OPTIMAL POLICIES FOR A MULTI-ECHELON INVENTORY PROBLEM*

ANDREW J. CLARK[1] AND HERBERT SCARF[2]

1. Introduction[3]

In the last several years there have been a number of papers (Reference 1) discussing optimal policies for the inventory problem. Almost without exception these papers are devoted to the determination of optimal purchasing quantities at a single installation faced with some pattern of demand. It has been customary to make the assumption that when the installation in question requests a shipment of stock, this shipment will be delivered in a fixed or perhaps random length of time, but at any rate with a time lag which is independent of the size of the order placed. There are, however, a number of situations met in practice in which this assumption is not a tenable one. An important example arises when there are several installations, say $1, 2, \cdots, N$, with installation 1 receiving stock from 2, with 2 receiving stock from 3, etc. In this example, if an order is placed by installation 1 for stock from installation 2, the length of time for delivery of this stock is determined not only by the natural lead time between these two sites, but also by the availability of stock at the second installation.

In this paper we shall consider the problem of determining optimal purchasing quantities in a multi-installation model of this type. First of all, let us remark that once the parameters of the model have been specified (lead times, purchase costs, demand distributions, holding and shortage costs, etc.), the optimal purchasing quantities may, in theory at least, be determined. The obvious way to proceed would be to define a cost function for each configuration of stock at the various installations, and in transit from one installation to another. We then remark that this cost function satisfies the type of functional equation which always appears in inventory theory, and from which the optimal provisioning policies may be determined by a recursive computation. It is clear, however, that this procedure is in general completely impractical since it necessitates the recursive computation of a sequence of functions of at least N variables.

The question is, therefore, whether the obvious recursive computation of optimal policies may be simplified for our multi-installation problem without compromising the optimality of the solution. The answer is that such a simplification may be obtained if several very plausible assumptions are incorporated in the model. With these assumptions, it will be demonstrated in this paper that the solution suggested by Clark in Reference 3 is indeed optimal. The solution will be described in detail below. It should be remarked here, however, that the virtue

* Received October 1959.

[1] Planning Research Corporation, Los Angeles, California.

[2] Stanford University, Stanford, California.

[3] This work was supported by the Bureau of Supplies and Accounts of the Department of the Navy.

of the solution given by Clark is that it permits the optimal levels to be computed separately by precisely those techniques which have been used in the past for the computation of optimal policies at a single installation.

In Section IV we shall discuss various applications of the multiple-installation technique to problems in which several installations have the same supplier. The type of complex discussed in Section III may be described by the scheme:

$$\underset{N}{[\ ]} \rightarrow [\ ] \rightarrow [\ ] \cdots \underset{2}{[\ ]} \rightarrow \underset{1}{[\ ]},$$

whereas the complex in Section IV has the scheme

$$\begin{tikzcd}[row sep=small, column sep=small]
 & & {[\ ]\cdots} \\
 & {[\ ]} \arrow[ur] \arrow[dr] & \\
{[\ ]} \arrow[ur] \arrow[dr] & & {[\ ]\cdots} . \\
 & {[\ ]} \arrow[r] & {[\ ]\cdots}
\end{tikzcd}$$

Unfortunately, the results for the latter type of complex are not as satisfactory as those for the former.

2. The Multiple-Installation Model and a Description of the Solution

Let us begin with a review of the model to be used for a single installation. An extensive discussion of this model is given in Reference 2, and we shall summarize here that material which will be of use to us.

A sequence of purchasing decisions is made at the beginning of a number of regularly spaced intervals. The cost of purchasing an amount z will initially be a general function $c(z)$, though we shall subsequently restrict ourselves to certain special cases. Delivery of an order occurs, say, λ periods after the order is placed, at which time the stock on hand is augmented by the amount of the order. During each period the stock on hand is depleted by an amount equal to the demand during the period, which is an observation from a distribution with density function $\phi(t)$, the demands being independent from period to period. (The demand distributions may actually differ from period to period.)

In addition to the purchase cost, it is customary to charge several other costs during each period. The first of these costs is a holding cost, proportional to the stock on hand at the beginning of the period if it is positive; and the second, a shortage cost proportional to the deficit of available stock at the end of the period if there is such a deficit. If the stock on hand at the beginning of the period is x, then the cost during the period, exclusive of purchasing costs, is given by

$$L(x) = \begin{Bmatrix} hx + p \displaystyle\int_x^\infty (t - x)\phi(t)\,dt; & x > 0 \\ p \displaystyle\int_0^\infty (t - x)\phi(t)\,dt; & x \leqq 0 \end{Bmatrix} \tag{1}$$

where h and p are the marginal holding and shortage costs, respectively. It is useful for us to introduce occasionally more general holding and shortage functions than the linear ones described in Equation (1), and for these functions there will be an analogous form for the one-period cost $L(x)$.

Any policy (sequence of purchasing decisions) produces a sequence of costs. Costs which occur n periods in the future are discounted by an amount α^n, so that we may form a total discounted cost as the result of any policy. The optimal purchasing policy is that one which minimizes the total discounted cost.

Let us consider an inventory problem in which there are n periods remaining, with x_1 units of stock on hand, w_1 units to be delivered one period in the future, and generally w_j units to be delivered j periods in the future, where $j = 0, 1, 2, \cdots, \lambda - 1$. Let $C_n(x_1, w_1, \cdots, w_{\lambda-1})$ represent the expectation of the discounted costs, beginning with such a configuration of stock and following an optimal provisioning scheme. Using the type of reasoning employed in Reference 2, this sequence of functions may be shown to satisfy the following functional equation:

$$C_n(x_1, w_1, \cdots, w_{\lambda-1}) = \underset{z \geq 0}{\text{Min}} \Big\{ c(z) + L(x_1) + \alpha \int_0^\infty C_{n-1}(x_1 + w_1 - t, w_2, \cdots, w_{\lambda-1}, z)\phi(t)\, dt \Big\}, \tag{2}$$

where the minimizing value of z is the optimal purchase quantity for the given stock configuration. In the writing of this equation we are explicitly assuming that all excess demand is backlogged until the necessary stock becomes available. This equation has been analyzed in considerable detail and we shall quote for future use those facts of relevance to us.

1. The optimal policy (i.e., the minimizing value of z) is a function of the total stock on hand plus on order, regardless of the dates of delivery. This property depends crucially on the assumption that excess demand is backlogged. Moreover, it may be shown that

$$\begin{aligned} C_n(x_1, w_1, \cdots, w_{\lambda-1}) &= L(x_1) + \alpha \int_0^\infty L(x_1 + w_1 - t)\phi(t)\, dt + \cdots \\ &+ \alpha^{\lambda-1} \int_0^\infty \cdots \int_0^\infty L(x_1 + w_1 + \cdots + w_{\lambda-1} - t_1 - \cdots - t_{\lambda-1})\phi(t_1) \\ &\cdots dt_1 \cdots + f_n(x_1 + \cdots + w_{\lambda-1}), \end{aligned} \tag{3}$$

and that f_n satisfies the functional equation

$$f_n(u) = \underset{y \geq u}{\text{Min}} \Big\{ c(y-u) + \alpha^\lambda \int_0^\infty \cdots \int_0^\infty L(y - t_1 - \cdots - t_\lambda)\phi(t_1) \cdots \phi(t_\lambda)\, dt_1 \cdots dt_\lambda + \alpha \int_0^\infty f_{n-1}(y-t)\phi(t)\, dt \Big\}. \tag{4}$$

If y^* is the minimizing value in Equation (4), then $y^* - u$ is the optimal purchase quantity, where $x_1 + w_1 + \cdots + w_{\lambda-1} = u$. (Obvious modifications in Equations (3) and (4) are required when n is less than the time lag.) These results (Reference 2) permit us to reduce the inventory problem with a time lag to one in which essentially no lag exists.

2. The results mentioned above are valid for any ordering function $c(z)$, whenever excess demand is backlogged. Now let us restrict our attention to the cost function

$$c(z) = \left\{ \begin{array}{cc} K + c \cdot z; & z > 0 \\ 0 \quad ; & z = 0 \end{array} \right\}. \tag{5}$$

(K is the setup cost and c the unit cost.) Let us also assume that the one-period costs

$$\alpha^{\lambda-1} \int_0^\infty \cdots \int_0^\infty L(y - t_1 - \cdots - t_\lambda)\phi(t_1) \cdots \phi(t_\lambda)\, dt_1 \cdots dt_\lambda$$

are convex. (This is certainly correct if the holding and shortage costs are linear, and in other cases also.) Then there exists a sequence of critical numbers (S_n, s_n) so that in period n it is optimal to order only if $x_1 + \cdots + w_{\lambda-1} < s_n$ and if we do order, we order an amount $S_n - (x_1 + \cdots + w_{\lambda-1})$ (Reference 4). The specific form of the one-period costs is irrelevant; we can require only that they be convex.

3. An additional simplification occurs if $K = 0$. The upper and lower critical numbers become the same and it is customary to denote their common value by $\bar{x}_n$. The optimal purchase quantity is given by

$$\text{Max } (0, \bar{x}_n - (x_1 + \cdots + w_{\lambda-1})).$$

In this case, somewhat more is known about the properties of the functions $f_n(u)$. First of all, $f_n(u)$ is always convex, and in addition, $f'_n(u) = -c$ for $u \leqq \bar{x}_n$ (Reference 2).

Now let us turn our attention to the description of the multiple-installation model. We shall make the following assumptions:

Assumption 1: Demand originates in the system at the lowest installation (installation 1), and at no other point in the system.

Assumption 2: The cost of purchasing and shipping an item from any installation to the next will be linear, without any setup cost. The only exception to this assumption will be at the highest installation, at which point a setup cost will be permitted.

Assumption 3: At the lowest installation (installation 1), a linear holding and shortage cost will be operative, in the same manner as the single-installation problem described above. We make the assumption that holding and shortage costs for the second installation do not depend only on the stock on hand at the second installation, but are functions of this stock, plus stock in transit to the first installation plus stock on hand at the first installation. Generally speaking, the holding and shortage costs at any level will be assumed to be functions of the stock at that level plus all other stock in the system which is actually at a lower level or in transit to a lower level. We shall call these costs the *natural* one-period costs at the level. They may, of course, be equal to zero.

Clark in Reference 3 has given the name "echelon" to the system consisting of the stock at any given installation plus stock in transit to or on hand at a lower installation. The echelons will be numbered according to the highest installation in the echelon. Our assumption may be stated as requiring that the one-period costs be functions of the echelon stock rather than installation stock. The simplifications described in this paper are very crucially tied to this assumption and assumption 2.

Assumption 4: Each echelon backlogs excess demand.

With these specifications in mind let us turn our attention to the determination of the optimal provisioning levels. The solution suggested by Clark is best described by means of an example. We consider the case of two installations. The natural lead time from installation 2 to installation 1 will be two periods in this example. Let us denote the stock on hand at installation 1 by x_1 ; the stock to be delivered one period in the future by w_1 ; and the stock on hand at installation 1, plus on hand at installation 2, plus in transit from 2 to 1, by x_2 (i.e., x_2 is echelon 2 stock). The one-period costs at installation 1 will be denoted by $L(x_1)$, and those at echelon 2 by $\tilde{L}(x_2)$. The unit shipping cost from 2 to 1 will be denoted by c_1 .

We begin by solving the problem (that is, determining the single critical numbers $\bar{x}_1 = 0$, $\bar{x}_2 = 0$, $\bar{x}_3$, $\bar{x}_4$, $\cdots$) for installation 1 without any reference to the remaining parts of the multiple-echelon system. In other words, we solve the single-installation problem for the lowest echelon, assuming that delivery of any order, regardless of its size, will be effected in two periods, and using in our calculations a unit cost equal to the transportation cost from the higher echelon, without any reference to the original purchase cost. This would suggest that if at the beginning of the n^{th} period the stock on hand plus on order at installation 1 is less than $\bar{x}_n$, we order the difference; and if the stock is larger than $\bar{x}_n$, we do not order. The problem is, of course, that there may not be adequate stock at installation 2 to fill such an order. In the solution given in this paper, it is shown that we ship only that part of the order for which there is available stock at the next highest echelon. This describes the optimal policy at the lowest installation (Theorem 1, below).

The next question is that of the optimal quantity of stock to bring in at echelon 2. It will be shown in the next section that the optimal purchase quantities at echelon 2 are functions only of x_2 , the stock at the two installations plus the stock in transit. Moreover, the optimal policies for this echelon may be computed by the standard single-installation model using the ordering cost appropriate to this echelon, and the natural one-period costs described above ($\tilde{L}(x_2)$). The important idea is that we must in some fashion introduce a penalty at this echelon for keeping a quantity of stock on hand which is insufficient to meet the normal requests from the lower installation (Theorem 2, below). The procedure for doing this is quite simple: We merely introduce an additional one-period cost at the second echelon which is precisely equal to the expected increment in total cost at installation 1, because the stock at echelon 2 is inadequate to bring the lower level's stock up to the required point $\bar{x}_n$.

In the example that we are discussing, the specific form for this additional one-period cost may be found as follows: We recall the definition of the functions $C_n(x_1, w_1)$ to be the minimum expected discounted cost at echelon 1 if there are n periods remaining and if the stock on hand is x_1; and the stock on order, w_1. (This function is to be computed on the basis of an ordering cost equal to c_1, the transportation cost.) For $n = 1$, $C_1(x_1, w_1) = L(x_1)$, and also $C_2(x_1, w_1) = L(x_1) + \alpha \int_0^\infty L(x_1 + w_1 - t)\phi(t)\, dt$. In this expression for C_2, the first term represents the expected one-period costs in the immediate period, and the second term represents similar costs for the next period. Inasmuch as delivery of any order takes two periods, there is no modification that can be made in these costs. For $n > 2$, we use the decomposition described in Equation (3); that is,

$$C_n(x_1, w_1) = L(x_1) + \alpha \int_0^\infty L(x_1 + w_1 - t)\phi(t)\, dt + f_n(x_1 + w_1). \tag{6}$$

The first two terms on the left-hand side are as described above; the third term represents the optimal cost exclusive of those costs which it is impossible to modify by a request for a shipment.

As in Equation (4), the functions $f_n(u)$ satisfy

$$\begin{aligned} f_n(u) = \operatorname*{Min}_{y \geqq u} \Big\{ & c_1(y - u) + \alpha^2 \iint L(y - t_1 - t_2)\phi(t_1)\phi(t_2)\, dt_1\, dt_2 \\ & + \alpha \int_0^\infty f_{n-1}(y - t)\phi(t)\, dt \Big\}, \end{aligned} \tag{7}$$

and the minimizing value is $\bar{x}_n$. In other words, if $x_1 + w_1 < \bar{x}_n$ so that ordering occurs, the minimum cost will be

$$\begin{aligned} & c_1(\bar{x}_n - u) + \alpha^2 \iint L(\bar{x}_n - t_1 - t_2)\phi(t_1)\phi(t_2)\, dt_1\, dt_2 \\ & \qquad + \alpha \int_0^\infty f_{n-1}(\bar{x}_n - t)\phi(t)\, dt. \end{aligned} \tag{8}$$

If, however, x_2, the stock at both installations, plus stock in transit, is less than $\bar{x}_n$, we will only be able to ship $x_2 - (x_1 + w_1)$ and therefore the minimum cost will be

$$\begin{aligned} & c_1(x_2 - u) + \alpha^2 \iint L(x_2 - t_1 - t_2)\phi(t_1)\phi(t_2)\, dt_1\, dt_2 \\ & \qquad + \alpha \int_0^\infty f_{n-1}(x_2 - t)\phi(t)\, dt. \end{aligned} \tag{9}$$

Expression (9) is of course larger than Expression (8), and the difference in cost is attributable exclusively to the insufficiency of stock at level 2. Therefore, the additional one-period loss to be charged to this echelon is given by Expression

(9) minus Expression (8), or

$$c_1(x_2 - \bar{x}_n) + \alpha^2 \iint [L(x_2 - t_1 - t_2) - L(\bar{x}_n - t_1 - t_2)]\phi(t_1)\phi(t_2)\, dt_1\, dt_2 \tag{10}$$
$$+ \alpha \int_0^\infty [f_{n-1}(x_2 - t) - f_{n-1}(\bar{x}_n - t)]\phi(t)\, dt,$$

if $x_2 < \bar{x}_n$ and zero if $x_2 > \bar{x}_n$. With this additional one-period loss to be charged to the second echelon, the optimal policy is then computed using standard techniques. Of course the specific values of Expression (10) involve the critical numbers $\bar{x}_n$ and the functions $f_n(u)$, but these will have been computed already for installation 1.

It is worth remarking that Expression (10) is a convex function of x_2, so that the optimal policy for the second echelon will be of the (S, s) type.

If there are more than two echelons, the same procedure is repeated, always augmenting the natural one-period loss at the echelon by the increment in total cost at the lower echelon due to the lack of available stock.

3. The Proof of Optimality

In this section we shall prove that the procedure suggested in the previous section is indeed optimal. Because of notational difficulties, we shall restrict our attention to the example described in the previous section although the ideas are quite general. In order to be specific we shall assume the time lag in delivery to installation 2 to be a single period.

Our approach will be to investigate the optimal solution for the entire system, and show that it reduces to the solution given by Clark. The first step is to write down a sequence of functional equations, analogous to Equation (3), but for the entire system rather than a single installation. We define $C_n(x_1, w_1, x_2)$ to be the minimum expected value of the discounted system costs if there are n periods remaining; if stock on hand at installation 1 is x_1; stock in transit, w_1; and system stock, x_2. At the beginning of the period two decisions are made: the first, a decision as to how much system stock to order for delivery next period; and the second, a decision as to the quantity of stock to be placed in transit to installation 1. The stock on hand plus in transit to installation 1 may be raised from $x_1 + w_1$ to y, where y is any number between $x_1 + w_1$ and x_2, at a cost of $c_1(y - x_1 - w_1)$; and if such a decision is taken, at the beginning of the next period stock on hand at installation 1 will be $x_1 + w_1 - t$ (t is the demand), and the stock in transit will be $y - x_1 - w_1$. The system stock is, of course, not modified by this decision; it can only be changed by a decision to introduce z units into the system (at a cost of $c(z)$), and will become $x_2 + z - t$. Therefore, if the two decisions described by y and z are taken, the inventories (x_1, w_1, x_2) become $(x_1 + w_1 - t, y - x_1 - w_1, x_2 + z - t)$, and the discounted value of expected future costs will be

$$\alpha \int_0^\infty C_{n-1}(x_1 + w_1 - t, y - x_1 - w_1, x_2 + z - t)\phi(t)\, dt. \tag{11}$$

In order to complete the accounting, we should consider the purchase (and transportation), holding and shortage costs. The purchase and transportation costs are given by

$$c(z) + c_1(y - x_1 - w_1). \tag{12}$$

The shortage and holding costs are given by

$$\bar{L}(x_2) + L(x_1), \tag{13}$$

the terms of which apply, respectively, to echelon 2 and installation 1.

$C_n(x_1, w_1, x_2)$ is, of course, equal to the minimum of Expressions (11) + (12) + (13), when y and z are chosen optimally, and we therefore obtain the following functional equation:

$$\begin{aligned} C_n(x_1, w_1, x_2) = \min_{\substack{x_1+w_1 \leqq y \leqq x_2 \\ 0 \leqq z}} \Big\{ & c(z) + c_1(y - x_1 - w_1) + \bar{L}(x_2) \\ & + L(x_1) + \alpha \int_0^\infty C_{n-1}(x_1 + w_1 - t, y - x_1 - w_1, x_2 + z - t)\phi(t)\, dt \Big\}, \end{aligned} \tag{14}$$

with the condition $C_0 \equiv 0$.

Let us also introduce the functional equation which would be used to compute optimal policies for installation 1 in *isolation*. Let $C_n(x_1, w_1)$ be the minimum expected value of the discounted costs for an n period problem at installation 1, which begins with x_1 units on hand and w_1 units in transit. We are assuming that the unit purchase price is the transportation cost and that *all* orders are delivered in two periods. C_n satisfies

$$\begin{aligned} C_n(x_1, w_1) = \min_{y \geqq x_1 + w_1} \Big\{ & c_1(y - x_1 - w_1) + L(x_1) \\ & + \alpha \int_0^\infty C_{n-1}(x_1 + w_1 - t, y - x_1 - w_1)\phi(t)\, dt \Big\}. \end{aligned} \tag{15}$$

Of course, the solution of Equation (15) is of no clear relevance to Equation (14) as yet.

Obviously $C_1(x_1, w_1) = L(x_1)$, and $C_1(x_1, w_1, x_2) = L(x_1) + \bar{L}(x_2)$. In other words, $C_1(x_1, w_1, x_2) = C_1(x_1, w_1) + g_1(x_2)$. We shall show that $C_n(x_1, w_1, x_2)$ may always be written as $C_n(x_1, w_1)$ + a function of x_2 alone, and this is the important step in verifying that Clark's solution is optimal.

Theorem 1. There is a sequence of functions $g_n(x_2)$, with $g_1(x_2) = \bar{L}(x_2)$, such that

$$C_n(x_1, w_1, x_2) = C_n(x_1, w_1) + g_n(x_2). \tag{16}$$

Moreover, it is optimal for installation 1 to provision without reference to installation 2, subject to the proviso that if insufficient stock is available at installation 2, then installation 1 will be content with getting as much as it can.

We shall demonstrate this theorem by induction. Let us suppose that Equation (16) is true for $(n - 1)$, and we shall then demonstrate its validity for n. Substituting in Equation (14), we obtain

$$\begin{aligned} C_n(x_1, w_1, x_2) = \underset{\substack{x_1+w_1 \leqq y \leqq x_2 \\ 0 \leqq z}}{\text{Min}} \Big\{ & c(z) + c_1(y - x_1 - w_1) + \bar{L}(x_2) \\ & + L(x_1) + \alpha \int_0^\infty C_{n-1}(x_1 + w_1 - t, y - x_1 - w_1)\phi(t)\, dt \\ & + \alpha \int_0^\infty g_{n-1}(x_2 + z - t)\phi(t)\, dt \Big\}. \end{aligned} \tag{17}$$

From Equation (17) we see that aside from the constraint that y be less than x_2, the optimal selection of y is such as to minimize

$$c_1(y - x_1 - w_1) + L(x_1) + \alpha \int_0^\infty C_{n-1}(x_1 + w_1 - t, y - x_1 - w_1)\phi(t)\, dt,$$

and this is, of course, the same as the single critical number $\bar{x}_n$ for the problem of installation 1 considered separately. If it turns out that $x_2 \geqq \bar{x}_n$, then the constraint $x_2 \geqq y$ is not operative, and we may therefore conclude that for $x_2 \geqq \bar{x}_n$,

$$\begin{aligned} C_n(x_1, w_1, x_2) = {} & C_n(x_1, w_1) \\ & + \underset{z \geqq 0}{\text{Min}} \Big\{ c(z) + \bar{L}(x_2) + \alpha \int_0^\infty g_{n-1}(x_2 + z - t)\phi(t)\, dt \Big\}. \end{aligned} \tag{18}$$

On the other hand, if $x_2 < \bar{x}_n$ (and therefore $x_1 + w_1 < \bar{x}_n$), installation 1 will be thwarted in its attempt to bring its stock level up to $\bar{x}_n$. Because of the convexity of the one-period costs, it is optimal to bring the stock level up as high as possible and therefore $y = x_2$. As a consequence, we see that for $x_2 < \bar{x}_n$,

$$\begin{aligned} C_n(x_1, w_1, x_2) = {} & c_1(x_2 - x_1 - w_1) + L(x_1) \\ & + \alpha \int_0^\infty C_{n-1}(x_1 + w_1 - t, x_2 - x_1 - w_1)\phi(t)\, dt \\ & + \underset{z \geqq 0}{\text{Min}} \Big\{ c(z) + \bar{L}(x_2) + \alpha \int_0^\infty g_{n-1}(x_2 + z - t)\phi(t)\, dt \Big\}. \end{aligned} \tag{19}$$

Now we are interested in showing that $C_n(x_1, w_1, x_2) - C_n(x_1, w_1)$ is a function of x_2 alone. From Equations (18) and (19), we see that this difference is equal to

$$\Lambda_n(x_1, w_1, x_2) + \underset{z \geqq 0}{\text{Min}} \Big\{ c(z) + \bar{L}(x_2) + \alpha \int_0^\infty g_{n-1}(x_2 + z - t)\phi(t)\, dt \Big\}, \tag{20}$$

where

$$\begin{aligned} \Lambda_n(x_1, w_1, x_2) = {} & c_1(x_2 - x_1 - w_1) + L(x_1) \\ & + \alpha \int_0^\infty C_{n-1}(x_1 + w_1 - t, x_2 - x_1 - w_1)\phi(t)\, dt - C_n(x_1, w_1), \end{aligned} \tag{21}$$

when $x_2 < \bar{x}_n$ and zero otherwise.

In order to demonstrate Theorem 1, it is therefore necessary to show that $\Lambda_n(x_1, w_1, x_2)$ is in reality a function of x_2 alone, and of course we need only consider the region $x_2 < \bar{x}_n$. In this region, however,

$$\begin{aligned} C_n(x_1, w_1) &= c_1(\bar{x}_n - x_1 - w_1) + L(x_1) \\ &\quad + \alpha \int_0^\infty C_{n-1}(x_1 + w_1 - t, \bar{x}_n - x_1 - w_1)\phi(t)\, dt, \end{aligned} \tag{22}$$

and therefore Equation (21) may be written as

$$\begin{aligned} \Lambda_n(x_1, w_1, x_2) &= c_1(x_2 - \bar{x}_n) \\ &\quad + \alpha \int_0^\infty [C_{n-1}(x_1 + w_1 - t, x_2 - x_1 - w_1) \\ &\qquad - C_{n-1}(x_1 + w_1 - t, \bar{x}_n - x_1 - w_1)]\phi(t)\, dt. \end{aligned} \tag{23}$$

Our theorem will be demonstrated if we can show that the integrand in Equation (23) is independent of x_1 and w_1. But by Equation (6),

$$C_{n-1}(x_1, w_1) = L(x_1) + \alpha \int_0^\infty L(x_1 + w_1 - y)\phi(y)\, dy + f_n(x_1 + w_1),$$

and therefore the integrand in Equation (23) is given by

$$\begin{aligned} &\alpha \int_0^\infty L(x_2 - t - y)\phi(y)\, dy + f_{n-1}(x_2 - t) \\ &\qquad - \alpha \int_0^\infty L(\bar{x}_n - t - y)\phi(y)\, dy - f_{n-1}(\bar{x}_n - t), \end{aligned} \tag{24}$$

which is a function of x_2 alone. We have therefore demonstrated Theorem 1.

We have, however, demonstrated somewhat more. Λ_n is now known to be a function of x_2, which may be written as

$$\begin{aligned} \Lambda_n(x_2) &= c_1(x_2 - \bar{x}_n) \\ &\quad + \alpha^2 \int_0^\infty \int_0^\infty [L(x_2 - t - y) - L(\bar{x}_n - t - y)]\phi(t)\phi(y)\, dt\, dy \\ &\quad + \alpha \int_0^\infty [f_{n-1}(x_2 - t) - f_{n-1}(\bar{x}_n - t)]\phi(t)\, dt, \end{aligned} \tag{25}$$

for $x_2 < \bar{x}_n$ and zero for $x_2 > \bar{x}_n$. But Equation (20), which represents $g_n(x_2)$, may be written as

$$g_n(x_2) = \operatorname*{Min}_{z \geq 0} \left\{ c(z) + \tilde{L}(x_2) + \Lambda(x_2) + \alpha \int_0^\infty g_{n-1}(x_2 + z - t)\phi(t)\, dt \right\}. \tag{26}$$

The solution of this equation provides us with the optimal policy for the entire system. As we see, all that is required is to augment the natural costs at echelon 2 by $\Lambda(x_2)$.

Theorem 2. The functions $g_n(x_2)$ satisfy Equation (26), by means of which the optimal system stock may be obtained.

4. Several Installations with the Same Supplier

In this section we shall generalize the model considered in Section III so as to include the possibility of several installations with the same supplier. We shall, however, retain the restriction that no installation has two different suppliers. All of the assumptions that have previously been mentioned, such as backlogging, no setup cost for transportation, etc., will be retained also in this model. The only point in need of clarification is our assumption that the natural losses should be functions of echelon stock, rather than installation stock. The notion of an echelon in this model will be as follows: We begin by selecting a specific installation, say installation I. Associated with I will be a number of other installations which receive, directly or indirectly, stock from installation I. The total stock at I, plus the stock in transit or on hand at these other installations will comprise the echelon associated with installation I. With this definition our assumption will again be that the natural one-period costs are associated with echelons, rather than installations.

Let us consider the following example of such a complex:

$$\begin{tikzcd}[row sep=small] & & & {[A_3]} \\ & & {[B_2]} \arrow[ur] \arrow[dr] & \\ {} \arrow[r] & {[C_1]} \arrow[uur] \arrow[dr] & & {[A_2]} \\ & & {[B_1]} \arrow[r] & {[A_1]} \end{tikzcd}$$

The procedure which we have shown in the previous section to be optimal for a simpler problem suggests the following procedure in this complex:

(1) For installations A_1, A_2, and A_3 (which are terminal installations) compute the optimal sequences of single critical numbers, assuming the installations to be in isolation; also that all requests for shipment are supplied during the natural lead time, and that the purchase cost is given by the transportation cost from the higher echelon.

(2) Augment the natural costs at echelon B_1 by the increment in cost at A_1 because of the inability to satisfy requests for stock at A_1, and do the same for B_2. Then compute the optimal stock levels at B_1 and B_2 separately, assuming the availability of infinite stock from C_1.

(3) Modify the natural costs at C_1 by the increment in cost at B_1 and B_2 because of the inability to satisfy requests for stock, and then compute the optimal policy at C_1.

If the directions given above are examined closely, it may be seen that they are ambiguous on a number of points. The clarification of these points will show where the Clark procedure departs from optimality in this model, whereas it was optimal for the model considered in Section III. Even though the procedure is not optimal, it has considerable merit, both in its ease of application and in its approximate validity.

Point 1. Shall we permit an arbitrary pair of installations to exchange stock; and if so, at what cost, and with what lags?

As an example, we are posing the question as to whether A_2 shall be permitted to ship excess stock to A_3. The desire to make this shipment might arise in two different ways. First of all, there may be insufficient stock at echelon B_2 to raise both A_2 and A_3 to their required critical levels, and the stocks left over at A_2 and A_3 may be out of balance by a sufficient amount so that it is wise to ship both B_2 and A_2 stock to A_3. Another possible cause of transshipment might be a substantial anticipated drop in demand at A_2 and an excess of carryover stock which might profitably be shipped to A_3.

In practice, however, transshipment of this sort would rarely take place. Moreover, if we permit this sort of transshipment to take place, the theoretical and computational aspects of the problem become quite complex. It would be meaningless for an installation to consider itself in isolation, inasmuch as its actual stock levels in the future would depend on the disposition of stock at all other installations. Since our primary aim is to be able to compute optimal policies at each installation separately, we shall assume that such transshipment is impossible. It is gratifying that such an assumption does not run contrary to what is done in practice.

Point 2. If all requests cannot be satisfied because of insufficient stock at a higher echelon, how is the available stock to be rationed among the requesting installations?

The answer to this question bears very heavily on the optimality of the procedure suggested above. We shall consider the following concrete case:

$$\begin{array}{ccc} & & [A_1] \\ & \nearrow & \\ \to [B] & & \\ & \searrow & \\ & & [A_2] \end{array},$$

and assume, for definiteness, that all routes have a time lag of one period, and that the transportation cost c_1 is the same from B to A_1 as from B to A_2. Let L^1 and L^2 represent the one-period costs at installations 1 and 2, respectively, and $\bar{L}$ the one-period cost at B. Let $C_n^1(x_1)$ and $C_n^2(x_2)$ represent the minimum costs for A_1 and A_2 computed separately, and let $\{\bar{x}_n^1\}$ and $\{\bar{x}_n^2\}$ be the sequence of single critical numbers for A_1 and A_2. The C functions satisfy the customary functional equations, for example:

$$C_n^1(x_1) = \underset{y_1 \geqq x_1}{\mathrm{Min}} \left\{ c_1(y_1 - x_1) + L^1(x_1) + \alpha \int_0^\infty C'_{n-1}(y_1 - t)\phi_1(t)\, dt \right\}, \tag{27}$$

and similarly for C_n^2.

In the spirit of Section III, we define $C_n(x_1, x_2, x_3)$ to be the minimum *system* cost if A_1 has x_1 units, A_2 has x_2 units, and the B echelon has x_3 units. These functions satisfy a functional equation analogous to Equation (17), i.e.,

$$\begin{aligned} C_n(x_1, x_2, x_3) = \mathrm{Min} \Big\{ & c(z) + c_1(y_1 - x_1) \\ & + c_1(y_2 - x_2) + \bar{L}(x_3) + L^1(x_1) + L^2(x_2) \\ & + \alpha \iint C_{n-1}(y_1 - t_1, y_2 - t_2, x_3 + z - t_1 - t_2)\phi_1(t_1)\phi_2(t_2)\, dt_1\, dt_2 \Big\}, \end{aligned} \tag{28}$$

where the minimization is over the region $y_1 \geqq x_1$, $y_2 \geqq x_2$, $x_3 \geqq y_1 + y_2$, $z \geqq 0$.

The crucial ideas behind the optimality results of the previous section were embodied in Theorems 1 and 2. The analogue of Theorem 1 for the model considered in this section would be that there exists a sequence of functions $g_n(x_3)$ with the property that

$$C_n(x_1, x_2, x_3) = C_n^1(x_1) + C_n^2(x_2) + g_n(x_3). \tag{29}$$

Does there exist such a sequence of functions? And if there does, what light is cast upon the question of allocation of stock (Point 2)?

Unfortunately the answer to the first of these questions is in the negative. The functions $C_n(x_1, x_2, x_3)$ *cannot* be broken down in the form of Equation (29). To see why this is so, let us assume that Equation (29) is valid for $n - 1$, and see what the consequences of Equation (28) and this assumption would be for $C_n(x_1, x_2, x_3)$. We would have

$$\begin{aligned} C_n(x_1, x_2, x_3) = \operatorname*{Min}_{\cdots} \Big\{ & c(z) + c_1(y_1 + y_2 - x_1 - x_2) \\ & + \bar{L}(x_3) + L^1(x_1) + L^2(x_2) + \alpha \int_0^\infty C_{n-1}^1(y_1 - t_1)\phi_1(t_1)\, dt_1 \\ & + \alpha \int_0^\infty C_{n-1}^2(y_2 - t_2)\phi_2(t_2)\, dt_2 \\ & + \alpha \iint g_{n-1}(x_3 + z - t_1 - t_2)\phi_1(t_1)\phi_2(t_2)\, dt_1\, dt_2 . \Big\} \end{aligned} \tag{30}$$

Aside from the constraint that $y_1 + y_2$ be less than x_3, the optimal selection of y_1 would be $\bar{x}_n^1$ and the optimal selection of y_2 would be $\bar{x}_n^2$. If $x_3 > \bar{x}_n^1 + \bar{x}_n^2$, the constraint is not operative, and from Equation (30) we would have

$$\begin{aligned} C_n(x_1, x_2, x_3) &= C_n^1(x_1) + C_n^2(x_2) \\ &+ \operatorname*{Min}_{z \geqq 0} \left\{ c(z) + \bar{L}(x_3) + \alpha \iint g_{n-1}(x_3 + z - t_1 - t_2)\phi_1(t_1)\phi_2(t_2)\, dt_1\, dt_2 \right\}. \end{aligned} \tag{31}$$

So far, so good. We run into a problem, however, when $x_3 < \bar{x}_n^1 + \bar{x}_n^2$. This is, of course, the problem raised by Point 2, and the answer is given by Equation (30). The numbers y_1 and y_2 should be selected according to the constraints

$$y_1 + y_2 = x_3, \; y_1 \geqq x_1, \; y_2 \geqq x_2 \tag{32}$$

and such as to minimize

$$\begin{aligned} & c_1(y_1 + y_2 - x_1 - x_2) + L^1(x_1) + L^2(x^2) \\ & \quad + \alpha \int C_{n-1}^1(y_1 - t_1)\phi_1(t_1)\, dt_1 + \alpha \int C_{n-1}^2(y_2 - t_2)\phi_2(t_2)\, dt_2. \end{aligned} \tag{33}$$

Therefore, in order to allocate properly we must solve the minimization problem (33) subject to the constraints (32). The problem is certainly solvable. The difficulty, however, is in the form of the answer. The answer *may* depend not only on x_3, but also on x_1 and x_2 (the stock at A_1 and A_2). It *may* depend on x_1 and x_2; generally it will not unless the stock levels x_1 and x_2 are seriously out of balance. But if the solution to the minimization problem does depend on x_1 and x_2, it will depend on them jointly and a factorization of the type given by Equation (29) would not be obtained.

Let us assume that such a lack of balance does not occur. Then y_1 and y_2 would be selected to minimize (33) subject to

$$y_1 + y_2 = x_3, \quad \text{alone.} \tag{34}$$

Call these solutions $\bar{x}_n^1(x_3)$ and $\bar{x}_n^2(x_3)$. Then Equation (30) would read

$$\begin{aligned} C_n(x_1, x_2, x_3) &= C_n^1(x_1) + C_n^2(x_2) + \Lambda_n(x_1, x_2, x_3) \\ &\quad + \operatorname*{Min}_{z \geqq 0} \left\{ c(z) + \bar{L}(x_3) + \alpha \iint g_{n-1}(x_3 + z - t_1 - t_2)\phi_1(t_1)\phi_2(t_2)\, dt_1\, dt_2 \right\}, \end{aligned} \tag{35}$$

where

$$\begin{aligned} \Lambda_n(x_1, x_2, x_3) &= c_1(\bar{x}_n^1(x_3) + \bar{x}_n^2(x_3) - x_1 - x_2) + L^1(x_1) + L^2(x_2) \\ &\quad + \alpha \int_0^\infty C_{n-1}^1(\bar{x}_n^1(x_3) - t_1)\phi_1(t_1)\, dt_1 - C_n^1(x_1) \\ &\quad + \alpha \int_0^\infty C_{n-1}^2(\bar{x}_n^2(x_3) - t_2)\phi_2(t_2)\, dt_2 - C_n^2(x_2), \end{aligned} \tag{36}$$

for $x_3 < \bar{x}_n^1 + \bar{x}_r^2$ and zero, otherwise. However, just as in Section III, if $x_1 < \bar{x}_n^1$ and $x_2 < \bar{x}_n^2$, this may be shown to be a function of x_3 alone, and this is the function that is to be taken to augment the natural costs at echelon B.

We repeat that Equations (35) and (36) are derivable only by means of the assumptions that the stock at installations A_1 and A_2 are not out of balance. Since this is expected to occur rather frequently, it suggests that Clark's approximation is an excellent one for this model.

5. Extensions

The discussion in Sections III and IV assumed that demand originates in the system at the lowest installation (echelon 1) and at no other point in the system. This, however, is not a necessary assumption and, in fact, the probability distributions used for the various echelons need have no relationship with each other. This may be demonstrated by considering the proof of optimality in Section III for the simple two-echelon example.

If $f_1(t_1)$ and $f_2(t_2)$ represent, respectively, the marginal demand distribution at echelon 1 and echelon 2, and $f(t_1, t_2)$ the joint distribution, then Equation (14) may be rewritten as follows:

$$\begin{aligned} C_n(x_1, w_1, x_2) &= \operatorname*{Min}_{\substack{x_1 + w_1 \leqq y \leqq x_2 \\ 0 \leqq z}} \Big\{ c(z) + c_1(y - x_1 - w_1) + \bar{L}(x_2) + L(x_1) \\ &\quad + \alpha \int_0^\infty \int_0^\infty C_{n-1}(x_1 + w_1 - t_1, y - x_1 - w_1, x_2 + z - t_2) f(t_1, t_2)\, dt_1\, dt_2 \Big\}. \end{aligned} \tag{37}$$

Following through the proof of Theorem 1 by substituting

$$C_n(x_1, w_1, x_2) = C_n(x_1, w_1) + g_n(x_2)$$

in Equation (37), we obtain

$$\begin{aligned} C_n(x_1, w_1, x_2) = \operatorname*{Min}_{\substack{x_1+w_1 \leq y \leq x_2 \\ 0 \leq z}} \Big\{ & c(z) + c_1(y - x_1 - w_1) + \bar{L}(x_2) \\ & + L(x_1) + \alpha \int_0^\infty C_{n-1}(x_1 + w_1 - t_1, y - x_1 - w_1) f_1(t_1)\, dt_1 \\ & + \int_0^\infty g_{n-1}(x_2 + z - t_2) f_2(t_2)\, dt_2 \Big\}, \end{aligned} \tag{38}$$

which is the same form as Equation (17). The remainder of the proof is the same as in Section III.

The ability to assign different distributions to the various echelons has several interesting applications. For example, the N-installation problem of Section III may be interpreted as N stages of production, where the time required for production in each stage is analogous to the delivery times in the inventory problem. The final stage of production (analogous to installation 1 in the inventory problem) is faced with an exogenous demand while each production stage may incur random losses through spoilage. The probability distribution used for the final production stage is the exogenous demand distribution augmented by losses during the stage. This distribution is successively augmented by losses in the other production stages to obtain distributions for these stages. The per unit ordering cost for each stage is the fabrication cost in the immediately prior stage. This example represents the case when the mean demand is an increasing funcion of the echelon number, i.e., the higher the echelon, the higher the mean.

An example of the opposite case is encountered in the inventory problem where items are regenerated through repair. Considering the problem of Section III again, suppose that items issued from installation 1 are exchanged for damaged items (on a one for one basis) which then undergo repair cycles of different durations according to the degree of damage. Thus, if t items are issued, then t reparable items are generated, with different portions, $t_1, t_2, \cdots$ ($t = \sum_i t_i$) being successively more remote, timewise, from being available for reissue. Here, the net demand faced by echelon k is given by $t - \sum_{i=1}^{k} t_i$ which is a decreasing function of k. If, throughout the repair cycle, items are scrapped as being uneconomically reparable, then the mean demand as a function of echelon number may be more general than the monotonically increasing or decreasing functions considered above.

Problems of the type described in Sections III and IV, together with the interpretations mentioned above, may be combined to portray almost any inventory and/or production structure. Such combinations may be used to make supply repair, and production decisions in an integrated fashion. Of course, in each application, the assumptions of the method must be analyzed with respect to their validity or effect.

References

1. K. J. Arrow, S. Karlin, H. Scarf, *Studies in the Mathematical Theory of Inventory and Production*, Stanford: Stanford University Press (1958), "Bibliography of Inventory Theory."
2. *Ibid.*, Chapter 10. (S. Karlin and H. Scarf)
3. A. J. Clark, "A Dynamic, Single-Item, Multi-Echelon Inventory Model," RM-2297, Santa Monica, California, The RAND Corporation (December, 1958).
4. H. Scarf, "The Optimality of (S, s) Policies for the Dynamic Inventory Problem," *Proceedings of the First Stanford Symposium on Mathematical Methods in the Social Sciences*, June, 1960.