J. Opl Res. Soc. Vol. 36, No. 1, pp. 43–47, 1985
Printed in Great Britain.
0160-5682/85 $3.00 + 0.00

CUSUM vs Smoothed-Error Forecast Monitoring Schemes: Some Simulation Results

EVERETTE S. GARDNER JR
U.S. Navy

This paper compares the performance of CUSUM and smoothed-error tracking signals for monitoring the adequacy of exponential smoothing forecasts. Previous research has favoured the CUSUM. However, there is some evidence that the performance of the smoothed-error signal can be improved by a simple modification in its application: the use of different smoothing parameters in the tracking signal and the forecasting model. The effects of this modification are tested using simulated time series. We conclude that the CUSUM is robust to the choice of forecasting parameter, while the smoothed-error signal is not. The CUSUM is also more responsive to small changes in the time series, regardless of the parameters used.

Key words: tracking signals, monitoring forecasts, CUSUM, exponential smoothing, simulation

INTRODUCTION

The aim of forecast monitoring is to detect biased errors as quickly as possible. This is an important consideration in large inventory systems where single exponential smoothing is widely used to generate forecasts. Single smoothing is simple and economical but will lag trends or changes in the level of the time series. If an increase in demand goes undetected, customer service will deteriorate. If a decrease goes undetected, excess stocks will build up. Further discussion of the need for monitoring the forecasts from exponential smoothing models may be found in Brown[1] or Montgomery and Johnson.[2]

This paper compares the performance of two forecast monitoring schemes or tracking signals for the single smoothing model. The single smoothing forecasts are generated by:

$$\hat{X}_t(1) = \alpha X_t + (1 - \alpha)\hat{X}_{t-1}(1). \tag{1}$$

α is the smoothing parameter, usually restricted in practice to the range 0 to 1.

The first monitoring scheme is based on the simple cumulative sum of the one-step-ahead errors, $e_t = X_t - \hat{X}_{t-1}(1)$. The CUSUM should fluctuate around zero if the forecasts are unbiased. Any significant departure from zero may indicate bias. The CUSUM signal, C_t, is revised after each observation of the series as follows:

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

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

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

The second scheme keeps watch over the smoothed error, which should also fluctuate around zero. The smoothed error is revised using:

$$E_t = \alpha e_t + (1 - \alpha)E_{t-1}. \tag{5}$$

Next we use equation (3) to revise the mean absolute deviation (MAD). The tracking signal is:

$$T_t = |E_t/\mathrm{MAD}_t|. \tag{6}$$

Gardner[3] compared the performance of these signals using the average run length (ARL) criterion—the number of time periods needed to detect a step change in the level of the series. Gardner's comparisons favoured the CUSUM. In these comparisons and most other research on forecast monitoring[4–6] equal α values were used in the forecasting model and the tracking signal.

However, McKenzie[7] showed that the ARL performance of the smoothed-error signal can be improved by allowing the α values used in the signal and the forecasting model to differ. McKenzie argued that α in the tracking signal should generally be smaller than that in the forecasting model. This reduces the variance of the signal and makes it easier to detect changes in the time series.

This paper compares the ARL performance of the CUSUM and smoothed-error signals, using simulated time series, with different α values in the signals and forecasting model. Conclusions are offered on the relative advantages of the two signals. To enable the research to be replicated, an appendix on computational details is included.

EXPERIMENTAL DESIGN

The signals were compared by simulating 1,000 time series, each with a constant level and variance. The initial value of the sum of errors and the smoothed-error were set equal to zero. The initial forecast and the MAD were set equal to expected values. At period 21, a step increase in level was added to each series, and the ARL to detect the step was measured. Step sizes ranged from 0.0 to 3.0σ in multiples of 0.5.

Control limits on each signal were selected to yield an ARL of 100 periods at a zero step-size. This run length may be interpreted as the average run between false alarms from each tracking signal. Other ARLs at a zero step were also measured but are not reported here because the relative differences between signals were about the same. Another way of interpreting the control limits is that they correspond to probabilities in the range of 0.01 to 0.03 than an exception report from the tracking signal is in fact a false alarm (a Type I error). Probability distributions of the signals can be found in Gardner[3] and McKenzie.[7].

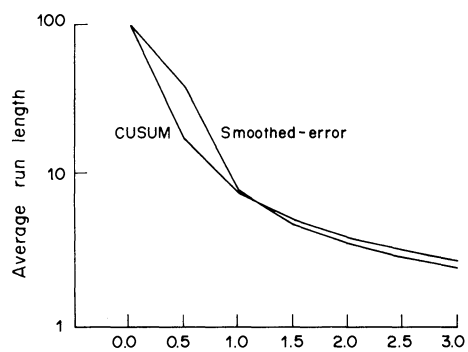
We shall use $\alpha(E)$ to refer to the parameter used to smooth the errors in each tracking signal. $\alpha(F)$ refers to the smoothing parameter used in the forecasting model. The parameters tested were $\alpha(E) = 0.05$, 0.10, coupled with $\alpha(F) = 0.1$, 0.2, 0.3.

COMPARISONS

Comparisons are shown in Figures 1–6, where step sizes are plotted on the horizontal axis and ARLs on the vertical (log scale). Owing to the large sample size, most ARL differences in the figures are statistically significant at the 0.01 level.

Figures 1 and 2 show the results of $\alpha(F) = 0.1$. If we smooth the errors at $\alpha(E) = 0.05$, the CUSUM is more responsive to a step of 0.5σ, while there is little difference between the signals beyond that point. When $\alpha(E)$ is increased to 0.1, the CUSUM is better up to a 1.0σ step, while the smoothed-error is better (about one period) for larger steps.

Figures 3 and 4 show the results when $\alpha(F)$ is increased to 0.2. The CUSUM has a large advantage at steps up to 1.5σ, with little practical difference between the signals beyond that point. Again, increasing $\alpha(E)$ from 0.05 to 0.1 increases the advantage of the CUSUM at small step-sizes.



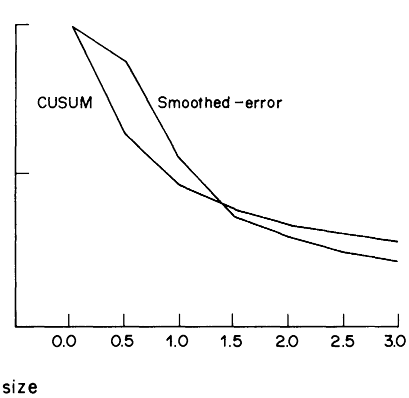


FIG. 1. *ARLs at* $\alpha(E) = 0.05$, $\alpha(F) = 0.10$.

FIG. 2. *ARLs at* $\alpha(E) = 0.10$, $\alpha(F) = 0.10$.

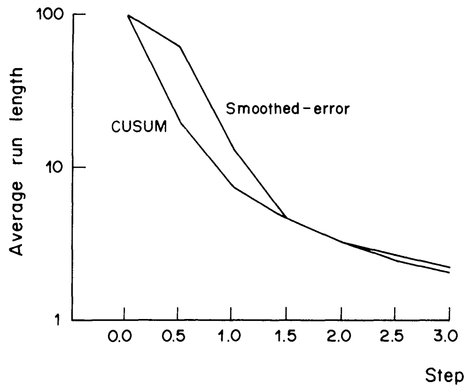


FIG. 3. *ARLs at* $\alpha(E) = 0.05$, $\alpha(F) = 0.20$.

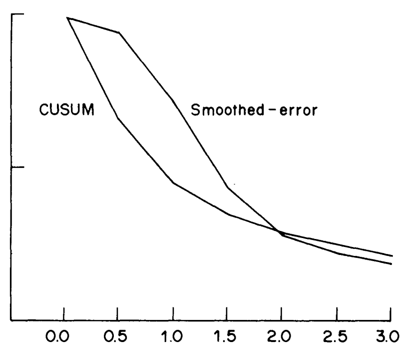


FIG. 4. *ARLs at* $\alpha(E) = 0.10$, $\alpha(F) = 0.20$.

Finally, Figures 5 and 6 show that increasing $\alpha(F)$ to 0.3 simply magnifies the comparisons at 0.2. For example, with $\alpha(E) = 0.1$, and $\alpha(F) = 0.3$, the CUSUM reacts an average of 10 periods faster to a 1.5σ step.

CONCLUSIONS

Smoothing with $\alpha(E) < \alpha(F)$ improves the performance of the smoothed-error tracking signal, compared to the results reported in previous research.[3,5] However, increasing either parameter makes it difficult for the signal to detect relatively small changes in the time series. The signal is reliable only when step changes are quite large, in the range of 3σ or greater.

The behaviour of the smoothed-error signal can be explained by the fact that it is only temporarily affected by a step change. The process of smoothing the errors often causes the signal to reset itself before an exception report can be issued. Usually, the signal's failure to report a step does not mean that the forecasts have caught up to the data. Ignoring the effects of noise, the impulse response function of single smoothing (see Brown[1]) demonstrates that the forecasts never reach the new level of the series after a step change in finite time.

The CUSUM, in contrast, is robust to parameter choice. There is little practical difference in ARL performance between different combinations of $\alpha(E)$ and $\alpha(F)$. The CUSUM also appears to be more sensitive to small changes in the time series. This can be explained by the fact that the CUSUM has a 'long memory'. That is, a step change in the time series results in a permanent change in the expected value of the sum of the errors. This is true, regardless of the size of the step or the smoothing parameters.

These findings have implications for forecast monitoring in inventory systems. For fast-moving or expensive inventory items, off-line action is required when there is a permanent change in

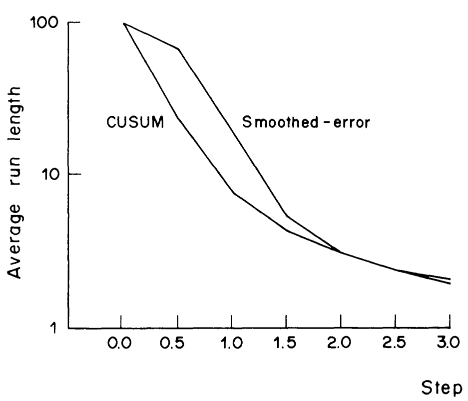


FIG. 5. *ARLs at* $\alpha(E) = 0.05$, $\alpha(F) = 0.30$.

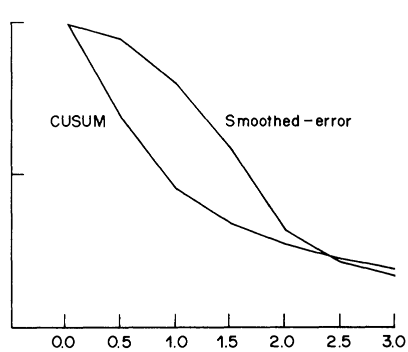


FIG. 6. *ARLs at* $\alpha(E) = 0.10$, $\alpha(F) = 0.30$.

demand. If demand goes up, new orders must be placed on a priority basis, while orders currently outstanding should be expedited into stock. Important customers must be notified if shipments will be delayed. If demand goes down, any unneeded orders should be cancelled to prevent excess inventory investment. These actions are necessary even for small changes in demand. To improve the chances of getting exception reports on small changes in demand, the CUSUM signal is recommended.

For slow-moving or inexpensive inventory items, tracking signal reports may be undesirable unless there has been a substantial change in demand. The work involved in reviewing these reports may not be justified by improvements in customer service. The smoothed-error tracking signal is recommended for these items. The signal is insensitive to small changes but does give timely reports when a large change occurs.

Both signals will have to be manually reset from time to time to remove the effects of outliers in the errors. Because of its long memory, the CUSUM is likely to need more frequent resets. This should not be a practical problem if use of the CUSUM is restricted to the most important segment of the inventory.

FURTHER RESEARCH

More research is needed on a problem not considered here, that of starting up the tracking signal in a volatile time-series or when the history of the series is limited. It is surely dangerous to smooth the errors with a very small α value unless a good initial estimate of the MAD is available. Just how good the initial estimate must be is not clear. Although little has been done to quantify the effects of the initial MAD, the following general consequences are obvious. So long as the true MAD exceeds the estimate, the ARLs to detect bias must increase for any given control limit. So long as the true MAD is less than the estimate, the number of false alarms must increase.

A poor initial MAD coupled with a small α value can have more serious consequences in inventory systems. The MAD computed in the tracking signal is often used to establish safety stocks. If the true MAD exceeds the estimate, safety stocks will be inadequate to achieve any given target level of customer service. If the true MAD is less than the estimate, safety stocks will be excessive.

APPENDIX

Simulation Details

Table 1 gives ARLs, standard errors and control limits for the tracking signals. Note that most run lengths at a zero step are not exactly 100 periods. The search routine used to find the control limits was stopped when a run length in the range 99 to 101 was obtained.

A 'run-in' period (the time from period 1 until the detection mechanism is turned on) of 20 observations was used. The same run-in was used in the previous work by Gardner[3] and Golder and Settle,[5] although McKenzie[7] used a run-in of 60 observations. It does not appear that the run-in period is an important factor in experimental design. The effects of the run-in were tested as shown in Table 2, which gives ARLs at run-ins of 20, 40, 60, 80 and 100 observations, step sizes of 0.0, 1.5 and 3.0σ, and parameter settings of $\alpha(E)$ and $\alpha(F) = 0.1$. The ARLs at a zero step vary by no more than one period as the run-in is increased, while the ARLs at 1.5 and 3.0σ vary by no more than 0.1 periods.

Another consideration in experimental design is that a tracking signal may exceed its control limit when the detection mechanism is turned on (at the end of the run-in). Thus an immediate report will occur, regardless of the size of the disturbance in the time series.

To assess the distortion caused by this problem, all runs in Table 1 were replicated, except that immediate reports were excluded from the statistics. The statistics were virtually unchanged. As should be expected from the probability distributions of the signals, the number of immediate reports was quite small, ranging from 10 to 26 time series in each sample of 1,000.

TABLE 1. *ARL statistics*

		Step	ARL		Std error		Control limit	
$\alpha(E)$	$\alpha(F)$	size	CUSUM	Sm. err.	CUSUM	Sm. err.	CUSUM	Sm. err.
0.05	0.10	0.0	100.5	100.4	3.10	3.09	6.185	0.266
		0.5	17.8	38.9	0.42	2.09		
		1.0	7.8	8.0	0.13	0.16		
		1.5	5.2	4.8	0.08	0.07		
		2.0	4.0	3.6	0.05	0.05		
		2.5	3.3	2.9	0.04	0.04		
		3.0	2.8	2.5	0.04	0.03		
0.10	0.10	0.0	100.0	99.5	3.07	3.14	6.325	0.466
		0.5	18.1	56.9	0.42	2.60		
		1.0	8.4	12.7	0.13	0.86		
		1.5	5.9	5.3	0.08	0.09		
		2.0	4.7	3.9	0.06	0.05		
		2.5	4.1	3.1	0.05	0.04		
		3.0	3.6	2.7	0.04	0.03		
0.05	0.20	0.0	100.2	99.7	3.14	3.11	4.775	0.220
		0.5	20.0	61.1	0.52	2.81		
		1.0	7.5	13.6	0.14	1.05		
		1.5	4.6	4.7	0.07	0.08		
		2.0	3.3	3.3	0.05	0.05		
		2.5	2.7	2.5	0.03	0.04		
		3.0	2.3	2.1	0.03	0.03		
0.10	0.20	0.0	100.7	100.0	3.02	3.15	4.830	0.405
		0.5	20.2	76.0	0.53	3.08		
		1.0	8.0	27.1	0.15	1.88		
		1.5	5.0	7.5	0.07	0.68		
		2.0	3.8	3.7	0.05	0.07		
		2.5	3.2	2.8	0.04	0.04		
		3.0	2.7	2.4	0.03	0.03		
0.05	0.30	0.0	100.0	100.8	3.05	3.02	4.045	0.192
		0.5	23.9	67.3	0.70	2.71		
		1.0	7.9	19.1	0.17	1.41		
		1.5	4.4	5.4	0.08	0.33		
		2.0	3.1	3.1	0.05	0.05		
		2.5	2.4	2.4	0.03	0.03		
		3.0	2.1	2.0	0.03	0.02		
0.10	0.30	0.0	99.6	99.8	3.07	3.15	4.075	0.362
		0.5	23.7	78.8	0.71	3.02		
		1.0	8.1	40.6	0.18	2.42		
		1.5	4.8	14.7	0.08	1.41		
		2.0	3.5	4.3	0.05	0.39		
		2.5	2.8	2.7	0.04	0.04		
		3.0	2.4	2.2	0.03	0.03		

TABLE 2. *Effects of the run-in period*: $\alpha(E) = 0.1$; $\alpha(F) = 0.1$

Run-in	CUSUM ARL*			Sm. err. ARL*		
period	0.0σ	1.5σ	3.0σ	0.0σ	1.5σ	3.0σ
20	100.0	5.9	3.6	99.5	5.3	2.7
40	100.3	5.9	3.6	99.0	5.2	2.7
60	100.4	5.8	3.6	98.6	5.3	2.6
80	99.4	5.9	3.6	99.1	5.3	2.7
100	100.9	5.8	3.6	100.1	5.2	2.7

*Control limits for each run: CUSUM 6.325; Sm. err. 0.466.

Acknowledgement—This research was supported by the Navy Regional Data Automation Center, Norfolk, Virginia.

REFERENCES

1. R. G. BROWN (1963) *Smoothing, Forecasting, and Prediction of Discrete Time Series.* Prentice–Hall, Englewood Cliffs, New Jersey.
2. D. C. MONTGOMERY and L. A. JOHNSON (1976) *Forecasting and Time Series Analysis.* McGraw–Hill, New York.
3. E. S. GARDNER (1983) Automatic monitoring of forecast errors. *J. Forecasting* **2,** 1–21.
4. M. BATTY (1969) Monitoring an exponential smoothing forecasting system. *Opl Res. Q.* **20,** 319–325.
5. E. R. GOLDER and J. G. SETTLE (1976) Monitoring schemes in short-term forecasting. *Opl Res. Q.* **27,** 489–501.
6. D. W. TRIGG (1964) Monitoring a forecasting system. *Opl Res. Q.* **15,** 272–274.
7. E. MCKENZIE (1978) The monitoring of exponentially weighted forecasts. *J. Opl Res. Soc.* **29,** 449–458.