Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Exponential Smoothing and Short-Term Sales Forecasting

P. J. Harrison,

To cite this article:
P. J. Harrison, (1967) Exponential Smoothing and Short-Term Sales Forecasting. Management Science 13(11):821-842. http://dx.doi.org/10.1287/mnsc.13.11.821

MANAGEMENT SCIENCE
Vol. 13, No. 11, July, 1967
Printed in U.S.A.

EXPONENTIAL SMOOTHING AND SHORT-TERM SALES FORECASTING*

P. J. HARRISON†

Imperial Chemical Industries Limited

Non-seasonal forecasting methods are examined by considering demand generating processes which are reasonable and general descriptions of customer demand and for which the popular predictors are shown to be optimal. Tests of the adequacy of the generating processes are described. The way in which the forecasting errors vary with the forecasting period is examined, and it is shown that this is dependent not only on the length of the period but also on the values of the forecasting parameters. The sensitivity of the predictors to departures from the optimal parameters is investigated, and the long debated comparison of Holt's linear growth predictor (1957) and Brown's linear growth predictor (1959) is examined. It is shown for the assumed generating model and for forecasting parameters lying within the usual limits, that even if there is an infinite amount of data available to establish the optimum forecasting parameters, the standard error of the one step ahead predictor exceeds that of the Holt predictor by no more than 1.6 percent. The generalized polynomial generating process is shown to have as its optimal least squares predictor the corresponding Box-Jenkins polynomial predictor (1962).

1. Introduction

In routine short-term sales forecasting and stock control systems, forecasts of customer demand are needed each week or month, often for hundreds or thousands of sales packs. These forecasts, which are usually for the immediate few periods, are used by Distribution Departments for stock control, by the Works for the production planning of intermediates and by Supply Departments for ordering raw materials and packages.

Complete routine forecasting systems should operate according to the principle of "Management by Exception". Forecasts produced by mathematical predictors are used unless exceptional circumstances occur. These exceptions happen when the assumed model for customer demand changes violently. In some cases, marked changes may be anticipated by Sales Intelligence; otherwise they can be detected by statistical control procedures such as the Backward Cumulative Sum procedure [5].

Comparisons of different mathematical forecasting methods have been made from the practitioner's point of view [6] and [7]. In this paper, non-seasonal forecasting methods are examined by considering demand generating processes which are reasonable and general descriptions of customer demand and for which the popular predictors are shown to be optimal.

* Received May 1965 and revised July 1966.

† The author would like to thank Mr. D. Oakes who wrote a number of computer programs needed for the preparation of this paper and also Professor P. Whittle who provided some most helpful comments.

The steady and the linear growth predictors are examined in detail, but the generalized polynomial predictors are just shown to be optimal for the general polynomial process, since, in the author's experience, routine short-term sales forecasting systems do not advantageously use quadratic or higher order polynomial predictors.

2. Predictors Applied in Routine Short-Term Sales Forecasting Systems

Various predictors have been applied in routine short-term sales forecasting systems. Amongst these, the most popular are the predictors proposed by Holt [8], Brown [2] and [3], Box and Jenkins [1], and Thiel, Nerlove and Wage [11] and [12]. All these predictors are particular forms of the Box-Jenkins Polynomial Predictor which for a polynomial of order n has the form:

$$d^*_{t+1} = d_t^* + \sum_{i=0}^{n-1} \eta_i S^i e_t \,, \tag{2.1}$$

where the η_i are constants, d^*_{t+1} is the one step ahead prediction made at time t; e_t is the one step ahead forecast error, and $S^j e_t$ denotes the j^{th} multiple sum of the past errors [i.e., $S^0 e_t = e_t$; $S^1 e_t = \sum_{j=0}^{\infty} e_{t-j}$; $S^2 e_t = \sum_{j=0} \sum_{k=0} e_{(t-j-k)}$].

To illustrate that the predictor includes other well known predictors as particular cases, it is convenient to consider (2.1) in the form:

$$d^*_{t+1} = \sum_{i=1}^{n} m_t^{(i)*}, \tag{2.2}$$

where

$$m_t^{(i)*} = \sum_{j=1}^{n} m_{t-1}^{(j)*} + A_i e_t \,, \tag{2.3}$$

$$e_t = d_t - d_t^*. \tag{2.4}$$

The A_i are the forecasting parameters. That the predictor can be written in this form can be established by expressing the m's in terms of the errors and substituting in (2.2).

The first order predictor ($n = 1$) reduces to the Simple Exponentially Weighted Average:

$$d^*_{t+1} = m_t^*,$$
$$m_t^* = m^*_{t-1} + Ae_t$$

which is the first order predictor of both Holt and Brown.

The second order predictor ($n = 2$) is the Holt growth predictor:

$$d^*_{t+1} = m_t^* + b_t^*$$
$$m_t^* = m^*_{t-1} + b^*_{t-1} + A_1 e_t$$
$$b_t^* = b^*_{t-1} + A_2 e_t$$

with $m_t^{(2)*}$ written as b_t^*. Brown's second order predictor is a particular form of Holt's, restricting the two parameters A_1 and A_2 so that

$$A_1 = 1 - \beta^2; \qquad A_2 = (1 - \beta)^2.$$

The predictor of Thiel, Nerlove, and Wage also restricts the parameters so that

$$A_1 = 2h/(1 - h); \qquad A_2 = hA_1 .$$

For the general n order predictor, Brown's method again restricts the forecasting parameters so that they are all functions of a discount parameter β.

3. A Generating Model for Non-Seasonal Customer Demand

3.1 The Steady Model

Customer demand for a steady selling product can often be reasonably described by a simple random walk process.

Essentially, the model states that the product is established and reasonably stable, and that the true customer need or usage at one period of time is equal to that at the previous period plus a slight random change. The observed demand then represents the true customer need with a superimposed error which reflects the customer ordering variations and which may be assumed to be random when there are many customers.

This can be represented mathematically by:

$$d_t = m_t + \epsilon_t \tag{3.1}$$

$$m_t = m_{t-1} + \gamma_t \tag{3.2}$$

where d_t is the observed demand for period t, m_t may be thought of as the true need of the product, or perhaps the customer usage during period t. ϵ_t and γ_t are uncorrelated random variables each with zero mean and variances $V(\epsilon)$, $V(\gamma)$.

3.2. The Linear Growth Model

Customer demand for most growing or declining products can be represented by a simple extension of the steady model to incorporate the slope. This model states that the true customer need at one period of time is equal to that at the previous time plus the slope plus a random component. In addition, the slope is subject to a random change and, as before, the observed customer demand is equal to the true customer need with a superimposed error. This is represented by:

$$d_t = m_t + \epsilon_t \tag{3.3}$$

$$m_t = m_{t-1} + b_t + \gamma_t \tag{3.4}$$

$$b_t = b_{t-1} + \delta_t \tag{3.5}$$

where b_t represents the slope or growth rate of the true need of customers at period t and δ_t is a random error with zero mean and variance $V(\delta)$.

3.3 The General Polynomial Growth Model

Extending the models of (3.1) and (3.2), the general polynomial model states that each "derivative" of the true customer need pattern undergoes a

random change; so that for the n order polynomial the model may be represented by:

$$d_t = m_t^{(1)} + \epsilon_t \tag{3.6}$$

$$m_t^{(i)} = m_{t-1}^{(i)} + m_t^{(i+1)} + \gamma_t^{(i)} \tag{3.7}$$

for $i = 1 \cdots n$, with $m_t^{(n+1)} = 0$, and the random errors ϵ and $\gamma^{(i)}$ having zero mean, and each set being uncorrelated and identically distributed.

4. An Important Result Derived By Box And Jenkins

For the models discussed in Section 3 it will be shown that the Box-Jenkins polynomial predictor is optimal in the linear least squares sense and the optimal values of the forecasting parameters A_i will be obtained in terms of the variances of the random elements ϵ and $\gamma^{(i)}$. A special case of a result derived by Box and Jenkins (1962, pp. 312 and 313) will be referred to. The relevant result may be shown as follows:

Consider a process generated by

$$d_{t+1} = \sum_{j=0}^{\infty} \eta_j d_{t-j} + \alpha_{t+1}\,, \tag{4.1}$$

where α_{t+1}, α_t, $\cdots$ are uncorrelated, identically distributed, random variables with zero mean, and η_j are constants. Suppose the series is predicted using

$$d_{t+1}^* = \sum_{j=0}^{\infty} \mu_j d_{t-j} \tag{4.2}$$

with the μ's as constants. Then, since,

$$\mathrm{Ex}\,(e_{t+1})^2 = \mathrm{Ex}\,(d_{t+1} - d_{t+1}^*)^2 = \mathrm{Ex}\,[\sum_{j=0}^{\infty}(\eta_j - \mu_j)d_{t-j}]^2 + \mathrm{Ex}\,(\alpha_{t+1})^2,$$

the predictor is optimal if $\mu_j = \eta_j$, and when this is so, the e's become the α's and are uncorrelated. By successive use of (4.2) the optimal predictor may be written as:

$$d_{t+1}^* = d_t^* + \sum_{j=0}^{\infty} W_j e_{t-j} \tag{4.3}$$

with the W's as constants. It follows that the equivalent predictor,

$$D_1\, d_{t+1}^* = \sum_{j=0}^{\infty} W_j e_{t-j}\,,$$

with D as the difference operator, is optimal for the equivalent stochastic process

$$D_1\, d_{t+1} = D_1 \alpha_{t+1} + \sum_{j=0}^{\infty} W_j \alpha_{t-j}\,.$$

In particular, the predictor is thus optimal for a series generated by

$$D_1\, d_{t+1} = D_1 \alpha_{t+1} + \sum_{j=0}^{n-1} \gamma_j S^j \alpha_t\,,$$

remembering that S^j denotes the j^{th} multiple sum. Differencing $n - 1$ times,

$$D_n\, d_{t+1} = D_n\, \alpha_{t+1} + \sum_{j=0}^{n-1} \gamma_j D_{n-j-1} \alpha_t\,.$$

This expression does not involve error terms beyond α_{t-n+1} and the relevant result is: If a stochastic variable d is such that its n^{th} difference can be repre-

sented as a moving average process of identically distributed independent variables α_t which have zero means and this process is of order $n + 1$, then the Box-Jenkins n order polynomial predictor is optimal and the α_t represent the one step ahead prediction errors for this optimal predictor.

5. A General Difference Equation for the Forecast and Random Errors

Consider the general n order demand generating process of (3.6). The optimal least squares predictor gives one step ahead forecasts such that

$$\text{Ex}\,(e_{t+1}^2) = \text{Ex}\,(d_{t+1} - d_{t+1}^*)^2$$

is minimized. It may readily be seen using (3.6) and (3.7) that the n^{th} difference of the observed demand series is such that

$$D_n\, d_{t+1} = D_n \epsilon_{t+1} + \textstyle\sum_{i=1}^{n} D_{n-i}\gamma_{t+1}^{(i)}. \tag{5.1}$$

The Box-Jenkins predictor defined by equations (2.2), (2.3), and (2.4) can be written in the form

$$m_t^{(i)*} = m_{t-1}^{(i)*} + m_t^{(i+1)*} + B_i e_t \tag{5.2}$$

where

$$A_k = \textstyle\sum_{i=k}^{n} B_i \tag{5.3}$$

so that

$$D_1 m_t^{(i)*} = m_t^{(i+1)*} + B_i e_t\,. \tag{5.4}$$

Differencing (5.4) $(n - 1)$ times and recurrently using the equation

$$D_n m_t^{(i)*} = \textstyle\sum_{j=0}^{n-i} B_{i+j} D_{n-1-j} e_t \tag{5.5}$$

From (2.2)

$$D_n\, d_{t+1}^* = \textstyle\sum_{i=1}^{n} D_n m_t^{(i)*} \tag{5.6}$$

$$= \textstyle\sum_{i=1}^{n} (\sum_{j=0}^{n-i} B_{i+j}) D_{n-i} e_t \tag{5.7}$$

$$= \textstyle\sum_{i=1}^{n} A_i D_{n-i} e_t\,. \tag{5.8}$$

Since

$$D_n\, d_{t+1} = D_n\, d_{t+1}^* + D_n e_{t+1}\,, \tag{5.9}$$

it follows that

$$D_n\, d_{t+1} = D_n e_{t+1} + \textstyle\sum_{j=1}^{n} A_j D_{n-j} e_t\,. \tag{5.10}$$

Using (5.1) and (5.10) a general difference equation is obtained, such that when the n order polynomial predictor is applied to an n order process, the n^{th} difference of the demands is:

$$D_n\, d_{t+1} = D_n e_{t+1} + \textstyle\sum_{i=1}^{n} A_i D_{n-i} e_t = D_n \epsilon_{t+1} + \sum_{i=1}^{n} D_{n-i}\gamma_{t+1}^{(i)}\,. \tag{5.11}$$

6. Predicting the Steady Model

The generating model for the steady model is from 3.1:

$$d_t = m_t + \epsilon_t \tag{6.1}$$

$$m_t = m_{t-1} + \gamma_t . \tag{6.2}$$

6.1 The Optimal Predictor and Parameters

The optimal predictor for this steady model will be shown to be the simple exponentially weighted average of past observations,

$$d^*_{t+1} = m_t^* \tag{6.3}$$

$$m_t^* = m^*_{t-1} + Ae_t \tag{6.4}$$

where e_t is the one step ahead forecast error.

The optimal value of A and the corresponding minimum variance V are

$$A = [(1 + 4R)^{1/2} - 1]/2R \tag{6.5}$$

$$V = V(\epsilon)/(1 - A) \tag{6.6}$$

where R is defined as

$$R = V(\epsilon)/V(\gamma). \tag{6.7}$$

For the trivial case in which $R = 0$ the optimal A is 1.

6.2 The General Difference Equation and Stability

The difference equation may be written as

$$e_{t+1} = (1 - A)e_t + \epsilon_{t+1} - \epsilon_t + \gamma_{t+1} . \tag{6.8}$$

Expanding the L.H.S.,

$$e_{t+1} = \epsilon_{t+1} - A\sum_{i=0}^{\infty}(1 - A)^i\epsilon_{t-i} - \sum_{i=0}^{\infty}(1 - A)^i\gamma_{t-i} ,$$

from which it is clear that the error variance and co-variances are convergent, provided

$$0 < A < 2. \tag{6.9}$$

In (6.7) as long as $R > 0$, the optimal parameter (6.5) satisfies (6.9).

6.3 Derivation of the Optimal Parameter

The difference equation (6.8) may be multipled in turn by e_{t+1} and e_t to give two equations. On taking expectations these equations are:

$$V = (1 - A)C + (1 + A)V(\epsilon) + V(\gamma) \tag{6.10}$$

$$C = (1 - A)V - V(\epsilon). \tag{6.11}$$

V and C are respectively the variance and the first co-variance of the one step ahead errors. Solving (6.10) and (6.11) for V gives

$$V = [2AV(\epsilon) + V(\gamma)]/A(2 - A). \tag{6.12}$$

Differentiating with respect to A to find the minimum variance parameter it is found that:

$$A = [(1 + 4R)^{1/2} - 1]/2R \tag{6.13}$$

with R as the ratio $V(\epsilon)/V(\gamma)$ defined in (6.7). The minimum variance is

$$V = V(\epsilon)/(1 - A), \tag{6.14}$$

and the co-variance C is zero. It is easily shown that all the error co-variances are zero simply by multiplying (6.8) by e_{t-i} starting with $i = 2$ and taking expectations. Since all the co-variances are zero, it follows that the predictor is optimal. This may also be seen by using the Box-Jenkins result (Section 4) since the first difference of the observations is a moving average process of the errors of order 2.

The conclusion is, as in [10], that for the assumed generating process the Exponentially Weighted Moving Average supplies the optimal linear least squares predictor. Furthermore, provided $0 < A \leqq 1$, it follows that any generating process for which the Exponentially Weighted Moving Average is optimal can be represented by the generating equation given by equations (6.1) and (6.2), although these equations do not necessarily describe the true generating process.

6.4 Verification of the Model and Derivation of the Optimal Parameter

The suitability of the generating process for describing the observation series may be assessed by examining the Serial Variation Curve of the data. The Serial Variation Functions of lag i is defined for a series of n observations as

$$S_i = \sum_{j=0}^{n-1-i} [d_{t-j} - d_{t-i-j}]^2/(n - i), \tag{6.15}$$

but from (6.1) and (6.2) it is seen that

$$d_t - d_{t-i} = \epsilon_t - \epsilon_{t-i} + \sum_{j=1}^{i} \gamma_{t-i+j}\,; \tag{6.16}$$

so that the expected value of S_i is

$$\text{Ex}\,[S_i] = \text{Ex}\,[d_t - d_{t-i}]^2 = 2V(\epsilon) + iV(\gamma). \tag{6.17}$$

Consequently, if the Serial Variation Function is a straight line, there is verification of the Model and therefore of the optimality of the Exponentially Weighted Moving Average predictor. Furthermore the slope of the line and the intercept for $i = 0$ supply estimates of $V(\gamma)$ and $V(\epsilon)$ and allow the calculation of the optimal value of the forecasting parameter.

An interesting example of the method occurred with the 471 observations on IBM Corporation Common Stock Closing Prices on the New York Stock Exchange. The data is tabulated in [3] (Table C.5, pp. 422 and 423). The first eight Serial Variation Functions were calculated as shown in Table 1. The Serial Variation Curve is represented extremely well by a straight line, but, unfortunately for would-be-speculators, the line passess through the origin showing that $V(\epsilon) = 0$ and $A = 1$; so that the generating process reduces to

$$d_t = d_{t-1} + \gamma_t$$

TABLE 1

The Serial Variation Function for the I.B.M. Stock Prices

	Lag							
	1	2	3	4	5	6	7	8
S.V.F.	3.3	6.7	9.9	13.2	16.4	19.6	23.4	27.5

with $V(\gamma_t) = 3.3$; and the optimal predictor states simply that the best estimate of future prices is the latest quoted price.

Good results showing the adequacy of the generating process have been obtained on a number of demand series, but, because of the usual shortage of data, general robustness, and precision, it is recommended that for short-term sales forecasting the optimal parameter is derived from a simulation of the predictor on the original demand data.

6.5 Sensitivity to Departures from the Optimal Predictor

Experiences of short-term sales forecasting have shown that the variance and co-variance of the forecasting errors are relatively insensitive to quite large departures from the optimal forecasting parameter. To illustrate this, equations (6.10) and (6.11) have been used to construct graphs 1A and 1B. The general conclusions reached from these graphs are:

a) Overestimation of A is less serious than the equivalent underestimation.
b) The variance and co-variance of the errors are insensitive to errors in determining the optimal value of the forecasting parameter.

Some routine sales forecasting systems operate with the parameter chosen to be either 0.1 or 0.2. That this is not a bad decision is confirmed by this sensitivity investigation and the fact that for this type of application the optimal parameter rarely exceeds 0.25.

6.6 The Variance of the n Step Ahead Prediction

The n step ahead predictor at time t is

$$F_n(t) = m_t^*.$$

The error E_n for n steps ahead is then

$$E_n = d_{t+n} - m_t^*,$$

which may be written as

$$E_n = d_{t+n} - m^*_{t+n-1} + \sum_{i=1}^{n-1} [m^*_{t+n-i} - m^*_{t+n-i-1}] \tag{6.18}$$

$$= e_{t+n} + A \sum_{i=1}^{n-1} e_{t+n-i};$$

so that for the optimal predictor

$$V(E_n) = [1 + (n-1)A^2]V(e). \tag{6.19}$$

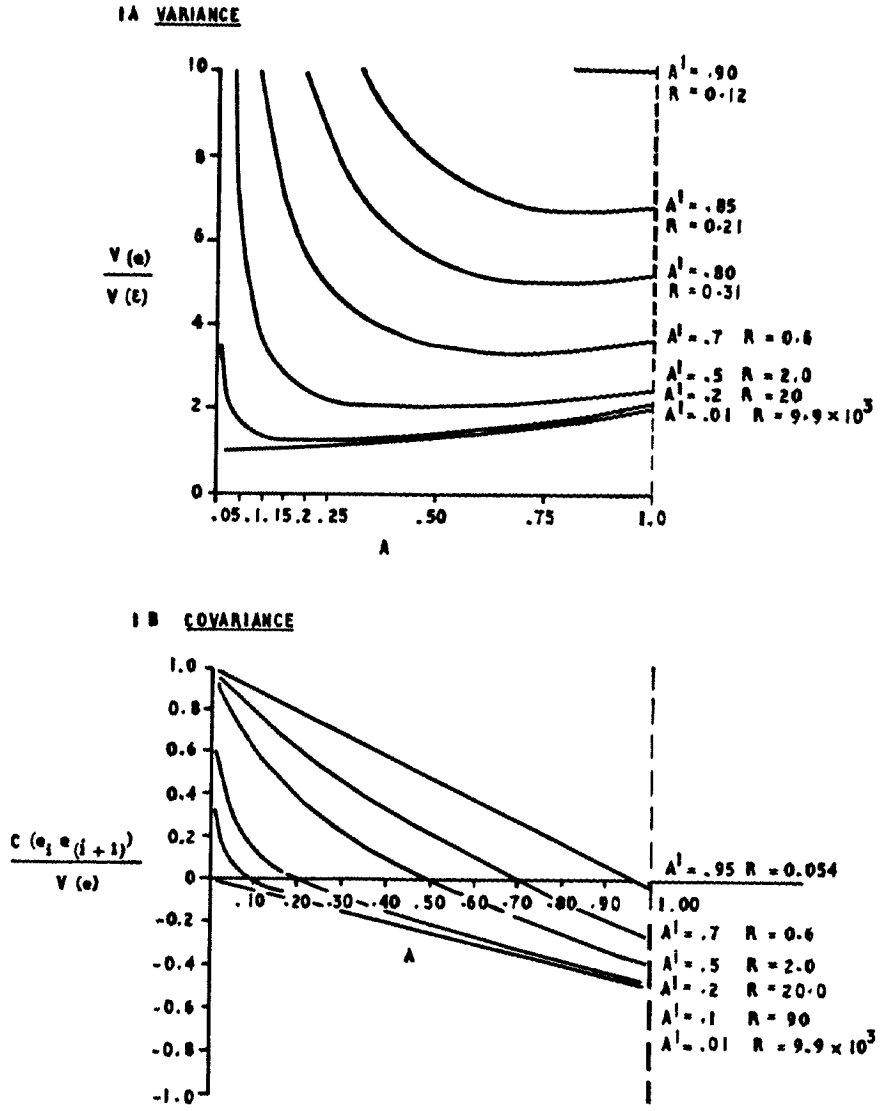

FIG. 1. Steady Random Walk (E.W.M.A. Predictor): The effect on the one step ahead forecast variance $V(e)$ and first covariance $C(e_i e_{(i+1)})$ of departures from the optimal A. (Optimal A^1 defined as $A^1 = ((1 + 4R)^{1/2} - 1)/2R$ where $R = V(\ )/V(\ )$).

Sales forecasts are often prepared over a period called the lead time. A lead time of n periods is defined as a cumulative forecast for the next n periods. For the steady model the prediction is simply nm_t^* so that using (6.18) the lead time forecast error is

$$\sum_{k=0}^{n-1} [1 + kA]e_{t+n-k}$$

which has a variance

$$n[1 + (n-1)A + (n-1)(2n-1)A^2/6]V(e). \tag{6.20}$$

This shows quite clearly that the lead time error depends on the value of the optimal parameter A and that the arguments which have taken place about this error just being a function of the lead time are false. Table 2 shows how the lead time variance depends on the lead time for $A = 0.1$ and for $A = 0.2$.

7. Predicting the Linear Growth Model

The generating model for the linear growth model is from 3.2

$$d_t = m_t + \epsilon_t \tag{7.1}$$

$$m_t = m_{t-1} + b_t + \gamma_t \tag{7.2}$$

$$b_t = b_{t-1} + \delta_t . \tag{7.3}$$

7.1 The Optimal Predictor and Parameters

The optimal predictor for this linear growth model will be shown to be the particular form of the Box and Jenkins predictor recommended by Holt as

$$d^*_{t+1} = m_t^* + b_t^* \tag{7.4}$$

$$m_t^* = m^*_{t-1} + b^*_{t-1} + A_1 e_t \tag{7.5}$$

$$b_t^* = b^*_{t-1} + A_2 e_t . \tag{7.6}$$

The optimal values of A_1 and A_2 and the minimum variance V are given by the equations

$$V(\epsilon) = (1 - A_1)V \tag{7.7}$$

$$V(\gamma) = (A_1^2 + A_1A_2 - 2A_2)V \tag{7.8}$$

$$V(\delta) = A_2^2 V. \tag{7.9}$$

7.2 The General Difference Equation and Stability

The difference equation may be written as

$$e_{t+1} = ae_t + be_{t-1} = E_{t+1} , \tag{7.10}$$

TABLE 2

Lead Time Error Variances as Multiples of the One Step Ahead Variance

Variance	Lead Time						
	1	2	3	4	6	12	24
$A = 0.1$	1	2.2	3.7	5.3	9.6	28	90
$A = 0.2$	1	2.4	4.4	7.0	14	58	302

with the definitions

$$a = 2 - A_1 - A_2 \tag{7.11}$$

$$b = -(1 - A_1) \tag{7.12}$$

$$E_{t+1} = \epsilon_{t+1} - 2\epsilon_t + \epsilon_{t-1} + \gamma_{t+1} - \gamma_t + \delta_{t+1}\,. \tag{7.13}$$

Expanding the R.H.S. of (7.10)

$$e_{t+1} = \sum_{i=0}^{\infty} w_i E_{t+1-i} \tag{7.14}$$

where

$$w_k = aw_{k-1} + bw_{k-2}\,,$$

and

$$w_0 = 1; \qquad w_1 = \alpha + \beta = a; \qquad w_2 = a^2 + b$$

where if α and β are roots of the equation

$$x^2 - ax - b = 0$$

in general

$$w_{k-1} = (\alpha^k - \beta^k)/(\alpha - \beta).$$

From (7.13) and (7.14)

$$V(e) = \sum_{k=0}^{\infty} [(D_2 w_k)^2 V(\epsilon) + (D_1 w_k)^2 V(\delta) + w_k^2 V(\delta)].$$

The error variance $V(e)$ tends to a limit provided $|\, w_k \,| < 1$ which implies $|\, \alpha \,|$ and $|\, \beta \,|$ are less than 1. It follows that for stability

$$|(a \pm (a^2 + 4b)^{1/2})/2\,| < 1,$$

and substituting for a and b using (7.11) and (7.12) the stability conditions are:

$$0 > A_2 \leqq (A_1 + A_2)^2/2 \tag{7.15}$$

$$2A_1 + A_2 < 4. \tag{7.16}$$

7.3 The Stability of the Optimal Parameters

The optimal parameters given by equations (7.7), (7.8) and (7.9) always satisfy the stability conditions since they are restricted by the fact:

$$0 < V(\epsilon); \qquad V(\gamma); \qquad V(\delta) < V;$$

so that the optimal parameters must lie within the subregion of the stability area defined by

$$0 < A_2 \leqq (A_1 + A_2)^2/(2 + A_1 + A_2) < 1 \tag{7.17}$$

$$0 < A_1 < 1. \tag{7.18}$$

It may be noted that some of the 'less' than restrictions of equation (7.15),

(7.17), (7.18) become 'less than or equal' for the set of optimal parameters in the particular cases when at least one of $V(\epsilon)$, $V(\gamma)$, $V(\delta)$ assumes a zero value.

7.4 Derivation of the Optimal Predictor

The difference equation

$$e_{t+1} = ae_t + be_{t-1} + E_{t+1} \tag{7.19}$$

may be multiplied in turn by e_{t+1} and e_t to give two equations. On taking expectations and defining

$$\mathrm{Ex}\,[e_{t-j}E_t] = y_j\,, \tag{7.20}$$

the equations are

$$\begin{aligned}(1 - b^2)S^2 &= a(1 + b)C_1 + by_2 + y_0 \\ aS^2 &= (1 - b)C_1 - y_1\,.\end{aligned} \tag{7.21}$$

S^2 and C_1 are the variance and first co-variance of the errors e. Eliminating S^2 from the equation:

$$(1 + b)[(1 - b)^2 - a^2]C_1 = (1 - b^2)y_1 + aby_2 + ay_0\,. \tag{7.22}$$

Define a, b and a constant V such that

$$bV = V(\epsilon) = y_2 \tag{7.23}$$

$$aV = (4 - a)V(\epsilon) + V(\gamma) = -y_1 \tag{7.24}$$

$$V = (6 - 4a + a^2 + b)V(\epsilon) + (2 - a)V(\gamma) + V(\delta) = y_0 \tag{7.25}$$

$$y_j = 0 \quad \text{for} \quad |j| > 2. \tag{7.26}$$

That such a set of values exists which satisfies the stability conditions is shown in Appendix 1. Then the R.H.S. of equation (7.22) reduces to

$$V[-a(1 - b^2) + ab^2 + a] = 0,$$

and using (7.11), (7.12), (7.17) and (7.18) the coefficient of C_1 is

$$A_1A_2[4 - 2A_1 - A_2] > 0.$$

It follows that $C_1 = 0$ and from (7.21) and (7.24) that $S^2 = V$. All the other co-variances C_i may be shown to be zero simply by multiplying (7.19) in turn by e_{t-i} $(i = 1 \cdots \infty)$ and taking expectations. It follows that since all the co-variances are zero the predictor is optimal. The Box-Jenkins result (section 4) may be applied to prove optimality since the second difference of the demand series for the above parameters may be expressed as a moving average process of the forecasting errors of order 3, the errors being uncorrelated with zero mean and constant variance.

7.5 Verifying the Model

It is possible to make some assessment of the suitability of the generating process for describing a data set by examining the Serial Variation Curve of the first differences of the data.

Defining $D(t)$ as the first difference of the demand series at time t so that:

$$D(t) = d_t - d_{t-1}$$

it follows from (7.1), (7.2) and (7.3) that

$$D(t) = \epsilon_t - \epsilon_{t-1} + \gamma_t + \sum_{j=0}^{\infty} \delta_{t-j},$$

and that

$$\begin{aligned} \text{Ex}\,[D(t) - D(t-i)]^2 &= 4V(\epsilon) + 2V(\gamma) + iV(\delta); \qquad i \geqq 2 \\ &= 6V(\epsilon) + 2V(\gamma) + V(\delta); \qquad i = 1. \end{aligned}$$

Thus the Serial Variation Curve of the first differences is expected to be a straight line with gradient $V(\delta)$ for $i \geqq 2$. In general Sales Forecasting applications, this method is not recommended for determining the variances $V(\epsilon)$, $V(\gamma)$ and $V(\delta)$, for the errors are usually large; and so this will only give a rough indication as to the adequacy of the model. The covariances of the second difference of the demand series may be examined to provide a test of adequacy, but in practise the parameters are best determined by simulation.

7.6 Sensitivity to Departures from the Optimal Parameters

In sales forecasting, the variance and covariance of the prediction errors are relatively insensitive to what might be regarded as quite large departures from the optimal forecasting parameters. To illustrate this graphs 2A and 2B have been constructed for typical values of the optimum parameters.

The main conclusions to be drawn from this investigation are that for sales products:

(a) The prediction variance $V(e)$ is more sensitive to departures from the optimum in A_2 than in A_1.

(b) The covariance between consecutive errors is much more sensitive to departures from the optimum in A_1 than in A_2.

(c) Generally, overestimation of the parameters is less serious than the equivalent underestimation of their values. (In practice, values of $A_1 > 0.25$ and $A_2 > 0.15$ are rarely found to be appropriate in routine short-term sales forecasting systems.)

7.7 The Variance of the n Step Ahead Predictor

The n step ahead predictor at time t is

$$F_n(t) = m_t^* + nb_t^*$$

so that the error E_n for n steps ahead is

$$E_n = d_{t+n} - m_t^* - nb_t^*,$$

which may be written as

$$\begin{aligned} E_n = d_{t+n} - m^*_{t+n-1} - b^*_{t+n-1} + \sum_{i=1}^{n-1} [m^*_{t+n-i} + ib^*_{t+n-i} \\ - m^*_{t+n-1-i}(i+1)b^*_{t+n+1-i}] \\ = e_{t+n} + \sum_{i=1}^{n-1} [A_1 + iA_2]e_{t+n-i} \end{aligned}$$

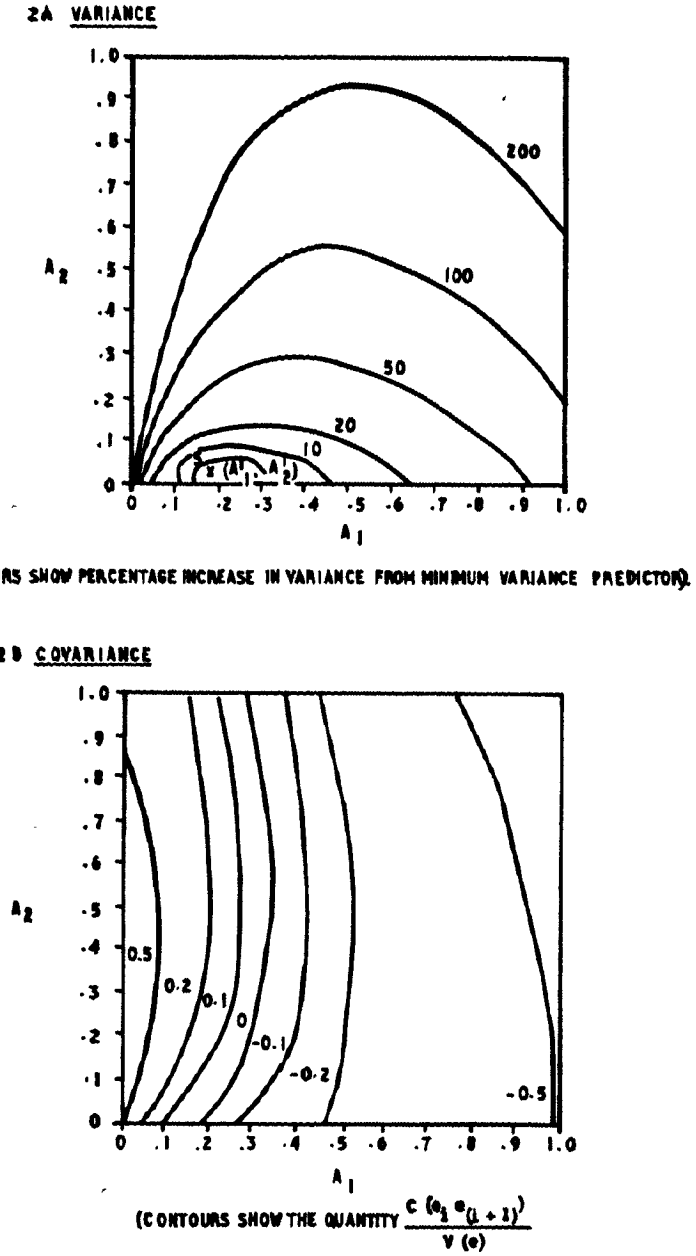

FIG. 2. Growth Model (Holt's Growth Predictor): The effect on the one step ahead forecast variance $V(e)$ and first covariance $C(e_t,e_{(t+1)})$ of departures from the optimal forecasting parameters A_1 and A_2. (For the illustration $A_1' = 0.2$; $A_2' = 0.022$).

so that for the optimal predictor

$$V(E_n) = [1 + (n-1)A_1^2 + n(n-1)A_1A_2 + n(n-1)(2n-1)A_2^2/6]V(e). \tag{7.27}$$

The lead time forecast error for the immediate n periods is:

$$\sum_{j=0}^{n-1} [1 + jA_1 + j(j+1)A_2/2]e_{t+n-j}$$

which has a variance

$$\sum_{j=0}^{n-1} [1 + jA_1 + j(j+1)A_2/2]^2 V(e). \tag{7.28}$$

This lead time variance is tabulated (see Table 3) for two typical sets of values of A_1 and A_2 when A_1 and A_2 are related so that

$$A_1 = 1 - \beta^2$$
$$A_2 = (1-\beta)^2$$

8. Brown's Predictor and the Linear Growth Model

In short-term sales forecasting systems growth products are nearly always predicted using the linear growth model. In the author's experience, which is associated with some thousands of products, higher order polynomials have never been needed for the short-term predictions.

Many practitioners prefer to predict growth products using Brown's method of exponentially weighted regression rather than Holt's method, although this has been criticized on theoretical grounds. The results of section 7 allow comparisons to be made between the two methods which will confirm the practitioners' experiences that there is nothing to gain by using Holt's method instead of Brown's.

8.1 Brown's Growth Predictor

The method of exponentially weighted regression can be regarded as a formalization of the natural approach to short-term sales forecasting. Generally, a salesman's approach to forecasting when faced with a graph of customer demand is to draw some curve, usually a straight line, through the observations and to derive forecasts by projecting the curve into the future. In drawing the curve he will make it fit the most recent observations the best.

TABLE 3

Lead Time Error Variances as Multiples of the One Step Ahead Variance

Variance	Lead Time						
	1	2	3	4	6	12	24
$\beta = 0.95$	1	2.2	3.7	5.4	9.7	32	153
$\beta = 0.9$	1	2.4	4.4	7.1	15	72	524

Brown's method formalizes this procedure in choosing the coefficients of the curve which minimize the exponentially weighted sum of squares of the differences between the observed demand and the curve.

For example, with the linear growth predictor

$$F_k(t) = m_t + kb_t$$

using a given discount factor β, m_t and b_t are chosen so that the quantity minimized is:

$$S_t = \sum_{i=0}^{\infty} \beta^i [d_{t-i} - m_t + ib_t]^2$$

so that m_t and b_t satisfy the recurrence relationships:

$$m_t = m_{t-1} + b_{t-1} + (1 - \beta^2)e_t$$

$$b_t = b_{t-1} + (1 - \beta)^2 e_t .$$

The forecasting parameter β is chosen to minimize the variance of the forecasting errors.

Intuitively it is expected that, for non-seasonal products which sell to a large number of customers, near optimal short-term forecasts will be derived.

8.2 A Comparison of Holt's and Brown's Methods

When Holt's method is used for prediction, it is inferred that this is the optimal predictor. If this is true and the forecasting parameters lie in the region defined by equations (7.17) and (7.18), the generating process of the demand sequence can be expressed as in (7.1), (7.2), and (7.3). The converse has also been shown true that if the generating process takes the form of (7.1), (7.2) and (7.3) which is the most reasonable model for a product to be predicted by the Holt growth predictor, then Holt's method is optimal and the forecasting parameters satisfy

$$0 < A_2 \leqq A_1 < 1$$

$$A_1^2 + A_1A_2 - 2A_2 \geqq 0.$$

The region in which these parameters lie is shown in Figure 1 together with the curve representing Brown's predictor, i.e.,

$$A_1 = 1 - \beta^2; \qquad A_2 = (1 - \beta)^2$$

$$0 < \beta < 1.$$

It will be noticed that the region is of extremely small area, particularly when A_1 is less than 0.25, and for sales products A_1 rarely exceeds this value. Using equations (7.21) it is possible to assess what additional errors will be incurred if instead of using the Holt method the best Brown parameters are used. Table 4 shows these additional errors in the two extreme cases when

$$A_1^2 + A_1A_2 - 2A_2 = 0$$

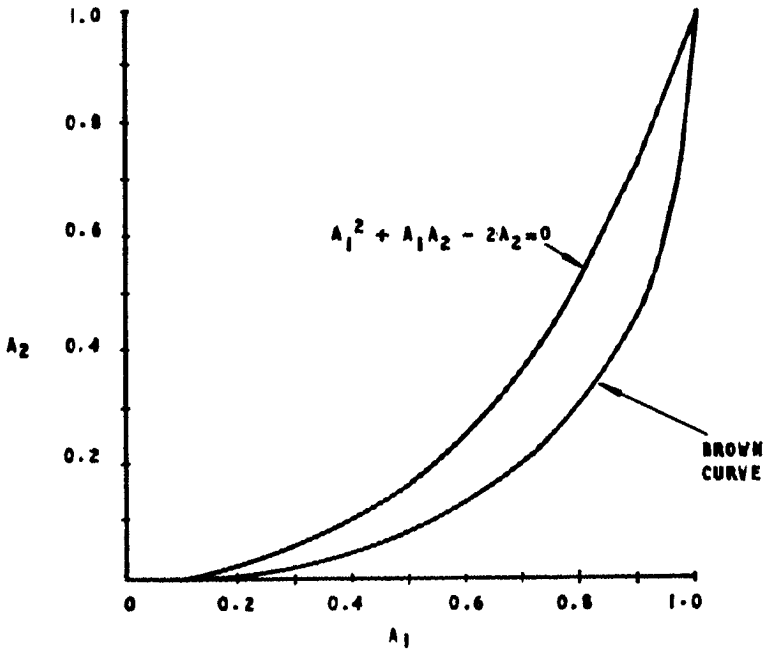

FIG. 3. The region in which the linear growth parameters lie (the region is defined by $0A_2A_11; A_1^2 + A_1A_2 - 2A_20$).

TABLE 4

The Percentage Increase in the Standard Error of the One Step Ahead Predictor When Brown's Method Is Used Instead of the Optimal Predictor

Optimal Predictor	A_1						
	0.05	0.10	0.15	0.20	0.25	0.50	0.90
Thiel et al.	0.01	0.15	0.27	0.35	0.44	0.75	0.09
$A_2 = 0$	0.2	0.6	0.9	1.2	1.6	3.5	7.2

which is equivalent to Thiel, Wage and Nerlove's version of the Holt predictor (i.e. $V(\gamma) = 0$) and when

$$A_2 = 0$$

when the predictor has a constant slope b_t. It is seen that in the region of interest

($A_1 < 0.25$) that, even if there were an infinite amount of data with which to establish the forecasting parameters of Holt's method, the maximum penalty would be an increase in the one step ahead standard error of less than 1.6 percent.

This result confirms short-term sales forecasting practitioners' experience, which is that

1) Sampling errors associated with the estimation of Holt's forecasting parameters outweigh any theoretical advantage which the predictor may be thought to possess, and, therefore, it is more convenient to work with Brown's predictor.
2) The concepts of Brown's predictor are more readily accepted by Sales and Distribution clerks who operate the forecasting systems.

9. Predicting the General Polynomial Model

The generating model for the general polynomial model of order n is from (3.3)

$$d_t = m_t^{(1)} + \epsilon_t \tag{9.1}$$

$$m_t^{(i)} = m_{t-1}^{(i)} + m_t^{(i+1)} + \gamma_t^{(i)} \qquad (i = 1 \cdots n) \tag{9.2}$$

with $m_t^{(n+1)}$ defined as zero; $V(\epsilon)$ 0 and $V(\gamma^{(n)}) > 0$.

9.1 The Optimal Predictor

The optimal predictor for this general model will be shown to be the general Box-Jenkins Polynomial Predictor which can be written as in equations (2.2), (2.3) and (2.4), i.e.,

$$d_{t+1}^* = \sum_{j=1}^{n} m_t^{(j)*} \tag{9.3}$$

$$m_t^{(i)*} = \sum_{j=i}^{n} m_{t-1}^{(j)*} + A_i e_t \tag{9.4}$$

$$e_t = d_t - d_t^*. \tag{9.5}$$

9.2 The General Difference Equation and Stability

The difference equation is, from (5.11),

$$D_n e_t + \sum_{i=1}^{n} A_i D_{n-i} e_{t-1} = E_t(n) \tag{9.6}$$

which may be written as

$$e_t = \sum_{i=1}^{n} a_i e_{t-i} + E_t, \tag{9.7}$$

where

$$a_i = (-1)^{i+1}\left[\binom{n}{i} - \sum_{j=1}^{n-i+1}\binom{n-j}{i-1} A_j\right], \tag{9.8}$$

$$a_0 = -1, \tag{9.9}$$

and

$$E_t = D_n \epsilon_t + \sum_{i=1}^{n} D_{n-i}\gamma_t^{(i)}. \tag{9.10}$$

Expanding (9.7)

$$e_t = \sum_{i=0}^{\infty} w_i E_{t-i} \tag{9.11}$$

where

$$w_k = \sum_{i=1}^{n} a_i w_{k-i} \tag{9.12}$$

$$w_0 = 1 \quad \text{and} \quad w_j = 0 \quad \text{for} \quad j < 0. \tag{9.13}$$

It follows from (9.10), (9.11) and (9.12) that if the roots of the equation

$$\sum_{i=0}^{n} a_i x^{n-i} = 0$$

lie within the unit circle, the predictor is stable.

9.3 The Optimum Parameters

The optimum parameters and the forecasting variance V will be shown to be defined by the following equations

$$a_i = -\mathrm{Ex}\,[E_t e_{t-i}]/V \qquad (i = 1 \cdots n) \tag{9.14}$$

$$V = \mathrm{Ex}\,[E_t e_t] \tag{9.15}$$

$$\mathrm{Ex}\,[E_t e_{t-i}] = \mathrm{Ex}\,[\sum_{j=0}^{n-i} w_i E_t E_{t-i-j}] \qquad (i = 1 \cdots n).$$

These parameters always exist and give a stable predictor with minimum variance

$$V = V(\epsilon)/(1 - A_1).$$

9.4 Proof of Optimality

For the n order process define y_t so that

$$y_t = D_n d_t = D_n \epsilon_t + \sum_{i=1}^{n} D_{n-i}\gamma_t^{(i)}; \tag{9.16}$$

then y_t represents a stationary process with covariances C_s such that

$$C_s = 0 \quad \text{for} \quad s > n.$$

Define $g_y(z)$ as

$$g_y(z) = \sum_{-\infty}^{\infty} C_s z^s = \sum_{-n}^{n} C_s z^s. \tag{9.17}$$

If $g(z)$ has a root α then it also has a root α^{-1}; so that δ_j can be defined as

$$g_y(z) = (\sum_0^n \delta_j z^j)(\sum_0^n \delta_j z^{-j}) \tag{9.18}$$

Defining variables e_j by

$$y_t = \sum_0^n \delta_j e_{t-j} = \delta(D) e_t \tag{9.19}$$

so that

$$e_t = y_t/\delta(D) = \sum_0^{\infty} \theta_j y_{t-j},$$

it follows that

$$g_y(z) = \delta(z)\delta(z^{-1}) g_\epsilon(z), \tag{9.20}$$

and comparing with (9.18)

$$g_e(z) = 1,$$

so that the variables e_j are uncorrelated. Furthermore since the characteristic function of y_t is the same for all t,

$$\phi(y) = \phi(y_t) = \prod_{j=0}^{n} \delta_j \phi(e_{t-j})$$

and

$$1 = \phi(y_t)/\phi(y_{t-1}) = \phi(e_t)/\phi(e_{t-n-1});$$

so that e_t and e_{t-n-1} have the same distribution function. Define

$$x_{t,k} = D_k y_t = D_{n+k}\epsilon_t + \sum_{i=1}^{n} D_{n+k-i}\gamma_t(i)$$

$$x_{t,k} = \sum_{0}^{n+k} \eta_j e_{t-j}$$

where the η_i are constants. Then

$$\phi(x_{t,k}) = \prod_{0}^{n+k} \eta_j \phi(e_{t-j})$$

and

$$1 = \phi(x_{t,k})/\phi(x_{t-1,k}) = \phi(e_t)/\phi(e_{t-n-k-1}),$$

which shows that the errors e_t are identically distributed. It follows from the Box-Jenkins result that the n order polynomial predictor is optimal for the n order generating process and that the forecasting errors are the e_t's so that a stable predictor always exists. If $a_i\ (i = \phi \cdots n)$ are the optimal parameters of the general difference equation then

$$e_t - \sum_{i=1}^{n} a_i e_{t-i} = E_t\ ;$$

so that multiplying in turn by e_{t-j} and taking expectations,

$$V(e) = \mathrm{Ex}\,[E_t e_t]$$

and

$$a_i V(e) = -E_t e_{t-i}\ .$$

In particular

$$a_n V(e) = -E_t e_{t-n} = V(e).$$

So that always

$$V(e) = V(e)/(1 - A_1).$$

Appendix 1. Linear Growth Model

Identities

$$y_2 = V(\epsilon)$$

$$-y_1 = (4 - a)V(\epsilon) + V(\gamma)$$

$$y_0 = (6 - 4a + a^2 + b)V(\epsilon) + (2 - a)V(\gamma) + V(\delta)$$

From Equation (7.13) it follows, defining

$$x_i = \mathrm{Ex}\,[E_t E_{t-j}],$$

that

$$x_2 = V(\epsilon)$$
$$x_1 = -[4V(\epsilon) + V(\gamma)]$$
$$x_0 = 6V(\epsilon) + 2V(\gamma) + V(\delta)$$
$$x_j = 0 \quad \text{for} \quad |j| > 2.$$

From 7.14

$$e_{t+1} = E_{t+1} + aE_t + (a^2 + b)E_{t-1} + \textstyle\sum_{i=3}^{\infty} w_i E_{t+1-i},$$

thus

$$y_0 = E_{t+1}e_{t+1} = x_0 + ax_1 + (a^2 + b)x_2$$
$$y_1 = E_{t+2}e_{t+1} = x_1 + ax_2$$
$$y_2 = E_{t+3}e_{t+1} = x_2$$

from which the identities follow.

Existence

Equations (7.23), (7.24), and (7.25) define A_1, A_2 and V which satisfy (7.7), (7.8), and (7.9). Denoting

$$A_3 = 1/V$$

the latter equation may be written as:

$$A_3V(\epsilon) = 1 - A_1$$
$$A_3V(\gamma) = A_1^2 + A_1A_2 - 2A_2$$
$$A_3V(\delta) = A_2^2.$$

So that, eliminating A_3,

$$A_2^2R_1 = 1 - A_1$$
$$A_2^2R_2 = A_1^2 + A_1A_2 - 2A_2$$

where

$$R_1 = V(\epsilon)/V(\delta) > 0; \qquad R_2 = V(\gamma)/V(\delta) > 0.$$

But the parabola and the hyperbola intersect in a point within the region

$$0 < A_1 < 1$$
$$0 < A_2 < 1.$$

This can easily be demonstrated since the parabola passes through the points [1,0] and [0, 1] with a segment of the curve joining these two points lying wholly

within the above region. The hyperbola passes through the points [0, 0] and $[1, ((1 + 4R_2)^{1/2} - 1)/2R_2]$ and again a segment of the curve lies wholly within the above region. It follows that $V > V(\epsilon); V(\gamma); V(\delta)$.

References

1. Box, G. E. P. and Jenkins, G. M., "Some Statistical Aspects of Adaptive Optimization and Control", *J.R. Statist. Soc., Sec. B*, Vol. 24, 1962, pp. 297–343.
2. Brown, R. G., *Statistical Forecasting for Inventory Control*, New York, McGraw Hill, 1959.
3. ——, *Smoothing Forecasting and Prediction*, Prentice Hall, 1963.
4. Cox, D. R., "Prediction by Exponentially Weighted Moving Averages and Related methods," *J.R. Statist. Soc., Sec. B*, Vol. 23, 1961, pp. 414–422.
5. Harrison, P. J. and Davies, O. L., "The Use of Cumulative Sum (Cusum) Techniques for the Control of Routine Forecasts of Product Demand," *Operations Research (J.O.R.S.A.)* Vol. 12, No. 2, 1964.
6. Harrison, P. J., "Short-Term Sales Forecasting," *Applied Statistics*, Vol. XIV, Nos. 2 and 3, 1965.
7. Harrison, P. J. and Scott, F. A., "A Development System for Use in Short-Term Sales Forecasting Investigations," a paper read at the Operational Research Societies Annual Conference at Shrivenham, 1965.
8. Holt, C. C., *Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages*. Carnegie Institute of Technology, Pittsburgh, Pennsylvania, 1957.
9. Holt, C. C., Modigliani, F., Muth, J. F., and Simon, H. A., *Planning Production Inventories and Work Force*. London, Prentice Hall, 1960.
10. Muth, J. F., "Optimal Properties of Exponentially Weighted Forecasts," *J. Amer. Statist. Soc.*, Vol. 55, 1960, pp. 299–306.
11. Nerlove, M. and Wage, S., "On the Optimality of Adaptive Forecasting," *Man Sc.*, Vol. 10, No. 2, 1964, pp. 207–229.
12. Thiel, H. and Wage, S., "On Some Observations on Adaptive Forecasting," *Management Science*, Vol. 10, No. 2, 1964, pp. 198–206.
13. Whittle, P., "Prediction and Regulation," E.U.P., 1963.