Monitoring Processes that Wander Using Integrated Moving Average Models

S. A. Vander Wiel

AT&T Bell Laboratories
Murray Hill, NJ 07974
15 April 1995

ABSTRACT

Often the least appropriate assumption in traditional control charting technology is that process data constitute a random sample. In reality most process data are correlated—either temporally, spatially, or due to nested sources of variation.

One approach to monitoring temporally correlated data uses a control chart on the forecast errors from a time series model of the process with, possibly, a transfer function term to model compensatory adjustments. If the time series term is an integrated moving average, then a sudden level shift in the process results in a patterned shift in the mean of forecast errors. Initially the mean shifts by the same amount as the process level but then it decays geometrically back to zero corresponding to the ability of the forecast to "recover" from the upset. We study 4 monitoring schemes: cumulative sums (CUSUMs), exponentially weighted moving averages, Shewhart individuals charts, and a likelihood ratio scheme. Comparisons of signaling probabilities and average run lengths show that CUSUMs can be designed to perform at least as good as, and often better than any of the other schemes. Shewhart individuals charts often perform much worse than the others. Graphical aids are provided for designing CUSUMs in this context.

KEYWORDS: autocorrelation, control charts, cumulative sum, exponentially weighted moving average, likelihood ratio

Monitoring Processes that Wander Using Integrated Moving Average Models

S. A. Vander Wiel

AT&T Bell Laboratories
Murray Hill, NJ 07974
15 April 1995

1 Monitoring Correlated Data

Too often a stream of manufacturing data that wanders about as in Figure 1 is subjected to a monitoring scheme that expects observations to behave like independent and identically distributed (iid) random variables. The "3-standard deviation control limits drawn on the figure are calculated by estimating the process variance using successive differences. Of course, this local measure of variance is smaller than the total variance because it does not include variability due to the meandering level and, thus, the control chart shows a lack of "statistical control." All standard control charts for continuous measurements are based on local measures of variability precisely so they will signal when presented with non-iid data as in Figure 1.

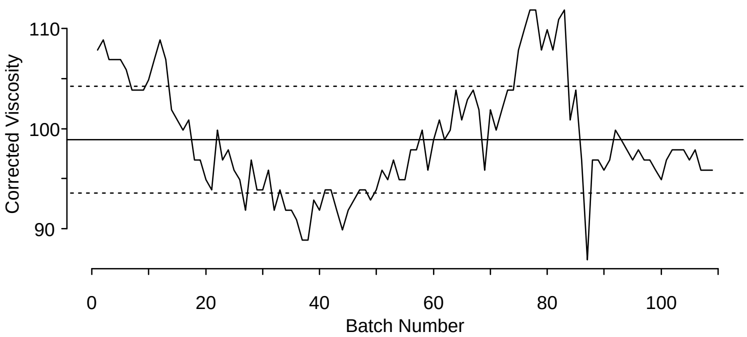

Figure 1: Viscosity from successive batches of polymer resin corrected for adjustments of the catalyst amount. The downward shift around batch 84 may have been preventable and it should have been detected by a monitoring scheme. A standard control chart with widened control limits, however, might not have detected the shift.

Engineers needing to monitor this kind of data are usually well aware of its wandering nature and they do not want a monitoring scheme to continually tell them what they already know. A common way to "fix" the problem is to widen the control limits until a standard control chart rarely signals. This clearly reduces the number of uninformative alarms, but it could also make the chart useless for the purpose of signaling unusual behavior that could lead to process improvement (Wardell, Moskowitz and Plante, 1992).

The data in Figure 1 represent consecutive batches of polymer from a process studied by Vander Wiel, Tucker, Faltin and Doganaksoy (1992). The plot shows viscosity measurements corrected for the effect of catalyst changes which are routinely made in this process. In other words, the plotted viscosities are (approximately) what would have resulted if the level of catalyst had remained fixed.

Clearly, the unadjusted process would wander and for this reason the amount of catalyst used in a given batch was determined on the basis of previous viscosity deviations from target through a combination of rules of thumb and operator judgement. Figure 2 shows actual viscosity measurements (top panel) and the corresponding catalyst adjustments (lower panel.) By comparing with Figure 1 it is obvious that the adjustment scheme transfers some (but not all) of the wandering from the viscosity measurements to the catalyst level, thus reducing variability in viscosity. "3-standard deviation control limits on the viscosity plot are again based on successive differences. The unusual behavior in period 84 stands out more than in Figure 1 but the chart continues to signal frequently because viscosity measurements still tend to wander.

Vander Wiel et al. (1992) developed a better catalyst adjustment scheme for this process to reduce viscosity

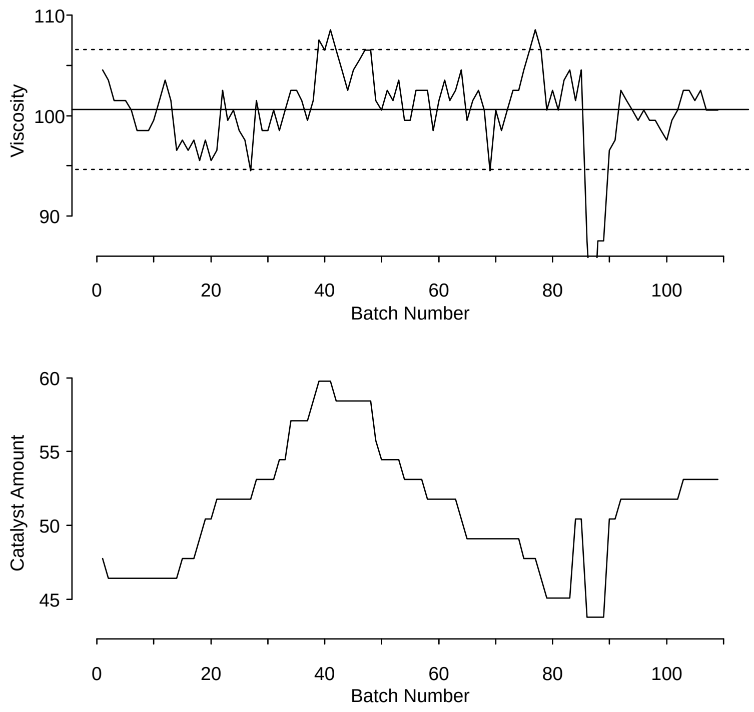

Figure 2: Actual viscosity measurements are shown in the top panel with the catalyst adjustments used to produce them in the lower panel.

variations below that shown in Figure 2. *Removing* sources of variation, however, rather than just *transferring* them to adjustment variables, is considered vital to the continued viability of this product. Reducing the process standard deviation by even 5% could save millions of dollars per year in waste and extra processing costs. Thus, it is important to monitor viscosity and to signal unexpected abrupt changes that might have "assignable causes." For example, the downward shift beginning with batch 84 could be the result of changing the feed stock from one silo to another or of a sudden drop in the ambient temperature at the plant. If the cause were known, it could possibly be removed or at least preemptive compensations could be made. Upsets would then occur less frequently.

The viscosity application is representative of a general problem. Processes that tend to wander are often subject to adjustments—either automatic or manual. These adjustments usually reduce both the process variability and autocorrelation. A disturbance, however, can knock the process off target until corrective feedback accommodates it. Detecting such events is the first step to understanding why they occur and preventing them in the future.

Although a process under feedback control will not usually be allowed to wander from target, any serial correlation not removed by the adjustment rule can greatly influence the false alarm rate of a standard control chart. Monitoring performance, however, can be isolated from the adjustment rule by applying control charts to forecast errors from a model of the entire input-output system. Such a model can include a deterministic term to describe the effect of control actions and a time series noise term to describe the underlying autocorrelated disturbance. Unusual events then manifest themselves in the forecast errors which are nominally iid. Linear transfer function models with ARIMA noise terms (Box and Jenkins, 1976) are a rich class of models suitable for this purpose. In cases where the unadjusted process tends to wander, first order integrated moving average (IMA) noise terms are often appropriate.

The purpose of this work is to compare the performance of several monitoring schemes applied to forecast errors from IMA processes. CUSUM charts of forecast errors are shown to provide good signaling performance in response to abrupt shifts in the process level. A second purpose is, therefore, to provide general advice and some graphical aids and for designing CUSUM charts in this context.

1.1 Monitoring viscosity forecast errors

To preview the results of this paper let us reconsider the polymer example. Vander Wiel et al. (1992) modeled viscosity measurements using a simple linear function of catalyst with an ARMA(1,1) noise term. Their estimated model for the noise term (based on much more data than shown in Figure 1) had a correlation structure similar to that of a first order IMA with parameter $\lambda = 0.8$ (see Section 2). In fact, the IMA model fits the Figure 1 data slightly better and we will adopt it for the remainder of the example. The top panel of Figure 3 shows one-step ahead forecast errors. Using a 2-batch moving range we estimate the error standard deviation as $\hat{\sigma} = 2.58$. The sharp drop in viscosities beginning with batch 84 appears as two large negative forecast errors that escape the $3\hat{\sigma}$ control limits. A large positive forecast error in period 87 reflects a sudden upward shift. The bottom panel of Figure 3 is a two-sided CUSUM chart [in Page's (1954) form ; see subsection 3.1] of the forecast errors. Both charts provide good reason to search for a "special cause" beginning with period 84.

In Figure 3 the CUSUM reference parameter $k = 1 \times \hat{\sigma}$ and the control limits $h = \pm 2.3\hat{\sigma}$ were selected using Figures 5 and 6 to give an ARL of 250 for iid forecast errors and to provide the best possible ARL for detecting a rather large (specifically, $4\hat{\sigma} = 10.3$) shift in the viscosity level. We will see (Figure 6), however, that even for a sustained 4σ shift the ARL is 20 which may seem large. The large ARL, however, is not due to a deficiency in the CUSUM chart. Rather, it reflects the limited information available in the data for detecting shifts. More insight into this is given in Section 2 which defines an IMA and shows how a shift in level affects forecast errors.

Section 3 discusses the design of CUSUM charts for IMA forecast errors. Making comparisons under 2 different performance criteria, the section also shows that CUSUMs dominate Shewhart individuals charts, a likelihood ratio scheme, and even exponentially weighted moving averages (EWMAs).

1.2 Relevant literature

Performance properties such as ARLs have been widely tabulated for CUSUM charts and EWMA charts applied to iid Gaussian sequences. See, for example, Goel and Wu (1971), Lucas and Crosier (1982), Crowder (1987), and Lucas and Saccucci (1990). Run lengths from Shewhart individuals and $\bar{X}$ charts on iid data are geometrically distributed and thus easy to analyze. Much less guidance, however, is available for choosing and designing

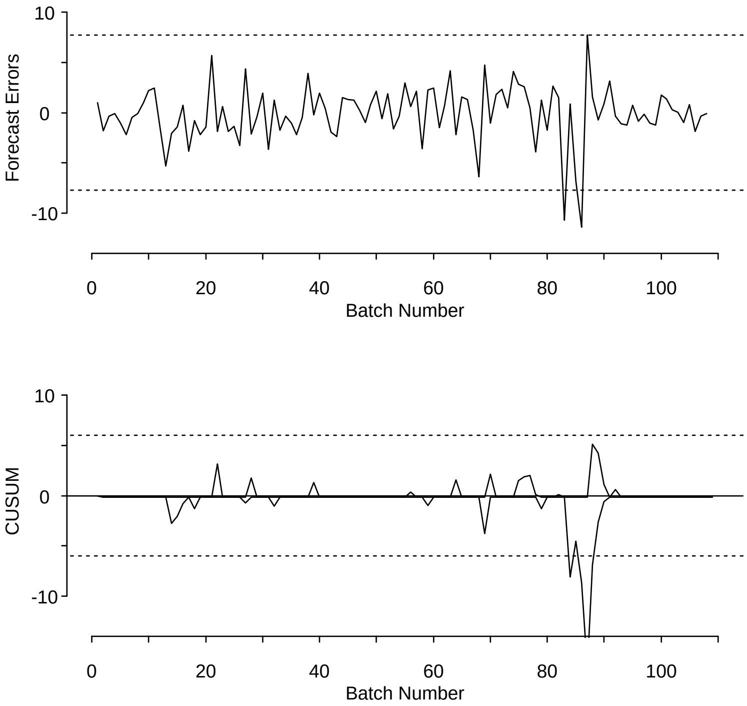

Figure 3: The top panel shows 1-step ahead forecast errors on a Shewhart individuals chart with $3\hat{\sigma}$ control limits. The lower panel is a CUSUM chart of the forecast errors. Both charts show strong evidence of an unusual event beginning in period 84.

monitoring schemes appropriate for autocorrelated data. Tracking signals have been used to monitor performance of forecasting systems for more than 30 years. Brown's (1962) tracking signal is the cumulative sum of forecast errors divided by an EWMA of their absolute values. Trigg (1964) replaced the cumulative sum in Brown's numerator with an EWMA. Golder and Settle (1976) simulated ARLs of these tracking signals. Gardner (1983) gave more extensive simulation results and introduced a tracking signal for detecting autocorrelation in the forecast errors—an indication that the forecasts can be improved.

The approach of monitoring forecast errors has reemerged recently in the quality improvement literature. Alwan and Roberts (1988) plot 1-step forecasts on a "common cause" chart with no control limits and plot forecast errors on a "special cause" chart with "3-standard deviation control limits. MacGregor (1988) outlines the essential concepts of process monitoring using control charts and process adjustment (control) using dynamic input-output models with time series errors. He suggests using control charts "for analyzing control system performance and as diagnostic tools in control schemes." Vander Wiel et al. (1992) successfully implemented this approach to control and monitor the batch polymerization process introduced above. After reducing the process variability using a minimum variance adjustment algorithm they monitored forecast errors using a CUSUM chart. Others who suggest monitoring forecast errors are Montgomery and Friedman (1989), Harris and Ross (1991), Montgomery and Mastrangelo (1991), and Box and Kramer (1992). Longnecker and Ryan (1992) study performance of Shewhart individuals charts on residuals from ARMA(1,1) and AR(2) processes. Superville and Adams (1994) compare individuals, CUSUM, and EWMA charts of forecast errors for AR(1) models and argue against using ARLs to select control charts for monitoring forecast errors. Instead they suggest using the probability of signaling by a fixed number of periods beyond the change point. The recommendation is based on the fact that forecasts "recover" from abrupt changes and thus leave only a limited "window of opportunity" for detection. Runger (1995) also gives ARLs for CUSUM charts of AR(1) forecast errors. Lu and Reynolds (1994) investigate EWMA and CUSUM charts applied to ARMA(1,1) processes and to their forecast errors.

When a process is stationary and autocorrelation is mild there may be no advantage to using forecast errors. For Gaussian AR(1) processes with a lag-1 correlation of no more than 0.5 (Yashchin, 1993) showed that ARL performance is virtually identical whether CUSUM-ing raw process measurements or their forecast errors. He also gave a means to determine how wide control limits should be to achieve a specified false alarm rate when CUSUM-ing autocorrelated data. Wardell, Moskowitz and Plante (1994) compared ARL performance of Shewhart individuals charts applied to forecast errors and to raw data from ARMA(1,1) processes. In many cases when the lag-1 correlation was positive, the raw data scheme performed better.

Finally, several early papers studied the affects of autocorrelation on various monitoring schemes. Goldsmith and Whitfield (1961) showed that negative autocorrelation can decrease false alarm rates for CUSUM charts. Conversely, positive autocorrelation increases rates. Additional studies have been reported by Johnson and Bagshaw (1974), Bagshaw and Johnson (1975), and Vasilopoulos and Stamboulis (1978).

2 Integrated moving averages and level shifts

Industrial data that would wander if no compensating actions were taken can often be modeled using a first order IMA noise term in a model with a deterministic component to describe the effects of adjustments. A linear regression of viscosity on catalyst amount with an IMA noise term provides a reasonable fit to the viscosity data in Figure 2. Box and Kramer (1992) and MacGregor (1988) place special importance on the IMA noise model because it sensibly fits data from a wide variety of industrial and economic processes. IMAs are often used to model stochastic disturbances in automatic control applications because the popular proportional-integral (PI) controller is optimal for first order input-output systems with IMA disturbances. A huge number of successful feedback loops under PI control in a wide range of applications is evidence that IMA approximations to correlated disturbances are useful (MacGregor, 1988).

Box and Kramer (1992) argue the appropriateness of IMAs based on the fact that the variance of lag-k differences increases linearly with k, even for large k. Often, however, only the short-lag autocorrelations have practical significance for forecasting, monitoring, and control. Thus, choosing between a non-stationary IMA model or a stationary ARMA model with a similar correlation structure is not too important.

We will define a process N_t $(t = 0, 1, 2, \ldots)$ as a first order IMA if

$$N_t = \alpha_t + \lambda \sum_{i=0}^{t-1} \alpha_i$$

where the α's are iid $N(0, \sigma^2)$ variates and $\lambda \in [0, 1]$. An IMA with $\lambda \neq 0$ is non-stationary with variance $\sigma^2(1 + \lambda^2 t)$ increasing linearly in t. The increments

$$N_t - N_{t-1} = \alpha_t - (1 - \lambda)\alpha_{t-1}$$

form a first order moving average. Special cases of the IMA family arise when $\lambda = 0$ giving an iid process and when $\lambda = 1$ giving a random walk. For $\lambda \in (0, 1)$ the IMA is equivalent to a random walk observed with iid measurement error (Box and Jenkins, 1976, Chapter 4).

The left column of plots in Figure 4 shows 250 observations from simulated IMAs with $\sigma = 1$ and λ varying from 0 to 1 in successive rows. Each panel uses the same set of αs. Notice how the IMAs wander more for larger values of λ. The remainder of Figure 4 is discussed below.

An estimate $\hat{\lambda}$ of λ can be used to form 1-step ahead forecasts of the IMA. The usual forecast of N_{t+1} based on $(N_0, \ldots, N_t)$ is

$$\hat{N}_{t+1|t} = \hat{\lambda} N_t + (1 - \hat{\lambda})\hat{N}_{t|t-1} \tag{1}$$

where the recursion is started from $\hat{N}_{0|-1} = 0$. $\hat{N}_{t+1|t}$ is an exponentially weighted moving average (EWMA) of the observed values of the IMA from N_0 through N_t. If $\hat{\lambda} = \lambda$, $\hat{N}_{t+1|t}$ is a minimum mean square error forecast of N_{t+1}. The forecast error in period t is

$$a_t \equiv N_t - \hat{N}_{t|t-1}. \tag{2}$$

This definition gives $a_t = \alpha_t - (\hat{\lambda} - \lambda) \sum_{i=1}^{t} (1 - \hat{\lambda})^{i-1} \alpha_{t-i}$ (for $t \geq 1$) so $a_t \approx \alpha_t$ if $\hat{\lambda} \approx \lambda$. The maximum likelihood estimator $\hat{\lambda}$ minimizes $\sum {a_t}^2$.

For processes under feedback control, N_t represents the actual process measurement minus an amount to account for the effects of previous control actions. Formulas for computing forecasts for transfer function models are simple linear recursions in the same spirit as (1). See Box and Jenkins (1976) for details. With this understanding of N_t, the development and results that follow apply to transfer function models with IMA noise terms as well as pure IMA models.

An important standard for comparing control charts has been how quickly they detect a sudden sustained shift in the process level. This is measured by the probabilistic characteristics of run lengths where a run length is the number of periods between a step shift and the first signal of the control chart. In particular, the average run length (ARL) has been emphasized. It is important to detect step shifts in processes that wander. If a sudden shift can be detected, it might be possible to remove the cause, eliminate a source of variability, and improve the process.

But a step change is more difficult to see when buried in an IMA than when buried in iid noise. The middle column of plots in Figure 4 shows the same IMAs as in the left column but with a step shift of 5σ beginning in period 150. The shift stands out in the iid ($\lambda = 0$) sequence but is less obvious as λ increases. For example, the random walk ($\lambda = 1$) drops abruptly around period 45 by an amount similar to the sharp increase at period 150. The first change, however is due to several consecutive negative α's while the second is a real shift in the level (or equivalently one huge α.)

One way to understand why shifts are easier to see for smaller λ is to consider how evidence of a shift builds as data accumulates. At period 151, each plot in the center column steps upward by about 5σ signifying a probable shift. In the iid plot all doubt about whether a shift really occurred is gone by, say, period 160, because each of the last 10 observations is about 5σ higher than the first 150. In the random walk plot, however, all the information about a shift comes in period 150 and new data contributes no new information because the random walk simply takes iid steps from its most recent position. For $\lambda = 0.5$ some evidence of a shift accumulates in the periods just after 150, but by, say, period 175 the process has wandered enough that new observations are not relevant to what happened at period 150. For middling values of λ it is tricky to mentally judge how much of what we see is due to the 5σ shift and how much is due to autocorrelation. Looking at the affect of a shift on the forecast errors a_t helps remove the ambiguity.

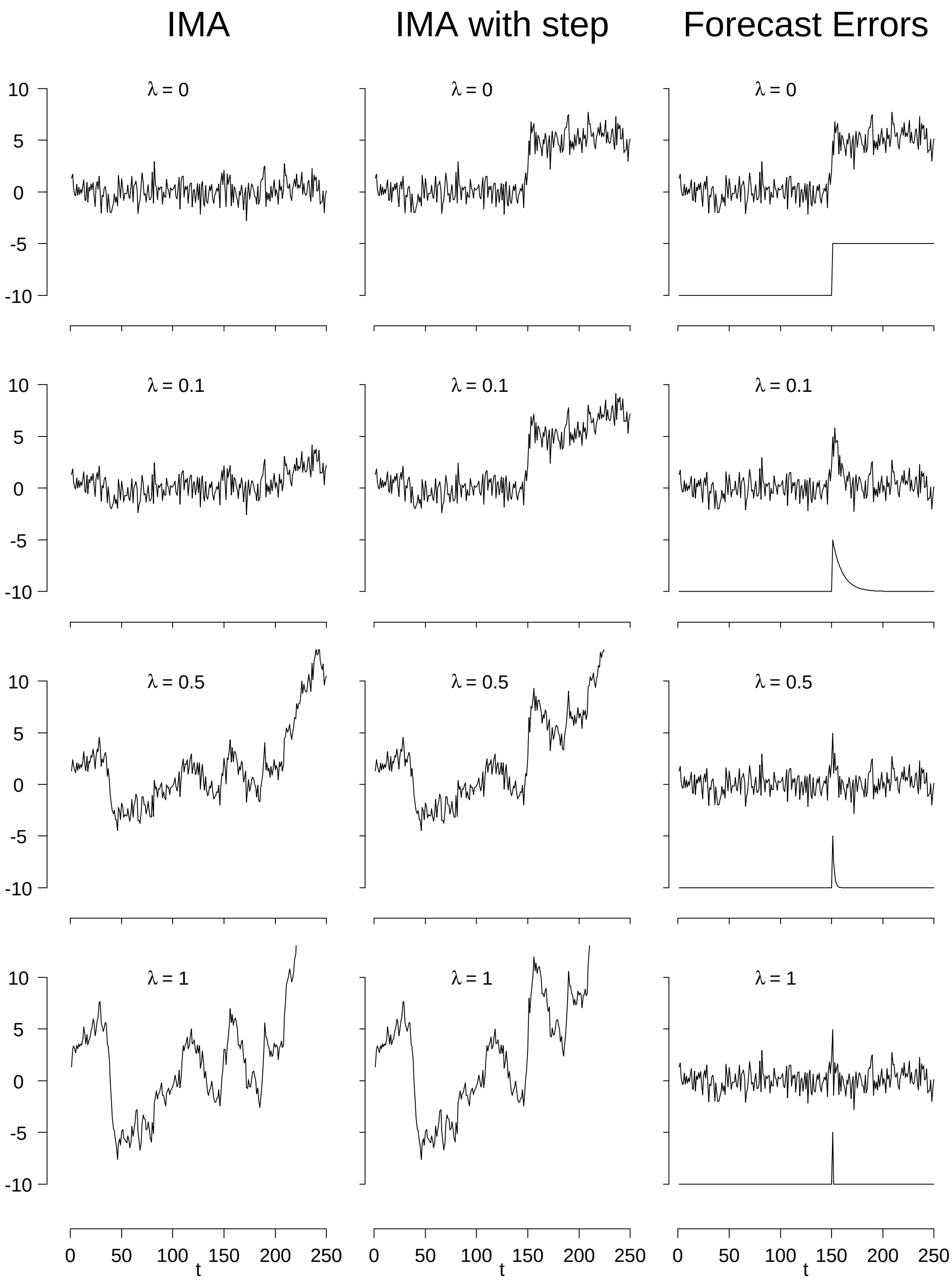

Figure 4: Simulated IMAs (left column), IMAs with a 5σ shift at period 150 (center column), and forecast errors (right column). IMAs (left) wander more as λ increases. The wandering makes step shifts in the level (center) more difficult to see as λ increases. Step shifts in the IMAs cause a patterned change in the forecast errors (right). The mean functions of the forecast errors are plotted in the bottom portion of these plots (shifted away from zero for clarity). It jumps by an amount equal to the size of the shift in the IMA and then decays back to zero at a geometric rate of $1 - \lambda$.

Suppose a process Z_t behaves like an IMA but experiences a step shift of size μ in its level at some period k. That is,

$$\begin{aligned} Z_t &= N_t, & t &= 1, \ldots, k-1 \\ &= N_t + \mu, & t &= k, k+1, \ldots \end{aligned}$$

where N_t is an IMA. If equation (2) (with $\hat{\lambda} = \lambda$ for simplicity) is applied to Z_t the forecast errors a_t will continue to be independent normal variates with variance σ^2 but their mean will not be zero in periods k and following. In fact,

$$\begin{aligned} Ea_t &= 0, & t &= 1, \ldots, k-1 \\ &= \mu(1-\lambda)^{t-k}, & t &= k, k+1, \ldots . \end{aligned} \tag{3}$$

See also Harris and Ross (1991). The mean shifts to μ in period k and returns exponentially to 0 thereafter. The return is faster for larger λ.

Evidence of a shift in level is easier to judge by searching for a geometric pattern in the sequence of forecast errors than by looking at a plot of raw data. This is because we do not have to mentally untangle the effects of autocorrelation from the effects of a possible shift. The forecast errors are not correlated and a level shift in the IMA creates a simple pattern in their mean. The pattern is seen in the third column of plots in Figure 4 which shows the forecast errors computed from the IMAs with shifts shown in the center column. The patterned mean is evident in the forecast errors and is shown (shifted downward for clarity) in the lower portion of each forecast error plot.

3 Comparisons of monitoring schemes

This section describes and compares several schemes for using forecast errors to monitor IMAs for step shifts. We study signaling performance of 4 different classes of monitoring schemes: CUSUMs, EWMAs, Shewhart individuals, and schemes that use likelihood ratio statistics. Each class is described below and some guidance is given for choosing a particular scheme from within a class. Monitoring schemes are compared based on two criteria: ARLs for shifts of various sizes, and the probability of signaling within 10 periods of the onset of a shift, which we denote by $P(10)$.

For the special cases of iid observations ($\lambda = 0$) and a random walk ($\lambda = 1$) good monitoring schemes seem obvious. In the iid case the forecast errors are identical to the process itself. A step in the level of the process is therefore a step in the mean of the forecast errors. This is the situation traditionally addressed in studies of control chart performance. The literature shows that for small and medium sized shifts (up to roughly 2.5σ), it is difficult to beat the ARL performance of properly designed CUSUM and EWMA charts. For large shifts Shewhart individuals charts perform best. Combining an EWMA or a CUSUM chart with a Shewhart individuals chart results in a control scheme with good ARL performance for both large and small shifts (Lucas, 1982).

In the case of a random walk ($\lambda = 1$) a step shift in the process results in a single forecast error with a non-zero mean. A scheme that uses more that the most recent forecast error will only weaken the evidence of a shift. The best choice for this case is an individuals chart.

If $\lambda \in (0, 1)$, the best approach to monitoring for step shifts is not obvious. Since the Shewhart individuals chart is a special case of both EWMAs and CUSUMs, it seems plausible that for each λ an EWMA chart or a CUSUM chart could be constructed to give good signaling performance. But it also seems plausible that better performance than both the EWMA and the CUSUM could be attained by a scheme (like the likelihood ratio scheme described below) which is sensitive to a specific pattern of geometric decay in the forecast errors. The comparisons in Subsection 3.2 show that the first of these hunches is true for the CUSUM. That is, for a given λ and shift size, a CUSUM chart can be designed to give good signaling performance in terms of either ARLs or $P(10)$'s. Properly designed CUSUMs are usually better than likelihood ratio schemes and sometimes outperform EWMAs. They are often much better than individuals charts.

3.1 Four classes of monitoring schemes

For each class of monitoring scheme we indicate how action limits can be selected to provide a given ARL or $P(10)$ value when no shift occurs. This determination is important for comparing several schemes because, to be fair, all schemes under consideration should have either the same ARL or the same $P(10)$ value when no shift occurs. We use ARL_0 and $P_0(10)$ to denote these so-called "in control" quantities.

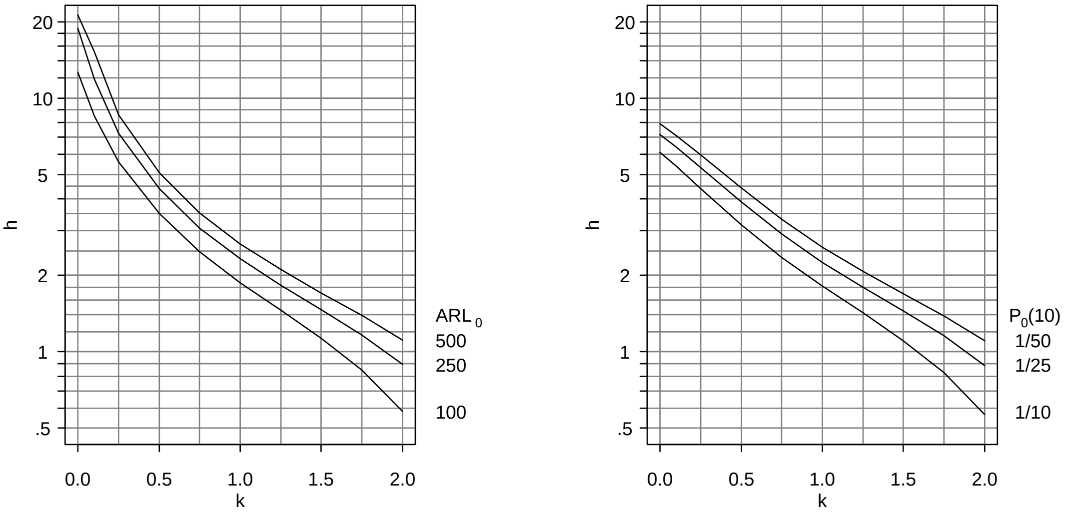

Figure 5: Plots to determine the action limit h yielding a given ARL_0 or $P_0(10)$ for various values of the CUSUM reference level k.

In the descriptions below a_t refers to forecast errors calculated using (2). Nominally, they should approximate an iid $N(0,\sigma^2)$ sequence. Since we do not consider the effects of estimating σ, we assume $\sigma = 1$. This is equivalent to assuming that the forecast errors have been scaled by dividing through by σ.

Subsection 3.3 outlines the computations we used for ARLs and signaling probabilities of the various schemes.

Shewhart individuals chart: This simply signals when $|a_t|$ exceeds an action limit h. The action limit can be set to give a desired ARL_0 or $P_0(10)$ by taking $-h$ to be respectively the $1/(2 \times \mathrm{ARL})$ quantile or the $[1 - (1 - P_0(10))^{1/10}]/2$ quantile of the standard normal distribution.

CUSUM: A (2-sided) CUSUM scheme is based on a high side statistic H_t and a low side statistic L_t:

$$\begin{aligned} H_t &= \max\{0, a_t - k + H_{t-1}\} \\ L_t &= \max\{0, -a_t - k + L_{t-1}\} \end{aligned}$$

where H_0 and L_0 are initialized at zero. H_t is sensitive to changes causing an increase in the mean of a_t while L_t is sensitive to changes causing a decrease. The scheme signals when $\max\{L_t, H_t\}$ exceeds an action limit h. The reference level k and the action limit h are design parameters. Typically k is set between 0.25σ and 1.5σ. For a given k, h can be selected to produce a desired ARL_0 or $P_0(10)$. The left panel of Figure 5 shows curves of h versus k for three values of ARL_0: 100, 250, and 500. The right panel shows h versus k for three values of $P_0(10)$: $1/10$, $1/25$, and $1/50$. Given k and either ARL_0 or $P_0(10)$, the appropriate curve can be used to find h. Setting $h = 0$ mimics a Shewhart individuals chart with "k-sigma" limits applied to the forecast errors. Gan (1991) gives more extensive versions of the ARL contours shown in Figure 5.

The choice of k gives some flexibility in ARL performance when shifts occur. Figure 6 is an aid for choosing a value of k for a particular application. The panels in Figure 6 show ARLs for various values of λ, μ (the shift size), and k with h selected (from Figure 5) to provide a given ARL_0. The panels are arranged so that columns represent values of λ from 0 to 0.5 and rows represent values of μ from 0.5 to 4. Each panel has 3 curves. The lower, middle and upper curves correspond to charts designed to have ARL_0 values of 100, 250 and 500 respectively. Each curve shows ARLs for steps of size μ for various values of k. A similar figure has been constructed based on the $P(10)$ criterion but it is not shown here because CUSUM designs using it would not be substantively different.

To choose a value for k in a particular application, look at the column of plots corresponding to λ nearest the estimated value. Now focus on the ARL curve in each panel of that column that represents a value of ARL_0 close to the one desired. Finally, visually choose a value of k that gives low ARLs for the sizes of shifts that are most important. This step may involve trading off performance for shifts of one size for better detection of shifts of another.

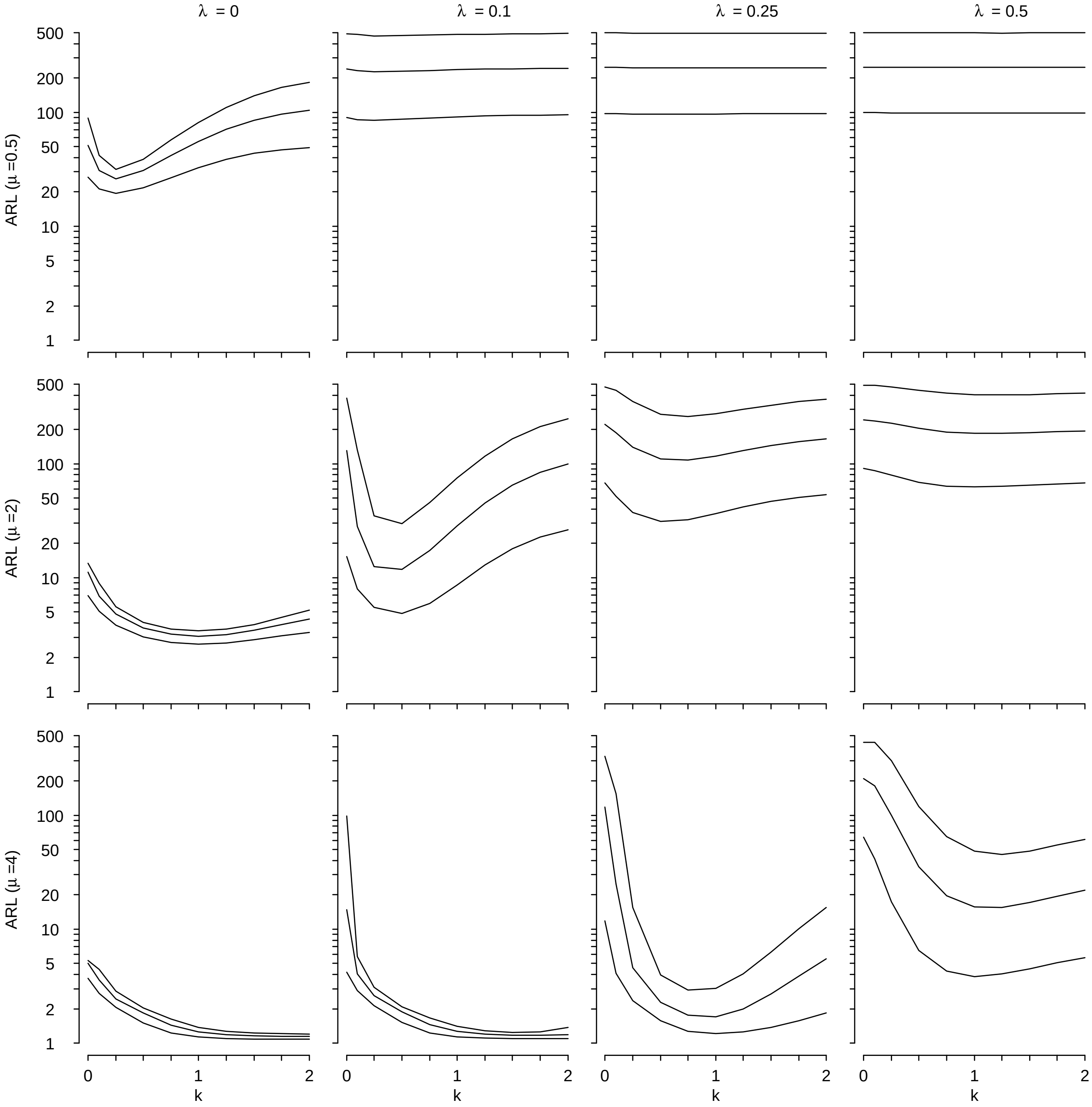

Figure 6: ARLs for CUSUM charts when an IMA(λ) process experiences a step shift of μ (innovation) standard deviations— μ increases down the columns; λ increases across the rows. The top, middle, and bottom curves in each panel correspond to CUSUMs with ARL_0 values of 500, 250, and 100 respectively. To choose a value for k, use the column with λ closest to the estimated value and focus on the curves corresponding to the desired ARL_0. Sometimes it will be necessary to sacrifice ARL performance for one value of μ to get better performance at another value.

As an example, suppose λ is estimated to be near 0.1. From the curves in the second column of plots, we see that the middle panel is most interesting. Shifts of 0.5σ (top panel) are nearly impossible to detect regardless of the value of k. Their ARL curves drop only slightly below ARL_0. On the other hand shifts of 4σ (bottom panel) are detected quickly as long as k is at least 0.5. For 2σ shifts, values of k between 0.25 and 0.5 give the smallest ARLs. Other values can result in substantially worse performance. A value of $k = 0.5$ is a reasonable choice. Referring to Figure 5, h should be set at about 3.5, 4.4, or 5.1 to give ARL_0 near 100, 250, or 500 respectively.

The other columns of Figure 6 can be used similarly. Usually the range of reasonable values for k is fairly obvious. Panels for values of λ exceeding 0.5 are not shown because they are very similar to those for $\lambda = 0.5$.

Several other aspects of Figure 6 are noteworthy. One striking feature is that the larger λ is, the more difficult it is to detect shifts of a given size. With $\lambda = 0.5$, we have virtually no power to detect even 2 standard deviation shifts. This is as we should expect from the discussion of Figure 4. The most interesting panels in Figure 6 lie near the top-left to bottom-right diagonal. In these panels the best value of k depends on λ. Getting small ARLs for small shifts in an iid process ($\lambda = 0$) requires k near 0.25 whereas getting reasonably small ARLs for large shifts when λ is 0.5 or higher requires k to be close to 1.0 or higher. This should be useful for control chart design and it follows the traditional wisdom for CUSUM charts that small values of k produce better ARLs for small shifts while large values of k give better ARLs for large shifts.

EWMA: An EWMA monitoring scheme is based on an exponentially weighted moving average of the forecast errors:

$$Q_t = \gamma a_t + (1-\gamma)Q_{t-1}$$

where Q_0 is initialized at zero. The scheme signals when $|Q_t|$ exceeds an action limit. The weight $\gamma \in (0, 1]$ and the action limit h are design parameters. Choosing γ and h for an EWMA scheme is similar to choosing k and h for a CUSUM scheme. Plots, of h versus γ (similar to the h versus k plots for the CUSUM) can be constructed to determine the EWMA control limit which yields a given ARL_0 or $P_0(10)$. Similarly, one could produce a figure for the EWMA, similar to Figure 6 for the CUSUM. In subsection 3.2, however, CUSUMs are shown to perform at least as well as and sometimes better than EWMAs. Thus, we provide design aids only for CUSUM schemes.

Likelihood ratio scheme: A monitoring scheme based on likelihood ratio statistics can take advantage of the patterned change that occurs in the mean of the forecast errors when an IMA undergoes a step shift. To make the scheme manageable we limit the data used in period t to the last $n + 1$ forecast errors $(a_{t-n}, \ldots, a_t)$. The forecast errors depend on *all* of the available data, however. For the comparisons in the next section we considered $n + 1 = 3$, 5 and 9 and found that the sample size made very little difference in ARL or $P(10)$ performance except in the iid case ($\lambda = 0$) for small shifts. Then larger values of n were helpful. Based on $(a_{t-n}, \ldots, a_t)$, the likelihood ratio statistic to test for a step shift occurring anywhere between period $t-n$ and period t is a monotone function of

$$U_t = \max_{k=0,\ldots,n} \left| Z_k^{(t)} \right| \tag{4}$$

where $Z_k^{(t)}$ is the "Z-statistic" in the regression of $(a_{t-k}, a_{t-k+1}, \ldots, a_t)$ on $(1, 1-\lambda, \ldots, (1-\lambda)^k)$, namely

$$Z_k^{(t)} = \frac{1}{\sigma} \sum_{i=0}^{k} (1-\lambda)^{k-i} a_{t-i} \Bigg/ \left[\sum_{i=0}^{k} (1-\lambda)^{2i} \right]^{1/2} .$$

$Z_k^{(t)}$ is sensitive to a geometrically decaying pattern in the forecast errors that starts in period $t-k$. Thus, U_t should be sensitive to a step shift in the IMA that begins in any one of the most recent $n+1$ periods. The likelihood ratio monitoring scheme signals when U_t exceeds an action limit h which is picked to yield a desired ARL_0 or $P_0(10)$. Vander Wiel (1994) studies the null distribution of U_t for the purpose of hypothesis testing.

3.2 Comparisons among schemes

Having introduced 4 classes of monitoring schemes, what can be said about their relative performances?

Figure 7 shows optimal ARL curves for IMA processes with parameters ranging from $\lambda = 0$ in the upper left panel to $\lambda = 1$ in the lower right one. Each curve is a function of the shift size μ and shows the minimum ARL that can be achieved within the given class of schemes subject to the constraint $\text{ARL}_0 = 500$. Generally, no single scheme within a class can attain the minimum ARL for several shift sizes so the optimal scheme changes with μ.

For example, within the CUSUM curves, k changes with μ to achieve the minimum ARL. Thus, Figure 6 should be used to judge how well a CUSUM chart optimized for one value of μ will perform for others.

Figure 8 shows optimal $1/P(10)$ curves constructed in a like manner. Each curve shows the minimum $1/P(10)$ value that can be achieved within the given class of schemes subject to the constraint $1/P_0(10) = 50$. The inverse of $P(10)$ is used to facilitate comparisons between Figures 7 and 8; it is the average number of independent trials it would take for a scheme to signal a shift within 10 periods of its onset.

The conclusions drawn from Figures 7 and 8 are remarkably simple. Namely, for a given sized shift, a CUSUM can be designed with respect to either the ARL criterion or the $P(10)$ criterion to perform at least as well as, and often better than, any of the other schemes. For the $P(10)$ criterion, our comparisons show a clear ordering of the 4 schemes. CUSUMs and EWMAs perform equally well, followed by likelihood ratio schemes and then by Shewhart individuals charts. For the ARL criterion CUSUMs always perform best and Shewhart individuals charts worst. The ordering between the EWMA and the likelihood ratio scheme, however, depends on λ.

Shewhart individuals charts sometimes preform miserably and never do better than the others. In the case of random walks ($\lambda = 1$), all schemes have equally poor performances. In fact, for random walks, the optimal member of each class reduces to a Shewhart individuals chart. A final broad observation from Figures 7 and 8 has already been made but bears repeating: it is substantially more difficult to detect level shifts in IMAs as λ increases.

3.3 Computing ARLs and signaling probabilities

Three good ways to studying the run length distribution of a monitoring scheme are (1) analytically deriving it; (2) approximating it by way of a discrete Markov Chain representation; and (3) building it up through Monte Carlo simulation. A fourth method is to derive and solve an integral equation satisfied by the ARL. That, however, is equivalent to the Markov Chain approximation up to a choice of an integration quadrature (Champ and Rigdon, 1991). We have used all three methods in this work. The S functions (Becker, Chambers and Wilks, 1988) used to generate the data for Figures 5 and 6 and comparable data for EWMA charts are available as a Unix compressed shar file `ima.arl.shar.Z` on the world wide web at URL `http://netlib.att.com/netlib/att/stat/prog/index.html` or by anonymous ftp from `netlib.att.com/netlib/att/stat/prog`. Shewhart individuals charts are simple enough to lend themselves to analytical methods even when applied to monitoring IMA forecast errors. Survival probabilities of the time T of the first signal after a shift, can be built up using $\Pr\{T > 0\} = 1$ and the recursion (for $t \geq 1$)

$$\Pr\{T > t\} = p_t \Pr\{T > t-1\}.$$

where $p_t = \Phi\left(h - \mu(1-\lambda)^{t-1}\right) - \Phi\left(-h - \mu(1-\lambda)^{t-1}\right)$ with Φ denoting the standard normal CDF. The ARL is the sum of these survivor probabilities. In the extreme cases of $\lambda = 0$ or 1 the recursion is trivial and computing the ARL involves summing a geometric sequence. For $\lambda \in (0,1)$, p_t quickly approaches a limit, p_∞, and the ARL can be approximated as

$$\text{ARL} \approx \sum_{t=0}^{\tau-1}(p_0 \cdots p_t) + \frac{p_1 \cdots p_\tau}{1 - p_\infty} \tag{5}$$

where τ is a large positive integer and $p_0 \equiv 1$.

Because p_t *increases* to p_∞, the approximation error is less than

$$(p_1 \cdots p_\tau)\left(\frac{1}{1-p_\infty} - \frac{1}{1-p_{\tau+1}}\right).$$

Thus, an easy method to obtain the ARL is to continue summing in in (5) until the error bound is as small as required.

Simple analytical results are not available for deriving run length distributions for CUSUM and EWMA schemes even when they are applied to iid Gaussian sequences. A computational technique based on Markov Chains is available, however. The basic idea is to discretize the "in control" region $(-h, h)$, representing it by a number of singletons. Properties of discretized CUSUMs or EWMAs are then derived from transition probability matrices P_t with entry (i, j) equal to the probability of moving to state j in period t conditional on being in state i in period $t-1$. The rows of P_t do not generally sum to 1 because the state space only represents $(-h, h)$ and there is always a probability of signaling.

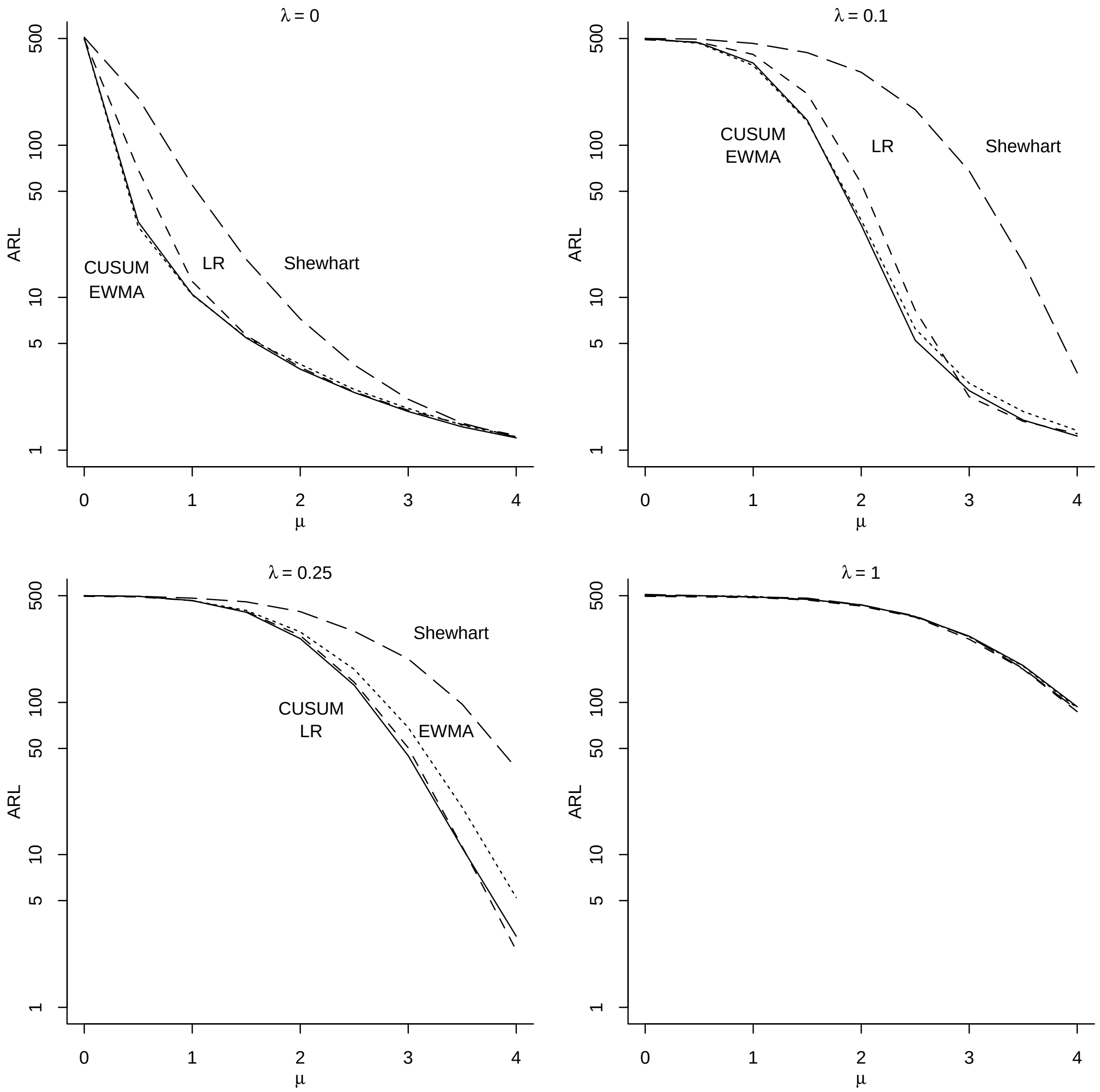

Figure 7: Optimal ARLs among 4 classes of monitoring schemes. CUSUMs dominate the other schemes and the Shewhart individuals chart is only competitive for random walks ($\lambda = 1$). Each curve shows the minimum ARL that can be achieved for various values of μ subject to $\text{ARL}_0 = 500$

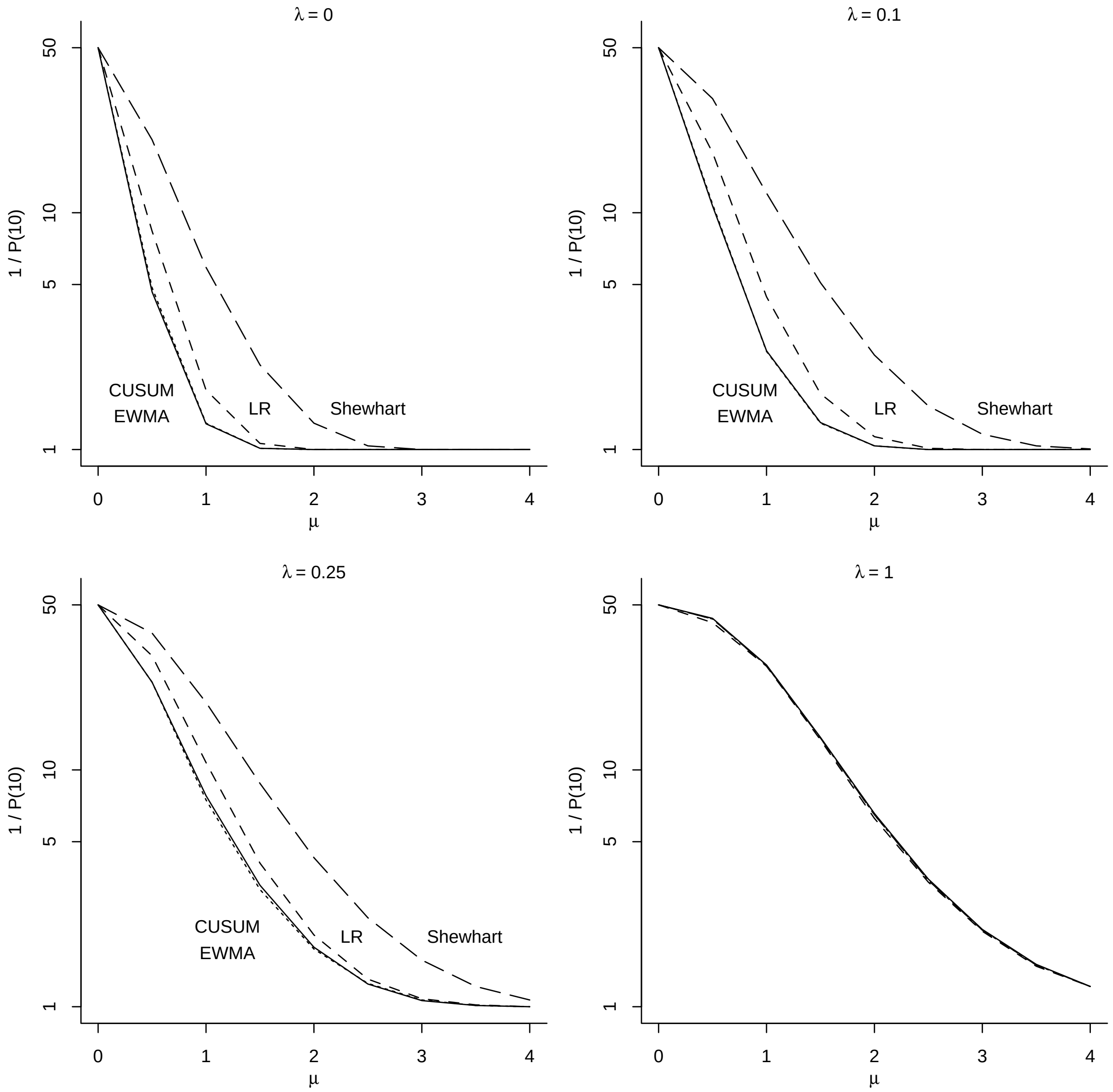

Figure 8: Optimal values of $1/P(10)$ among 4 classes of monitoring schemes. CUSUMs and EWMAs dominate the likelihood ratio scheme which dominates the Shewhart individuals chart. Each curve shows the minimum $1/P(10)$ that can be achieved for various values of μ subject to $1/P_0(10) = 50$

Brook and Evans (1972) were the first to analyze CUSUMs by Markov Chains. Woodall (1984) gives details for efficiently representing the state space for 2-sided CUSUMs and for calculating transition probability matrices. Lucas and Saccucci (1990) give details for EWMA schemes. In these studies, however, the transition matrices P_t did not depend on t because the observations following a shift in the mean were iid. For IMA monitoring we have seen that the forecast errors will be independent with constant variance and a patterned mean that converges geometrically to zero. Thus, P_t changes with t, but converges as t becomes large.

If $\boldsymbol{\pi}$ is a prior probability vector for the state of the Markov chain in period 0, then survival probabilities of the discretized scheme are given by

$$\Pr\{T > t\} = \boldsymbol{\pi}' P_0 \cdots P_t \mathbf{1}$$

where P_0 is the identity matrix. The ARL is the sum of these probabilities and can be approximated by

$$\text{ARL} \approx \boldsymbol{\pi}' \left[\sum_{t=0}^{\tau-1} (P_0 \cdots P_t) + (P_0 \cdots P_\tau)(I - P)^{-1} \right] \mathbf{1}. \tag{6}$$

where P is either $P_{\tau+1}$ or the limit matrix P_∞. Assuming that the two choices of P give ARLs that bracket that of the discretized scheme, we increased τ until either the relative error was less than 0.1% or the absolute error was less than 0.01 or the mean of the forecast errors was less than 0.001. [The bracketing result is not difficult to prove for (continuous) 1-sided EWMA and CUSUM schemes using the fact that the plotted quantities are increasing functions of the forecast errors; 2-sided schemes, however, are more difficult to analyze and discretization further complicates matters.]

For EWMA calculations we used 96 states in $(-h, h)$. For 2-sided CUSUMs the state space is 2-dimensional, and thus grows quickly as the number, d, of discrete values in each dimension is increased. For this reason, we follow the procedure of Brook and Evans (1972) of computing the ARL for several values of d and reporting the ARL of the continuous-state procedure as the least squares intercept in a regression of ARL on $1/d$ and $1/d^2$. For the 2-sided CUSUM, we used $d = 4, 7, 8, 9, 10$.

It still remains to specify the prior probabilities, $\boldsymbol{\pi}$, on the initial state. Two reasonable alternatives are (1) to let $\boldsymbol{\pi}$ be an indicator vector pointing to the state with the CUSUM or EWMA equal to zero; or (2) to let $\boldsymbol{\pi}$ be the vector of steady state probabilities of the Markov chain with $\mu = 0$ and conditional on not signaling. We refer to the first alternative as a cold start and the second as a warm start. All figures in this work are for cold starts but we have calculated some ARLs for warm starts and found that they differ by immaterial amounts from the cold start values.

To find h corresponding to a given ARL_0 [or $P_0(10)$] for CUSUM and EWMA schemes we first found two values of h whose ARL_0's [or $P_0(10)$'s] bracketed the target value. Then we used the bisection method (for finding a root) to home in on the desired value of h.

The likelihood ratio scheme lends itself to neither analytical nor numerical analysis using Markov chain methods. In this case ARLs and signaling probabilities were approximated using Monte Carlo simulation. For the simulation, we used simple FORTRAN code linked to control functions written in S (Becker et al., 1988). We also used simulation to verify Markov chain calculations for EWMA and CUSUM schemes.

For ARLs we averaged 10,000 run lengths where each run started from a level shift beginning in the first period and where forecast errors for previous periods were taken to be zero. This is analogous to computing "cold start" ARLs for Markovian schemes. If the run length distribution is crudely approximated as geometric with mean and standard deviation equal to the ARL, then 10,000 runs produces estimated ARLs with standard errors of 1%. (Simulation data shows that the equal mean and standard deviation assumption is roughly correct.) For $P(10)$ values we used the fraction of signals by period 10 in 40,000 runs. Using the binomial variance, standard errors for estimated signaling probabilities are less than 0.15% when the nominal probability is no greater than $1/10$, as is the case for all of our comparisons. All Markov chain computations were verified to agree to within 3.5 standard errors of estimates obtained from the simulations.

4 Summary and Discussion

4.1 Summary

For many years the process monitoring field has not dealt squarely with the fact that most time-ordered data are autocorrelated. In fact, most monitoring schemes still use only the most basic statistical models assuming that data will constitute simple random samples from a specified distribution. For data with substantial positive autocorrelation, standard control charts will signal much too frequently.

This paper has addressed statistical process monitoring for the case when the stochastic component of process data is well modeled as an integrated moving average (IMA) process. Typically such a process will be operated under feedback control. A sustained shift in the underlying level of the process leaves forecast errors independent with constant variance but causes a patterned response in their mean consisting of an initial jump followed by a geometric decay to zero.

Comparisons among 4 classes of monitoring schemes applied to forecast errors showed that properly designed CUSUMs perform as well as and often better than any of the other schemes. Likelihood ratio schemes and especially EWMA charts are also competitive but can sometimes be beaten by CUSUMs. Shewhart individuals charts are often much less sensitive to level shifts than the others. Further studies of CUSUM and EWMA performance and design sensitivity can be conducted using S functions (Becker et al., 1988) available from the author as described in Subsection 3.3.

4.2 Generalizing to ARIMA Processes

This work has focused on performance of forecast monitoring schemes for the important case in which the underlying stochastic nature of a process can be modeled as an IMA. The methods, however, can also be applied to assess performance for detecting abrupt changes in the level of a general ARIMA model. Consider monitoring a series $Z_t = \mu I[t \geq k] + N_t$ where N_t is an ARIMA(p,d,q) process satisfying $\phi(B)(1-B)^d N_t = \theta(B)\alpha_t$ where, according to standard notation and assumptions, a_t is an iid N$(0, \sigma^2)$ sequence, B is the backshift operator, and $\phi(B)$ and $\theta(B)$ are polynomials of respective degrees p and q with roots lying outside the unit disc. In this case, 1-step ahead forecast errors of Z_t are given by $Z_t - \hat{Z}_{t|t-1} = \mu\delta_{t-k} + \alpha_t$ where

$$\delta_{t-k} = \frac{\phi(B)(1-B)^d}{\theta(B)} I[t \geq k].$$

The sequence of deterministic means $\mu\delta_{t-k}$ is zero until until period k when the sequence responds in a predetermined manner to the abrupt shift in level. Based on the theory of finite difference equations (for example, Goldberg (1958) or Fuller (1976)), the pattern will tend to an asymptotic level exponentially fast. For stationary models ($d = 0$) the new level will be nonzero while for nonstationary models ($d \geq 1$) the pattern tends to zero. This behavior allows analyses of run length behavior for Shewhart individuals, CUSUM and EWMA charts to be carried out using the analytic and Markov Chain techniques presented here for the IMA case. In computing ARLs, for example, once the mean is suitably close to its asymptote, the remainder of the scheme can be treated as if new observations are iid. Wardell et al. (1994) give details for computing the run length distribution for individuals charts of forecast errors from stationary ARMA models.

It does not seem profitable, to tabulate ARLs for various monitoring schemes under a large number of different ARIMA models. It would be more useful to provide a tool for making the appropriate computations as they are required. Another useful approach for general ARIMA models would be to determine how broadly the recommendations given here apply in the larger class.

Do CUSUM charts dominate in run length performance when applied to other ARIMA models? Intuition suggests they will dominate for models that produce mean patterns similar to the exponential decays coming from IMA models. In some cases, however, the mean pattern will not remain on one side of zero. If the sign alternates, one would expect better performance by CUSUM-ing forecast errors with alternating signs. For other kinds of sign changes it is not clear that a CUSUM or EWMA would be appropriate. For example, if the mean pattern were a slowly damped sinusoid, perhaps the likelihood ratio scheme would dominate the others. Perhaps the CUSCORE statistics of Box and Ramirez (1992) could be used in this case.

4.3 Further Commentary

The likelihood ratio scheme might seem to be too oriented toward detecting a step shift when compared to the individuals, CUSUM, and EWMA charts which are usually thought of as "general purpose" charts for monitoring a mean. All forms of deterministic level change, however, such as a ramp or spike will manifest themselves (in filtered form) in the forecast errors means. Forecast errors persistently to one side of zero affect the likelihood ratio scheme in (more or less) the same manner they affect the other schemes—namely, they push the plotted statistic toward an action limit. The likelihood ratio scheme is tuned to the particular case of an underlying step shift. If an application required, the scheme could be derived assuming a ramp, spike, sinusoid or any other form of process upset.

All of the control charts presented here might be criticized for sometimes having huge ARLs when measured against typical ARLs for monitoring the mean of uncorrelated data. We have already argued that these differences do not necessarily reflect weaknesses of the monitoring schemes but rather reflect the limited information content of the data. Simply stated, step shifts are more difficult to see within wandering processes. Also, a mitigating factor to the high ARLs is that wandering processes are often under some form of active feedback control. Thus, abrupt changes are compensated for even if they are never explicitly detected. The effect of any particular upset is typically short lived. Nevertheless, if an upset can be detected and diagnosed, possibly the source can be eliminated resulting in a process with fewer problems in the future.

It is important to admit that the quantitative ARL and $P(10)$ results presented here assume that the wandering disturbance affecting a process is known to be an IMA with known parameters. In reality, model forms are almost never known, parameters are never constant, and their estimates are never perfect. A small study of CUSUM ARLs when λ is misestimated shows that they are most sensitive to misestimation for small λ and small shift size μ. For example consider an estimated value $\hat{\lambda} = 0.5$ and a CUSUM designed with $\text{ARL}_0 = 500$ and $k = 1.25$. The actual ARL_0 varies from -40% to $+35\%$ as the true λ varies from 0.4 to 0.6. ARLs for 4σ shifts vary from -20% to $+10\%$. Analogous results are more extreme for smaller $\hat{\lambda}$ and less extreme for larger $\hat{\lambda}$. Non-Gaussian data and imprecise knowledge of σ can also greatly affect ARL performance as they do in standard iid process monitoring settings.

In light of the sensitivity to model assumptions, ARL and $P(10)$ calculations should not be taken too seriously. It is the qualitative results of this work that are most usable: namely, CUSUM charts of forecast errors perform competitively for processes that wander. The closer to a random walk the wandering becomes, the more like a Shewhart individuals chart the CUSUM should be designed. And finally, smaller values of k work better for detecting smaller shifts and vice versa.

In this regard the levels of μ and λ displayed in Figure 6 were selected so as to cover the qualitatively interesting ARL features in the ranges $\mu \in [0, 4]$ and $\lambda \in [0, 1]$. Plots could be constructed to allow more precise determination of the k that minimizes ARL for a given shift size. The important practical use of Figures 5 and 6, however, is to choose values of h and k that are in the right "ball park" and to get a rough indication of how readily the chart can detect shifts of various sizes. Since ARLs are notoriously sensitive to assumptions which, in practice, are never true, precise choice of CUSUM parameters is not necessary for practical problems.

5 Acknowledgement

Thanks are due to the editor, associate editor, and especially the referees for helpful suggestions that have improved both the content and presentation of this work.

References

Alwan, L. C. and Roberts, H. V. (1988). Time-series modeling for statistical process control, *Journal of Business and Economic Statistics* **6**: 87–95.

Bagshaw, M. and Johnson, R. A. (1975). The effect of serial correlation on the performance of Cusum tests, II, *Technometrics* **17**: 73–80.

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988). *The New S Language: A Programming Environment for Data Analysis and Graphics*, Wadsworth Publishing Co (Belmont CA).

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control (rev. Ed.)*, Holden-Day, San Francisco.

Box, G. E. P. and Kramer, T. (1992). Statistical process monitoring and feedback adjustment—a discussion, *Technometrics* **34**: 393–399.

Box, G. E. P. and Ramirez, J. (1992). Cumulative score charts, *Quality and Reliability Engineering International* **8**: 17–27.

Brook, D. and Evans, D. A. (1972). An approach to the probability distribution of cusum run lengths, *Biometrika* **59**: 539–549.

Brown, R. G. (1962). *Smoothing, Forecasting and Prediction of Discrete Time Series*, Prentice-Hall, Englewood Cliffs, NJ.

Champ, C. W. and Rigdon, S. E. (1991). A comparison of the Markov chain and the integral equation approaches for evaluating the run length distribution of quality control charts, *Communications in Statistics, Part B–Simulation and Computation* **20**: 191–204.

Crowder, S. V. (1987). A simple method for studying run-length distributions of exponentially weighted moving average charts, *Technometrics* **29**: 401–407.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*, Wiley, New York.

Gan, F. F. (1991). An optimal design of CUSUM quality control charts, *Journal of Quality Technology* **23**: 279–286.

Gardner, E. S. J. (1983). The trade-offs in choosing a time series method, *Journal of Forecasting* **2**: 263–267.

Goel, A. L. and Wu, S. M. (1971). Determination of A.R.L. and a contour nomogram for cusum charts to control normal mean, *Technometrics* **13**: 221–230.

Goldberg, S. (1958). *Introduction to Difference Equations*, Wiley, New York.

Golder, E. R. and Settle, J. G. (1976). Monitoring schemes in short-term forecasting, *Operations Research Quarterly* **27**: 489–501.

Goldsmith, P. L. and Whitfield, H. (1961). Average run lengths in cumulative chart quality control schemes, *Technometrics* **3**: 11–20.

Harris, T. J. and Ross, W. H. (1991). Statistical process control procedures for correlated observations, *The Canadian Journal of Chemical Engineering* **69**: 48–57.

Johnson, R. A. and Bagshaw, M. (1974). The effect of serial correlation on the performance of cusum tests, *Technometrics* **16**: 103–112.

Longnecker, M. T. and Ryan, T. P. (1992). Charting correlated process data, *Technical Report 166*, Texas A&M University, Department of Statistics.

Lu, C.-W. and Reynolds, Marion R., J. (1994). Control charts based on residuals for monitoring autocorrelated processes, *Technical Report 94-8*, Virginia Polytechnic Institute and State University, Blacksbur, VA 24060.

Lucas, J. M. (1982). Combined Shewhart-CUSUM quality control schemes, *Journal of Quality Technology* **14**: 51–59.

Lucas, J. M. and Crosier, R. B. (1982). Fast initial response for Cusum quality-control schemes: Give your Cusum a head start, *Technometrics* **24**: 199–205.

Lucas, J. M. and Saccucci, M. S. (1990). Exponentially weighted moving average control schemes: Properties and enhancements, *Technometrics* **32**: 1–12.

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

Montgomery, D. C. and Friedman, D. J. (1989). Statistical process control in a computer-integrated manufacturing environment, *in* J. B. Keats and N. F. Hubele (eds), *Statistical Process Control in Automated Manufacturing*, Marcel-Dekker, New York, pp. 67 – 87.

Montgomery, D. C. and Mastrangelo, C. M. (1991). Some statistical process control methods for autocorrelated data, *Journal of Quality Technology* **23**: 179–193.

Page, E. (1954). Continuous inspection schemes, *Biometrika* **41**: 100–115.

Runger, G. C. (1995). Average run lengths for CUSUM control charts applied to residuals, *Communications in Statistics—Theory and Methods* **5**(1): 273.

Superville, C. R. and Adams, B. M. (1994). An evaluation of forecast-based quality control schemes, *Communications in Statatistics—Simulation and Computation* **23**: 645–661.

Trigg, D. W. (1964). Monitoring a forecasting system, *Operations Research Quarterly* **15**: 271–325.

Vander Wiel, S. A. (1994). Statistical process monitoring using integrated moving averages, *Technical report*, AT&T Bell Laboratories, Murray Hill, NJ.

Vander Wiel, S. A., Tucker, W. T., Faltin, F. W. and Doganaksoy, N. (1992). Algorithmic statistical process control: Concepts and an application, *Technometrics* **34**: 286–297.

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. (1992). Control charts in the presence of data correlation, *Management Science* **38**: 1084–1105.

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.

Woodall, W. H. (1984). On the markov chain approach to the two-sided cusum procedure, *Technometrics* **26**: 41–46.

Yashchin, E. (1993). Performance of cusum control schemes for serially correlated observations, *Technometrics* **35**(1): 37–52.