Journal of Forecasting, Vol. 2, 1–21 (1983)

Automatic Monitoring of Forecast Errors

EVERETTE S. GARDNER, JR.
U.S. Navy, Norfolk, Va., U.S.A.

ABSTRACT

This paper evaluates a variety of automatic monitoring schemes to detect biased forecast errors. Backward cumulative sum (cusum) tracking signals have been recommended in previous research to monitor exponential smoothing models. This research shows that identical performance can be had with much simpler tracking signals. The smoothed-error signal is recommended for $\alpha = 0.1$, although its performance deteriorates badly as α is increased. For higher α values, the simple cusum signal is recommended. A tracking signal based on the autocorrelation in errors is recommended for forecasting models other than exponential smoothing, with one exception. If the time series has a constant variance, the backward cusum should give better results.

KEY WORDS Tracking signals Monitoring forecasts Quality control Cusum Exponential smoothing Simulation Autocorrelation

In most forecasting systems, it is highly desirable to automatically monitor forecast errors to ensure that the system remains in control. For example, if a non-seasonal forecasting model is applied to a time series with unsuspected seasonality, biased errors will occur. When a trend develops in a time series being forecasted by simple exponential smoothing, the forecasts will lag. If the trend remains constant, the simple exponential smoothing model will lag the time series to infinity. Most forecasting models with fixed parameters will lag step changes in the mean, trend, or seasonality components of a time series. These problems need to be detected as quickly as possible to enable the forecasting model to be refitted to the data or changed to a more appropriate model.

There are at least three warning signs when a forecasting system goes out of control. The first indicator is the cumulative sum (cusum) of the forecast errors, which can be computed and tested in several different ways. The cusum should fluctuate around zero when the system is in control. If biased errors occur, the cusum will depart from zero. The second indicator is an estimate of the mean forecast error, which will also depart from zero when biased errors occur. The third indicator is the first-order autocorrelation in forecast errors. Since biased errors tend to have the same sign, the existence of any significant positive autocorrelation indicates lack of control.

Readability: The Gunning fog index for this paper is about 13.

0277-6693/83/010001-21$02.10

Received March 1982
Revised September 1982

This paper evaluates tracking signals to monitor each of these warning signs. Two of the signals are based on cusums—Brown's simple cusum and the sophisticated backward cusum system developed by Harrison and Davies. Trigg's smoothed-error signal, which is widely used in practice, and an autocorrelation signal developed in this paper are also evaluated. One section of the paper is devoted to an analysis of each signal, including updating equations, probability distributions, and response rates to biased errors.

Detailed performance comparisons among these signals are deferred to the last section of the paper. Comparisons are made on the basis of the number of time periods required to detect bias in two types of errors: the errors from simple exponential smoothing, and randomly generated $N(0, 1)$ deviates. Exponential smoothing errors are autocorrelated by the nature of the forecasting process. This distortion has a significant effect on the ability of each tracking signal to detect bias. The $N(0, 1)$ errors are independent and represent the type of errors which should be expected from other forecasting models.

This research can be replicated. Details of the simulation work are discussed in the Appendix.

THE SIMPLE CUSUM TRACKING SIGNAL

The first tracking signal for forecast control was proposed by Brown (1959). This signal compares the cumulative sum of the errors at the end of each period to the smoothed MAD. The updating equations as each new error is observed are:

$$e_t = X_t - F_t \tag{1}$$

$$\text{SUM}_t = e_t + \text{SUM}_{t-1} \tag{2}$$

$$\text{MAD}_t = \alpha|e_t| + (1-\alpha)\text{MAD}_{t-1} \tag{3}$$

$$C_t = |\text{SUM}_t/\text{MAD}_t| \tag{4}$$

The forecast error, e_t, is the actual time series value, X_t, minus the forecast, F_t. As usual, the smoothing parameter is restricted to $0 \leq \alpha \leq 1$. If C_t exceeds a significant multiple of the smoothed MAD, forecasts may be biased.

There are two approaches to the problem of selecting control limits for C_t as well as other tracking signals. First, we can use a control limit that yields some desired probability of getting a 'false trip', defined as a Type I error, or a case where the control limit is exceeded due to chance rather than biased forecasts. Second, we can base the control limit on the number of time periods required to detect biased forecast errors of any given size.

Using probabilities to select control limits for the simple cusum

In the first approach, the normal distribution is often used to find the probability of a false trip. For example, Brown (1963, pp. 288–289) shows that the simple cusum for exponential smoothing is approximately normally distributed, with standard deviation equal to

$$\sigma_c = 0.884/(\alpha)^{1/2} \tag{5}$$

Extensive tables in Montgomery and Johnson (1976) can be used to find control limits based on similar approximations for a variety of other forecasting models.

The validity of a normal approximation for C_t based on (5) was tested by simulation. Table 1 gives the simulated distributions of C_t from simple exponential smoothing forecasts (one-ahead) for $\alpha = 0.1$, 0.2, and 0.3. The distributions were compiled using 1000 time series. Each series was

Table 1. Cumulative distribution of the simple cusum tracking signal (exponential smoothing)

Cumulative probability	C_t		
	$\alpha = 01$	$\alpha = 0.2$	$\alpha = 0.3$
0.80	3.6	2.6	2.1
0.90	4.7	2.9	2.8
0.95	5.6	4.1	3.5
0.96	5.9	4.3	3.7
0.97	6.3	4.6	3.9
0.98	6.8	5.0	4.3
0.99	7.5	5.6	4.9

120 periods in length, with a constant mean and random normal noise about the mean. The tracking signal was started with $SUM_0 = 0$. MAD_0 was set equal to its expected value, given the noise in the series and the smoothing parameter used. The smoothing parameter in the forecasting model was set equal to the parameter used to smooth the MAD. Comparisons showed there was no advantage to using different parameters. A frequency distribution was compiled of the value of C_t at the end of periods 21–120, giving 100,000 observations on C_t across all series. The first 20 periods were used as a 'run-in' to wash out the effects of initial conditions.

A normal approximation for C_t with standard deviation equal to (5) agrees very well with the simulated distribution for $\alpha = 0.1$ in Table 1. However, there are some differences at higher α values. This should be expected, since the assumptions used to develop (5) hold only for small α values. For example, at $\alpha = 0.2$, the 90 per cent control limit (double sided) should be 1.65σ or about 3.3 using the approximation. However, Table 1 gives a control limit of 2.9. Since Table 1 is based on a large sample size, it should be a better tool for estimating probabilities.

Notice that the control limits at any probability level decrease as α is increased. An obvious reason for this effect is that the increase in α increases the MAD. A more subtle reason is that exponential smoothing models have a tendency to induce negative autocorrelation in the errors from one-ahead forecasts. To illustrate, suppose a time series has a constant mean with random noise about the mean. The autocovariance of successive errors from simple exponential smoothing in this case is (Brown, 1963, p. 310):

$$R = \frac{\alpha(1-\alpha)}{2-\alpha}\sigma_x^2[1-(2-\alpha)(1-\alpha)^{-1}] \tag{6}$$

where σ_x^2 is the variance of the time series. R decreases as α increases, ranging from about $-0.05\sigma_x^2$ at $\alpha = 0.1$ to $-0.17\sigma_x^2$ at $\alpha = 0.3$. Thus exponential smoothing errors tend to alternate in sign, which helps reduce the sum of errors.

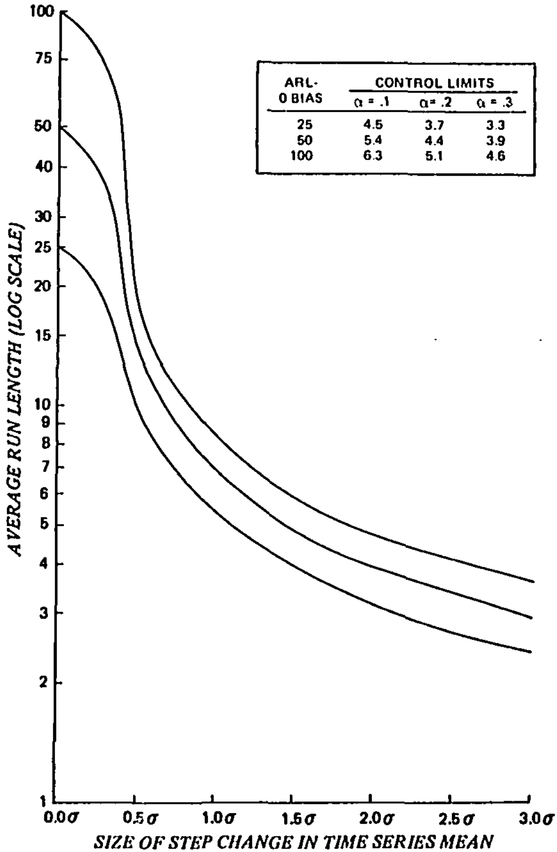
Using average run lengths to select control limits for the simple cusum

The second approach to the problem of selecting control limits is based on the concept of the average run length (ARL) to detect biased forecast errors. The run length is defined by convention in industrial quality control as the number of periods required to detect a change in the process being monitored. It is computed as the period number in which a quality control scheme first breaks its control limit minus the period number immediately before a change occurred. Hence, the minimum run length is one period. This definition of the run length should not be confused with the theory of runs in non-parametric statistics (Siegel, 1956).

Figure 1 illustrates the ARL approach. The response curves show the ARL to detect a step change in the time series mean, again using the simple exponential smoothing model. ARLs are plotted on a log scale versus step changes (expressed as a multiple of the standard deviation of the time series) on a linear scale.

The response curves were generated by simulation, using a sample of 1000 time series similar to those described above. The first step in the simulation was to use an iterative search procedure to find control limits yielding ARLs of 25, 50, and 100 periods on unbiased errors (no change in the mean) at $\alpha = 0.1$, 0.2, and 0.3. These ARLs estimate the run length until the first trip at the control limits shown. They were computed after a run-in of twenty periods. The second step in the simulation was to measure the run lengths at those control limits to detect step changes at period 21 in the time series mean. Step changes ranged from 0.5 to 3.0σ in increments of 0.5.

To illustrate how the table of control limits in Figure 1 is used to select a response curve, suppose we want a target ARL on unbiased errors of 50 periods. The smoothing parameter is 0.10. The



ARL-0 BIAS	CONTROL LIMITS α = .1	α = .2	α = .3
25	4.5	3.7	3.3
50	5.4	4.4	3.9
100	6.3	5.1	4.6

Figure 1. Simple cusum ARL performance (exponential smoothing)

control limit on C_t should be set at 5.4, which yields the middle response curve. A target ARL of 100 periods corresponds to a control limit of 6.3 and the upper response curve.

Several generalizations are important in interpreting Figure 1. First, only one set of response curves is given, since response rates to biased errors were about the same at each α value tested. Second, the control limits in Table 1 cannot be used to predict ARLs on unbiased errors. For example, the 95 per cent control limit at $\alpha = 0.1$ is 5.6, which might be expected to give a run length of about twenty periods of unbiased errors. But the control limits in Figure 1 show that C_t will run for more than fifty periods at a limit of 5.6.

The reason for this difference is that the distribution of C_t on unbiased errors in Table 1 is based on uninterrupted sequences of observations—the runs were not stopped when any control limit was broken. In Figure 1, the runs were stopped when the control limits shown were broken. As should be expected, successive values of C_t are highly autocorrelated when the signal is uninterrupted. The behaviour of the signal includes some extremely long runs between trips at any given control limit. Once the signal breaks a control limit, it tends to stay outside that limit for a number of periods.

Another generalization is that the ARLs on unbiased errors in Figure 1 are highly variable. Standard deviations of the ARLs on unbiased errors are roughly equal to the ARLs. (Standard errors, of course, are negligible due to the sample size.) Standard deviations decline exponentially as bias is introduced, to about 50 per cent of the ARL at a 1.5σ step and 30 per cent at 3.0σ. These results are consistent with Golder and Settle's simulation study (1976), and hold for every tracking signal tested in this research. Because of this variability, control limits yielding ARLs less than about twenty-five periods on unbiased errors are not recommended. Otherwise, the number of false trips could be unmanageable.

Interpolation in Figure 1 should produce control limits suitable for most practical applications of the simple exponential smoothing model. For other types of smoothing models, the control limits shown will give comparable ARLs on unbiased errors, provided that the same α values are used to smooth the mean. Response rates to bias vary according to the type of model and the other smoothing parameters used. However, any desired response curves can easily be established by simulation, using Figure 1 as a starting point.

Advantages and disadvantages of the simple cusum

Compared to other tracking signals, the simple cusum has two advantages. The most important is that the simple cusum performance is independent of the smoothing parameter used. As discussed below, the performance of tracking signals based on the smoothed error or autocorrelation deteriorates badly as the smoothing parameter is increased. This surprising problem with the smoothed-error signal is also discussed by Brown (1982a), who recommends the simple cusum over the smoothed error.

Unlike the more complex cusum schemes discussed below, the performance of the simple cusum is also independent of the variance of the time series. If the variance of the series changes, this is estimated implicitly by MAD_t. Both SUM_t and MAD_t will change proportionately, leaving the ARL performance unchanged. This conclusion was confirmed by simulation tests.

The simple cusum also has some disadvantages. As Trigg (1964) and numerous others have pointed out, the simple cusum never forgets large errors. To see the effect of this memory, suppose that an isolated random error moves the signal from near-zero to a point close to the control limit. From then on, suppose that the forecast errors are small for a long period of time. The signal will wiggle back and forth around the level of the isolated error. Perhaps much later, a second isolated error in the same direction as the first can finally trip the signal, although the system is in perfect control.

Another disadvantage is ironic. If the forecasting system starts to give exceptionally accurate forecasts, the signal may trip. Suppose that near-perfect forecasts begin to occur. This will cause the MAD to tend to zero, leaving the cusum unchanged. Thus C_t will tend to infinity.

The final disadvantage may also be ironic. The discussion above is concerned only with exponential smoothing, since other tracking signals give better results when the forecast errors are independent. This point is illustrated in the section on performance comparisons below.

Modifications to the simple cusum

The simple cusum signal can be used with a smoothed estimate of the standard deviation of the forecast errors in the denominator rather than the MAD. Brown (1982a) recommends this procedure, since the standard deviation of the errors is needed to compute safety stocks in inventory control applications. The standard deviation is often estimated by 1.25MAD, but this is correct only for normally distributed errors (as in this research). In applications where an estimate of the standard deviation of the errors is needed, the safest course is to replace equation (3) by

$$\text{MSE}_t = \alpha e_t^2 + (1-\alpha)\text{MSE}_{t-1} \tag{7}$$

The square root of (7) can then be used in place of MAD_t in (4). Control limits in Table 1 and Figure 1 can be adjusted by multiplying each limit by 1.25.

For time series with a long history, the variance of the forecast errors can be estimated and used as a fixed quantity in the denominator of the simple cusum signal. If the variance remains unchanged, this procedure improves the performance of the signal. ARLs to detect any level of bias above 1.0σ are reduced by 1–2 periods, while maintaining the same ARL on unbiased errors. Control limits to use the simple cusum in this fashion are given in the Appendix. The limits are biased on the expected value of MAD as a fixed quantity in the signal.

Whether this procedure is advisable for any given time series is a difficult question. If the true variance exceeds the estimate, ARLs to detect bias generally increase (the entire response curve shifts upward). If true variance is less than the estimate, the number of false trips increases, since the entire response curve shifts downward.

BACKWARD CUSUM CONTROL SYSTEMS

More elaborate cusum control systems based on industrial quality control theory have been proposed by numerous researchers. Examples include the work of Barnard (1959), Brown (1971), Coutie *et al.* (1966), Ewan (1963), Ewan and Kemp (1960), Harrison and Davies (1963), Van Dobben De Bruyn (1968), and Woodward and Goldsmith (1964). The control system developed by Harrison and Davies appears to be the most practical for routine forecasting applications, since it is simpler and requires less data storage than the other systems. The Harrison and Davies system is outlined and extended below.

Basic calculations

The theory of backward cusums is based on the following idea. If we could guess when a past change in the time series occurred, then the sum of all the errors since the change would be the best tracking signal available today. If a change in the series occurred in period 50, and the current period is number 55, then the sum of the errors from periods 50–55 would be more sensitive to the change than the sum from periods 1–55.

Since we have no way of knowing the number of past periods to sum in advance, it appears to be necessary to maintain a battery of all possible cusums. The first is just the last error. The second is the sum of the last two errors, and so on to the beginning of the series. Each of these sums should be the most sensitive signal available to detect a change in the time series that occurred i periods ago, where i is the number of periods used in each cusum.

The number of cusums required to implement this idea quickly gets out of hand in any time series, but the basic idea can be modified to a more practical scheme. One modification is to maintain only the last six to twelve cusums since information on older changes is not likely to be of much value at present.

An example of a six period backward cusum control system is shown in Table 2. Control limits are set up in linear form, as

$$L_i = \sigma w(i + h). \tag{8}$$

L_i is the limit (+ or −) on the ith backward cusum, σ is the standard deviation of forecast errors, computed during a period when the system is in control, and w and h are constants selected by simulation. In this case, $\sigma = 10$, $w = 1$, and $h = 2$. Linear limits are used to allow for larger random cusums as more errors are summed.

Table 2. Example of the backward cusum method: $L_i = \sigma w(i + h) = 10(i + 2)$

Period	Forecast error	S_1	S_2	S_3	S_4	S_5	S_6
1	−10	−10					
2	20	20	10				
3	15	15	35	25			
4	5	5	20	40	30		
5	−25	−25	−20	−5	15	5	
6	−25	−25	−50*	−45	−30	−10	−20
Control Limits: L_1 to L_6		±30	±40	±50	±60	±70	±80

The cusums are computed backward in time as follows:

$$S_1 = e_t$$

$$S_2 = e_t + e_{t-1}$$

$$S_3 = e_t + e_{t-1} + e_{t-2}, \text{ etc.}$$

In the example, the sixth error causes S_2 to break its lower control limit, signalling apparent negative bias in forecast errors.

Implicit tests for bias

The storage requirements for six cusums and their control limits would still be burdensome in most forecasting systems. To reduce the storage requirements, Harrison and Davies (1964) devised a system whereby all cusums can be implicitly tested against their control limits by storing only four quantities. Proof of this result may be found in Harrison and Davies (1964) or Coutie *et al.* (1966). The proof given by Harrison and Davies contains an error in notation, so Coutie's is recommended.

Two of the four quantities stored are constants: σw and $L_0 = \sigma w h$. Since the limits increase by σw each period, these two quantities are sufficient to compute each L_i. The other two quantities are moving parameters. The parameter used to check for positive bias is the minimum of the differences between the positive control limits and their corresponding cusums, $L_i - S_i$, from the last period. The other parameter, used to check for negative bias, is the maximum of the differences between the negative control limits and their cusums, $-L_i - S_i$.

An example shows that testing these two differences each period is equivalent to testing all cusums. The test for positive bias uses:

$$D_t^+ = \text{MIN}[D_{t-1}^+, L_0] + \sigma w - e_t \tag{9}$$

D_0^+ is set equal to L_0. If $D_t^+ < 0$, lack of control is signalled. The test for negative bias is:

$$D_t^- = \text{MAX}[D_{t-1}^-, -L_0] - \sigma w - e_t \tag{10}$$

D_0^- is set equal to $-L_0$. If $D_t^- > 0$, lack of control is signalled.

Table 3. The implicit test for bias: $L_0 = 20$, $\sigma w = 10$

Period	Forecast error	D_t^+	D_t^-
1	−10	40	−20
2	20	10	−50
3	15	5	−45
4	5	10	−35
5	−25	45	−5
6	−25	55	10*

Table 3 repeats the example in Table 2 using these tests. The control system is started with $D_0^+ = 20$ and $D_0^- = -20$. The equations below demonstrate that each value of D may be traced back to Table 2 as the minimum or maximum difference between the control limits and their corresponding cusums:

Minimum difference	*Maximum difference*
$D_1^+ = L_1 - S_1$	$D_1^- = -L_1 - S_1$
$D_2^+ = L_1 - S_1$	$D_2^- = -L_1 - S_1$ or $-L_2 - S_2$
$D_3^+ = L_2 - S_2$	$D_3^- = -L_1 - S_1$
$D_4^+ = L_3 - S_3$	$D_4^- = -L_1 - S_1$
$D_5^+ = L_4 - S_4$	$D_5^- = -L_1 - S_1$
$D_6^+ = L_1 - S_1$	$D_6^- = -L_2 - S_2$

Control limits for the backward cusum system

Finding control limits (w and h) for the backward cusum system is a tedious process. The only published control limits for forecast errors that are based on a large sample size, with a run-in period to wash out initial conditions, are those given by Golder and Settle (1976). Unfortunately, Golder and Settle's control limits were not very useful in this research. Their limits do not yield even multiples of ARLs on unbiased errors, making it difficult to compare the backward cusum to other tracking signals. For reasons explained in their paper, Golder and Settle show different sets of control limits to detect different levels of bias, making it impossible to study the effects of varying the w and h parameters. Their results were replicated in this research, however, and no significant differences in ARLs were found.

Table 4. Backward cusum response comparisons, exponential smoothing, $\alpha = 0.1$

Control limits (w, h)	ARL: unbiased errors	ARL to detect bias	
		1.5σ	3.0σ
0.1, 52.4	50	3.9	1.9
0.2, 21.2	50	3.7	1.8
0.3, 11.7	50	3.7	1.7
0.4, 7.6	50	4.0	1.6

Table 4 illustrates the search process used to find control limits in this research. The w parameter was varied by increments of 0.1, starting at 0.1, since Golder and Settle's control limits indicate that the best w values are usually fractional. At each value of w, the h parameter was varied to find the limits yielding an ARL = 50 periods on unbiased errors. Using those w and h values, the ARLs to detect 1.5σ and 3.0σ step changes in the mean were measured. This procedure was continued until the ARL at 1.5σ passed through a minimum. If there was a tie for the best limits at 1.5σ, the tie was broken with the minimum ARL at 3.0σ. Given the best w value using these criteria, h was varied to find control limits for ARL = 25 and 100 periods on unbiased errors.

The final control limits are given in Table 5 for independent errors and for exponential smoothing at $\alpha = 0.1$ and 0.2. The procedure used to generate bias in the $N(0, 1)$ errors was similar to that used in other quality control studies (Goldsmith and Whitfield, 1961). The mean error was shifted from zero in increments of 0.5, while holding the variance constant.

Although choosing the control limits in this fashion was somewhat arbitrary, it seemed to be the best way to deal with the complexities of the response rates. Bias equal to 1.5σ is about the smallest level that any tracking signal can detect in a reasonable number of periods. Since the ARL at 3.0σ was always two periods or less for virtually any set of limits, it seemed best to minimize the ARL at 1.5σ rather than 3.0σ. In view of the prodigious amount of computer time necessary to conduct these searches, no attempt was made to estimate ARLs at other levels of bias or at higher α values.

Given the σ of the forecast errors, σw and $L_0 = \sigma w h$ can be computed directly from Table 5. If

Table 5. Backward cusum control limits and average run lengths

Type of errors	α	Control limits (w, h)	ARL: unbiased errors	ARL to detect bias	
				1.5σ	3.0σ
Independent $N(0, 1)$	—	0.6, 3.4	25	2.8	1.3
		0.6, 4.2	50	3.2	1.5
		0.6, 5.1	100	3.8	1.7
Exponential smoothing	0.1	0.3, 9.7	25	2.9	1.4
		0.3, 11.7	50	3.7	1.7
		0.3, 13.9	100	5.0	1.9
	0.2	0.1, 37.9	25	3.1	1.5
		0.1, 44.1	50	4.0	1.7
		0.1, 50.0	100	5.3	2.0

the limits in Table 5 are compared to Golder and Settle's limits, it should be emphasized that they are expressed differently. Golder and Settle show σw and h for exponential smoothing of $N(0, 1)$ time series. One must divide σw by the expected value of σ to get w suitable for use with any other time series.

Advantages and disadvantages of the backward cusum system

Although the backward cusum is the most thorough tracking signal available, it has not been widely used in practice for a number of reasons. The first reason is the lack of published control limits. The second is the lack of comparisons to other tracking signals. Third, although the backward cusum system can be operated by storing four quantities, the other tracking signals considered in this research require storage of only two quantities. The additional storage requirements for the backward cusum are a considerable disadvantage in large forecasting systems. Fourth, the backward cusum requires the assumption of a constant variance in the time series. This is a disadvantage in setting up a control system for a time series with a short history. Also, as discussed above for the simple cusum, the performance of the backward cusum can be erratic if the true variance of the series differs significantly from the estimated variance.

THE SMOOTHED-ERROR TRACKING SIGNAL

The smoothed-error tracking signal is widely used in practice. This signal was developed by Trigg (1964) in an attempt to overcome the disadvantages of the long memory of the simple cusum. The updating equations each period are

$$E_t = \alpha e_t + (1 - \alpha)E_{t-1} \tag{11}$$

$$\text{MAD}_t = \alpha|e_t| + (1 - \alpha)\text{MAD}_{t-1} \tag{3}$$

$$T_t = |E_t/\text{MAD}_t| \tag{12}$$

Simplifications

These equations are standard in the literature and in practice, but they can be simplified by recognizing that exponential smoothing recurrence relations are basically approximations to exact discounted averages (Gilchrist, 1967, 1976). The exact discounted form for (11) is

$$E_t = \sum_{i=0}^{t-1} \beta^i e_{t-i} \Big/ \sum_{i=0}^{t-1} \beta^i \tag{13}$$

The exact form for (3) is

$$\text{MAD}_t = \sum_{i=0}^{t-1} \beta^i |e_{t-i}| \Big/ \sum_{i=0}^{t-1} \beta^i \tag{14}$$

β is the discount factor, restricted to $0 \leq \beta \leq 1$.

After some uninteresting algebra, these forms reduce to new recurrence relations:

$$E_t = e_t + \beta E_{t-1} \tag{15}$$

$$\text{MAD}_t = |e_t| + \beta \text{MAD}_{t-1} \tag{16}$$

$$T_t = |E_t/\text{MAD}_t| \tag{12}$$

Surprisingly enough, the exact discounted averages turn out to be simpler than the exponential smoothing approach. The value of T_t is exactly the same using either exponential smoothing or discounted averages, provided that the same starting values are used and $\beta = 1 - \alpha$.

Control limits for the smoothed-error signal

A great deal of conflicting analytical and simulation work has been done on the probability distribution of T_t. Trigg's original distribution (1964) is strictly applicable only to forecasting models other than exponential smoothing, since he ignores the autocorrelation in errors caused by exponential smoothing. Brown (1967) took this autocorrelation into account in developing distributions for exponential smoothing. Batty (1969), working independently of Brown, developed somewhat different distributions for exponential smoothing. The tables given by Montgomery and Johnson (1976) extended Brown's work.

The simulated distributions in this research generally agree with Trigg for independent errors and Batty for exponential smoothing. Differences are small enough to attribute to chance. Good correspondence was obtained with Brown's results at $\alpha = 0.1$, but there were significant differences at higher α values. Table 6 gives probability distributions computed like those for the simple cusum. The distribution for independent errors is given only for $\alpha = 0.1$. No advantage was found for higher α values. The distributions for exponential smoothing set α in the forecasting model equal to $\alpha = 1 - \beta$ in the tracking signal, since there was no advantage for doing otherwise. Response curves and control limits for $\alpha = 0.1$ with exponential smoothing are shown in Figure 2.

Table 6. Cumulative distribution of the smoothed-error tracking signal

Cumulative probability	T_t			
	$N(0, 1)$ errors	Exponential smoothing		
	$\alpha = 0.1$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$
0.80	0.37	0.28	0.40	0.51
0.90	0.47	0.36	0.49	0.61
0.95	0.54	0.42	0.57	0.69
0.96	0.56	0.44	0.59	0.71
0.97	0.59	0.46	0.61	0.73
0.98	0.62	0.50	0.65	0.77
0.99	0.67	0.55	0.69	0.81

Advantages and disadvantages of the smoothed-error signal

Figure 3 shows why the smoothed-error signal is not recommended for higher α values. Three response curves are plotted in Figure 3, originating at ARL = 50 periods on unbiased errors. The curves correspond to $\alpha = 0.1$, 0.2, and 0.3. The ARLs at $\alpha = 0.2$ are significantly larger than at $\alpha = 0.1$ for levels of bias up to 2.0σ. The curve for $\alpha = 0.3$ is useless for error detection.

The reasons for deterioration in performance are complex. Increasing α causes the smoothed error to increase faster than the smoothed MAD. This effect can be seen most clearly in Table 6, where the distribution of T_t expands as α increases. The result is that the tracking signal finds it harder to distinguish between bias and purely random fluctuations in the time series. Control limits

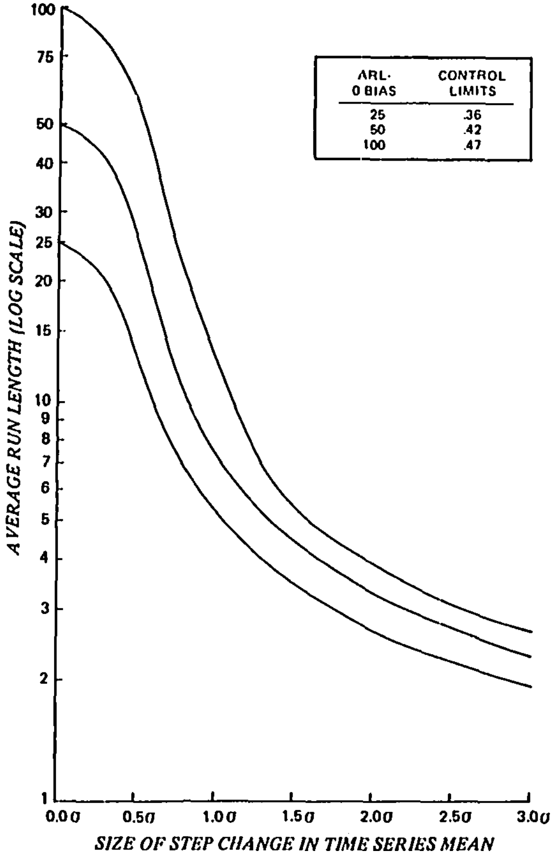


Figure 2. Smoothed-error signal ARL performance (exponential smoothing, $\alpha = 1 - \beta = 0.1$)

must be expanded to allow for larger random fluctuations, but this delays the reaction to bias. These conclusions support Brown (1982a).

Response curves are not given for independent errors. Although the smoothed-error signal works well on independent errors, the autocorrelation signal is a better choice. Like the simple cusum, the performance of the smoothed-error signal based on the smoothed MAD is independent of the variance of the time series. Control limits to use the expected value of MAD as a fixed quantity in the smoothed-error signal are listed in the Appendix.

THE AUTOCORRELATION TRACKING SIGNAL

The easiest autocorrelation pattern to track is the first-order autoregression on successive errors, $e_t = \phi e_{t-1}$, where ϕ is the true autoregressive parameter. Letting r be the estimate of ϕ, ordinary

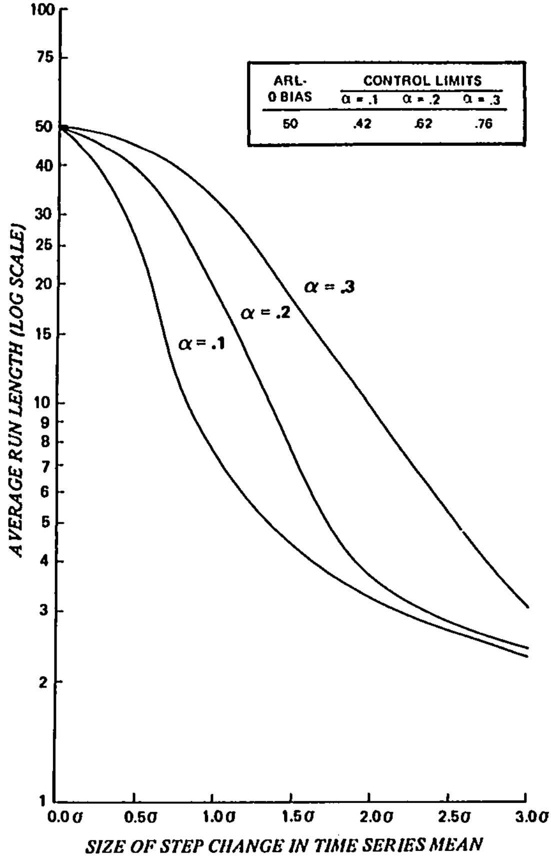


Figure 3. Effects of increasing α on the smoothed-error signal

least squares gives

$$r = \frac{\sum e_t e_{t-1}}{\sum e_{t-1}^2} \tag{17}$$

The sums in (17) are equally weighted. To turn r into a tracking signal, the sums can be discounted with β. Following a similar development to that for the discounted version of T_t, the discounted least squares estimate of ϕ is

$$\text{COV}_t = e_t e_{t-1} + \beta \text{COV}_{t-1} \tag{18}$$

$$\text{MSE}_t = e_{t-1}^2 + \beta \text{MSE}_{t-1} \tag{19}$$

$$r_t = \text{COV}_t / \text{MSE}_t \tag{20}$$

Table 7. Cumulative distribution of the autocorrelation tracking signal: $N(0, 1)$ errors, $\alpha = 1 - \beta = 0.1$.

Cumulative probability	r_t
0.80	0.17
0.90	0.27
0.95	0.35
0.96	0.37
0.97	0.39
0.98	0.43
0.99	0.48

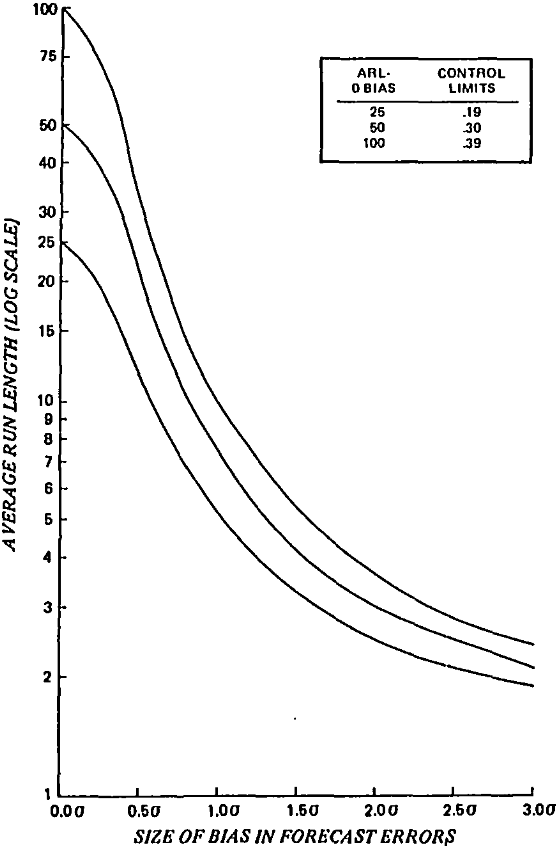


Figure 4. Autocorrelation signal ARL performance [$N(0, 1)$ errors, $\alpha = 1 - \beta = 0.1$]

Control limits for the autocorrelation signal

Table 7 lists the simulated probability distribution of r_t on independent errors. The distribution is symmetric about zero. Thus control limits can be set for both positive and negative autocorrelation or for positive autocorrelation only, which corresponds to bias. Response curves to detect bias in independent errors are shown in Figure 4 for $\alpha = 1 - \beta = 0.1$. Again, there was no advantage for higher α values. Control limits based on the expected value of MSE as a fixed quantity are given in the Appendix. If the MSE is smoothed, r_t is independent of the variance of the series.

Advantages and disadvantages of the autocorrelation signal

The autocorrelation signal is the only signal which can be used to track both positive and negative autocorrelation in forecast errors. This ability is useful in fitting adaptive filtering (Makridakis and Wheelwright, 1978) or adaptive estimation procedure [AEP] (Carbone and Longini, 1977) models to historical data. The parameters in these forecasting models are usually trained on historical data until the MSE is minimized or the change in MSE between training cycles is less than some minimum. If the autocorrelation signal is used to track the errors during each training cycle, it provides an alternative criterion for stopping the training. The auto correlation signal gives a good indication of the fit of the model to the last part of the time series. This may be a better criterion for stopping the training than the MSE or autocorrelation computed with equal weights over all the data.

The autocorrelation signal is not recommended for exponential smoothing models because of the autocorrelation induced by the forecasting process. r_t spends most of its time below zero on exponential smoothing errors, uselessly chasing negative autocorrelation. If r_t is negative when a step change in the time series occurs, it will lag behind other signals in sounding the alarm unless the step is quite large.

Another disadvantage of the autocorrelation signal applies to independent forecast errors as well as errors from exponential smoothing. The product $e_t e_{t-1}$ is in the numerator of r_t. Suppose e_t is the first in a run of large biased errors and e_{t-1} has an opposite algebraic sign. r_t will not begin to track the positive autocorrelation until the second error with the same sign has been observed.

PERFORMANCE COMPARISONS

All tracking signals were compared at control limits yielding ARLs of 25, 50 and 100 periods on unbiased errors, both independent and from simple exponential smoothing. Comparisons were made using both smoothed and fixed values of the variance (MAD, MSE, or σ) of forecast errors. Parametric tests for differences between means were used to test the ARLs at bias equal to 1.5 and 3.0σ. A simple non-parameteric test, the sign test (Siegel, 1956) on the direction of differences between matched pairs of the ARLs, was also used. Both the parametric and the non-parametric tests gave the same conclusions. It is unnecessary to review all the comparisons, since the ranking of each signal was the same at all the ARLs on unbiased errors. Comparisons are given in Table 8 only for an ARL = 50 periods on unbiased errors.

All differences are statistically significant at the 0.01 level in Table 7 except among those marked with an asterisk. The signals are identified using the notation above, except that B indicates the backward cusum. Control limits not available in the tables above may be found in the Appendix.

In category I of the table, the autocorrelation signal is the best choice if the errors are independent and tracking signals must rely on a smoothed variance. If all signals use the expected value of the variance of errors, the backward cusum is superior. Notice the effects of using the

Table 8. Performance comparisons at ARL = 50 periods on unbiased errors ($n = 1000$ time series)

	Type of errors	Variance	Signal[1]	ARL to detect bias[2] 1.5σ	3.0σ
I.	Independent $N(0, 1)$	Smoothed	r	4.2	2.1
			T	4.8	2.9
			C	8.6	6.6
II.		Expected value	r	3.7	2.0*
			T	3.9	2.0*
			C	5.6	3.1
			B	3.2	1.5
III.	Exponential smoothing $\alpha = 0.1$	Smoothed	r	11.5	2·2*
			T	4.4	2.3*
			C	4.9	3.0
IV.		Expected value	r	7.7	2.0*
			T	3.6**	1.8**
			C	3.9	2.0*
			B	3.7**	1.7**
V.	Exponential smoothing $\alpha = 0.2$	Smoothed	r	26.0	5.1
			T	7.7	2.4
			C	4.9	3.3
VI.		Expected value	r	17.5	2.1
			T	5.5	1.6*
			C	3.5	1.8*
			B	4.0	1.7*

[1] Signal definitions: r = Autocorrelation Signal; T = Smoothed Error; C = Simple Cusum; B = Backward Cusum.
[2] All ARL differences are statistically significant at the 0.01 level, except among those marked with a *.

expected value of variance. The performance of the smoothed-error signal improves to a point where it is identical to the autocorrelation signal at 3.0σ. Although the difference between the autocorrelation and smoothed-error signals is only 0.2 periods at 1.5σ, this is still significant due to the sample size.

Categories III and IV compare the signals for exponential smoothing at $\alpha = 0.1$. The smoothed-error signal is the best choice with a smoothed variance. With the expected value of variance, there is no difference between the smoothed error and backward cusum at either level of bias. The asterisks should be interpreted to mean that the simple cusum and the autocorrelation signal are tied at 3.0σ, and that both significantly exceed the smoothed-error and backward cusum.

Categories V and VI show the effects of increasing α to 0.2. The simple cusum is almost three periods better than the smoothed-error at 1.5σ with a smoothed variance. But the smoothed-error is almost one period better at 3.0σ. The large difference at 1.5σ would probably be more important to most users, making the simple cusum the best choice. The difference at 1.5σ increases as α increases. Although not shown, with $\alpha = 0.3$ the simple cusum is thirteen periods better at 1.5σ.

Using the expected value of variance at $\alpha = 0.2$, the simple cusum is superior at 1.5σ. There is no significant difference among the three contenders at 3.0σ. The advantage of the simple cusum at 1.5σ is illogical, since the simple cusum is a special case of the backward cusum. The difference may be due to chance or the procedure used to find control limits. More precise limits would probably reduce the difference.

It is interesting to contrast these results to the only previous research involving tracking signal comparisons, by Golder and Settle. They compared the backward cusum to the smoothed-error signal for exponential smoothing. They concluded that the backward cusum should produce smaller ARLs at all levels of bias than the smoothed-error signal. However, they used a fixed expected value of variance in the backward cusum, while using the smoothed variance in the smoothed-error signal. This is an unfair comparison. If both signals are based on the expected value of variance, there is no difference at $\alpha = 0.1$ at any level of bias.

CONCLUSIONS

On the basis of ARL performance, the most convincing conclusion from this research is that backward cusums do not appear to be worth the additional complexity and storage to monitor exponential smoothing models. Identical performance can be had with such simpler tracking signals. The simpler tracking signals can also be used without being forced to assume that the variance of the time series is constant.

Less convincing is the conclusion that the smoothed-error signal is better than the simple cusum at $\alpha = 0.1$. Regardless of the statistical significance of the comparisons, the differences in ARLs are still less than one period. The stability of the simple cusum over all smoothing parameters is an important advantage. This stability may persuade some forecasters to use the simple cusum at $\alpha = 0.1$.

At higher α values, the performance of the smoothed-error signal deteriorates badly. At $\alpha = 0.2$, the smoothed-error signal is at best a marginal choice. At $\alpha = 0.3$, the smoothed-error signal is useless. These results are surprising, since the smoothed-error signal is widely used in practice. Users of the smoothed-error signal at higher α values should consider switching to the simple cusum.

If the time series has a constant variance, backward cusums should be the best choice for monitoring independent forecast errors. However, the advantage of the backward cusum over the autocorrelation signal is only about 1/2 period at all levels of bias. Furthermore, the autocorrelation signal is simpler and can be used without assuming a constant variance.

IMPLEMENTATION SUGGESTIONS

The simple cusum, smoothed-error, and autocorrelation signals are cheap and easy to use. The only implementation problem is in determining starting values. Numerators should logically be started at zero; denominators must never be started at zero. Estimates of the MAD or MSE, however rough, should be used to start denominators. Otherwise, simulation results show that the signals misbehave, giving numerous false trips until denominators build up to a reasonable value. This observation supports Montgomery and Johnson (1976).

Once the tracking signals get under way and a trip occurs, numerators should be reset to zero after adjusting the forecasting model. Unless numerators are reset, the signals will tend to stay in a tripped position. It seems best to leave denominators alone after a trip and allow the smoothing or discounting process to wash out any bias.

FURTHER RESEARCH

A follow-on study is planned to evaluate several other quality control models for forecasting, including the parabolic mask control system developed by Brown (1971, 1982a). Although this system requires storage of at least eight data points (previous simple cusum values), it uses a more powerful statistical test for bias than the simpler tracking signals. Whether the additional storage and complexity will pay off in improved ARL performance is an important question.

ACKNOWLEDGEMENTS

This research was supported by the Burroughs Wellcome Company, Research Triangle Park, North Carolina, and by Golden Gate University, San Francisco, California. A portion of this paper was presented at the Second International Symposium on Forecasting, Istanbul, Turkey, July, 1982. Helpful suggestions on an earlier draft of this paper were provided by Douglas Blazer of the University of North Carolina at Chapel Hill.

During the refereeing process, R. G. Brown of Materials Management Systems, Inc., Norwich, Vermont, provided several private papers covering the same ground as part of this research (simulation comparisons among the simple cusum, smoothed-error, and autocorrelation signals, all with a smoothed variance estimate). Although Brown used a different experimental design, his conclusions agree with this research. This was a great help in validating the simulation models described above.

Finally, thanks are due to Mary Ann Gardner of the Stanwick Corporation, Norfolk, Virginia, who prepared the tables and graphs for this paper.

APPENDIX: SIMULATION DETAILS

The random number generator used was the 'RANDN' function, a standard multiplicative congruential generator supplied with the UNIVAC 1100 series computers. This generator has given excellent results on benchmark tests for independence and uniformity of random numbers. Documentation for this generator is available from Sperry UNIVAC, Systems Publications Department, Post Office Box 500, Blue Bell, Pennsylvania 19422.

The probability distributions were tabulated using 1000 time series of $N(0, 1)$ deviates. Each series was 120 periods in length. All signals were started with numerators equal to zero and denominators equal to expected values. The first forecast (for period 1) was set equal to zero for the exponential smoothing distributions. Tracking signal values were recorded at the end of periods 21–120 and summed across all series. The random number generator was reseeded every 100 time series or 12,000 observations. Seeds were obtained from the 'IRAND' random integer generator, also available on the UNIVAC 1100 series.

Average run lengths were computed using the same type of series, except each was 500 periods in length. Starting values and the run-in period were the same as above. A constant bias was added to periods 21–500 of each series. Run lengths were computed as the period number when the signal first exceeded its control limit minus 20, giving a minimum run of 1 period and a maximum of 480. Some signals occasionally ran for more than 400 periods on unbiased errors (at an ARL = 100), although none ran up to the maximum. The random number generator was reseeded every 20 series or 10,000 observations.

The control limits for each tracking signal were searched until ARLs on unbiased errors within 1/2 period of 25, 50, and 100 periods were found. This was a conservative policy for stopping the search. Comparisons showed that the ARL's on unbiased errors could vary by up to 3 periods with little effect on the ARL's to detect bias.

Table 9. Control limits based on expected values of MAD or MSE

Type of errors	α	ARL: unbiased errors	Control limits for		
			r	T	C
Independent $N(0, 1)$	0.1	25	0.17	0.45	7.2
		50	0.29	0.54	9.5
		100	0.40	0.62	13.0
Exponential smoothing	0.1	25	—	0.37	4.5
		50	0.38	0.43	5.3
		100	—	0.49	6.1
	0.2	25	—	0.61	3.5
		50	0.22	0.69	4.1
		100	—	0.78	4.7

The run-in length of 20 periods was selected in order to replicate Golder and Settle's results. Samples using a run-in of 30 periods produced no significant differences in ARL's.

The expected values needed to start each tracking signal or to use a fixed variance in the signals can be computed from Batty's work (1969). The expected values of MAD for exponential smoothing of $N(0, 1)$ deviates are 0.818, 0.841 and 0.865 for $\alpha = 0.1$, 0.2 and 0.3, respectively. Expected values of the standard deviation of forecast errors are 1.026, 1.054, and 1.085. Control limits based on expected values are given in Table 9.

REFERENCES

Barnard, G. A., 'Control charts and stochastic processes', *Journal of the Royal Statistical Society*, **21** (1959), 239–71.

Batty, M., 'Monitoring an exponential smoothing forecasting system', *Operational Research Quarterly*, **20** (1969), 319–325.

Brown, Robert G., *Statistical Forecasting for Inventory Control*, New York: McGraw-Hill, 1959.

Brown, Robert G., *Smoothing, Forecasting and Prediction of Discrete Time Series*, Englewood Cliffs, N.J.: Prentice-Hall, 1963.

Brown, Robert G., *Decision Rules for Inventory Management*, New York: Holt, Rinehart, and Winston, 1967.

Brown, Robert G., 'Detection of turning points in a time series', *Decision Sciences*, **2** (1971), 383–403.

Brown, Robert G., *Advanced Service Parts Inventory Control*, Norwich, Vermont: Materials Management Systems, Inc., 1982a.

Carbone, Robert and Longini, R. L., 'A feedback model for automated real estate assessment', *Management Science*, **24** (1977), 241–248.

Coutie, G. A., *et al.*, *Short-Term Forecasting*, Edinburgh: Oliver and Boyd, 1966.

Ewan, W. D. and Kemp, K. W., 'Sampling inspection of continuous processes with no autocorrelation between successive results', *Biometrika*, **47** (1960), 363–380.

Ewan, W. D., 'When and how to use cu-sum charts', *Technometrics*, **5** (1963), 1–22.

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

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

Gilchrist, Warren G., 'Methods of estimation using discounting', *Journal of the Royal Statistical Society* (B), **29** (1967), 355–369.

Gilchrist, Warren G., *Stastical Forecasting*, London: Wiley, 1976.

Harrison, P. J. and Davies, O. L. 'The use of cumulative sum (CUSUM) techniques for the control of routine forecasts of product demand', *Operations Research*, 12 (1963), 325–333.
Makridakis, S. and Wheelwright, S. C., *Forecasting: Methods and Applications*, Santa Barbara, California: Wiley/Hamilton, 1978.
Montgomery, Douglas C. and Johnson, Lynwood A., *Forecasting and Time Series Analysis*, New York: McGraw-Hill, 1976.
Page, E. S., 'Continuous inspection schemes', *Biometrika*, 41 (1954), 100–115.
Siegel, Sidney, *Nonparametric Statistics*, New York: McGraw-Hill, 1956.
Trigg, D. W., 'Monitoring a forecasting system', *Operational Research Quarterly*, 15 (1964), 271–274.
Van Dobben De Bruyn, C. S., *Cumulative Sum Tests: Theory and Practice*, London: Griffin, 1968.
Woodward, R. H. and Goldsmith, P. L., *Cumulative Sum Techniques*, London: Oliver and Boyd, 1964.

Author's biography:
Everette S. Gardner, Jr. is a Commander in the U. S. Navy and a lecturer in management science for Golden Gate University. He has experience in forecasting and inventory control in several military physical distribution systems. He holds the Ph.D. in Business Administration (Production and Operations Management) from the University of North Carolina at Chapel Hill.

Author's address:
Everette S. Gardner, Jr., 6348 Dartmouth Way, Virginia Beach, VA 23464, U.S.A.

REFEREES' COMMENTS

Editor's note: When this paper was sent for review, we asked the referees if they would be willing to be listed as referees for the paper and, in addition, would they be willing to have comments from their review published along with the paper. Five referees were contacted. Four referees completed the review by the time deadline. All rated the paper highly, all were willing to list their names, and all were agreeable to having some of their comments published. Everette Gardner was able to use many of the referees' comments to revise the paper. From the remaining comments I have listed some that may be helpful to the readers in assessing this study. The referees' comments are presented here in alphabetical order. (JSA)

Everett Adam, Jr., School of Business, University of Missouri, Columbia, Missouri, 65211, U.S.A.

The author correctly judges the importance of the issue of changing forecasting model parameters once underlying demand conditions change. Automatic monitoring of forecast errors is an approach to this issue and is most useful in computerized forecasting systems where the technology of forecasting is well understood by management and operations research specialists. There are, however, many simple forecasting applications when forecast errors will not be automatically monitored—a point the author fails to recognize. This research has important implications for such users as they can be encouraged to use the simplicity of a cumulative sum error measure, a measure that could be readily understood by most users regardless of their forecasting and statistical sophistication.

The strength of this paper lies in (1) the relationship of this research to previous studies in formulating the experimental design and (2) the attention given to proper simulation design. The author carefully stated the main research hypotheses: the evaluation of a variety of automatic forecasting monitoring schemes for detecting biased forecasting errors. The independent variables of interest are forecasting tracking signals, shown in Table 8. Various dependent variables are used, but most primary results are interpreted using the average run length (ARL), such as in Table 8.

The details of the computer simulation—as well as side issues of a technical nature—can easily detract the reader from the main hypotheses: selection of the 'better' tracking signal. The author is to be commended in the reporting of such detail for those interested in technical issues and replication. However, this also detracts from the main purpose of the manuscript. The author does pause for interpretation—as in comparison to Golder and Settle's simulation and discussion of advantages and disadvantages of each tracking signal—a nice feature of the paper.

This paper is well thought out in design, properly conducted and clearly written. Statistical analysis is appropriate for the study and is correctly applied. The results favour simplicity in automatic monitoring of forecast errors—results that are intuitively appealing to managers.

Robert G. Brown, Materials Management Systems, P. O. Box 332, Norwich, Vt. 05055, U.S.A.

The paper is similar to work that I did a few years ago, but never polished for publication, so I am happy to see these results. I enclose a few non-published papers to forward to the author and to supplement some of his work, which is quite a bit more thorough than mine was.

The importance of fast detection is that it takes some lead time to react to a detection. If it takes a long time to detect a change, that delays the response.

One problem of coupling the forecast revisions with the tracking signal is that when there is a step or ramp change in the data, the forecasting process will increase the standard deviation quite rapidly.

Saleha B. Khumawala, College of Business, U. of Houston, Houston, Texas 77004, U.S.A.

The objective of the paper is to evaluate tracking signals and the author has accomplished this objective by doing simulation. In doing so, he has done a good survey of the literature.

Carlos A. Valenzuela, Management Sciences Section, Air Products and Chemicals, Inc., P.O. Box 538, Allentown, Pa. 18105, U.S.A.

The paper met the standard of replicability. We managed to replicate part of the simulation work to compare two of the tracking signals analysed by the author, i.e. the simple cusum and smoothed-error tracking signals for values of alpha in the range 0.1 to 0.3 in the case of exponential smoothing and using the expected value of the variance. Our results were in close agreement with those of the author. As a practitioner, I found this paper to be highly valuable.