Journal of the Operations Research
Society of Japan
Vol. 35, No. 2, June 1992

A DESIGN OF ADAPTIVE EXPONENTIAL SMOOTHING USING A CHANGE DETECTION STATISTIC

Duk B. Jun
Korea Advanced Institute of Science and Technology

(Received March 12, 1990; Revised February 4, 1991)

Abstract An adaptive exponential smoothing method is proposed by use of the change detection statistic for random level change in exponential smoothing, where the mean, the variance, and the limiting distribution of the statistic are derived. The comparison results with Trigg and Leach's and the exponential smoothing method by simulation show that the proposed method gives the smaller mean square error than the exponential smoothing for almost all cases, while it performs better than Trigg and Leach's for frequently occurred large level changes.

1. Introduction

Exponential smoothing is one of the simplest and most well known, forecasting methods having been successfully applied for many years. There have been several articles concerning the method (Brown[2], [3]), the theory behind the method (Box and Jenkins[1], Harvey[9]), and the application of the method(Brown[3], Box and Jenkins[1]).

It has been long known, however, that one of the deficiencies of the method is its inability to respond quickly to interventions, to interruptions, or to large changes in level of the underlying process. To overcome such problems, there have been several methods proposing that the proportionality constant used in exponential smoothing depend on forecast error or recent realizations of the process (Van Dobben de Bruyn[4], Chow[5], Roberts and Reed[13], Jun and Oliver[11] and others). One such method and probably the most successful to date is to use a tracking signal defined as the absolute value of the ratio of exponentially smoothed errors to the exponentially smoothed absolute errors (Trigg and Leach[15]).

In this paper a new adaptive proportionality constant is proposed to monitor exponential smoothing in case of repeated occurrences of random level change, using the test statistic derived in the previous paper of Jun[9] to detect an occurrence of random level change in application of the exponential smoothing method. The statistic is the quadratic function of the one-step-ahead forecast errors obtained by exponential smoothing, where the same discount factor used in the exponential smoothing method is applied to the forecast errors. The proposed proportionality constant is based on the ratio of the change detection statistic using the exponential smoothing forecast errors to those using the absolute values of the errors. A simulation comparing this method with both Trigg and Leach's[15] method and with the original exponential smoothing method shows the suggested method performs better than Trigg and Leach's as the random level changes occur more frequently and the size of the change becomes larger.

2. Change Detection Statistic

The well known exponential smoothing forecast updating equation is given by

$$f_t = f_{t-1} + (1-\alpha)e_{t-1}; \; e_{t-1} = Z_t - f_{t-1} \tag{1}$$

where the one-step-ahead forecast, f_{t-1}, for the random variable, Z_t, made at time $t-1$ and e_{t-1} defines the corresponding one-step-ahead forecast error. According to the updating equation, the exponential smoothing forecast geometrically discounts the previous observations by their ages, i.e.

$$f_t = (1-\alpha)\sum_{i=0}^{\infty} \alpha^i Z_{t-i} \tag{2}$$

Jun[9] formulates an occurrence of random level change in the underlying process following the simple dynamic linear model of Harrison and Stevens[8] by inclusion of a dummy variable, Δ, representing a random level change. The equations of motion for the random observation, Z_t, the unknown random level, L_t, a major level change, Δ, and the random noise terms, a_t, b_t, are given by

$$Z_{t+1} = L_{t+1} + a_{t+1}, \quad t = 0, 1, \cdots, n-1 \tag{3a}$$

$$L_{t+1} = L_t + b_{t+1}, \quad t \neq M \tag{3b}$$

$$= L_t + \Delta + b_{t+1}, \quad t = M \tag{3c}$$

$\{a_t\}$ and $\{b_t\}$ are assumed to be serially uncorrelated Gaussian inputs with mean 0 and known variances σ_a^2 and σ_b^2, respectively. L_0, Δ, $\{a_t\}$, $\{b_t\}$ and the random variable of the change point, M, are also assumed mutually independent.

Harrison[7] shows that the updating equation (1) gives optimal forecasts, meaning minimum mean square error, in the above model when there is no major level change i.e. $\Delta = 0$. In Jun[10] a test statistic is derived to detect a major level change, Δ, occurring at unknown time M, after n observations for any given prior distributions of L_0, Δ and M. Assuming the diffuse priors of Δ and M, it is also shown that the test statistic to detect an abrupt level change in exponential smoothing becomes

$$S_n = e_{n-1}^2 + \frac{(e_{n-2} + \alpha e_{n-1})^2}{1+\alpha^2} + \cdots + \frac{(e_1 + \alpha e_2 + \cdots + \alpha^{n-2} e_{n-1})^2}{1 + \alpha^2 + \cdots + \alpha^{2n-4}}, \tag{4}$$

where $\{e_t\}$ are the one-step-ahead forecast errors from the exponential smoothing method.

Exponential smoothing supplies accurate forecasts with uncorrelated small forecast error when random levels fluctuate within the usual limits of the perturbation noise. In this case, the change detection statistic of equation (4) becomes small. When a large level change occurs, however, the exponential smoothing forecasts take time to adapt to the new level causing large forecast errors with a run after the change. For example, suppose that a large level change occurs between Z_t and Z_{t+1}. Then e_t, the first forecast error following the change becomes very large and the next forecast errors decrease exponentially until the exponential smoothing forecasts adapt to the new level. Thus the quadratic term including $e_t + \alpha e_{t+1} + \cdots + \alpha^{n-t-1}e_{n-1}$ in equation (4) greatly affects S_n, allowing for detection of the change. Note that the same discount factor used in the exponential smoothing is applied to the forecast errors and each term in the change detection statistic represents the log-likelihood ratio of an occurrence of level change to no change, i.e. $\Delta \neq 0$ v.s. $\Delta = 0$ in equation (3) at every possible change point.

The null distribution of the statistic S_n in the above equation is not simple, except that $S_2 = e_1^2$ with the chi-square distribution. The exact mean and variance of the statistic, however, can be calculated by equation (4). The derivation and the proof of the limiting distribution following the normality are shown in Appendix I. In addition, S_n can be transformed using a linear combination of *i.i.d.* chi-square random variables. The critical values of the null distribution of the statistic might be calculated by numerical integration as in Sen and Srivastava[14].

3. Adaptive Exponential Smoothing

The use of the data-independent proportionality constant in the forecast updating equation (1) of exponential smoothing gives appropriate forecasts when the random levels fluctuate within the usual limits. However, once an unusual large level change occurs, it can hardly adapt to the new level.

Thus the construction of a truly adaptive proportionality constant using the change detection statistic is proposed as follows:

$$f_t = f_{t-1} + \kappa_t(Z_t - f_{t-1}) \quad (5a)$$

$$\kappa_t = \frac{S_t}{T_t} \quad (5b)$$

$$S_t = e_{t-1}^2 + \frac{(e_{t-2} + \alpha e_{t-1})^2}{1+\alpha^2} + \cdots + \frac{(e_1 + \alpha e_2 + \cdots + \alpha^{t-2}e_{t-1})^2}{1+\alpha^2+\cdots+\alpha^{2t-4}} \quad (5c)$$

$$T_t = e_{t-1}^2 + \frac{(|e_{t-2}| + \alpha|e_{t-1}|)^2}{1+\alpha^2} + \cdots + \frac{(|e_1| + \alpha|e_2| + \cdots + \alpha^{t-2}|e_{t-1}|)^2}{1+\alpha^2+\cdots+\alpha^{2t-4}}, \quad (5d)$$

where $\{e_t\}$ are the one-step-ahead forecast errors obtained by the original exponential smoothing. The adaptive proportionality constant consists of the ratio of the change detection statistic, S_t, the change detection statistic, has the same form as in equation (4). T_t has the same functional form as S_t but uses the absolute values of one-step-ahead forecast errors obtained by equation (1). The adaptive proportionality constant, κ_t, is the ratio of S_t to T_t and has a value between zero and one. When the exponential smoothing produces accurate forecasts with uncorrelated small forecast errors, κ_t becomes small so that the next forecast has a value close to the current forecast. On the other hand, once a large level change has occurred, the exponential smoothing needs enough time to adapt to the new level. Thus $\{e_t\}$ after the change become unusually large with a run until the exponential smoothing forecasts increase to the new level. The subsequent large value of the change detection statistic makes κ_t large and the next forecast reflects much of the current forecast error.

4. Comparison with Trigg and Leach

Trigg and Leach[15] recognized the need for a truly adaptive proportionality constant which depended not only on the time but also on the values of observations as they occurred. The algorithm can be summarized as follows. Construct a tracking signal based on the absolute ratio of exponentially smoothed one-step-ahead forecast errors to exponentially smoothed absolute values of one-step-ahead forecast errors. That is,

$$f_t = f_{t-1} + \kappa_t(Z_t - f_{t-1}) \tag{6a}$$

$$\kappa_t = \frac{|P_t|}{|Q_t|} \tag{6b}$$

$$P_t = (1-\xi)(Z_t - f_{t-1}) + \xi P_{t-1}, \quad 0 < \xi < 1, \quad P_0 \text{ given} \tag{6c}$$

$$Q_t = (1-\xi) \mid Z_t - f_{t-1} \mid + \xi Q_{t-1}, \quad Q_0 \text{ given} \tag{6c}$$

The parameter κ_t is continuously adjusted and is data-dependent. Although the underlying theory of this model is not well understood, it has successfully stood the test of time.

5. Simulation Results

For purposes of simulation, we modified Series A of Box and Jenkins. Each value of Box and Jenkins Series A is multiplied by $\sqrt{5}$ making the variance of series in this paper 5 times that of the one given by Box and Jenkins. The series used in this paper consists of observations 1 through 100 of the original Series A.

Parameter estimation was based on a grid search using the first 60 observations (see Jun and Oliver[11]; Jun[10]). The estimates yielded $\bar{\alpha} = 0.775$ for exponential smoothing and $\bar{\xi} = 0.9, P_0 = Q_0 = 0.1$ as the best fit for Trigg and Leach's method. $f_{60} = 37.6$ was used for initial forecast for the last 40 observations (61 through 100) as used in Jun[10], Jun and Oliver[11].

The following modification of the last 40 observations allowed a comparison. First, change points were randomly chosen with equal probability from the 62nd to the 100th and the number of changes occurring in the 39 periods was given by an odd number from 1 to 9. Second, the different sizes of the level change were generated from Gaussian random variables with mean 0 and variances of 1, 5, 10, 15 and 20. These values were added to each observation from the change point to the end of the series.

For each set of values of the number of the changes and the change size, one hundred modified series were generated and forecasted by the proposed method, Trigg and Leach's and exponential smoothing. The average ratio of sum of squared one-step-ahead forecast errors from one hundred experiments for each case is given in tables 1 and 2. In table 1, the ratios of mean square error(MSE) from the exponential smoothing to that from the proposed method are presented. From table 1, we see that, with the exception of the case of one level change of the smallest size, the proposed method performs better than exponential smoothing. From table 2 we see the proposed method performs better than Trigg and Leach's method as level changes occur more frequently and the size of change increases.

The paired comparison one-sided test for mean difference of MSE obtained by the three methods showed statistically significant differences among the three methods for most cases. For each set of values of the number of change and the change size given, the differences of the MSE obtained by each method for each trial of 100 experiments were calculated and the sample mean and standard deviation of those values were obtained, as shown on table 3 and 4.

For illustration of the proposed method, the series of Jun[10] was used. This is the same series used for the above comparison except that it includes one level change with size of 1.5 at the 71st observation. Figure 1 shows the series and the one-step-ahead forecasts made by the proposed method and by Trigg and Leach's. The proposed method seems to supply forecasts more quickly adapting than Trigg and Leach's to the level change artificially made at the 71st observation and to the suspected level change at the 95th, and returning to the previous level after the suspected outlier occurred on the 64th observation.

Appendix I : Limiting distribution of S_n

For the simplicity of notation, we set $E(e_1^2) = 1$.

Rearranging terms in equation (4) yields

$$S_n = \sum_{k=1}^{n-1} \frac{(e_{n-k} + \alpha e_{n-k+1} + \cdots + \alpha^{k-1} e_{n-1})^2}{1 + \alpha^2 + \cdots \alpha^{2k-2}} \tag{1}$$

$$= \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} a_{ij}(\alpha) e_i e_j, \tag{2}$$

where

$$a_{ij}(\alpha) = \sum_{k=1}^{min(i,j)} \frac{\alpha^{i+j-2k}}{1 + \alpha^2 + \cdots + \alpha^{2n-2k-2}}. \tag{3}$$

Noting that

$$S_n = \sum_{i=1}^{n-1} a_{ii}(\alpha) e_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=1}^{i-1} a_{ij}(\alpha) e_i e_j$$

and, the uncorrelatedness of e_i^2 and $e_i e_j$ yields

$$E(S_n) = (n-1),$$

and

$$Var(S_n) = \sum_{i=1}^{n-1} (a_{ii}(\alpha))^2(\tau - 1) + 4\sum_{i=1}^{n-1}\sum_{j=1}^{i-1}\{a_{ij}(\alpha)\}^2, \tag{4}$$

where $\tau = E(e_1^4)$.

We can find $M < \infty$ such that

$$| a_{ij}(\alpha) - \alpha^{i-j} | \leq M(\alpha^{i+j} + \alpha^{2n+i-3j}) \quad , i \geq j. \tag{5}$$

From (5), we have

$$\sum_{i=1}^{n-1} a_{ij}(\alpha)e_i^2 = \sum_{i=1}^{n-1} e_i^2 + O_p(1)$$

and

$$\sum_{i=1}^{n-1}\sum_{j=1}^{i-1} a_{ij}(\alpha)e_ie_j = \sum_{i=1}^{n-1}\sum_{j=1}^{i-1} \alpha^{i-j}e_ie_j + O_p(1).$$

Therefore,

$$S_n = \sum_{i=1}^{n-1}\sum_{j=1}^{n-1} \alpha^{|i-j|}e_ie_j + O_p(1). \tag{6}$$

Derivation of limiting distribution of S_n

Observe that $S_n = \underline{e}_n' A_n \underline{e}_n$, where $\underline{e}_n = (e_1, \cdots, e_{n-1})'$ and A_n is an $(n-1)\times(n-1)$ matrix whose (i,j) element is $\alpha^{|i-j|}$. Letting $\underline{U}_n = A_n^{1/2}\underline{e}_n = (U_1, \cdots, U_{n-1})'$, where $A_n^{1/2}$ is an $(n-1)\times(n-1)$ matrix satisfying $A_n^{1/2}A_n^{1/2} = A_n$. The covariance matrix of $\underline{U}_n$ is A_n, which is the covariance matrix of an AR(1) process with correlation coefficient α and *i.i.d.* $(0, 1-\alpha^2)$ sequence of noise.

Therefore, there is a sequence of *i.i.d.* $N(0, 1-\alpha^2)$ sequence $\{\epsilon_i\}_{i=-\infty}^{\infty}$ of noise such that

$$U_n = \alpha U_{n-1} + \epsilon_n, \quad n = \cdots -2, -1, 0, 1, 2, \cdots.$$

Therefore,

$$n^{-1}S_n = n^{-1}\underline{U}_n'\underline{U}_n + O_p(n^{-1})$$

and the limiting distribution of $n^{-1}S_n$ is the same as the limiting distribution of unadjusted sample variance of a first order AR(1) process. Hence, under normality of errors $\{e_n\}$,

$$n^{-1/2}(S_n - (n-1)) \quad \Rightarrow \quad N(0, 2(1+\alpha^2)/(1-\alpha^2)).$$

See Fuller([6], p.254). Also from (4) and (5), we can show

$$\lim_{n\to\infty} \frac{Var(S_n)}{\{2n(1+\alpha^2)/(1-\alpha^2)\}} = 1.$$

Therefore,

$$\frac{S_n - (n-1)}{\sqrt{Var(S_n)}} \quad \Rightarrow \quad N(0,1).$$

Acknowledgement

This research has been supported by the Korea Science and Engineering Foundation.

References

[1] G. E. P. Box and G. M. Jenkins, *Time series analysis, forecasting and control*, Holden-Day, San Francisco, 1976.

[2] R. G. Brown, *Smoothing, Forecasting and Prediction of Discrete Time Series*, Prentice-Hall: New Jersey, 1962.

[3] R. G. Brown and R. F. Meyer, "The Fundamental Theorem of Exponential Smoothing", *Operations Research*, 9, 673-685, 1961.

[4] C. S. Van Dobben de Bruyn, "Prediction by Progressive Correction", *Journal of the Royal Statistical Society (B)*, 26, 113-122, 1964.

[5] W. M. Chow, "Adaptive Control of the Exponential Smoothing Constant", *AIIE. 16*, 314-317, 1965.

[6] W. A. Fuller, *Introduction to Statistical Time Series*, John Wiley & Sons, 1976.

[7] P. J. Harrison, "Exponential Smoothing and Short-Term Sales Forecasting", *Management Science(A)*, 13(11), 821-42, 1967.

[8] P. J. Harrison and C. F. Stevens, "Bayesian Forecasting", *Journal of Royal Statistical Society(B)*, 38(3), 205-47, 1976.

[9] A. C. Harvey, "A Unified View of Statistical Forecasting Procedures", *Journal of Forecasting*, 3, 245-275, 1984.

[10] D. B. Jun, "On Detecting and Estimating a Major Level or Slope Change in General Exponential Smoothing", *Journal of Forecasting*, 8(1), 55-64, 1989.

[11] D. B. Jun and R. M. Oliver, "An Adaptive Response Rate to Monitor Exponential Smoothing", *Operations Research Center Report*, UCB/ORC-85/3, University of California, Berkeley; *Asian-Pacific Journal of Operational Research*, 4(2), 112-121, 1987.

[12] R. M. Oliver, "Exponential Smoothing-I", *Operations Research Letters*, 3(3), 111-117, 1984.

[13] S. D. Roberts and R. Reed, Jr., "The Development of a Self-Adaptive Forecasting Technique," *AIIE, Trans.* 1, 314-322,1969.

[14] A. Sen and M. S. Srivastava, "On Tests for Detecting Change in Mean", *Annals of Statistics*, 3(1), 98-108, 1975.

[15] D. W. Trigg and A. C. Leach, "Exponential Smoothing with an Adaptive Response Rate", *Operations Research Quarterly*, 18, 53-59, 1967.

Duk Bin Jun
Dept. of Management Science
Korea Advanced Institute of Science and Technology
373-1 Kusong-dong Yusong-gu Taejon
305-701 Korea

Table 1: The Ratio of MSE from One Hundred Experiments by Exponential Smoothing to That by the Proposed Method

# of changes / change size	1	3	5	7	9
N(0,1)	0.910	0.997	1.038	1.131	1.142
N(0,5)	1.097	1.373	1.378	1.559	1.565
N(0,10)	1.161	1.533	1.672	1.691	1.734
N(0,15)	1.264	1.675	1.782	1.785	1.805
N(0,20)	1.530	1.885	1.962	1.880	1.942

Table 2: The Ratio of MSE from One Hundred Experiments by Trigg and Leach's to That by the Proposed Method

# of changes / change size	1	3	5	7	9
N(0,1)	0.896	0.933	0.984	0.997	1.023
N(0,5)	0.921	1.026	1.089	1.133	1.163
N(0,10)	0.943	1.056	1.135	1.182	1.204
N(0,15)	0.952	1.078	1.152	1.201	1.220
N(0,20)	0.969	1.115	1.172	1.199	1.250

Table 3: The Sample Mean and Standard Deviation of Each MSE by the Exponential Smoothing Minus the Corresponding MSE by the Proposed Method for One Hundred Experiments

# of changes / change size	1	3	5	7	9
N(0,1)	−2.3822* (2.5255)	−0.0280 (5.1650)	1.1161° (5.9357)	4.6092* (10.3457)	4.9681* (8.8117)
N(0,5)	3.2365* (10.7284)	17.2040* (23.3656)	21.6841* (24.1472)	39.1881* (41.5836)	42.6417* (38.2683)
N(0,10)	5.9402* (16.1254)	33.4479* (34.1657)	57.9092* (63.2040)	73.9488* (77.9015)	101.3740* (111.0934)
N(0,15)	11.0847* (24.1604)	53.6636* (51.3773)	89.8398* (94.8549)	114.4028* (116.2482)	155.6021* (165.6708)
N(0,20)	28.5673* (50.0816)	86.3778* (90.5187)	133.6199* (151.4268)	168.1504* (164.6463)	225.6369* (219.6084)

* denotes statistical difference at 1% significance level
° denotes statistical difference at 5% significance level
( ) denotes standard deviation

Table 4: The Sample Mean and Standard Deviation of Each MSE by Trigg and Leach's Minus the Corresponding MSE by the Proposed Method for One Hundred Experiments

# of changes / change size	1	3	5	7	9
N(0,1)	−2.8510* (1.4656)	−1.9283* (3.0673)	−0.5533° (3.1354)	−0.0937 (5.1973)	0.8149 (5.0564)
N(0,5)	−2.5294* (2.6912)	1.1993° (6.4699)	5.0797* (9.7060)	9.2847* (15.7323)	12.2914* (14.6952)
N(0,10)	−2.0290* (3.2957)	3.5489* (9.9033)	11.6159* (19.0098)	19.4891* (23.3178)	28.1838* (30.0127)
N(0,15)	−1.9401* (3.6995)	6.1369* (12.9862)	17.4680* (26.7668)	29.2695* (34.0890)	42.6142* (43.4612)
N(0,20)	−1.7426* (4.7433)	11.2205* (17.8283)	23.8374* (29.9112)	37.5854* (45.7776)	60.0328* (53.8075)

* denotes statistical difference at 1% significance level
° denotes statistical difference at 5% significance level
( ) denotes standard deviation

Figure 1. One-step-ahead forecasts by Trigg and Leach's
and the proposed method