch number $t-1$ and applied throughout batch number t. The process (20), together with a feedback control law, is referred to as the closed-loop system and is illustrated in Fig. 5. The constant T denotes a fixed set point, or target value, for the output variable.

The most common type of feedback controller for run-to-run processes is the proportional-integral (PI) controller, of which the so-called EWMA controller of [25] is a special case [19]. The PI control law is of the form

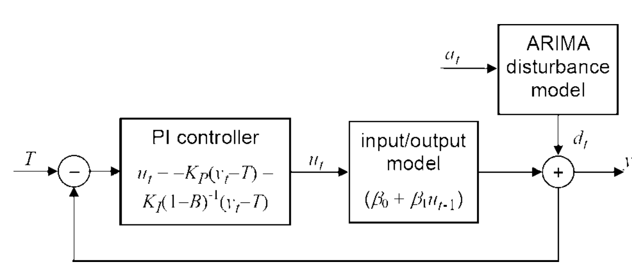

Figure 5. Block diagram of a closed-loop system consisting of a run-to-run process and a PI feedback control law.

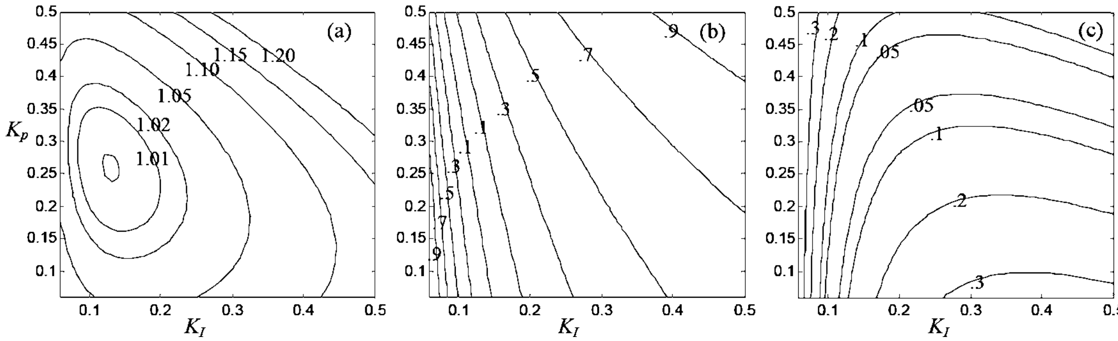

Figure 6. Contour plots of the closed-loop variance σ_z^2 (a), the magnitude $|S_{\theta,1}|$ of the sensitivity with respect to θ_1 (b), and the magnitude $|S_{\theta,2}|$ of the sensitivity with respect to θ_2 (c) as functions of K_p and K_I for the feedback control example.

$$u_t = -K_\mathrm{P}(y_t - T) - K_\mathrm{I}(1 - B)^{-1}(y_t - T), \qquad (21)$$

where K_P and K_I are the proportional and integral constants, respectively, and are controller design parameters that an user must select. Note that the $(1 - B)^{-1}$ operator represents an integrator that sums the errors $y_t - T$ over all past t.

Let $z_t = y_t - T$ denote the output deviation from target, and suppose the disturbance is modeled as the ARIMA process $d_t = (1 - B)^{-1}x_t$, where x_t follows the ARMA model (1). Combining this with Eqs. (20) and (21), one can show that in the steady-state the closed-loop system follows the model

$$z_t = \frac{1}{1 + (k_P\beta_1 + k_I\beta_1 - 1)B - k_P\beta_1 B^2} x_t. \qquad (22)$$

The variance of z_t is of particular interest in run-to-run feedback-controlled processes because it represents the output variance about the desired target value. Eq. (22) implies that we can use the results in Sections 2 and 3 to study the sensitivity of the closed-loop output variance with respect to deviations in the parameters of the disturbance model $d_t = (1 - B)^{-1}\Phi^{-1}(B)\Theta(B)a_t$. In this case, ρ_k represents the autocorrelation function of the closed-loop output.

6.1. Example

We illustrate the use of the sensitivity results for feedback control of run-to-run processes with an example from [33], originally considered in [14]. The output variable is the fill-weight of a powdered food product, and the adjustable input variable is a valve that directly determines the volume of powder dispensed per unit time. The input was transformed so that $\beta_0 = 0$, and $\beta_1 = 1$, and the disturbance was modeled as the IMA(1,2) process $d_t = (1 - B)^{-1}(1 - 0.61B - 0.26B^2)a_t$. Hence, $\Phi(B) = 1$, and $\Theta(B) = (1 - 0.61B - 0.26B^2)$.

Eq. (8) implies that the lowest sensitivity is achieved when the autocorrelation of z_t is zero. This coincides with the so-called minimum variance control [8, 13] in which one chooses the control law that results in $z_t = a_t$. From Eq. (22), this is achieved if we choose $K_\mathrm{P}\beta_1 + K_\mathrm{I}\beta_1 - 1 = -0.61$ and $K_\mathrm{P}\beta_1 = 0.26$, i.e., $K_\mathrm{P} = 0.26$ and $K_\mathrm{I} = 0.13$. There are many other considerations that may preclude the use of these PI controller parameters, such as the ability to react quickly to step disturbances or a necessity to keep the control adjustments small. In this case, achieving low output variance and low sensitivity to the disturbance parameters must be balanced with the other considerations, for which one could construct plots such as those in Fig. 6 to serve as guidance. Figure 6 shows the closed-loop output variance (σ_z^2) and the absolute values of the sensitivities ($|S_{\theta,1}|$ and $|S_{\theta,2}|$), calculated via Eqs. (4) and (8), respectively. Figures 6(b) and 6(c) indicate that overcompensation relative to minimum variance control (i.e., K_P and K_I larger than the minimum variance values) is better than undercompensation from a robustness perspective. In particular, using a value for K_I smaller than 0.13 results in much higher sensitivity to the disturbance parameters θ_1 and θ_2.

7. CONCLUSIONS

In this article, we have considered a rather general problem in which one applies a linear filter to an ARMA time series and is interested in the variance of the filter output. Filters are typically designed based on an estimated ARMA model, and deviations in the true parameters from their assumed values will cause the filter output variance to differ from its assumed value. To quantify this, we have derived analytical expressions for the sensitivity of the output variance with respect to deviations in the ARMA model parameters. These were used to develop an approximate confidence interval or Bayesian credible region for the output variance. The sensitivity expressions also have a relatively simple form that provides intuition into when one can expect the variance to be strongly affected by parameter errors. The sensitivities to the AR parameters were shown to depend on only the nominal AR polynomial and autocorrelation function of the filter output. Likewise, the sensitivities to the MA parameters depend on only the nominal MA polynomial and autocorrelation function of the filter output. We have used three examples—EWMA forecasting, SPC with autocorrelated process data, and feedback control of run-to-run manufacturing processes—to illustrate how the sensitivity results can aid in understanding the impact of ARMA parameter estimation errors and designing more robust statistical filtering procedures.

APPENDIX

In this appendix, we derive Eqs. (7) and (8), the expressions for the sensitivities. Differentiating (4) gives

$$\frac{\partial \sigma_z^2}{\partial \phi_i} = 2\sigma_a^2 \sum_{j=0}^{\infty} g_j d_j^{\phi_i}, \text{ and} \tag{A1}$$

$$\frac{\partial \sigma_z^2}{\partial \theta_i} = 2\sigma_a^2 \sum_{j=0}^{\infty} g_j d_j^{\theta_i}. \tag{A2}$$

where $d_j^{\phi_i} = \partial g_j/\partial \phi_i$ and $d_j^{\theta_i} = \partial g_j/\partial \theta_i$. From the relationship $G(B) = \Phi^{-1}(B)\Theta(B)H(B)$, we have $g_j - \phi_1 g_{j-1} - \phi_2 g_{j-2} - \cdots - \phi_p g_{j-p} = h_j - \theta_1 h_{j-1} - \theta_2 h_{j-2} - \cdots - \theta_q h_{j-q}$. Differentiating both sides with respect to ϕ_i and θ_i gives

$$d_j^{\phi_i} - \phi_1 d_{j-1}^{\phi_i} - \cdots - \phi_p d_{j-p}^{\phi_i} - g_{j-i} = 0, \text{ and} \tag{A3}$$

$$d_j^{\theta_i} - \phi_1 d_{j-1}^{\theta_i} - \cdots - \phi_p d_{j-p}^{\theta_i} = -h_{j-i}, \tag{A4}$$

where it is understood that $g_j = h_j = d_j^{\phi_i} = d_j^{\theta_i} = 0$ for $j < 0$, and we have used the fact that $H(B)$ does not depend on the true ARMA parameters. If we view $d_j^{\phi_i}$ and $d_j^{\theta_i}$ as sequences in the index j (so that $Bd_j^{\phi_i} \equiv d_{j-1}^{\phi_i}$), we can write (A3) as

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

which in turn we can write as

$$d_j^{\phi_i} = \Phi^{-1}(B) g_{j-i} = \sum_{k=0}^{\infty} P_k g_{j-i-k}.$$

Proceeding similarly from (A4) gives

$$d_j^{\theta_i} = -\Phi^{-1}(B) h_{j-i} = -\Phi^{-1}(B)\Theta^{-1}(B)\Phi(B) g_{j-i} = -\Theta^{-1}(B) g_{j-i} = -\sum_{k=0}^{\infty} Q_k g_{j-i-k}.$$

Combining these with (4), (5), (6), (A1), and (A2), the sensitivity measures become

$$S_{\phi,i} = \frac{2\sigma_a^2 \sum_{j=0}^{\infty} g_j \sum_{k=0}^{\infty} P_k g_{j-i-k}}{\sigma_a^2 \sum_{j=0}^{\infty} g_j^2} = 2\sum_{k=0}^{\infty} P_k \frac{\sigma_a^2 \sum_{j=0}^{\infty} g_j g_{j-(i+k)}}{\sigma_a^2 \sum_{j=0}^{\infty} g_j^2}, \text{ and}$$

$$S_{\theta,i} = \frac{-2\sigma_a^2 \sum_{j=0}^{\infty} g_j \sum_{k=0}^{\infty} Q_k g_{j-i-k}}{\sigma_a^2 \sum_{j=0}^{\infty} g_j^2} = -2\sum_{k=0}^{\infty} Q_k \frac{\sigma_a^2 \sum_{j=0}^{\infty} g_j g_{j-(i+k)}}{\sigma_a^2 \sum_{j=0}^{\infty} g_j^2}.$$

Recognizing that the denominator and numerator in the far right expressions are the variance and the lag-$(k+i)$ autocovariance of $z_t = \sum_{j=0}^{\infty} g_j a_{t-j}$ [12] completes the derivation of Eqs. (7) and (8).

ACKNOWLEDGMENT

This work was supported by the National Science Foundation under grants #CMMI-0354824 and #CMMI-0758557.

REFERENCES

[1] B.M. Adams and I.T. Tseng, Robustness of forecast-based monitoring schemes, J Qual Technol 30 (1998), 328–339.
[2] L.C. Alwan and H.V. Roberts, Time-series modeling for statistical process control, J Business Econ Stat 6 (1988), 87–95.
[3] C.F. Ansley and R. Kohn, Prediction mean squared error for state space models with estimated parameters, Biometrika 73 (1986), 467–473.
[4] D.W. Apley and J.B. Kim, Cautious control of industrial process variability with uncertain input and disturbance model parameters, Technometrics 46 (2004), 188–199.
[5] D.W. Apley and H.C. Lee, Design of exponentially weighted moving average control charts for autocorrelated processes with model uncertainty, Technometrics 45 (2003), 187–198.
[6] D.W. Apley and H.C. Lee, Robustness comparison of exponentially weighted moving average charts on autocorrelated data and on residuals, J Qual Technol 40 (2008), 428–447.
[7] D.W. Apley and J. Shi, The GLRT for statistical process control of autocorrelated processes, IIE Trans 31 (1999), 1123–1134.
[8] K.J. Åström and B. Wittenmark, Computer controlled systems: Theory and design, 2nd ed., Prentice Hall, Englewood Cliffs, NJ, 1990.
[9] J.M. Bernardo and A.F.M. Smith, Bayesian theory, Wiley, New York, 2002.
[10] P.M. Berthouex, W.G. Hunter, and L. Pallesen, Monitoring sewage treatment plants: Some quality control aspects, J Qual Technol 10 (1978), 139–149.
[11] G.E.P. Box and G. Jenkins, Time series analysis, forecasting, and control, Holden-Day, San Francisco, 1990.
[12] G.E.P. Box, G. Jenkins, and G. Reinsel, Time series analysis, forecasting, and control, 3rd ed., Prentice-Hall, Englewood Cliffs, NJ, 1994.
[13] G.E.P. Box and A. Luceño, Statistical control by monitoring and feedback adjustment, Wiley, New York, 1997.
[14] R.A. Boyles, Phase I analysis of autocorrelated processes, J Qual Technol 32 (2000), 395–409.
[15] C. Chatfield, Calculating interval forecasts, J Business Econ Stat 11 (1993), 121–135.
[16] C.H. Chin and D.W. Apley, Optimal design of second-order linear filters for control charting, Technometrics 48 (2006), 337–348.
[17] D.N. DeJong, B.F. Ingram, and C.H. Whiteman, A bayesian approach to dynamic macroeconomics, J Econ 98 (2000), 203–223.
[18] X. de Luna, Prediction intervals based on autoregression forecasts, J R Stat Soc Ser D (Stat) 49 (2000), 87–93.
[19] E. del Castillo, Statistical process adjustment methods for quality control, Wiley, New York, 2002.
[20] E. del Castillo and A.M. Hurwitz, Run-to-Run process control: Literature review and extensions, J Qual Technol 29 (1997), 184–196.
[21] W.A. Fuller and D.P. Hasza, Properties of predictors for autoregressive time series, J Amer Stat Assoc 76 (1981), 155–161.

[22] D. Han and F. Tsung, A generalized EWMA control chart and its comparison with the optimal EWMA, CUSUM and GLR schemes, Ann Stat 32 (2004), 316–339.
[23] D. Han and F. Tsung, A reference-free cuscore chart for dynamic mean change detection and a unified framework for charting performance comparison, J Am Stat Assoc 101 (2006), 368–386.
[24] B.E. Hansen, Interval forecasts and parameter uncertainty, J Econ 135 (2006), 377–398.
[25] A. Ingolfsson and E. Sachs, Stability and sensitivity of an EWMA controller, J Qual Technol 25 (1993), 271–287.
[26] W.A. Jensen, L.A. Jones-Farmer, C.W. Champ, and W.H. Woodall, Effects of parameter estimation on control chart properties: A literature review, J Qual Technol 38 (2006), 349–364.
[27] P. Kabaila and Z. He, The adjustment of prediction intervals to account for errors in parameter estimation, J Time Ser Anal 25 (2004), 351–358.
[28] C.W. Lu and M.R. Reynolds, EWMA control charts for monitoring the mean of autocorrelated processes, J Qual Technol 31 (1999), 166–188.
[29] J.M. Lucas and M.S. Saccucci, Exponentially weighted moving average control schemes: Properties and enhancements, Technometrics 32 (1990), 1–12.
[30] D.C. Montgomery, Introduction to statistical quality control, 5th ed., Wiley, New York, NY., 2005.
[31] D.C. Montgomery and C.M. Mastrangelo, Some statistical process control methods for autocorrelated data, J Qual Technol 23 (1991), 179–193.
[32] D.C. Montgomery and W.H. Woodall, A discussion on statistically-based process monitoring and control, J Qual Technol 29 (1997), 121–162.
[33] R. Pan and E. Del Castillo, Identification and fine-tuning of closed-loop processes under discrete EWMA and PI adjustments, Qual Reliability Eng Int 17 (2001), 419–427.
[34] P.C.B. Phillips, The sampling distribution of forecasts from a first-order autoregression, J Econ 9 (1979), 241–261.
[35] G.C. Runger, T.R. Willemain, and S. Prabhu, Average run lengths for cusum control charts applied to residuals, Commun Stat: Theory Methods 24 (1995), 273–282.
[36] S.W. Roberts, Control chart tests based on geometric moving averages, Technometrics 1 (1959), 239–250.
[37] A. Rodriguez and E. Ruiz, Bootstrap prediction intervals in state-space models, J Time Ser Analysis 30 (2009), 167–178.
[38] M. Sampson, The effect of parameter uncertainty on forecast variances and confidence intervals for unit root and trend stationary time-series models, J Appl Econ 6 (1991), 67–76.
[39] W. Schmid, "On EWMA charts for time series," In: Frontiers of statistical quality control, Vol. 5, H.J. Lenz and P.-Th. Wilrich (Editors), Physica-Verlag, Heidelberg, 1997, pp. 115–137.
[40] W. Schmid and A. Schöne, Some properties of the EWMA control chart in the presence of autocorrelation, Ann Stat 25 (1997), 1277–1283.
[41] D.S. Stoffer and K.D. Wall, Bootstrapping state-space models: Gaussian maximum likelihood estimation and the kalman filter, J Am Stat Assoc 86 (1991), 1024–1033.
[42] Z.G. Stoumbos, M.R. Reynolds Jr., T.P. Ryan, and W.H. Woodall, The state of statistical process control as we proceed into the 21st century, J Am Stat Assoc 95 (2000), 992–998.
[43] C.R. Superville and B.M. Adams, An evaluation of forecast-based quality control schemes, Commun Stat Simul Comput 23 (1994), 645–661.
[44] L.A. Thombs and W.R. Schucany, Bootstrap prediction intervals for autoregression, J Am Stat Assoc 85 (1990), 486–492.
[45] K.D. Wall and D. Stoffer, A state space approach to bootstrapping conditional forecasts in ARMA models, J Time Ser Anal 23 (2002), 733–751.
[46] D.G. Wardell, H. Moskowitz, and R.D. Plante, Run-length distributions of special-cause control charts for correlated processes, Technometrics 36 (1994), 3–17.
[47] W.H. Woodall and D.C. Montgomery, Research issues and ideas in statistical process control, J Qual Technol 31 (1999), 376–386.
[48] T. Yamamoto, Asymptotic mean square prediction error for an autoregressive model with estimated coefficients, J Royal Stat Soc. Ser C (Appl Stat) 25 (1976), 123–127.
[49] T. Yamamoto, Predictions of multivariate autoregressive-moving average models, Biometrika 68 (1981), 485–492.
[50] N.F. Zhang, A statistical control chart for stationary process data, Technometrics 40 (1998), 24–38.