Forecasting and Stock Control for Intermittent Demands

J. D. CROSTON
P. & O. Steam Navigation Co., London

Exponential smoothing is frequently used for the forecasts in stock control systems. The analysis given shows that intermittent demands almost always produce inappropriate stock levels. Demand for constant quantities at fixed intervals may generate stock levels of up to double the quantity really needed. A method of overcoming these difficulties is described, using separate estimates of the size of demand, and of the demand frequency. The rules for setting the safety stock levels have also to be adjusted before consistent protection can be obtained against being out of stock.

INTRODUCTION

EXPONENTIAL smoothing is frequently used for forecasting demand in a routine stock control system (Brown[4]) when large numbers of products may be involved. The basic model for the algorithms used is given in a subsequent section.

An audit of one such system, which had been operating for over a year revealed that for some low demand items the stock levels of the system appeared to be excessive. Historical records showed that for some series of demands which had been stable over a considerable period the stocks were appreciably higher than the maximum demand which had occurred. Further investigation showed that these errors appeared associated with items for which the demand was intermittent, and usually for several items at a time.

An analysis of the response of the forecasting elements of the system has therefore been carried out, to determine the types of error which arise if there are many review intervals in which no demand occurs.

Methods of avoiding these errors are described which also allow a specific protection to be provided against being out of stock, whether the demands occur every interval or not. A single system can therefore be employed for all types of demand.

BASIC MODEL

Consider a routine stock control system in which updating occurs at fixed, unit time intervals, which are much shorter than the times between successive demands for the product. There will therefore be frequent occasions on which the demand will be zero, although the average demand may be for several units.

Although the inter-arrival time and the size of demand are usually both random variables let us first consider uniform demands of magnitude μ occurring every p review intervals, where p is an integer. If the first demand occurs at time $t = 1$, the demand y_t is represented by:

$$y_t = \begin{cases} \mu, & t = np+1, \quad n = 0, 1, 2, \ldots, \\ 0, & \text{otherwise,} \end{cases}$$

where n indexes the non-zero demands.

The rules for stock replenishment are frequently based upon a linear combination of the estimates of average demand, and the mean absolute deviation of the one period ahead forecasting errors. A typical procedure using single-stage exponential smoothing would be:

$$e_t = y_t - \hat{y}_{t-1}, \tag{1.1}$$

$$\hat{y}_t = \hat{y}_{t-1} + \alpha e_t, \tag{1.2}$$

$$m_t = (1-\alpha)\, m_{t-1} + \alpha |e_t|, \tag{1.3}$$

$$R_t = \hat{y}_t + k m_t, \tag{1.4}$$

where y_t is the demand at time t, $\hat{y}_t$ the forecast of the average demand made at time t, and used as a one step ahead predictor of the demand at time $t+1$, e_t is the error of the predictor, m_t the estimated mean absolute deviation of the errors, R_t the replenishment level to which the stock is raised and k is a constant for all products in the system.

The forecast given by (1.1) and (1.2) is a weighted average of the past observations, and is the simplest form of those systems described by Holt,[2] Brown,[4] Winter,[6] Ward,[7] Harrison[8] and Burgin,[9] all of whom suggest the use of small values of α, of the order of 0·1–0·2. Box and Jenkins[1] have shown that for a number of series, particularly the IBM Stock Prices examples, their methods of analysis may result in a parsimonious model of this simple form, but they emphasize that the optimal weighting constant may correspond to a value of α close to 1·0.

The behaviour of the systems is generally robust against changes in the pattern of demand, but it is shown below that serious errors arise if the demand is intermittent, and there are review intervals in which no demand occurs. If the forecasts are used to control a stock system the level required to give a specified protection against being out of stock cannot be stipulated, unless the inter-arrival interval is known.

The starting values of the estimators $\hat{y}_t, m_t$ for $t = 0$ are based on the previous demands, and the effect of the initial assumptions decrease with time, so that for regular intermittent demands the pattern of forecast, error, and mean absolute deviation will become that shown in Figure 1. The estimates in equations (1.1)–(1.3) will be updated every review interval but as the highest values of replenishment R_t occur immediately following a demand, at which time the stock is at a minimum, replenishment will only be made following each demand, and it is only the values of the estimators $\hat{y}$ and m at such times which are relevant. The values so used are designated $\hat{y}^*$ and m^*.

The safety stock to protect against the variability of demand is regulated by the arbitrary constant k, which is normally given a value of about 3 in a system with a one period lead time, giving approximately 1 per cent probability of stock shortage if the demand distribution is normal.

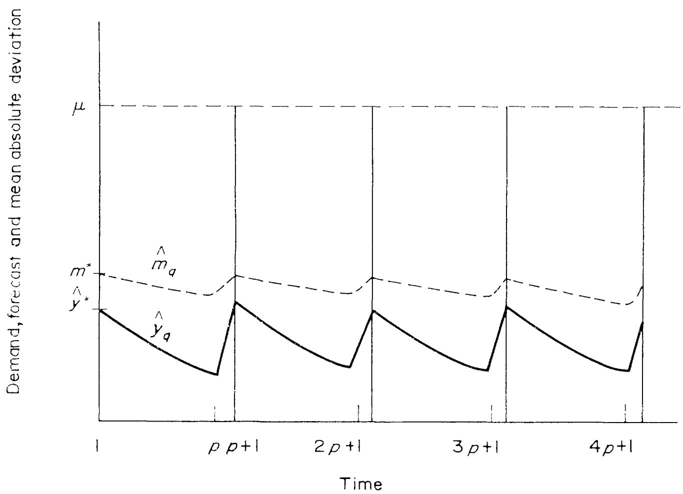

FIG. 1. *Response to demand occurring at time $t = 1$ and thereafter at intervals, p.*

INTERMITTENT DEMANDS AT REGULAR INTERVALS

Let the demand consist of regular orders for μ units recieved every p review intervals, as shown in Figure 1. Consider the behaviour of the estimators over a single cycle as shown in Figure 2.

It is shown in Appendix A that the forecasts $\hat{y}^*$ and mean absolute deviation m^* on which replenishment will be based are:

$$\hat{y}^* = \frac{\mu\alpha}{1-\beta^p}, \quad m^* = \frac{\alpha\mu\{1-\beta^{p-1}[1-\alpha(p-1)]\}}{(1-\beta^p)^2}, \tag{2}$$

where $\beta = 1-\alpha$. The replenishment level is then obtained from 1·4. The values have been calculated for a range of inter-arrival times of 1–15 review intervals, with smoothing constants α between 0·05 and 1·0 in Table 1, and are shown graphically in Figure 3 for $\mu = 10$.

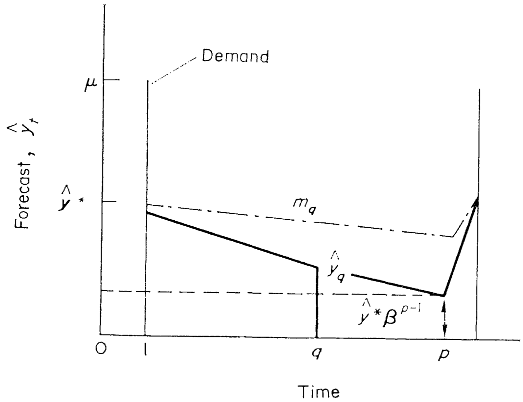

FIG. 2. *One cycle for demand of μ units every p intervals when equilibrium reached.*

TABLE 1. FORECAST OF AVERAGE y^* AND ESTIMATE OF MEAN ABSOLUTE DEVIATION m^* USED FOR CALCULATING REPLENISHMENT LEVEL R, FOR REGULAR DEMANDS FOR 10 UNITS EVERY p REVIEW INTERVALS, WITH VARIOUS SMOOTHING CONSTANTS α

α		0·05			0·1			0·3			0·6			1·0		
p	$\bar{y}$	$\hat{y}^*$	m^*	R	$\hat{y}^*$	m^*	R	$\hat{y}^*$	m^*	R	$\hat{y}^*$	m^*	R	$\hat{y}^*$	m^*	R
1	10	10	0	10	10	0	10	10	0	10	10	0	10	10	0	10
2	5·0	5·1	5·1	21	5·3	5·3	22	5·0	5·9	24	7·1	7·1	29	10	10	40
3	3·3	3·5	4·6	18	3·7	4·8	19	4·6	5·6	22	6·4	7·1	28	10	10	40
4	2·5	2·7	3·9	15	2·9	4·1	16	3·9	5·0	19	6·2	6·6	26	10	10	40
5	2·0	2·2	3·4	13	2·4	3·6	14	3·6	4·5	18	6·1	6·3	25	10	10	40
10	1·0	1·2	2·0	8	1·5	2·3	9	3·1	3·4	14	6·0	6·0	24	10	10	40
15	0·67	0·9	1·5	6	1·3	1·7	7	3·0	3·1	13	6·0	6·0	24	10	10	40

It is seen that the forecasts of demand $\hat{y}^*$ underestimate the size of the demands which occur, as would be expected. However, they also overestimate the long term average demand $\bar{y}$ where $\bar{y} = \mu/p$. The average demand appears in many stock replenishment and batch size formulae, and the percentage errors implicit in using $\hat{y}^*$ for this estimate are $100(\alpha p/(1-\beta^p)-1)$ and are shown in Figure 4. The error decreases with decreasing α, giving further justification for the empirical recommendation of reducing α with demand suggested by Burgin and Wild.[9]

The maximum replenishment level for uniform demands occur if orders occur once every two review intervals. For the commonly used range of smoothing constants of 0·05–0·20 the level of replenishment is more than twice the ideal replenishment level, and therefore considerable excess stock would be carried.

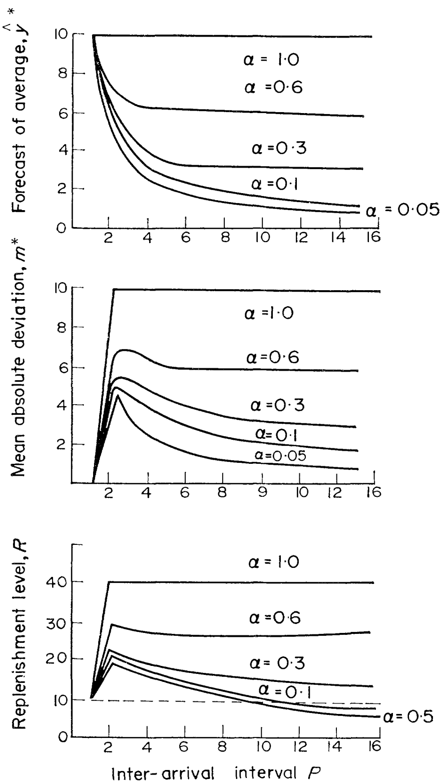

FIG. 3. *Forecasts, mean absolute deviations and replenishment levels for demands of* 10 *units every p review intervals.*

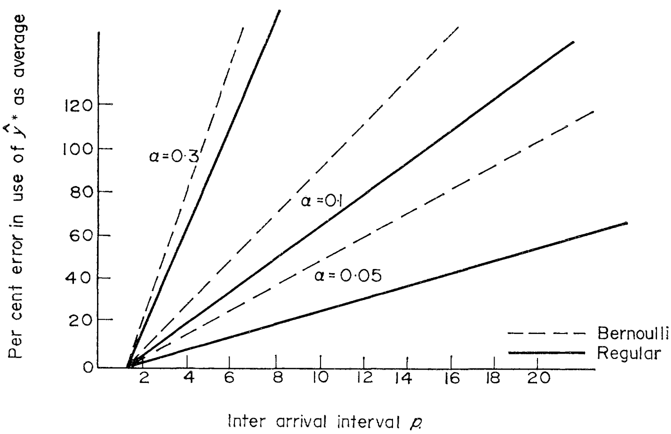

FIG. 4. *Percentage error in forecast of average rate of demand when demands are for constant quantities and* (*a*) *arrive at uniform time intervals;* (*b*) *arrivals have Bernoulli distribution.*

STOCHASTIC ARRIVAL AND SIZE OF DEMAND

Having established the type of errors generated by uniform arrivals, the model is now extended to cover stochatic arrival and size of demand. Let the occurrence of a demand in any specific review interval be generated by a Bernoulli process, with a constant probability $1/p$ that a demand will occur. The average inter-arrival interval will therefore again be p review intervals. Let the size of the demands when they occur be independently distributed with a normal distribution $N(\mu, \sigma^2)$. Let the control and replenishment system be the same as that specified in Section 2 above.

The exponentially weighted moving average of the process, $\hat{y}_t$ from (1.1) and (1.2) is:

$$\hat{y}_t = \alpha \sum_{k=0}^{\infty} \beta^k y_{t-k} \tag{3}$$

and if y_t has a normal distribution, the expected value of $\hat{y}_t$ is μ with variance:

$$\operatorname{var}(\hat{y}_t) = \alpha\sigma^2/(2-\alpha) \tag{4}$$

as shown by Brown.[4]

However the process we are considering is intermittent, hence substituting for y_{t-k} in (3) the values:

$$\left.\begin{array}{llll} y_{t-k} = 0, & & & \text{prob}\,(1-1/p), \\ y_{t-k} = \mu + e_t & \begin{cases} e_t \sim N(0, \sigma^2), \\ E(e_j\, e_k) = 0 \end{cases} & \text{all } j \neq k, & \text{prob}\,(1/p). \end{array}\right\} \tag{5}$$

the first and second moments are:

$$E(\hat{y}_t) = \mu/p$$

$$\operatorname{var}(\hat{y}_t) = \frac{\alpha}{2-\alpha}\left(\frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p}\right). \tag{6}$$

The first term, in μ^2, represents the inflation of the variance due to the intermittent demand, and is a maximum for $p = 2$.

When the system is being used for stock replenishment, or batch size ordering, the replenishment will almost certainly be triggered by a demand which has occurred in the most recent review interval.

As a consequnce the demand pattern given in (5) is therefore modified to give $y_t = \mu + e_t$ for $k = 0$ with probability 1, and equation (5) is only valid for $k \neq 0$.

Substituting in (3), the estimate $\hat{y}^*$ used for forecasting and control has the expected value:

$$E(\hat{y}^*) = \mu(\alpha + \beta/p), \tag{7}$$

with variance:

$$\operatorname{var}\hat{y}^* = \alpha^2\sigma^2 + \frac{\alpha^2\beta^2}{2-\alpha}\left(\frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p}\right). \tag{8}$$

The average demand is again inflated by the fact that replenishment immediately follows a demand, and if considerable errors are to be avoided α must be small compared with β/p. The error is conveniently expressed as a percentage of the average demand, and is easily shown to be $100\alpha(p-1)$ as given in Fig. 4, which shows the increase in estimating error produced by the Bernoulli arrival of demand as compared with constant inter-arrival intervals.

ESTIMATION OF INTER-ARRIVAL INTERVAL

A more useful and flexible forecasting model is obtained if a separate estimate is made of the probability of occurrence of demand, and of the size of demand when it occurs, giving the process:

$$y_t = x_t . z_t,$$

where

$$x_t = \begin{cases} 1, & \text{prob}\,(1/p) \quad \text{(when demand occurs)}, \\ 0, & \text{prob}\,(1-1/p) \end{cases}$$

with size of demand $z_t \sim N(\mu, \sigma^2)$ and

$$E[(x_t - 1/p)(z_t - \mu)] = 0,$$

where p is the average inter-arrival interval.

The expected value of x_t is $1/p$, and if it is assumed that the demand process is non-stationary, it would appear convenient to obtain a weighted average of x_t using the normal methods of exponential smoothing. However the method suffers from the same bias as estimating any intermittent demand shown in (3) above, and therefore would be unsuitable for control.

An unbiased estimate is obtained if a new variable q_t is introduced to measure the time elapsed since the last non-zero observation, and which is used to update the estimated inter-arrival time each time a demand occurs. Practical experience with a large number of intermittent demand series has given satisfactory estimates for $\hat{p}$ with a value of α between 0·1 and 0·2, using:

$$\left.\begin{aligned} \hat{p}_t &= \hat{p}_{t-1}, \quad y_t = 0, \\ \hat{p}_t &= \beta\hat{p}_{t-1} + \alpha q_t, \quad y_t \neq 0, \end{aligned}\right\} \tag{9}$$

the distribution of q_t will be approximately geometric, with mean p and variance $(p-1)^2$. If we denote the value of $\hat{p}_t$ used in the replenishment system as p^*, then using the relationship between variance of input and output of a simple exponential smoothing forecasting system as given by Brown[5] (p. 110) and putting $\mu = p$, $\sigma^2 = (p-1)^2$ we obtain:

$$E(\hat{p}^*) = p, \quad \operatorname{var} \hat{p}^* = \frac{\alpha(p-1)^2}{2-\alpha}.$$

If the estimate of demand $\hat{p}^*$ is updated only when a demand occurs, and it is assumed that the size of demand is independent of the lapsed time q, the estimate of the average demand will be $\hat{y}^*/\hat{p}^*$, with variance given by:

$$\begin{aligned}\operatorname{var}(\hat{y}^*/\hat{p}^*) &= \left[\frac{E(\hat{y}^*)}{E(\hat{p}^*)}\right]^2 \left[\frac{\operatorname{var}\hat{y}^*}{E^2(\hat{y}^*)} + \frac{\operatorname{var}\hat{p}^*}{E^2(\hat{p}^*)}\right] \\ &= \frac{\mu^2 \alpha}{p^2(2-\alpha)} \left[\frac{\sigma^2}{\mu^2} + \frac{(p-1)^2}{p^2}\right] \\ &= \frac{\alpha}{2-\alpha}\left[\frac{(p-1)^2}{p^4}\mu^2 + \frac{\sigma^2}{p^2}\right]. \end{aligned} \tag{10}$$

Comparison of equations (10) and (6) shows that the variance of the estimate of the average is reduced by making the separate estimate $\hat{p}^*$ of the arrival interval, for all intermittent arrivals $(p > 1)$. In addition $\hat{y}/\hat{p}^*$ gives an unbiased estimate of the rate of demand.

MODIFIED ALGORITHM FOR INTERMITTENT DEMANDS

The original algorithm given in (1.1)–(1.4) was modified to give a separate estimate of the inter-arrival time $\hat{p}^*$, and of the demand when it occurs, z^*. The deviation of the non-zero demands from z^* is used to obtain an estimate of the mean absolute deviation of demand m, which then refers only to the non-zero demands. All estimators are updated only when a demand occurs.

The final algorithm used is given in Appendix B, and it is seen that the number of programme steps has been rather more than doubled, and that the vector of retained data has been increased from the two variables ($\hat{y}$ and m) to four (p, q, z and m). The calculation time is however hardly increased at all, as the number of steps carried out when the demand is zero is reduced.

A comparison of the performance of the two systems is given in Table 2 when applied to an intermittent demand occurring on the average every six review intervals, of mean size 3·5 units, standard deviation 1·0 units.

The most striking feature is the reduction in bias of the long-term average rate of demand. The replenishment level R is increased, and now gives approximately the specified 95 per cent protection against out of stock, whereas previously shortage occurred for 20 per cent of demands.

It should be noted however that although two major sources of error have been eliminated the response time has been increased by a factor of p as compared with systems which are updated every review interval. Higher values of α, in the range of 0·2–0·3, may therefore be found necessary if there is a high proportion of items with non-stationary intermittent demand. The response to a change in the frequency of demands is asymmetrical, being faster for increased than for a decreased frequency. Similarly, although the immediate response of new and old systems to an impulse is identical, the subsequent decay is slower for the proposed system.

TABLE 2. COMPARISON OF STANDARD AND REVISED SYSTEMS

Input: 180 observations. Average inter-arrival interval 6
Average size of demand 3·5 units
Standard deviation 1·6 units
Time average of demands 0·58 per review interval

Occurrence of non-zero demands

Time	1	5	6	7	12	36	39	42	47	56	58	60	67	71	
Demand	5	3	2	5	3	6	6	5	2	1	2	1	3	2	
Time	72	73	78	84	89	94	98	104	110	130	133	155	159	170	176
Demand	4	6	2	3	2	6	3	4	2	6	3	6	2	4	5

Response of estimators

Characteristic	Standard method	Revised method
Inter-arrival interval	Not estimated	6·1 (0·7)
Size of non-zero demands	separately	3·45 (0·17)
Average demand	0·84 (0·24)	0·62 (0·06)
Bias in average (%)	+45	+7
Mean absolute deviation, m^*	1·19	1·22
Replenishment stock level		
($k = 3$) to give theoretical probability of stock	4·4	—
($k = 2$) shortage of 75 per cent	—	6·9
No. of stock shortages	6	0
[Standard errors shown ( )]		

It is therefore much more important to have a group of control signals to indicate deviation of demand from the expected values, both in respect of magnitude and inter-arrival interval than with the standard system.

With some systems the declared aim may be to maintain a specified probability that each line of the product range will be in stock. The level of protection may well vary with demand. Frequently such levels are set heuristically from observation of the system performance, but if they are set from the estimated standard deviation of demand, assuming a normal distribution, allowance must be made for the occurrence of the zero demands. Thus if the safety stock is set as some multiple of the standard deviation of demand (or of mean absolute deviation as in (1.4)) the probability of running out of stock in a given time interval will be $1/p$ times that which would occur if there was demand in every review interval.

A further improvement in stock effectiveness may therefore be obtained by making the safety factor a function of the inter-arrival interval.

In the example given in Table 2 it was therefore necessary to reduce the value of the safety stock constant k from 3·0 to 2·0 to maintain equivalent probability of being out of stock with the proposed system. The value of k is conveniently set by a table look-up based on the estimated inter-arrival interval p.

CONCLUSIONS

The analysis above has shown that intermittent arrivals of demand can increase the stock replenishment levels and bias the estimates of average demand. The effect is most marked for large smoothing constants, and if demands of uniform size occur regularly in alternate review intervals, when the excess stock carried may exceed the minimum stock required by 100 per cent.

The errors in performance, due to their complex structure, are ignored in general, and are not detected because they result in a high stock of slow moving items—a result which also commonly occurs as a result of reduction in demand rate. The use of such biased estimators to establish safety stock levels and average demands for infrequently ordered lines is fraught with inconsistency.

The improved system described makes separate estimates of demand size, and arrival of demand, thus eliminating the bias, but the system controlling the stock has to be modified to take account of the characteristics of the new model of demand. It has the considerable advantage that the same forecasting system operates for intermittent and frequent demands as the system will behave in the standard way if a demand occurs every review interval.

The system gives a much more explicit representation of the demand pattern for the considerable proportion of slow moving items which exist in most stock holdings. The proposed system avoids the errors inherent when exponential smoothing is applied to intermittent demands. Excess stock levels are avoided, and in those cases in which infrequent demands produced too low a stock level to supply the required protection, specified out of stock probability is achieved.

The additional statistic of demand frequency enables the production, stock-holding, order processing and delivery costs to be determined much more accurately, and as a consequence it is a simple matter to identify uneconomic products, and improve the profitability of a range of products, using the data provided by the intermittent demand model.

REFERENCES

[1] G. E. P. Box and G. M. Jenkins (1970) *Time Series Analysis Forecasting and Control.* Holden-Day, San Francisco.

[2] C. C. Holt (1957) Forecasting Seasonal and Trends by Exponentially Weighted Moving Averages. Carnegie Institute of Technology, Pittsburg, Pennsylvania.

[3] C. C. Holt, F. Modigliani, J. F. Muth and H. A. Simon (1960) *Planning Production Inventories and Workforce.* Prentice-Hall, New Jersey.

[4] R. G. Brown (1959) *Statistical Forecasting for Inventory Control.* McGraw-Hill, New York.

[5] R. G. Brown (1962) *Smoothing Forecasting and Prediction of Discrete Time Series*. Prentice-Hall, New Jersey.
[6] P. R. Winters (1960) Forecasting sales by exponentially weighted moving averages. *Mgmt Sci.* **6**, 324.
[7] D. H. Ward (1963) Comparison of different systems of exponentially weighted prediction. *The Statistician* **13**, 173.
[8] P. J. Harrison (1965) Short term sales forecasting. *Appl. Stat.* **14**, 102.
[9] T. A. Burgin and A. R. Wild (1967) Stock control—experience and usuable theory. *Opl Res. Q.* **18**, 35.
[10] C. S. Beightler, L. G. Mitten and G. L. Nemhauser (1961) A short table of z-transforms and generating functions. *Ops Res.* **9**, 574.
[11] D. W. Trigg (1964) Monitoring a forecasting system. *Opl Res. Q.* **15**, 271.
[12] M. L. Shone (1967) Exponential smoothing with an adaptive response rate. *Opl Res. Q.* **18**, 318.

APPENDIX A

BEHAVIOUR OF SINGLE-STAGE EXPONENTIAL SMOOTHING ESTIMATORS WITH REGULAR INTERMITTENT DEMANDS OF UNIFORM SIZE

Consider demands of uniform magnitude μ occurring at time $t = 1$ and every p review intervals thereafter. The pattern of forecasts obtained is shown in Figure 2 and repeats every p intervals. Consider the single cycle from $t = 1$ to $t = p$. Let y^* be the forecast upon which replenishment is based, using the algorithm given in equations (1.1)–(1.4). It is assumed that such replenishment will occur immediately following a demand.

Let $Y(z)$ be the generating function of demand, where $Y(z)$ is a polynomial in z, the coefficient of z^n being the demand at time n and z being so chosen that $Y(z)$ converges:

$$Y(z) = \mu z(1+z^p+z^{2p}+\ldots)$$

$$= \frac{\mu z}{1-z^p}. \tag{11}$$

Taking transforms of (1.1) and (1.2) we obtain the generating function for the forecast $\hat{Y}(z)$:

$$\hat{Y}(z) = \alpha Y(z)+\beta z\hat{Y}(Z)$$

$$= \frac{\alpha}{1-\beta z} Y(Z), \tag{12}$$

where $\beta = 1-\alpha$ and substituting for $Y(z)$ from (11), the generating function of the forecasts is:

$$\hat{Y}(z) = \frac{\mu\alpha z}{(1-\beta z)(1-z^p)}. \tag{13}$$

Expanding in powers of z gives the values of the forecasts for time $(np+q$

$$\hat{y}_{np+q} = \frac{\mu\alpha\beta^{q-1}}{1-\beta^p}, \quad q = 1,2,\ldots,p; \quad n = 0,1,2,\ldots,\infty. \tag{14}$$

The one-step-ahead forecasts for the review intervals in which the demands occur are made at time np and are given by:

$$\hat{y}_{np} = \frac{\mu\alpha\beta^{p-1}}{1-\beta^p}. \tag{15}$$

The demands occur at times $t = np+1$ and replenishment occurs immediately, and it is therefore the values of the estimated parameters of the model as updated at times $np+1$ which will be used in the replenishment formula. When equilibrium is reached this estimate of the demand, designated by $\hat{y}^*$, will be the coefficient of z^{np} as $n \to \infty$ in equation (13). Using the standard formulae for inversion of the generating function (e.g. Beightler[10]):

$$\hat{y}^* = \frac{\mu\alpha}{1-\beta^p}. \tag{16}$$

If $\hat{y}^*$ is used to estimate the long-term average demand, its correct value should be μ/p. If used for estimating the size of demand which will occur, it should have the value μ. Thus although $\hat{y}^* = \mu$ for $p = 1$ for all values of α, and $\hat{y}^* \to \mu/p$ for small α, it provides a biased estimate for either or both the average demand over time, $\bar{y}$, and also of the demand occurring, μ. The percentage errors are given in Table 3 for a number of values of p and α.

TABLE 3. PERCENTAGE ERROR WHEN STANDARD FORECASTING PROCEDURE IS USED FOR INTERMITTENT DEMAND OF PERIOD p

Periodicity	Percentage error in forecasting average demand			
	Smoothing constant			
p	0·05	0·1	0·3	1·0
1	0	0	0	0
2	2·5	5·2	17·6	100
3	4·2	10·8	37·1	200
5	10·3	22·0	80·5	400
10	24·6	53·5	208·7	900
15	39·7	88·7	352·0	1400

The generating function of the errors $E(z)$ is defined by equation (1.1) and is given by:

$$E(z) = Y(z) - z\hat{Y}(z). \tag{17}$$

$\hat{Y}(z)$ is only defined by equation (14) for $q = 1,2,\ldots,p$, and hence $E(z)$ has to be obtained directly from (11) and (14) rather than by substitution of (11) and (13) in (17):

$$e_{np+q} = \begin{cases} \mu\left(\dfrac{1-\beta^{p-1}}{1-\beta^p}\right), & q = 1, \\ -\dfrac{\mu\alpha\beta^{q-1}}{1-\beta^p}, & q = 2, \ldots, p. \end{cases} \tag{18}$$

If the value of β lies within the range $0<\beta<1$ the errors will be positive for $q = 1$, and negative otherwise. The generating function of the modulus of the error $|E|(z)$ may therefore be written directly from (18), using the fact that the demand is cyclical with interval p:

$$|E|(z) = \frac{1}{1-z^p}\left(\mu z \frac{1-\beta^{p-1}}{1-\beta^p} + \sum_{q=2}^{p} \frac{\mu\alpha\beta^{q-2} z^q}{1-\beta^p}\right)$$

$$= \frac{\mu z}{(1-z^p)(1-\beta^p)}\left(1-\beta^{p-1}+\alpha\sum_{q=2}^{p}\beta^{q-2} z^{q-1}\right). \tag{19}$$

The mean absolute deviation is obtained from the modulus of the errors by (1.3) and its generating function is therefore:

$$M(z) = \beta z M(z) + |E|(z)$$

$$= \frac{\alpha}{1-\beta z}|E|(z) \tag{20}$$

and from (19):

$$M(z) = \frac{\mu\alpha z}{(1-z^p)(1-\beta^p)(1-\beta z)}\left(1-\beta^{p-1}+\alpha\sum_{q=2}^{p}\beta^{q-2} z^{q-1}\right). \tag{21}$$

Denoting the value of m_t used for forecasting and stock control in equation (1.4) as m^*, it is obtained as the coefficient of z^{np+1} for large integer n in (21). For $p>2$

$$m^* = \frac{\mu\alpha}{(1-\beta^p)^2}[1-\beta^{p-1}+(p-1)\,\alpha\beta^{p-1}]. \tag{22}$$

$$= \frac{\mu\alpha\{1-\beta^{p-1}[1-\alpha\beta(p-1)]\}}{(1-\beta^p)^2}.$$

The values of m^* and $\hat{y}$ given in (16) and (22) are those used for replenishment, in a standard exponential smoothing system, and enable the bias of the estimates to be calculated for specified values of p and α.

APPENDIX B

MODIFIED ALGORITHM FOR INTERMITTENT DEMANDS

Model:

$$y_t = x_t(\bar{z}_{\eta-1} + e_\eta),$$

where the index t refers to the review interval, and the index η to the serial number of the non-zero demands. x_t has a Bernoulli distribution:

$$x_t = \begin{cases} 1, & \text{prob}\,(1/p), \\ 0, & \text{prob}\,(1-1/p), \end{cases}$$

$$E[(x_i - 1/p)(x_j - 1/p)] = 0 \quad \text{all } i, j, \quad i \neq j.$$

z_η is the non-zero observation of the process which is assumed to be an exponentially weighted moving average (EWMA) (Box and Jenkins,[1] p. 106):

$$\bar{z}_\eta = z_{\eta-1}(\lambda) + e_\eta,$$

$$\bar{z}_{\eta-1}(\lambda) = \lambda \sum_{j-1}^{\infty} (1-\lambda)^{j-1} z_{\eta-j}.$$

It is an IMA process of order (0, 1, 1).

The value of λ in the above model should be estimated from past data, but if the series is short it may have to be chosen arbitrarily from experience.[5] The value so chosen is referred to as α.

e_η is assumed to have a normal distribution with zero mean:

$$e_\eta \sim N(0, \sigma^2), \quad E(e_j e_k) = 0 \quad \text{all } j, k, \quad j \neq k.$$

Updating procedure:

$$\left.\begin{aligned} e_\eta &= y_t - \bar{z}_{\eta-1}, \\ \bar{z}_\eta &= \bar{z}_\eta - \alpha e_n, \\ m_\eta &= (1-\alpha)\, m_{\eta-1} + \alpha |e_\eta|, \\ R_t &= \bar{z}_\eta + k m_\eta, \\ \bar{p}_\eta &= \bar{p}_{\eta-1}(1-\alpha) + \alpha q, \\ \bar{y}_t &= \bar{z}_\eta / \bar{p}_t, \\ q &= 1, \end{aligned}\right\}, \quad y_t \neq 0,$$

$$q = q + 1, \quad y_t = 0.$$

In addition it is advisable to incorporate a number of exception report signals such as the following, before updating:

(1) For all values of t:

$$(1 - 1/\bar{p}_\eta)^q < k_1, \quad k_1 \text{ say } 0{\cdot}01$$

implies that the time lapse since the last non-zero demand is significantly greater than expected.

(2) For all non-zero demands:

(a) $$q/\bar{p}_\eta < k_2, \quad k_2 \text{ say } 0{\cdot}2$$

implies that the demand occurred earlier than would be expected.

(b) Size of individual non-zero demands out of control if:

$$|e_\eta| > k_3 m_\eta, \quad k_3 \text{ say } 3{\cdot}0\text{–}5{\cdot}0.$$

(c) Tracking signal γ out of control implying bias or inadequate model. Let s_η be the smoothed error where:

$$s_\eta = s_{\eta-1}(1-\alpha)+\alpha e_\eta,$$

$$\gamma_\eta = s_\eta/m_\eta,$$

$$|\gamma_\eta| > k_4, \quad k_4 \text{ say } 0{\cdot}50\text{–}0{\cdot}70.$$

(See Trigg,[11] Shone.[12])