OMEGA The Int. Jl of Mgmt Sci., Vol. 11, No. 6, pp. 621-625, 1983
Printed in Great Britain.
0305-0483/83 $3.00 + 0.00

Evolutionary Operation of the Exponential Smoothing Parameter: Revisited

INTRODUCTION

CHOOSING PARAMETERS for exponential smoothing models can be a difficult problem. Simulation with historical data (*ex ante* testing) is the most objective method of parameter selection. However, simulation may be infeasible in large forecasting systems. Simulation may also be of little value for time series with limited history or for nonstationary series.

Because of these problems, many practitioners choose parameters subjectively, considering the tradeoff between model stability and the response rate to changes in the time series. For example, Montgomery and Johnson [19] suggest a range of 0.1 to 0.3 for the parameter α in the single smoothing model. An α of 0.1 gives stable forecasts but a poor response to sudden changes in the level of the series. Increasing α to 0.3 makes the forecasts more sensitive to noise but gives a better response to changes in the series.

Recommended values for double smoothing are $\alpha = 0.1$ to 0.2 (see Brown [4]). Suggestions for other types of smoothing models can be found in Gardner and Dannenbring [12], McClain [17] and McClain and Thomas [18]. Although the validity of these 'standard' parameters is questionable (Chatfield [6]), they are widely used.

The problem of parameter selection can be avoided by using an adaptive smoothing model. These models are designed to improve performance by letting the smoothing parameter vary automatically, as a function of recent forecast accuracy. The parameter in an adaptive model should be moderate during periods of stability, but should increase in response to changes in the series.

There have been at least five empirical studies claiming accuracy advantages for adaptive smoothing models. Ekern [10, 11] reexamined four of these studies—by Bunn [5], Dennis [9], Hollier *et al.* [13] and Whybark [21]. Ekern concluded that these studies did not present any convincing evidence in favour of adaptive smoothing.

The fifth study in favor of adaptive smoothing was by Chow [7], who used the evolutionary operation methodology developed by Box [2] to control the smoothing parameter. Chow reported that the adaptive model was superior to a model with a fixed parameter on 68 of 69 time series. Chow's study has been cited by numerous authors. Examples of citations can be found in the forecasting texts by Armstrong [1], Makridakis and Wheelwright [15] and Montgomery and Johnson [19].

This paper reexamines Chow's study and points out an error in the model formulation for smoothing a linear trend. The error biased the performance comparisons in favor of the adaptive model.

In the next section, the smoothing formulations for a linear trend are reviewed. The following sections discuss Chow's model formulations, adaptive control system, and performance comparisons. Computational results are given using correct model formulations on four time series still available from the Chow study.

EXPONENTIAL SMOOTHING WITH A LINEAR TREND

If the time series has a relatively constant mean, single exponential smoothing is an appropriate forecasting model. The single-smoothed average is

$$S_t = \alpha X_t + (1 - \alpha)S_{t-1}, \quad (1)$$

where X_t is the time series value in t. The smoothing parameter is α, where $0 \leq \alpha \leq 1$. The forecast for $t + 1$ is simply S_t:

$$F_{t+1} = S_t. \quad (2)$$

If single smoothing is applied to a series containing a linear trend, the forecasts will lag the data. The generating process for a linear trend is

$$X_t = a + bt + \epsilon_t. \quad (3)$$

The forecasts from equations (1) and (2) will lag X_t by an expected value of

$$\left(\frac{1-\alpha}{\alpha}\right)b.$$

Proofs are given in Brown [3, 4].

There are several equivalent models which compensate for the lag in single smoothing and give unbiased forecasts. The first model was developed by Brown [3]:

$$S_t = \alpha X_t + (1 - \alpha)S_{t-1}, \quad (4)$$

$$T_t = \alpha(S_t - S_{t-1}) + (1 - \alpha)T_{t-1}, \quad (5)$$

$$F_{t+1} = S_t + \left(\frac{1-\alpha}{\alpha}\right)T_t + T_t. \quad (6)$$

The first two terms in equation (6) give a local estimate of the intercept of the trend line (the origin of time is shifted to the end of the current period, t):

$$\hat{a} = S_t + \left(\frac{1-\alpha}{\alpha}\right)T_t. \quad (7)$$

The last term in equation (6) is a local estimate of the slope of the trend line:

$$\hat{b} = T_t. \quad (8)$$

Brown's double smoothing [4] is equivalent to the 1959 model and is encountered more frequently in the literature. The double smoothing model is written:

$$S_t = \alpha X_t + (1 - \alpha)S_{t-1}, \quad (9)$$

$$S_t'' = \alpha S_t + (1 - \alpha)S_t'' - 1, \quad (10)$$

$$F_{t+1} = 2S_t - S_t'' + \frac{\alpha}{1-\alpha}(S_t - S_t''). \quad (11)$$

The local estimates of the slope and intercept are:

$$\hat{a} = 2S_t - S_t'', \quad (12)$$

$$\hat{b} = \frac{\alpha}{1-\alpha}(S_t - S_t''). \quad (13)$$

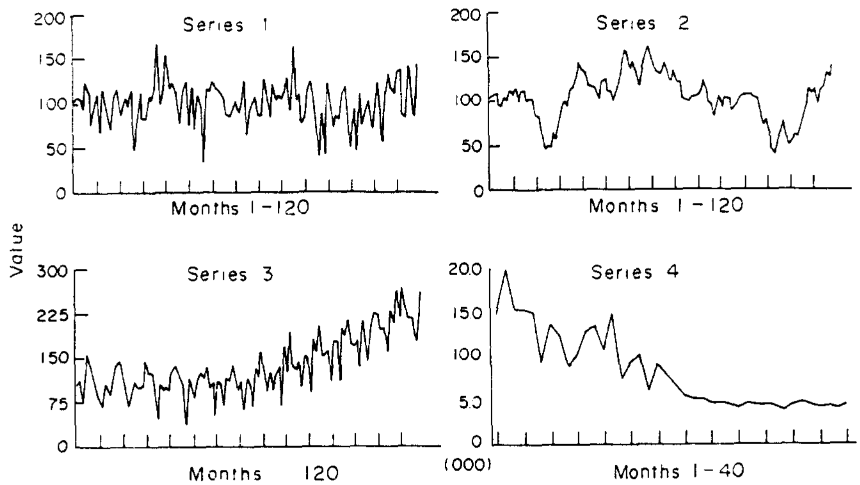

FIG. 1. *Time series used in replication.*

Brown's 1959 and 1963 models can be rearranged to yield another equivalent formulation, which is somewhat easier to use:

$$S_t = S_{t-1} + T_{t-1} + \alpha(2-\alpha)e_t, \quad (14)$$

$$T_t = T_{t-1} + \alpha^2 e_t, \quad (15)$$

$$F_{t+1} = S_t + T_t. \quad (16)$$

Other model formulations for smoothing a linear trend are reviewed in McClain and Thomas [18] and Montgomery and Johnson [19].

CHOW'S MODEL FORMULATIONS

Chow compared two versions of Brown's 1959 model—one with a fixed α and one with an adaptive α. However, he omitted the last term (T_t) from the forecast equation (6).

Evolutionary operation of the smoothing parameter

The evolutionary operation system used by Chow requires that three forecasts be computed each time period. The first is computed with a base or 'normal' value of α, and is the forecast used by management. The other forecasts are computed with $\alpha_H = \alpha + 0.05$ and $\alpha_L = \alpha - 0.05$. If the current absolute error using α is less than the absolute error for α_H and α_L, no change is made. If the absolute error from α_H or α_L is lower, α is reset to α_H or α_L. New values for α_H and α_L are computed around α and the process begins anew.

Chow recommends a starting base $\alpha = 0.10$. Thereafter, constraints are set on α such that $0.10 \le \alpha \le 0.90$. Thus the minimum $\alpha_L = 0.05$ and the maximum $\alpha_H = 0.95$.

The smoothed mean absolute deviation (MAD) of errors can be used instead of the current absolute error as the criterion for changing α. Chow found no difference in forecasting performance between the MAD and current error criteria.

Evaluation of Chow's performance comparisons

The adaptive α was compared to a fixed $\alpha = 0.1$ on 69 time series. The performance measure used appears to have been the mean squared error (MSE) (the paper discusses the 'mean forecasting variance'). The adaptive model was superior on 68 of 69 series—on the one exception, the difference was small.

These results seem too good to be accepted without serious questioning on several points. First, consider stable time series with a constant mean, which were among the series tested by Chow. An adaptive α should never do better than a moderate fixed α, since there is simply no change in the data for the model to follow.

Second, consider a series with a trend. Both of Chow's models are biased low in the presence of any trend, regardless of the smoothing parameter. The adaptive α, using Chow's incorrect formulation, will increase and offset some of this bias. But the correct fixed-parameter model will offset all of the bias.

Third, Chow's choice of parameters for the fixed α model might be improved. To get a better response to changing conditions in the series, one could choose to smooth with $\alpha = 0.2$ or more, rather than $\alpha = 0.1$.

Replication

Four time series remain from those tested by Chow. The four series are plotted in Fig. 1. The first three series are available in Brown ([3], Tables 1.1–1.3). The last series is from Chow ([7], Table 2).

Series 1 is random noise about a constant mean. Series 2 is highly autocorrelated, and therefore difficult to forecast with an exponential smoothing model. Series 3 follows a horizontal path for the first half of the series and then trends upward sharply. Series 4 fluctuates wildly for the first half of the series and then drops off rapidly to a constant level with little noise.

MSE results for these series are given in Table 1. The first two columns repeat Chow's results for the adaptive model based on the current absolute error and the fixed $\alpha = 0.1$ model. The last three columns list the MSE's for the correct model formulations. In the replication, all models were started under the most naive assumptions, with $S_0 = X_1 = F_1$, and $T_0 = 0$. These starting conditions should favor the adaptive model.

The adaptive model did worse in the replication than in Chow's study on every time series. This may be suprising except on series 1, where the extrapolation of any trend is unwarranted. The difficulty may be in starting conditions, the number of periods used to compute the MSE, or in the way Chow's model was interpreted in the replication. In some cases, the adaptive model indicated a shift to $\alpha = 0.05$ was desirable. This shift was not made (the MSE was computed using the

TABLE 1. MSE COMPARISONS

Series	Chow results		Replication—correct models		
	Adaptive[1]	Fixed $\alpha = 0.1$	Adaptive[1]	Fixed $\alpha = 0.1$	Fixed $\alpha = 0.2$
1	534	547	692	587	648
2	130	371	156	356	224
3	798	1068	979	851	975
4	3.9×10^6	8.9×10^6	4.4×10^6	4.8×10^6	4.1×10^6

[1]Adaptive model based on current absolute error.

errors from $\alpha = 0.1$), since α_L would become zero. This appears to be the way Chow used the model.

On Series 1, both fixed α models do better than the adaptive model, as should be expected. On Series 2, the adaptive model has a large advantage, again as should be expected. The adaptive α stays in the range of 0.7 to 0.9 during most of this series, indicating the need to consider an ARIMA model.

On Series 3, the fixed $\alpha = 0.1$ model has an advantage. This is due to the long period of stability during the first half of the time series. Both the adaptive α and the fixed $\alpha = 0.2$ models do better over the second half of the series, after the trend begins.

On Series 4, Notice that the use of the correct model for $\alpha = 0.1$ reduces the MSE by almost half. Smoothing with a fixed $\alpha = 0.2$ does better than the adaptive model.

CONCLUSIONS

It is impossible to generalize from only four time series. Furthermore, the results for the four series could change with more analysis. The performance of the fixed α models should improve with better starting values and more refined parameters. The adaptive model might be improved with the MAD rather than the current error as the criterion for changing α. However, the results with the four series do make it seem unlikely that the adaptive model would have the overwhelming advantages reported by Chow on the rest of his sample.

Three large empirical studies have found little difference between adaptive and fixed-parameter smoothing models—see Dancer and Gray [8], Makridakis and Hibon [14] and Makridakis *et al.* [16]. Considering this review of Chow's study and the reviews by Ekern [10, 11], we conclude that there is still no empirical evidence that adaptive models are more accurate than models with fixed parameters.

These comments should not be taken to mean that adaptive models are of no value. Considerations other than accuracy are important in most forecasting systems. As Chow points out, adaptive models may reduce the need for manual intervention in the forecasting system.

Adaptive models also eliminate the need to bother with parameter selection. It should be noted that Chow's adaptive method is easier to use than most such methods, since it requires no prior information about the time series. This is an important consideration for time series with limited history. The Trigg and Leach method requires some estimate of the MAD to start up the forecasting model, while the Whybark method requires an estimate of the standard deviation of the forecast errors.

REFERENCES

1. ARMSTRONG JS (1978) *Long-Range Forecasting.* Wiley, New York.
2. BOX GEP (1957) Evolutionary operation: A method for increasing industrial productivity. *Appl. Statist.* **6**, 3–23.
3. BROWN RG (1959) *Statistical Forecasting For Inventory Control.* McGraw-Hill, New York.
4. BROWN RG (1963) *Smoothing, Forecasting and Prediction of Discrete Time Series.* Prentice-Hall, Englewood Cliffs, NJ.
5. BUNN DW (1980) A comparison of several adaptive forecasting procedures. *Omega* **8**, 485–491.
6. CHATFIELD C (1978) The Holt-Winters forecasting procedure. *Appl. Statist.* **27**, 264–279.
7. CHOW WM (1965) Adaptive control of the exponential smoothing constant. *Jl Ind. Engng.* **16**, 314–317.
8. DANCER RE & GRAY CF (1972) An empirical evaluation of constant and adaptive computer forecasting models for inventory control. *Decis. Sci.* **8**, 228–238.
9. DENNIS JD (1978) A performance test of a run-based adaptive exponential forecasting technique. *Prodn Inventory Mgmt* **19**, 43–46.
10. EKERN S (1981) Adaptive exponential smoothing revisited. *Jl Opl Res. Soc.* **32**, 775–782.
11. EKERN S (1982) On simulation studies of adaptive forecasts. *Omega* **10**, 91–93.
12. GARDNER ES & DANNENBRING DG (1980) Forecasting with exponential smoothing: Some guidelines for model selection. *Decis Sci.* **11**, 370–383.
13. HOLLIER RH, KHIR M & STOREY RR (1981) A comparison of short-term adaptive forecasting methods *Omega* **9**, 96–98.
14. MAKRIDAKIS S & HIBON M (1979) Accuracy of forecasting: An empirical investigation. *Jl R. Statist. Soc. (A)* **142**, 97–145.
15. MAKRIDAKIS S & WHEELWRIGHT SC (1978) *Forecasting: Methods and Applications.* Wiley/Hamilton, Santa Barbara, CA.
16. MAKRIDAKIS S *et al.* (1982) The accuracy of extrapolation (time series) methods: results of a forecasting competition. *Jl Forecasting* **1**, 111–153.
17. MCCLAIN JO (1974) Dynamics of exponential smoothing with trend and seasonal terms. *Mgmt Sci.* **20**, 1300–1304.
18. MCCLAIN JO & THOMAS LJ (1973) Response-Variance tradeoffs in adaptive forecasting. *Ops Res.* **21**, 554–568.
19. MONTGOMERY DC & JOHNSON LW (1976) *Forecasting and Time Series Analysis.* McGraw-Hill New York.
20. TRIGG DW & LEACH AG (1967) Exponential smoothing with an adaptive response rate. *Ops Res. Q.* **18**, 53–59.
21. WHYBARK DC (1973) A comparison of adaptive forecasting techniques. *Logistics Transpn Rev* **8**, 13–26.

Everette S. Gardner Jr
(February 1983)

Commander
Supply Corps
US Navy
US Atlantic Fleet Headquarters
Norfolk
VA 23511
USA