AD645144

UNITED STATES NAVAL POSTGRADUATE SCHOOL

CLEARINGHOUSE
FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION

Hardcopy	Microfiche		
$3.00	$.65	33 pp	

ARCHIVE COPY

SOME REMARKS ON EXPONENTIAL SMOOTHING

by

Peter W. Zehna

December 1966
Technical Report/Research Paper No. 72

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

SOME REMARKS ON EXPONENTIAL SMOOTHING

by

Peter W. Zehna

December 1966
Technical Report/Research Paper No. 72

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral E. J. O'Donnell, USN
Superintendent

Dr. R. F. Rinehart
Academic Dean

ABSTRACT:

A critical analysis of the technique of exponential smoothing as a demand forecasting tool in inventory theory. Certain standard formulas which have been developed for this technique are shown to be only asymptotically valid and therefore suspect when the number of demand periods is small. Alternate formulas, valid for any number of time periods, are derived for one special case that is commonly treated. Certain statistical weaknesses of this forecasting technique are then analyzed and, in particular, the use of mean absolute deviation to estimate variability is criticized.

This task was supported by: Naval Supply Systems Command, Code 13

Prepared by: P. W. Zehna

Approved by:
J. R. Borsting
Chairman, Department of
Operations Analysis

Released by:
C. E. Menneken
Dean of
Research Administration

Naval Postgraduate School Technical Report/Research Paper No. 72
December 1966

UNCLASSIFIED

1. Introduction

The term "exponential smoothing" seems to have been coined for the first time by R. G. Brown [1] in 1959 for a particular time series forecasting technique (or a statistical estimation technique, depending on one's paint of view.) Basically, the technique involves weighting each bit of past history with geometrically decreasing weights, less and less weight being given to the older part of the history. Certainly such a procedure has a great deal of intuitive appeal and, moreover, it has been shown that exponential smoothing entails less computer storage than some of the classical techniques such as forecasting by a moving average. These and other advantages are well documented in the book [4] on smoothing by Brown, a book almost entirely devoted to the exponential smoothing technique. Since an inventory system, particularly under a periodic review model, so often entails basing decisions for the future on past demand history, forecasting techniques are of considerable interest to the inventory manager.

It is quite evident that exponential smoothing has been widely adopted by Naval Supply Systems Command as a basic forecasting technique. A review of almost any document, such as various ALRAND reports and PAR documents which involve forecasting or estimation makes it quite clear that this is the case. And, since the book [4] by Brown is practically a sole source of information on the subject, it is not surprising to find said book extensively referenced throughout such documents. The writer has not been able to find any other text materials in which anything beyond a cursory treatment of exponential smoothing is

given. And yet this textbook by Brown, Chapter 9 in particular, is replete with errors of both a typographical and a conceptual nature. Some added difficulty is created by the use of notation which is not consistent with the meaning usually given such symbols in related scientific literature. For example, the notation $\hat{a}$, $\hat{b}$, $\hat{c}$, does not always denote estimates of the corresponding parameters a, b, c as they are normally is used. In other cases, the same symbol has been used ambiguously for two different quantities which certainly leads to confusion.

One of the biggest indictments of the material presented in Chapter 9 of Brown's book is the fact that his so-called Fundamental Theorem, which hardly qualifies a theorem to begin with, is only an asymptotic (with time) result but is presented, used and discussed in such a way as to lead the reader to believe otherwise. Indeed, since the entire book rests basically on this Fundamental Theorem, it is not surprising that nearly every result in the book is an asymptotic result. This includes claims for statistical unbiasedness which is weak enough in itself without holding only asymptotically. Yet, except for an occasional and casual use of the phrase, "after the initial transient becomes negligible," the reader is never made aware of this fact.

Another fundamental criticism from a statistical point of view is Brown's constant use of mean absolute deviation (MAD) to estimate statistical variation. For the futility of using MAD to account for variability has been well documented in the statistical literature for

years. Its use by Brown seems to be justified mainly, and not surprisingly, because of its amenability to the exponential smoothing technique. Out of curiosity, the writer did a quick survey of the recent literature on the subject of variability and has been unable to find any significant result that would change one's attitude toward MAD. And yet, the disadvantages associated with this measure of variability is not mentioned once in Brown's book. But there is no hesitation in mentioning (p. 282) the computational disadvantage in using the standard deviation as a measure of variability. And of course computational convenience is but one of a list of criteria to be considered in selecting a model and it is a real disservice to ignore other, perhaps even more important, criteria.

The purpose of this report, then, is to clarify some of the results given in Brown's book and to emphasize, much more strongly than does the author himself, the assumptions, tacit and otherwise, that yield these results. In this way, it is hoped that the reader will be more aware of the restrictive nature of some of the formulas derived in Brown's book and will thereby exercise some caution in their application. For a special case where Brown's formulas are only asymptotically (in time) valid, alternative forms are presented which are valid for finite values of time parameters.

2. Initial Conditions

The first matter to be discussed in this report concerns the very definition of exponential smoothing. In the first place, Brown seems to be inconsistent in the definition employed in his early papers

[1] and [2], and the one adopted later in his textbook [4]. In the former, single exponential smoothing of the sequence $x_0, x_1, x_2, \ldots, x_t$ is defined by,

$$\bar{x}_t = \alpha \sum_{j=0}^{t-1} (1 - \alpha)^j x_{t-j} + \alpha (1 - \alpha)^t x_0$$

which may as well be written

$$\bar{x}_t = \alpha \sum_{j=0}^{t} (1 - \alpha)^j x_{t-j}$$

since it is identically the same. (The parameter α is a number in the interval [0, 1], called the smoothing constant.) This is equation (3) page 675 of [2]. Yet, on page 101 of [4] we find the symbol $S_t(x)$ used to denote the same quantity and this time is defined to be,

$$S_t(x) = \alpha \sum_{j=0}^{t-1} (1 - \alpha)^j x_{t-j} + (1 - \alpha)^t x_0 .$$

The difference, of course, is in the coefficient of $(1 - \alpha)^t$ in both expressions or, viewed another way, the difference lies in the weight to be given the observation x_0. In any case, both formulas are claimed to be derived from the basic recursion relation,

$$S_t(x) = \alpha x_t + (1 - \alpha) S_{t-1}(x),$$

presumably valid for $t = 1, 2, 3, \ldots$. But successive substitution in this recursion relation only yields

$$S_t(x) = \alpha \sum_{j=0}^{t-1} (1 - \alpha)^j x_{t-j} + (1 - \alpha)^t S_0(x) \ .$$

Clearly, then, the question of compatibility of these two forms of the definition of the exponential smoothing operator depends upon how one defines the initial condition $S_0(x)$. If the first formula is to be valid then we must have $S_0(x) = \alpha\, x_0$ while if the textbook form is used then it must be the case that $S_0(x) = x_0$. Since Brown is not explicit on this point we can only postulate what was intended. In either case, the resulting definition depends somewhat on how x_0 is treated since in one case x_0 is given weight α initially and unit weight in the other case. In the first case, given in Brown's paper, in viewing exponential smoothing as a variation of averaging so that the result is a weighted sum of the observations, then the sum of the weights is not unity which is awkward statistically speaking.

Of course, how one defines the initial condition is of little consequence when only asymptotic results are considered since the effect of the initial condition eventually becomes negligible in either of the above cases. And, for this reason, the inconsistency in defining $S_0(x)$ (actually the utter lack of any explicit mention of same) never appears to be a problem because, as we have said, Brown's results are, by and large, only asymptotically valid hence applicable only to a steady state condition. Yet, the point is more than merely academic. The formula is a result of a recursion relation and, to apply such a relation in a model requires an initial condition as does any application of a mathematical recursion. Moreover, statistical properties, notably

unbiasedness, definitely depend upon how one treats the initial condition. Finally, there are many realistic situations in which there is simply not enough past history to justify the application of an asymptotic result in which case the initial condition becomes a very important factor and can considerably influence the consequences.

Several points of view regarding the meaning to be attached to x_0 in the sequence x_t, $t = 0, 1, 2, \ldots$, can be justified. If x_t represents the demand occurring in the tth time period of an inventory model, then it is quite natural to define $x_0 \equiv 0$ since initially, that is before we begin operating the system, there is no demand. In that case, it does not matter which of the above forms we use for $S_0(x)$ since, in either case we obtain $S_0(x) \equiv 0$ also. But then we may as well write

$$S_t(x) = \alpha \sum_{j=0}^{t-1} (1 - \alpha)^j x_{t-j},$$

in which case, writing β for $1 - \alpha$, the sum of the weights is

$$\alpha \sum_{j=0}^{t-1} \beta^j = 1 - \beta^t,$$

which is not unity. One of the consequences of this result is that if we are observing a process with constant mean then the smoothing operator $S_t(x)$ is not unbiased as is often claimed in such circumstances. This is precisely one of the problems encountered by Bessler and the writer

[8] in attempting to apply exponential smoothing to a dynamic inventory model originally developed by Vassian in 1955. This led them to define a modified version of smoothing which they call <u>finite exponential smoothing</u>. Denoting this modification by $\tilde{S}_t(x)$, it is defined in [8] by

$$\tilde{S}_t(x) = \alpha_t \sum_{j=0}^{t} \beta^j x_{t-j}$$

where

$$\alpha_t = \frac{\alpha}{1 - \beta^t} .$$

With the coefficients thus normalized, the sum of the corresponding weights is unity as desired. Further properties of this modified version of smoothing and some of its applications may be found in [8].

Another point of view that might be taken regarding the initial condition applies when the assumption in the model is that $x_t = \xi_t + e_t$ where ξ_t is a deterministic function of t and e_t is a random variable with mean zero and constant variance σ^2. In that case, it is natural to suppose that $x_0 = \xi_0 + e_0$ to be consistent with the rest of the model. Whether or not such an assumption is suitable depends upon further considerations in the model. For example, suppose it is assumed that $\xi_t \equiv a$, where $a \neq 0$. In that case, $S_t(x)$ is unbiased if we use the version $S_0(x) = x_0$ but is not unbiased if we use $S_0(x) = \alpha x_0$ instead.

In many of the applications which Brown discusses in his book [4], he speaks of x_0 as representing some initial -- any initial -- estimate of, say demand, up to the time the process is to be observed. In some cases, such an estimate may be sheer judgment, or rather guess, as to what the, say constant mean demand will be. In other cases, it may be obtained from the manner in which it is hoped that the process will behave. In still other cases, x_0 may be a number which depends upon some related process whose behavior has been previously observed. In any case we are then considering x_0 as being an estimate from a separate distribution, one not necessarily related to the assumption $x_t = \xi_t + e_t$. Then $S_t(x)$ is or is not unbiased depending upon both the distribution that does represent x_0 as well as which form of $S_0(x)$ we use. For example, if $\xi_t = a$ for $t = 1,2,\ldots$ then

$$E[\ S_t(x)\ ] = a - a\ \beta^t + \beta^t\ E[\ x_0\ ]$$

if we take $S_0(x) = x_0$ while

$$E[\ S_t(x)\ ] = a - a\ \beta^t + \alpha\ \beta^t\ E[\ x_0]$$

if we take $S_0(x) = \alpha\ x_0$. In either case, whether or not $E[\ S_t(x)\ ] \equiv a$ depends upon $E[\ x_0\ ]$ and certainly in general it will be the case that $E[\ S_t(x)\ ] \not\equiv a$.

3. Fundamental Theorem

As indicated earlier, most of the mathematics of exponential smoothing is summarized in what Brown calls his Fundamental Theorem of Exponential Smoothing, the statement and "proof" of which is given on page 133 of [4]. Using the model $x_t = \xi_t + e_t$ where, in general,

$$\xi_t = a_0 + a_1 t + \frac{a_2}{2} t^2 + \ldots + \frac{a_n}{n!} t^n \text{ and } \{ e_t \}_{t=0}^{n}$$

represent independent random variables, identically distributed with zero means and constant variance σ^2, Brown asserts that his fundamental theorem 'proves that it is possible to estimate the n + 1 coefficients in an nth order polynomial model by linear combinations of the first (n + 1) orders of exponential smoothing." The general k^{th}-order smoothing operator is defined inductively by

$$S_t^{[k]}(x) = \alpha S_t^{[k-1]}(x) + (1 - \alpha) S_{t-1}^{[k]}(x) \text{ for } t = 1,2,3,\ldots$$

In the first place, the fundamental theorem is not really a theorem at all but simply an observation that the pth-order smoothing operator can be written explicitly in terms of the coefficients of the model. But worse, what is stated as the fundamental theorem is simply not true. Thus, even for p = 1 it is just not true that

$$S_t(x) = \sum_{k=0}^{n} (-1)^k \frac{x_t^{(k)}}{k!} \alpha \sum_{j=0}^{\infty} j^k \beta^j$$

as asserted by the theorem. Later in this section, we will derive the correct expression for $S_t(x)$ and show that what is given here is an approximation.

Secondly, even if one were to call the result a theorem in a broad sense, the proof that is given is not a proof of the statement of the theorem at all. Indeed, the opening line of the proof on page 133 asks the reader to "Think of the infinite sequence of observations,...,x_t for $t = -\infty,\ldots,-1,0,1,\ldots,\infty$." But one is not given an infinite sequence of observations. In fact, all that is given for any application are the observations $x_0, x_1, x_2, \ldots, x_t$. Giving the author the benefit of the doubt, however, let us suppose that the "extra variables," are simply being used as surplus variables to generate a proof. Certainly the observations $x_{t+1}, x_{t+2}, \ldots$ turn out to be redundant for we find, reading further, that a new sequence is introduced by the definition

$$S_t = \begin{cases} 0 & \text{if } t < 0 \\ \alpha\beta^t & \text{if } t \geq 0 \end{cases}$$

whereupon it is asserted that

$$S_t(x) = \sum_{j=0}^{\infty} x_{t-j} S_j$$

found by the convolution of $\{x_j\}_{j=-\infty}^{\infty}$ and $\{S_k\}_{k=-\infty}^{\infty}$.

Thus, the effect of defining $S_{-1}, S_{-2}, \ldots$ to be zero is to cancel out the

observations $x_{t+1}, x_{t+2}, \ldots$ in writing the convolution product given in the text. But what remains is, after correcting a misprint on page 133, given by

$$S_t(x) = \alpha \sum_{j=0}^{\infty} \beta^j x_{t-j}$$

and this is not the definition of $S_t(x)$ although the author certainly uses the same symbol and refers to this as the single exponential smoothing operator.

What possible points of view can be taken to resolve this apparent inconsistency? One approach would be to assume the author intended to define S_t by means of

$$S_t = \begin{cases} \alpha \beta^j & \text{if } 0 \le j \le t \\ 0 & \text{otherwise} \end{cases} .$$

Or, we might assume that the extra variables are all zero, that is, $x_n \equiv 0$ if $n < 0$. In either case, convolution would then yield the formula

$$S_t(x) = \alpha \sum_{j=0}^{t} \beta^j x_{t-j}$$

which is consistent with the fact that we will be estimating with observations $x_0, x_1, \ldots, x_t$. Unfortunately, this formula is still not quite the same as that given previously in the text on page 101 where $S_t(x)$ is defined. There, the coefficient of x_0 is given as β^t whereas

here in the fundamental theorem, the coefficient of x_0 is $\alpha \beta^t$ under any of the above versions.

A third criticism is that the theorem does not prove (even if it were valid) that the coefficients in the model can be estimated by linear combinations of $S_t^{[1]}(x)$, $S_t^{[2]}(x), \ldots, S_t^{[n+1]}(x)$ as quoted above. There is still the question of solving the system of equations given by the theorem for the coefficients. The author proceeds to do this for two special cases in the remainder of the chapter. But even so, we are compelled to remark that, of course it is <u>possible</u> to estimate the coefficients this way. Indeed one <u>can</u> use any function of the observations to estimate them. But for any estimates to be meaningful they should satisfy some criteria, at least from a statistical point of view. Are the estimates presented by the author unbiased? We have seen that in general they are not. For the special case $S_t = a_1 + a_2 t$, the estimates given are certainly not least squares nor, if normality is assumed, maximum likelihood since these estimates are well known and are not the same. One of the few criteria claimed to be satisfied and shown by D'Esopo [3] is that the estimates, not surprisingly, minimize "exponentially discounted least squares," i.e., minimizes the quantity

$$\alpha \sum_{j=0}^{\infty} \beta^j (x_{t-j} - p_{t-j})^2$$

at least among polynomial fits. Such a ground rule for deriving estimates is not conventional, however, and is tantamount to selecting an estimate by fiat.

It might be instructive to see, in contrast to what appears in Brown's fundamental theorem, what the precise results are at least for the special case of a linear model. In order to maintain the same notation as Brown we will assume a deterministic model at first so that we suppose $x_t = a + bt$, $t = 0,1,2,\ldots$, Brown is not explicit on this point, continually confounding the original random model with the deterministic version whenever it suits his purpose. We will be careful to always make this distinction, however, so that estimation can be discussed in its proper contexts while analytic operations are only performed on deterministic quantities to which they should be restricted. We then have, in Brown's notation, $x_t^{(0)} = a + bt$ and $x_t^{(1)} = b$. Since two versions of $S_t(x)$ exist even in the same context for finite t, we will have to make a choice of definitions. Here we will assume that the definition $S_0(x) = x_0$ is to be preferred since, then, the sum of the weights will be unity in the version

$$S_t(x) = \sum_{k=0}^{t-1} \beta^k x_{t-k} + \beta^t x_0 .$$

Also, double smoothing can then be written

$$S_t^{[2]}(x) = \alpha \sum_{k=0}^{t-1} \beta^k S_{t-k}(x) + \beta^t S_0(x).$$

Here we have made the natural assumption that

$$S_0^{[2]}(x) = S_0(x).$$

In order to derive the finite analogues of Brown's fundamental theorem, it is only necessary to substitute in these formulas and simplify the resulting algebra. The simplification is assisted by a knowledge of finite expansions functions of the basic geometric progression $\sum_{k=0}^{t} \beta^k$. For the record, the first three of these expansions are given below. They, and others, can easily be derived by successively differentiating with respect to the continuous variable β $(0 < \beta < 1)$ and simplifying the resulting algebra.

$$\sum_{k=0}^{t} \beta^k = \frac{1 - \beta^{t+1}}{\alpha}$$

(3-1) $$\sum_{k=0}^{t} k\,\beta^k = \frac{\beta - (t+1)\,\beta^{t+1} + t\,\beta^{t+2}}{\alpha^2}$$

$$\sum_{k=0}^{t} k^2\,\beta^k = \frac{\beta + \beta^2 - (t+1)^2\,\beta^{t+1} + (2t^2 + 2t-1)\,\beta^{t+2} - t^2\,\beta^{t+3}}{\alpha^3}$$

From the above definition and assumptions we then have

$$S_t(x) = \alpha \sum_{k=0}^{t-1} \beta^k\,(a + b\,(t-k)) + a\,\beta^t = \alpha\,(a + bt) \sum_{k=0}^{t-1} \beta^k$$

$$- \alpha\,\beta \sum_{k=0}^{t-1} k\,\beta^k + a\,\beta^t.$$

After some simplification, we obtain,

$$(3\text{-}2) \qquad S_t(x) = x_t^{(0)} - b\frac{\beta}{\alpha} + \frac{b}{\alpha}\beta^{t+1} .$$

Likewise, substituting in the formula for double smoothing yields,

$$(3\text{-}3) \qquad S_t^{[2]}(x) = x_t^{(0)} - 2b\frac{\beta}{\alpha} + 2\frac{b}{\alpha}\beta^{t+1} + bt\,\beta^{t+1}$$

These are the exact formulas for $S_t(x)$ and $S_t^{[2]}(x)$, valid for all finite t, and of course they differ from those given by Brown.

It is now apparent how one can derive Brown's results as asymptotic versions of the exact cases. Since $0 < \beta < 1$, we have $\beta^{t+1} \to 0$ and $t\,\beta^{t+1} \to 0$, as $t \to \infty$. Then we may say that, for sufficiently large values of t, we may approximate $S_t(x)$ and $S_t^{[2]}(x)$ by,

$$(3\text{-}4) \qquad \begin{aligned} S_t(x) &\doteq x_t^{(0)} - \frac{\beta}{\alpha} x_t^{(1)} \\ S_t^{[2]}(x) &\doteq x_t^{(0)} - 2\frac{\beta}{\alpha} x_t^{(1)} \end{aligned}$$

These are the formulas one would obtain from substituting into the Fundamental Theorem of page 133.

To actually apply these results and evaluate them statistically, we would want to consider the model $x_t = \xi_t + \epsilon_t$ where $\xi_t = a + bt$ and,

as before, ϵ_t has mean zero and variance σ^2. Brown would have us use as estimates based on the data $x_0, x_1, \ldots, x_t$, the quantities,

$$\hat{x}_t^{(0)} = 2S_t(x) - S_t^{[2]}(x)$$

(3-5)

$$\hat{x}_t^{(1)} = \frac{\alpha}{\beta}\left[S_t(x) - S_t^{[2]}(x)\right]$$

These are easily obtained by solving (3-4) as though they were equations and then replacing $x_t^{(0)}$ and $x_t^{(1)}$ by the symbols $\hat{x}_t^{(0)}$ and $\hat{x}_t^{(1)}$ since they involve or are themselves unknown parameters. Whatever means they are arrived at, certainly they are properly called estimates since they are functions of the data $x_0, x_1, \ldots, x_t$. They are not, however, unbiased as Brown claims if one uses, as one should, the precise formulas for $S_t(x)$ and $S_t^{[2]}(x)$.

To see that the estimates are biased, we notice first that

$$E[\hat{x}_t^{(0)}] = 2E[S_t(x)] - E[S_t^{[2]}(x)].$$

But,

$$S_t(x) = \alpha \sum_{k=0}^{t-1} \beta^k x_{t-k} + \beta^t x_0$$

and, since $E[x_{t-k}] = a + b(t-k)$, we have,

$$E[S_t(x)] = \alpha \sum_{k=0}^{t-1} \beta^k (a + b(t-k)) + a\beta^t$$

which is the same expression we dealt with in the deterministic model (the $S_t(x)$ of that model). From that result, we have

$$E[\ S_t(x)\ ] = a + bt - b\frac{\beta}{\alpha} + \frac{b}{\alpha}\ \beta^{t+1}.$$

Similarly,

$$E[\ S_t^{[2]}(x)\ ] = a + bt - 2b\frac{\beta}{\alpha} + 2\frac{b}{\alpha}\ \beta^{t+1} + bt\ \beta^{t+1}.$$

Putting these facts together we thus obtain,

$$E[\ \hat{x}_t^{(0)}\ ] = a + bt - bt\ \beta^{t+1}$$

(3-6)

$$E[\ \hat{x}_t^{(1)}\ ] = b - b\ \beta^t - \alpha\ bt\ \beta^t$$

In both cases, the estimates are biased downward, with a bias that is a function of the "trend" b. Since b is unknown, the bias may be serious depending of course on the magnitude of b. The bias factors do converge to zero as time increases beyond bounds however, and we may say that the estimators Brown gives are thereby asymptotically unbiased.

For the case n = 2, that is for a quadratic model $a_t = a_0 + a_1 t + \frac{a_2}{2} t^2$, similar conclusions can be reached. The algebra involved is somewhat burdensome, however, and will not be repeated here. Suffice it to say that the exact formulas for

$S_t(x)$, $S_t^{[2]}(x)$ and $S_t^{[3]}(x)$ are such that for t sufficiently large, Brown's versions of these expressions hold. Again, if these approximations are treated as equations, one can solve the resulting system for the derivatives $x_t^{(0)}$, $x_t^{(1)}$ and $x_t^{(2)}$ to obtain Brown's results. When treated as estimates they are not, of course, unbiased any more than the linear case. Also, the unsuspecting reader should be warned that the results, published on pages 140 through 144 should be read and interpreted with caution even after correcting some obvious misprints. Thus, on page 140 for example, $\hat{a}_0(t)$ and $\hat{a}_1(t)$ are not, as one might presume from the model, estimates of a_0 and a_1 but rather estimates of $x^{(0)}(t) = a_0 + a_1 t + \frac{a_2}{2} t^2$ and $x^{(1)}(t) = a_1 + a_2 t$, respectively. Happily, of course, $\hat{a}_2(t)$ does happen to be an estimate of a_2 since, for this case, $x^{(2)}(t) = a_2$.

No attempt was made to examine the results for higher order polynomials. Based on the quadratic model, it is clear that the algebra involved would be too unwieldy to make the task practical. Perhaps this is as good a justification for resorting to asymptotic results as any. And it should be stated that there is no serious objection to deriving asymptotic results and considering estimators with only asymptotic properties. The objection is to the inordinate use of the same notation for the finite case and the asymptotic case in formula after formula. Together with a complete lack of any discussion of the difference, it leads the unsuspecting reader to believe that the results are stronger than they really are.

4. Mean Absolute Deviation

In inventory applications of random demand models, safety levels are often determined in terms of some measure of variability, usually the common standard deviation of the demand distribution. As was mentioned in the introduction, Brown prefers to use mean absolute deviation, or MAD for short. This in spite of the statistical grounds for not using this particular measure. As he points out (page 275) the mean absolute deviation is proportional to the standard deviation in any probability distribution. Both are, after all, functions of the parameters of the distribution. But finding an appropriate estimate for MAD and deriving the corresponding distribution theory to guarantee the required probability for safety levels is quite another matter. Brown has not done this and, to make matters worse, never distinguishes between a population or true MAD and an estimate thereof, even to the point of using the same symbol and name for them.

In the first place, the definition adopted by Brown for MAD, denoted Δ, reduces to $\Delta = E[\,|x - \mu|\,]$ where x is any random variable having mean μ. As he himself points out on page 283 it would be better to define Δ as $E[\,|x-m|\,]$ where m is any median of the distribution of x. This is because $E[\,|x-c|\,]$ is minimized by choosing $c = m$. Yet he ignores this criterion and uses μ instead of m, justifying his choice on the basis that forecasts estimate means rather than medians. But if one can justify computing Δ instead of σ because Δ is proportional to σ, surely the same argument can be used to estimate m instead of μ.

This is hardly a convincing reason but we will pass this point and use Brown's definition. Of course, in a symmetric distribution $\mu = m$ as he brings out. But it is precisely in the applications to random demand that skewed distributions such as the Poisson and Negative Binomial families arise in practice. This is especially pertinent to standard assumptions in Naval supply systems.

Brown quite aptly shows that the ratio of Δ to σ is approximately 0.8 for the Normal, Exponential, Uniform and Triangular families of probability distributions. Yet, except for the normal family, the interest must be primarily academic so far as inventory applications are concerned. It would be far more interesting, and quite instructive, to see what the situation is for other distributions. In particular, an examination of the Poisson family reveals that 0.8 can be a very poor approximation. In the Poisson mass function

$$p(x;\ \lambda) = e^{-\lambda} \frac{\lambda^x}{x!} \quad ,$$

$x = 0,1,2,\ldots$ with $0 < \lambda < 1$,

we have

$$\Delta = \sum_{x=0}^{\infty} \left| x - \lambda \right| p(x;\ \lambda)$$

$$= \lambda e^{-\lambda} + \sum_{x=1}^{\infty} (x - \lambda) e^{-\lambda} \frac{\lambda^x}{x!} = \lambda e^{-\lambda} + \sum_{x=1}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} - \lambda \sum_{x=1}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!}$$

$$= \lambda e^{-\lambda} + \lambda - \lambda (1 - e^{-\lambda}) = 2 \lambda e^{-\lambda} .$$

Since $\sigma = \sqrt{\lambda}$, we have $\frac{\Delta}{\sigma} = 2\sqrt{\lambda}\, e^{-\lambda}$. Values of this ratio are shown for a variety of values of λ in Table 1.

λ	0.01	0.05	0.10	0.25	0.50	0.75	0.90	0.99
$\frac{\Delta}{\sigma}$	0.198	0.425	0.572	0.779	0.858	0.818	0.771	0.739

TABLE 1. Ratio $\frac{\Delta}{\sigma}$ for Poisson family

As is evident from the table, the approximation 0.8 is extremely poor for slow moving items where the Poisson with small mean λ is a typical assumption. For values of $\lambda > 1$ in the Poisson family and the geometric distribution with mean greater than unity, a similar analysis shows that the approximation 0.8 is not bad, however.

This may appear to be a minor academic point until one finds that the same ratio of $\sqrt{\frac{2}{\pi}}$ is used in the applications of Chapter 20 quite independent of any assumption as to the underlying probability distribution of demand. Also we might point out that even though Δ is proportional to σ in the population, it does not follow that the estimates $\tilde{\Delta}$ and $\tilde{\sigma}$ enjoy the same sort of relationship. This would imply a type of invariance principle such as that enjoyed by maximum likelihood estimates, and is, in general, not true when the estimates are not maximum likelihood.

This brings up another matter concerning MAD estimates. Brown uses error forecasts to estimate Δ. In fact, for the particular data

$x_0, x_1, \ldots, x_t$, the error forecast, $e(t)$ is defined by $e(t) = x_t - \hat{x}_{t-1}$ where $\hat{x}_{t-1}$ is taken to be the forecast at time t-1 of the demand at time t. Now in our basic model with constant mean, $\xi_t = a$, and exponential smoothing used to estimate the mean, we have

$$\hat{x}_{t-1} = \alpha \sum_{k=0}^{t-2} \beta^k x_{t-1-k} + \beta^{t-1} x_0$$

and if $E[\, x_0 \,] = a$, $E[\, \hat{x}_{t-1} \,] = a$. It then follows that $E[\, e(t) \,] = 0$ and, from independence, the variance $\sigma_e^2(t)$ of the error forecast becomes

$$\sigma_\varepsilon^2(t) = \sigma^2 + \frac{\alpha}{1+\beta}(1 - \beta^{2t-2})\sigma^2 + \beta^{2t-2}\sigma^2.$$

as can be easily verified. Letting $t \to \infty$ we observe that the limiting variance σ_e^2 is given by

$$\sigma_e^2 = (1 + \frac{\alpha}{1+\beta})\,\sigma^2 = \frac{2}{2-\alpha}\,\sigma^2,$$

a formula which is used throughout the text by Brown as though it were valid for all t. Incidentally, if there is a possibility of trend present so that the assumption of constant mean is suspect, not even this asymptotic formula should be used to describe the variance of forecast error.

Granted that t is sufficiently large so that the above asymptotic variance applies, it would follow that the true MAD for e_t, say Δ_e, would be defined by $E[\,|e_t|\,]$ since $E[e_t] = 0$. Then <u>if</u> it were true that $\Delta_e = \sqrt{\frac{2}{\pi}}\,\sigma_e$, as for a normal distribution, it would then follow that $\Delta_e = \sqrt{\frac{2}{\pi}}\,\frac{2}{2-\alpha}\,\sigma$ as Brown claims. Then of course

$$\sigma = \sqrt{\frac{\pi}{2}}\sqrt{\frac{2-\alpha}{2}}\;\Delta_e$$

and if we can estimate Δ_e, we could then estimate σ by invoking an (unproved) invariance principle obtaining

$$\tilde{\sigma} = \sqrt{\frac{\pi}{2}}\sqrt{\frac{2-\alpha}{2}}\;\tilde{\Delta}_e\,.$$

In other words, if $\hat{\sigma}$ is the usual maximum likelihood estimate of σ for the present assumption, it follows from the invariance principle that

$$\hat{\Delta}_e = \sqrt{\frac{2}{\pi}}\sqrt{\frac{2}{2-\alpha}}\;\hat{\sigma}$$

is the maximum likelihood estimate of Δ_e. We are on safe grounds, statistically speaking. Now, a reasonable estimate of Δ_e based on the sample $e_1, e_2, \ldots, e_t$ and the fact that $E[e_t] = 0$ would be the sample analogue of $E[\,|e_t|\,]$, namely, $\frac{1}{t}\sum_{i=1}^{t}|e_i|$. Brown, however, guided by exponential smoothing, uses instead the estimate

$$\tilde{\Delta}_e = \alpha \sum_{k=0}^{t-1} \beta^k \, | e_{t-k} | + \beta^t \, \Delta \, (0).$$

Thus, apart from the initial condition, $\tilde{\Delta}_e$ is an exponentially weighted average of the same variables $|e_1|$, $|e_2|, \ldots,$ $|e_t|$, which makes it about twice removed from any known distribution theory. If $\tilde{\Delta}_e$ is used in the above formula for $\tilde{\sigma}$, what can be said about the resulting estimate? It is definitely not maximum likelihood. Neither is it unbiased nor likely to be minimum variance. In truth, without some knowledge of the distribution of $\tilde{\Delta}_e$, even under normality assumptions, very little can be said about $\tilde{\sigma}$.

In summary, then, there is a definite need for more distribution theory before a strong case can be made for exponentially smoothed estimates of MAD. Brown claims on page 286 that, "If one can estimate the mean absolute deviation of the forecast errors, it is quite simple to infer the probability that any given multiple of the estimated value will be exceeded." Quite the contrary, however, it is not only difficult but practically impossible to infer such probability statements without a knowledge of the distributions involved. For example, even if x is normal with mean μ and variance σ^2 so that for any $0 < \gamma < 1$ we can compute the value of K such that

$$\gamma = P_r \, [\, x > \mu + K \, \sigma]$$

it does <u>not</u> follow that when we estimate μ by exponential smoothing, say

$\tilde{\mu}$, and σ by $\sqrt{\frac{\pi}{2}}\ \tilde{\Delta}$, that $P_r\ [x \geq \tilde{\mu} + K\sqrt{\frac{\pi}{2}}\ \tilde{\Delta}\ ]$ is still γ. Yet this seems to be tacitly implied at several points of the book. At the very least, one should have some simulation results for the distribution of $\tilde{\mu} + K\sqrt{\frac{\pi}{2}}\ \tilde{\Delta}$ to make the result more plausible, as recommended by Asher and Wallace [6]. As they point out, if the usual Gauss-Markov assumptions are made, MAD or any estimator other than least squares will come off second best. The results of their study show that MAD is about 20% efficient compared to minimum variance estimators and also displayed greater bias.

5. Conclusions and Recommendations

Lest this report be taken as a total indictment of exponential smoothing as a forecasting technique, let it be said that it is freely admitted that this idea of weighting the past with ever-decreasing weights has a great deal of intuitive appeal. And it is granted that the technique has a computational advantage in requiring less computer storage than more standard techniques. Carried to its extreme, however, one could equally well justify using only the current observation for estimation purposes and ignore the past completely. At least such an estimator would possess some well known statistical properties.

And this is one of the points we wish to stress. An estimator, to be valuable, must satisfy various criteria that have been used to judge such estimators. Exponential smoothing, regardless of its intuitive appeal, must be able to stand the test alongside other alternatives. Invariably, this involves some knowledge of the probability distribution of estimators. Without such a knowledge, it is difficult to approve

or disapprove heartily of exponential smoothing. Certainly Brown has not developed such theory and neither, apparently, has anyone else to any extent. Lacking such a theory, a recent study by Astrachan and Sherbrooke [7] involved an empirical test of exponential smoothing. The results showed that exponential smoothing was not significantly better than techniques currently being used.

But even if these statistical points were resolved we would have to object to the way in which the results are presented in Brown's book for reasons clearly detailed in this report. To this end we are inclined to agree with the review of the book done for Operations Research (Vol. 13, No. 2) by Fishman who says, "In assessing the over-all contribution of this book to the forecasting literature, I would argue that it confuses rather than enlightens the well-informed as well as the mathematically unsophisticated reader." The writer would add that even the mathematically sophisticated reader may have considerable difficulty unravelling some of the ambiguity present in various formulae as well as justifying several claims to mathematical rigor. In any case, the user of this book should be aware of the asymptotic nature of the results and apply them with this restriction in mind.

Finally, we have seen that the indiscriminate use of mean absolute deviation as a measure of statistical variation creates the same theoretical problems that have caused it to be abandoned by statisticians these many years. As Asher and Wallace [6] put it, "... one should be prepared to give up considerable efficiency." The difficulties of obtaining probability distributions for MAD estimators introduced by

Brown appear to be extremely difficult at best. We re-emphasize the fact that such estimators, as well as any exponential smoothing estimators, must be more than a means of arriving at a number, ease of computation notwithstanding. Perhaps the variance estimation techniques we have criticized in this report are fruitful. But without some knowledge of the theory, and their probability distributions in particular, there simply is no way to pass judgment on them.

As for further research, the areas we have been discussing offer rich opportunities indeed. Since this report has essentially been devoted to a critique of Brown's book, it is perforce, negative in its spirit and conclusions. A more positive approach would be to define alternative procedures which would be as appealing as smoothing for computing purposes and would admit a statistical theory at the same time. This is especially needed for statistical variation to replace MAD as a means of determining safety levels. It is strongly recommended that further research in this specific direction be undertaken. It may very well turn out that the smoothing procedures are actually close to optimal in some sense. But it needs to be established that they are.

It does not appear feasible to develop formulas for exponential smoothing beyond the quadratic model. The algebra involved is simply too unwieldy. Perhaps it might be wise to reiterate at this point that we have no objection to asymptotic results as long as they are clearly labeled such. Indeed, for higher order polynomials it appears necessary to resort to such limiting results. Another possible area of research would thus be to investigate further the statistical properties of Brown's asymptotic formulae.

BIBLIOGRAPHY

[1.] Brown, Robert G., "Less Risk in Inventory Estimates," Harvard Business Review, July-August 1959, pp. 104-16.

[2.] Brown, Robert G. and R. F. Meyer, "The Fundamental Theorem of Exponential Smoothing," Operations Research, Vol. 9, No. 5, September-October 1961, pp. 673-85.

[3.] D'Esopo, D. A., "A Note on Forecasting by the Exponential Smoothing Operator," Operations Research, Vol. 9, No. 5, September-October 1961, pp. 686-7.

[4.] Brown, Robert G., Smoothing, Forecasting and Prediction of Discrete Time Series, Prentice-Hall, Inc., 1963.

[5.] Dobbie, J. M., "A Simple Proof of a Theorem in Exponential Smoothing," Operations Research, Vol. 11, No. 3, May-June 1963, pp. 303-476.

[6.] Asher, V. G., and T. D Wallace, "A Sampling Study of Minimum Absolute Deviations," Operations Research, Vol. 11, No. 5, September-October 1963, pp. 747-58.

[7.] Astrachan, Max and C. C. Sherbrooke, "An Empirical Test of Exponential Smoothing," RAND Corporation, RM-3938-PR, March 1964.

[8.] Bessler, S. A. and P. W. Zehna, "Servomechanisms, Exponential Smoothing, and a Multiechelon Inventory Problem," Decision Studies Group, September 1966.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*

U.S. Naval Postgraduate School
Monterey, California 93940

2a. REPORT SECURITY CLASSIFICATION

UNCLASSIFIED

2b. GROUP

3. REPORT TITLE

SOME REMARKS ON EXPONENTIAL SMOOTHING

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Task Progress, August 1966 - October 1966

5. AUTHOR(S) *(Last name, first name, initial)*

Zehna, Peter W.

6. REPORT DATE

December 1966

7a. TOTAL NO. OF PAGES

28

7b. NO. OF REFS

8

8a. CONTRACT OR GRANT NO.

b. PROJECT NO. 14212

c. Task No. RDT&E 015-02-100

d.

9a. ORIGINATOR'S REPORT NUMBER(S)

TR-72

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. AVAILABILITY/LIMITATION NOTICES

Qualified requesters may obtain copies of this report from DDC.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY

Naval Supply Systems Command, Code 13

13. ABSTRACT

A critical analysis of the technique of exponential smoothing as a demand forecasting tool in inventory theory. Certain standard formulas which have been developed for this technique are shown to be only asymptotically valid and therefore suspect when the number of demand periods is small. Alternate formulas, valid for any number of time periods, are derived for one special case that is commonly treated. Certain statistical weaknesses of this forecasting technique are then analyzed and, in particular, the use of mean absolute deviation to estimate variability is criticized.

DD FORM 1 JAN 64 1473

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
Demand Forecasting						
Exponential Smoothing						
Mean Absolute Deviation						
Statistical Estimation						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.