Some Properties of a Simple Moving Average when Applied to Forecasting a Time Series
Author(s): F. R. Johnston, J. E. Boyland, M. Meadows and E. Shale
Source: *The Journal of the Operational Research Society*, Vol. 50, No. 12 (Dec., 1999), pp. 1267-1271
Published by: Palgrave Macmillan Journals on behalf of the Operational Research Society
Stable URL: http://www.jstor.org/stable/3010636
Accessed: 14-11-2015 11:38 UTC

Journal of the Operational Research Society (1999) 50, 1267–1271 0160-5682/99 $15.00

http://www.stockton-press.co.uk/jors

Some properties of a simple moving average when applied to forecasting a time series

FR Johnston[1]*, JE Boyland[2], M Meadows[1] and E Shale[1]

[1]*University of Warwick and* [2]*Buckinghamshire University College, UK*

Simple (equally weighted) moving averages are frequently used to estimate the current level of a time series, with this value being projected as a forecast for future observations. A key measure of the effectiveness of the method is the sampling error of the estimator, which this paper defines in terms of characteristics of the data. This enables the optimal length of the average for any steady state model to be established and the lead time forecast error derived. A comparison of the performance of a simple moving average (SMA) with an exponentially weighted moving average (EWMA) is made. It is shown that, for a steady state model, the variance of the forecast error is typically less than 3% higher than the appropriate EWMA. This relatively small difference may explain the inconclusive results from the empirical studies about the relative predictive performance of the two methods.

Keywords: forecasting; time series; moving averages; exponentially weighted moving averages

Introduction

The basic idea underlying all scientific forecasting is that the series of interest can be represented as being generated by an identifiable model. For situations without growth or seasonality, the simplest realistic form is the steady state model (SSM)[1–3] The observations are represented as being random perturbations around an unknown mean, which, through time, undergoes a random walk.

Having accepted (at least temporarily) a specific model as a reasonable representation of reality, the forecaster's problem is to estimate any relevant parameters, which for the SSM is just one, namely the current value of the underlying level. Many approaches have been suggested, but exponentially weighted moving averages (EWMA) is frequently used. It has been the subject of many reviews[4] and theoretical analysis and development.[1,5,6] However, a resurgence in interest and use of the simple moving average (SMA) has become apparent. Sanders and Manrodt[7] conducted a survey of forecasting practices in US corporations and found moving averages to be the most familiar and most used quantitative technique. Sani and Kingsman[8] found that SMA may result in lower inventory costs than EWMA. Despite the simplicity of SMA (and its frequent usage), the basic theoretical underpinnings for employment with a SSM have not been developed.

This paper commenses by defining the SSM, then examines the sampling properties of the SMA which lead to the conditions necessary to minimise forecast error. From this, the true variance of lead time demand can be calculated. Finally, a comparison of the relative performance of the two averaging systems is made. The theoretical results are illustrated through a simulation study.

*Correspondence: Dr FR Johnston, The University of Warwick, Coventry CV4 7AL, UK
E-mail: ORSJ@wbs.warwick.ac.uk

The steady state model

The SSM model[1] can be defined by two equations. The first, the observation equation, defines how the data point results from a random disturbance about a true but unknown level, whilst the second, the system equation, represents the movement of this level through time as a random walk process.

A specific variant of the model, which corresponds to an ARIMA (0, 1, 1) process is where the stochastic elements are treated as being independent drawings from distributions with zero means and variances which are constant through time. Using the notation in West and Harrison,[2] this can be written as

Observation equation

$$y_t = \theta_t + v_t \qquad v_t \sim [0; V] \tag{1}$$

System equation

$$\theta_t = \theta_{t-1} + w_t \qquad w_t \sim [0; W] \tag{2}$$

where y_t is the observation at time t; θ_t is the underlying level at time t; v_t and w_t are stochastic terms from distributions with zero means and fixed variances V and W respectively.

If EWMA is employed as an estimator of the unknown current level θ_t, then the optimal value of the smoothing constant, denoted as α, depends on the ratio of V to W, and

$$\alpha = \left(-1 + \sqrt{1 + 4V/W}/(2V/W)\right) \tag{3}$$

Using this optimal value minimises the variance of the estimate of the level (C_α) and the one-step ahead forecast (Q_α). (The suffix on these symbols is used to indicate that the estimates have resulted from an EWMA.) The value for C_α is given by

$$C_\alpha = \alpha V \tag{4a}$$

However, the observational noise variance is difficult to estimate, whereas the variance of one-step-ahead forecast errors can be estimated from observations. Should it be required, V could be estimated using the relationship $V = Q_\alpha/(1-\alpha)$ and W is then found using (3). C_α and Q_α are linked by the equation

$$C_\alpha = \alpha(1-\alpha)Q_\alpha \tag{4b}$$

From these relationships, a number of results follow, namely the variance of lead time demand,[5] and the quantification of limits for the safe projection of a forecast.[9]

Similar results have not been defined for the SMA, nor has a theoretical comparison been made of the relative effectiveness of the two averaging systems. The following section defines the optimal length for a SMA, followed by a comparison of performance with an EWMA.

The optimal length for a SMA

Let the simple moving average of n values, made after the knowledge of the observation y_t, be m, defined as

$$m = (1/n)\sum_{i=0}^{n-1} y_{t-i} \qquad (i = 0 \text{ to } n-1)$$

The steady state model enables earlier observations to be expressed in terms of the current level as

$$y_{t-i} = \theta_t - \Sigma w_{t-j} + v_{t-i} \qquad (j = 1 \text{ to } i)$$

From the definition of a SMA, it is clear that the estimate will be unbiased, as all error terms have an expected value of zero.

To calculate the variance of the average, it is necessary to recognise the covariance introduced by the cumulation of the disturbance terms,[5] whereupon the variance of m, here designated C_n, (to indicate it has come from a SMA of length n) is

$$\begin{aligned} C_n &= \left(\frac{1}{n^2}\right)[nV + W\Sigma j^2] \qquad (j = 1 \text{ to } n-1) \\ &= \left(\frac{1}{n^2}\right)[nV + W(n-1)n(2n-1)/6] \\ &= \frac{V}{n} + \frac{W(2n^2 - 3n + 1)}{6n} \end{aligned} \tag{5}$$

To minimise the uncertainty on the estimate of the level, it is necessary to select a value of n which minimises (5). However, the above expression is only defined for positive integer values of n. Therefore, to identify the minimum value of the function, it will now be treated as if it were defined for all values of n on the real line. When defined, the function is clearly differentiable and the minimum is given by

$$N^2 = \frac{3V}{W} + \frac{1}{2} \tag{6}$$

where the 'optimal' value of n is denoted by N.

Substituting back into (5) gives the complement of (4a), namely

$$C_N = \frac{V(2N - \frac{3}{2})}{N^2 - \frac{1}{2}} \tag{7a}$$

It should be noted that this is the minimum possible variance attainable by SMA. If the optimal value of n is non-integer, then the variance obtained using the best integer value will be higher than that given above. Expressing the sampling error variance in terms of the more measurable forecast error variance, Q_N results in

$$C_N = \frac{Q_N(2N - \frac{3}{2})}{(N+1)^2} \tag{7b}$$

Lead time demand

For a model without growth or seasonality, the mean forecast for all future periods is the estimate of the current level. The uncertainty associated with these forecasts comprises three elements,[9] the uncertainty of the estimate of the level (C_N) the uncertainty caused by the stochastic development of the mean (W) and the residual observational noise (V). The uncertainty in the estimate of the underlying level introduced by the first two components is present across all future forecasts, which are therefore correlated. The variance of the cumulative forecast over a lead time must recognise this factor, or else understate the true uncertainty. The expressions have been calculated for an EWMA process,[5] and similar expressions are derived here for the SMA.

Let $q(k)$ be the expected value of the forecast k periods ahead, and $Q(k)$ be the corresponding variance.

Using the observation and system (1) and (2), leads to

$$q(k) = m$$

and

$$Q(k) = V + C_N + kW \tag{8}$$

Let the cumulative lead time forecast for h periods have a mean $s(h)$ where

$$\begin{aligned} s(h) &= \Sigma q(k) \qquad (k = 1 \text{ to } h) \\ &= hm \end{aligned}$$

and a variance

$$S(h) = \Sigma Q(k) \qquad (k = 1 \text{ to } h)$$

The covariance terms in the summation of (8) yield

$$S(h) = hV + h^2 C_N + h(h+1)(2h+1)W/6 \tag{9}$$

Expressing this in terms of N and W gives

$$S(h) = (hW/3)(N+h)^2$$

The one-step ahead variance is

$$S(1) = Q(1) = (W/3)(N+1)^2$$

It is often erroneously assumed that the lead time variance is just h times the one-step ahead variance. If a factor (f) is defined as the adjustment necessary to correct h times the one-step ahead variance, then from the previous two equations

$$f = \frac{(N+h)^2}{(N+1)^2}$$

The correction factor increases with the horizon h, and as the optimal length of the average decreases. For a frequently used average of length 10, with a typical lead time of say, 6 periods the correction factor is greater than two, illustrating the substantial underestimate of variance in some systems. Table 1 lists the correction factor for some sample values.

Table 1 The correction factor for the variance of lead time demand, being the ratio of the true variance to the one-step ahead figure multiplied by the lead time.

Lead time	*Length of moving average*		
	20	*15*	*10*
1	1.00	1.00	1.00
2	1.10	1.13	1.19
4	1.31	1.41	1.62
6	1.53	1.72	2.12
8	1.78	2.07	2.68
10	2.04	2.44	3.31

A comparison with EWMA

Harrison[1] demonstrated than an EWMA was the optimal estimator for the mean of a SSM. It will have the minimum variance, and therefore should lead to the best forecast. The equations derived in an earlier section enable a direct comparison between the two averaging methods. Different SSM are characterised by different ratios of V/W, and each would necessitate a specific optimal averaging length and smoothing constant leading to sampling and forecast error variances. Table 2 presents some representative comparisons, in which all the variances are expressed as a percentage of the one step ahead forecast error for the best EWMA system. Two key columns are the ratios giving the relative performance of the two averaging methods.

The sampling variance of the SMA is some 15 to 20% higher than the corresponding EWMA estimate. It is shown in the Appendix that the asymptotic value is $2/\sqrt{3}$. Whilst the difference is appreciable, it makes only a small contribution to the one-step ahead forecast error (Q_N). Using normal numbers of terms in an average, the forecast error from a SMA is less than 3% higher than using an EWMA. This factor roughly halves when the standard deviation, and therefore, resulting decisions, are considered. Given that the theoretical difference in forecast error between the two methods is small, it is not surprising that practical comparative studies have reached inconclusive results as to the optimal method.[10]

For any SSM, the value of the smoothing constant (α) and the length of the simple average (N) to minimise variance can be found. Relating these can be used to explain the characteristics of EWMA to a non numerate manager, therefore providing a link and comparison of EWMA with the behaviour of the more intuitive SMA.

From (3) and (6), the equivalences are

$$N^2 = \frac{3(1-\alpha)}{\alpha^2} + \frac{1}{2}$$

Table 2 The ratio of the variance of the forecast error on employing SAM compared with EWMA. Different SSM are compared, having different optimal length (N) and smoothing constants (α)

SSM resulting in optimal		*Ratio of variances*		
N	α	*Noise to signal* V/W	*Variance mean* $C_N/C\alpha$	*Forecast error* $Q_N/Q\alpha$
3	0.443	2.8	1.194	1.048
6	0.252	11.8	1.176	1.033
8	0.195	21.2	1.171	1.027
10	0.159	33.2	1.168	1.022
12	0.135	47.8	1.166	1.019
16	0.103	85.2	1.163	1.015
20	0.083	133.2	1.161	1.012
24	0.070	191.8	1.160	1.010
50	0.034	833.2	1.157	1.005
∞	0.000	∞	1.155	1.000

or

$$\alpha = \frac{(\sqrt{(12N^2+3)} - 3)}{2N^2 - 1} \tag{10}$$

Brown[11] suggested relating the constants so that the estimates have the same average age, or lag, resulting in

$$\alpha = \frac{2}{N+1} \tag{11}$$

Brown considered a model of demand with a static mean. For this model the smoothing constant and the value of N which give the same variance for the estimate of the level are again given by (11). This equivalence was formulated for explanatory purposes for those who did not have a 'feel' for EWMA and smoothing constants. However, Brown's linking equation is only valid for a model with a fixed and unchanging mean, under which circumstances EWMA will confer disbenefits over SMA. For most practical situations, where the underlying level of the mean does change through time, (10) should be used to determine an equivalent smoothing constant.

Simulation

Most comparisons of SMA and EWMA are conducted over a genuine data set, which will of necessity be relatively short. In practice, for EWMA, the relation between the mean square forecast error and the smoothing constant is always found to be a smooth continuous function, facilitating the selection of the optimal parameter. For SMA, the relationship is usually not smooth and the minimum is not always obvious. However, when exploring the properties of the model, it is possible to simulate over a very large number of data points, when a smooth relationship is obtained and the results are then easier to compare. In the experiments, the theoretical optimal value for the length of the moving average always yields the minimum observed variance. Some results from one experiment are illustrated in Figure 1. The SSM had values of V and W set at 88.4 and 1.04 respectively, when the length of the best average is 16 points, and yields a one-step ahead forecast error of 100. The ordinate is the sampling error variance (SEV) for this particular model. However, the figure generalises, and the ordinate also yields the SEV expressed as a percentage of the one-step ahead forecast error for all SSM having a best SMA of 16 points. That is, for a SSM whose mean is best estimated by a SMA of length 16, the SEV will be 10.6% of the one-step ahead forecast error. This will rise to 11.0% if the average is taken over 12 points.

On the one hand, the slight increase in uncertainty introduced by using a non optimal averaging length demonstrates the robustness of SMA. However the difference does become significant when lead time forecasts are considered due to the presence of the squared term in (9). The SE variance contributes only 10.6% to the one-step ahead forecast error but represents 45.1% of the variance for a ten period lead time, and any unnecessary error due to non optimal length is correspondingly magnified.

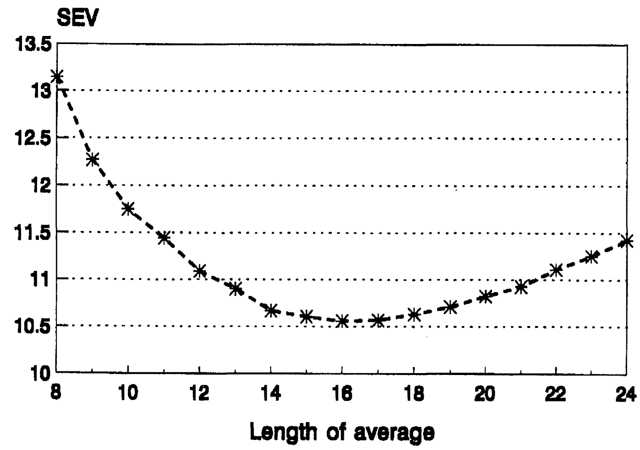

Figure 1 The sampling error variance for SMA.

Conclusions

The sampling error has been determined for a simple moving average when used on a time series (SSM) This is important in its own right, but also leads to the following practical points:

(1) If the generating model is known, the optimal number of terms for inclusion in the average can be determined.
(2) An expression for the true variance of lead time demand can be obtained, and the correction factor calculated, which is of major significance in inventory service level calculations.
(3) A theoretical comparison can be made of SMA with EWMA. The minuscule advantage of the latter explains the conflicting experimental results of many workers.
(4) An equivalence between the two averaging methods has been determined based on selecting the parameters which give the lowest sample and forecast error variances. This seems a more preferable criteria to equating the average age of the data employed in the estimates of the current level.

Appendix

This Appendix derives the limiting value of the ratio of an optimal SMA estimate to the variance of an optimal EWMA. By equations (7*a*) and (4*a*)

$$\frac{C_N}{C_\alpha} = \frac{2N - \frac{3}{2}}{\alpha(N^2 - \frac{1}{2})}$$

Substituting values of N and α from (6) and (3) and denoting the ratio V/W by r,

$$\frac{C_N}{C_\alpha} = \frac{2\sqrt{(3r+1/2)} - 3/2}{3r[(-1+\sqrt{(1+4r)})/2r]}$$

$$\frac{C_N}{C_\alpha} = \frac{\frac{4}{3}\sqrt{(3r+1/2)} - 2}{\sqrt{(4r+1)} - 1}$$

Letting $r \to \infty$,

$$\frac{C_N}{C_\alpha} \to \frac{2}{\sqrt{3}} = 1.1547.$$

References

1 Harrison PJ (1967). Exponential smoothing and short-term sales forecasting. *Mgmt Sci* **13**: 821–842.

2 West M and Harrison PJ (1989). *Bayesian Forecasting and Dynamic Models*. Springer-Verlag: New York.

3 Harvey AC (1989). *Forecasting, Structural Time Series Models, and the Kalman Filter*. Cambridge University Press; Cambridge.

4 Gardner ES (1985). Exponential smoothing: the state of the art *J of Forecasting* **4**: 1–28.

5 Johnston FR and Harrison PJ (1986). The variance of lead-time demand. *J Opl Res Soc* **37**: 303–309.

6 Muth JF (1960). Optimal properties of exponentially weighted forecasts. *J Amer Statist Assoc* **55**: 299–306.

7 Sanders NR and Manrodt KB (1994). Forecasting practices in US corporations: survey results. *Interfaces* **24**(2): 92–100.

8 Sani B and Kingsman BG (1997). Selecting the best periodic inventory control and demand forecasting methods for low demand items. *J Opl Res Soc* **48**: 700–713.

9 Johnston FR and Boylan JE (1994). How far ahead can an EWMA model be extrapolated. *J Opl Res Soc* **45**: 710–714.

10 Makridakis S *et al.* (1982). The accuracy of extrapolation (time series) methods. *J of Forecasting* **1**: 111–154.

11 Brown RG (1963). *Smoothing, Forecasting and Prediction* pp 107–108. Prentice-Hall: Englewood Cliffs, NJ.

Received May 1998;
accepted July 1999 after one revision