ARE MULTI-ECHELON INVENTORY METHODS WORTH IMPLEMENTING IN SYSTEMS WITH LOW-DEMAND-RATE ITEMS?*

JOHN A. MUCKSTADT† AND L. JOSEPH THOMAS†

In practice, most multi-echelon inventory systems are managed using adaptations of single location methods. This paper shows that such methods can be dramatically inferior to methods designed to take advantage of a system’s structure. This is especially true in repair parts inventory systems where most items have low demand rates.

In this paper we describe a multi-echelon method adapted to a particular situation. Data taken from a large (over $100 million investment) industrial inventory system are used to provide comparative results on overall inventory required to meet a given service level. The multi-echelon method used involves an $(s - 1, s)$ ordering policy and Poisson demand, both of which are appropriate for low demand items; this method requires much lower total investment than the single-location method requires to achieve the same average level of performance. Thus it appears that multi-echelon methods are worthwhile in many situations.

(INVENTORY/PRODUCTION—MULTI-ECHELON SYSTEMS)

1. Introduction

Multi-item, multi-echelon inventory systems frequently involve tens of thousands of stock-keeping units (SKUs) and investments of hundreds of millions of dollars. They are often managed using adaptations of single-location models. However, as we will show, these models can perform poorly in comparison with methods designed to take advantage of the system structure. Taking advantage of system structure is important when there are many items with low demand and high relative cost, since these items may be best managed by having stock only at a central supply point. (This is discussed further in several places in the paper, including §5.) In nearly all large scale inventory systems, the majority of the items have very low demand rates. In this paper we use data from a large spare parts supply system (with over $100 million total investment) in which low demand rate items far outnumber high demand rate items.

The system studied is a two-echelon inventory system with no lateral resupply. A distribution center (DC) is at the upper level or echelon and customer-service warehouses are on the lower level. The system examined has several DC’s; but, since lateral resupply is not allowed, we can examine one DC-system. We assume a DC can be resupplied from outside vendors or the company-owned plant in a known constant amount of time. As we will discuss, the lead time for resupplying the warehouses depends on the availability of stock at the DC at the time a warehouse places an order.

Throughout the paper we use the continuous review $(s - 1, s)$ inventory policy at each location; that is, after each demand at a location an order is placed for one unit, bringing on-hand plus on-order inventory at the location up to $s$ units. This model is appropriate for low demand items at the warehouse level and for most items at the DC. (In the application we examined the marginal ordering cost at both the warehouse

*Accepted by Warren H. Hausman; received January 20, 1978. This paper has been with the authors 9 months for 2 revisions.

†Cornell University.
and the DC is very low since there are frequent scheduled shipments independent of the individual item demand and since real-time inventory status is maintained at each location.) The demand process for each item at the warehouse (lower) level is assumed to be a Poisson process. We also assume the demand processes among warehouses to be independent. Since each demand is immediately forwarded to the DC, the DC also faces a Poisson demand process with a mean equal to the sum of the means at the warehouse level.

One main objective in this paper is to investigate the potential worth of applying multi-echelon inventory models in certain practical situations. In §2 we discuss in qualitative terms some different procedures that can be and have been used to control inventories in multi-echelon systems. A multi-echelon model that is appropriate for the particular system we are studying is described in §3. §4 contains the numerical results we obtained to compare the different control alternatives. These results show that multi-echelon methods can be of significant practical value. Some concluding remarks are given in the final section.

2. Alternative Methods of Control: A Qualitative Discussion

Objectives (Measures of Performance)

In choosing a performance measure we must remember that in the setting examined here, the sole function of the inventories is to provide customer service. Thus only out-of-stock situations at the warehouses will be measured; DC stock levels and stockouts are important only in that they influence warehouse stockouts. To best meet the service goals, we propose using a time-weighted performance measure at each warehouse rather than a number-of-incidents measure. To see why, consider the items that have zero stock at the warehouse level. If the warehouse has no stock, then, unless otherwise constrained, a multi-echelon model will also select zero stock for the DC when an incident-oriented performance measure is used. The reason for this is that there is no chance of improving warehouse customer service by carrying DC-level inventory. The probability of immediately satisfying a customer demand (and avoiding a stockout incident) is zero no matter what the DC stock level is. A main purpose of the DC-level inventory is to inexpensively (relatively) provide increased range and depth of stock over that carried by the warehouse. If a time-weighted objective is used, a multi-echelon model has an incentive to increase the range of stock so as to provide quick replenishment following a stockout. Thus, we propose using a time-weighted objective as the system performance measure for a given level of total investment. (As we will discuss, the precise measure depends on the situation.) Some of the above points were made by [7].

Alternative Methods of Control

We will discuss two methods for making stock level decisions: (i) "Level Decomposition," and a (ii) Multi-Echelon Method ("Item Decomposition"). A third system often used in practice, in which a certain number of expected "Days of Supply" are kept at each location in the system, performed so poorly in our study that it was eliminated as being a benchmark against which to test the other models.

In "Level Decomposition" an aggregate service level goal is set for each echelon. Each echelon allocates stock among items to minimize investment subject to meeting a constraint on the average service level for all items. As we will show, each pair of
warehouse/DC service levels implies a system performance level experienced by customers and a system investment that can be used to form tradeoff curves. This method is studied here for two reasons. First, it properly allocates money within a level, and thus it is a good method among the set of single-level methods. Second, it was in use at the firm we studied.

The idea behind "Item Decomposition" is essentially the same as the notion of a "dual distribution system" as proposed by Heskett et al. [4]. Some low demand items should be carried only at the DC so that system investment can be reduced while providing adequate support. The mathematical techniques used here simultaneously set the stock levels for all items at each location. Thus, the decision about which items stock only at the DC is made automatically using the optimization model which we developed.

The "Level Decomposition" method cannot appropriately choose to have different levels of service at the two echelons for different items, because there is no information flow regarding one level's decisions to the other level when the stock levels are established. In general the same items will have good service or bad service at both levels. This is true because the cheapest way to obtain an average service level objective at any location is to carry the higher demand, lower cost items and to ignore some low demand, higher cost ones. Thus safety stock at the warehouse level tends to be duplicated at the DC level for some items and is unavailable for others at both locations.

3. A Description of the Models

System Description

Recall that the system we studied contains several DC systems—one DC with several warehouses in correspondence with the DC—and that since lateral resupply is not allowed, each DC system can be examined separately. When a demand occurs at warehouse \( j \) several possible events can take place. If warehouse \( j \) has on-hand stock for item \( i \), there is no delay in satisfying the demand. A normal resupply request is made. If the DC has stock on hand, the resupply time is denoted by \( B_{ij} \). If the DC does not have stock on hand, the resupply time is \( B_{ij} + \) delay.
available at the DC to send to the warehouse. The normal DC resupply time is \( T_{r0} \). If warehouse \( j \) does not have on-hand stock for item \( i \), an emergency resupply request is made to the DC. If the DC has stock on-hand, the item is immediately shipped to the warehouse and the resupply time is \( ET_j \). The DC places a normal resupply request with its supplier and receives a replacement unit \( T_{r0} \) time units later. When the DC does not have the required item in stock, an emergency order is placed with its supplier and the item is shipped directly to the warehouse and arrives \( ET_j \) time units later. We assume that \( ET_2 > ET_1 \). Note that all emergency orders are satisfied outside the normal resupply channels. The system's operation is depicted in Figure 1.

In addition to \( B_j \), \( ET_2 \) and \( ET_1 \), we need the following notation.
- \( \lambda_i \) = the expected daily demand rate for item \( i \) at warehouse \( j \).
- \( \lambda_{r0} = \sum_j \lambda_i \) = the expected daily demand rate for item \( i \) at the DC.
- \( s_j \) = the stock level for item \( i \) at warehouse \( j \).
- \( s_{r0} \) = the stock level for item \( i \) at the DC.
- \( c_i \) = the cost of item \( i \) at location \( j \).
- \( p(x | \lambda_i B_j) \) = the probability distribution of demand during a resupply time of length \( B_j \). The distribution is assumed to be Poisson. This assumption is warranted for the low-demand parts studied.

**Level Decomposition**

The Level Decomposition method determines stock levels separately for each echelon and each location. Thus the location \( (j) \) subscript is suppressed in \( \lambda_i \), \( c_i \), \( B_j \), and \( s_i \) in this subsection. At each location the item stock levels are calculated so that an aggregate performance goal (a service level measure) is achieved at a minimum level of investment in inventory. One common way of measuring the service level at a location is to determine the fraction of customers whose demands can be satisfied immediately. This measure is sometimes referred to as the fill rate. (Other measures could be used as well.) We define the fill rate to be \( \alpha \) if the probability an arriving customer can be immediately satisfied is \( \alpha \). Based on the analysis in [2], when the demand process is a Poisson process, \( \sum_{x=0}^{s_i-1} p(x | \lambda_i B_j) \) measures the probability that a customer demand can be immediately satisfied for item \( i \) given \( B_j \) is the resupply time;

\[
p(x | \lambda_i B_j) = e^{-\lambda_i B_j} \frac{(\lambda_i B_j)^x}{x!},
\]

which represents the probability that the actual number of units on order for the item at the location is \( x \) when the expected number on order is \( \lambda_i B_j \). Then the fill rate for a location is

\[
\sum_i \frac{\lambda_i}{(\sum_j \lambda_j)} \cdot \sum_{x=0}^{s_i-1} p(x | \lambda_i B_j).
\]

The model used to calculate the stock levels at any location when Level Decomposition is used is:

\[
\begin{align*}
\text{min} & \quad \sum_i c_i s_i \\
\text{subject to} & \quad \sum_i \left( \frac{\lambda_i}{\sum_j \lambda_j} \right) \cdot \sum_{x=0}^{s_i-1} p(x | \lambda_i B_j) \geq \alpha \quad \text{and} \\
& \quad s_i = \lfloor \lambda_i B_j \rfloor, \lfloor \lambda_i B_j \rfloor + 1, \ldots,
\end{align*}
\]
where \([y]\) represents the integer part of \(y\). Thus, each \(s_i\) is constrained to be at least as large as the integer part of the expected lead time demand. Note that the objective function contains only inventory investment, which is to be minimized subject to the fill-rate constraint. Since an \((s - 1, s)\) ordering policy is used, the number of orders placed and, thus, the total ordering cost, is constant. Problem (1) must be solved for each location to find the total system inventory requirements. To solve this problem we can use a Lagrangian technique as described in reference [3]. The Lagrangian form of (1) is separable by item so the problem is easy to solve.

**Item Decomposition**

All shortages in the actual system are supplied by emergency shipments and not by normal replenishment. In Item Decomposition (the multi-echelon model), the probability distributions reflect this operating policy. Let \(h(x \mid \lambda_y T_y(s_{\theta}))\) represent the probability distribution for the number of units in normal resupply for item \(i\) at warehouse \(j\), where

\[ T_y(s_{\theta}) = \text{the average normal resupply time for warehouse } j \text{ for item } i \text{ given the DC stock level for item } i \text{ is } s_{\theta}. \quad T_{\theta} \text{ is the normal resupply time for item } i \text{ at the DC.} \]

Furthermore, Sherbrooke [7] shows that

\[ T_y(s_{\theta}) = B_y + (1/\lambda_{\theta}) \sum_{x > s_{\theta}} (x - s_{\theta})p(x \mid \lambda_{\theta} T_{\theta}) \]

where, as indicated above:

- \(B_y\) = the normal resupply time from the DC to warehouse \(j\) for item \(i\) when stock is available at the DC to satisfy a demand.

Then

\[ h(x \mid \lambda_y T_y(s_{\theta})) = p(x \mid \lambda_y T_y(s_{\theta}))/ \sum_{w=0}^{s_y} p(w \mid \lambda_y T_y(s_{\theta})). \]

The preceding expression for \(h(x \mid \lambda_y T_y(s_{\theta}))\) is only an approximation. This approximation is based on a result due to Feeney and Sherbrooke [2] who extended Palm’s theorem to the “lost sales” case. Note that since emergency orders are all satisfied outside the normal resupply channel, they are analogous to lost sales. The results developed in reference [2] depend on the normal resupply times being independent and identically distributed random variables. These random variables are not independent, however, so that our representation for \(h(x \mid \lambda_y T_y(s_{\theta}))\) is only an approximation. To see that the resupply times are dependent, consider the following example. Suppose a normal resupply request is made by a warehouse and the DC currently has no stock on hand. Then the resupply request will be delayed until the DC has available stock. The next resupply request by a warehouse is more likely to be delayed given the fact that the DC was out of stock than if the DC had stock on hand to satisfy the preceding request. Consequently, the values of these two resupply time random variables are dependent. However, when the DC stock level exceeds the integer part of the expected number of units on order by the DC in its normal resupply system, the exact and approximate distribution for the number of units in resupply at a warehouse are very similar (see Shanker [6] for an extensive analysis). For low demand items \(\sum \lambda_y T_y(s_{\theta}) \ll 1\) the approximation is almost always accurate to two decimal places. For higher demand rate items, say \(\sum \lambda_y T_y(s_{\theta}) > 5\), the approximation is excellent (normally accurate to two decimal places) even when the DC stock level is at or slightly below...
the expected DC demand over the normal lead time. The exact expression for
\( h(x | \lambda_y T_y(s_0)) \) is quite complicated and can be developed using the methods presented in reference [3].

The distribution for the number of orders in normal resupply at the DC at a point in
time must also reflect the fact that not all demands on the DC are satisfied through
normal resupply channels. Let \( k_i(s_0) \) represent the probability that an emergency
demand placed by a warehouse for item \( i \) can be immediately satisfied at the DC. We
approximate this probability by \( \sum_{i=0}^{s_0-1} p(x | \lambda_0 T_0) \), the exact probability when all
excess DC demand is handled through regular resupply. This approximation is in one
sense conservative, since excess demand is removed from the normal resupply system.
On the other hand, if an emergency order is placed by a warehouse for item \( i \), then at
least \( s_{ij} \) units are currently on order by that warehouse for the item. Thus, if the
fraction of expected system demand for item \( i \) originating at that warehouse is large
(over 20% of the total expected system demand), then \( \sum_{i=0}^{s_0-1} p(x | \lambda_0 T_0) \) over-states the
probability that an emergency demand placed by that warehouse can be immediately satisfied. This is true because \( s_0 \) may be equal to a large fraction of \( s_0 \), and stockout conditions at the warehouse and the DC will be highly correlated. However, this
approximation is very close for systems in which \( s_0 \) is greater than or equal to the
integer part of \( \lambda_0 T_0 \) and no more than 10% of expected system demand for an item
occurs at one warehouse.

The objective function can now be stated as the sum over all items and warehouse
locations of the product of three terms: (i) the probability the next system demand
is for a type \( i \) item at warehouse \( j \), (ii) the probability of having to use the emergency
resupply system, \( h(s_{ij} | \cdot ) \) (this is the probability there are no items in stock at the
warehouse, since there can never be more than \( s_{ij} \) units in normal resupply at a
warehouse), and (iii) the expected time of emergency resupply. That is, the objective is

\[
\sum_i \sum_j \frac{\lambda_j}{\sum_j \lambda_j} h(s_{ij} | \lambda_j T_j(s_0)) \cdot \left[ ET_1 \cdot \sum_{x=0}^{s_0-1} p(x | \lambda_0 T_0) + ET_2 \cdot \left[ 1 - \sum_{x=0}^{s_0-1} p(x | \lambda_0 T_0) \right] \right].
\]

(2)

Thus the objective function measures the expected time needed to satisfy a customer
demand. This model represents the operation of a particular industrial spare parts
distribution system. The model does not include an additional dollar cost for emergency
resupply; in the application we studied the company considered the customer
waiting time cost to exceed the emergency resupply costs by a substantial amount.

The Item Decomposition (multi-echelon) formulation can now be stated. The
problem is to:

\[
\text{minimize } (2)
\]

subject to \( \sum_i \sum_j c_{ij} s_{ij} \leq \text{Budget Constraint} \) and

\( s_{ij} = \left[ \lambda_j T_j(s_0) \right], \left[ \lambda_j T_j(s_0) \right] + 1, \ldots \).

The problem was solved using the Lagrangian technique described in reference [3], in
which each item problem is each solved separately once \( s_0 \) and the Lagrangian
multiplier are known. (The \( j \)-subscript in the budget constraint must go from zero to \( m \)
to include the DC expense.)
4. Computational Comparisons

Data

Daily demand rates for a sample of 418 items were obtained for a particular industrial system. We assumed, to reduce computational cost and without loss of generality, that there were 15 identical warehouses with one DC in each system. The calculations are no more difficult if 15 different warehouses are used; there are just more calculations. The composite warehouse is an average of data for several warehouses. The items chosen represented high and low demand rates and unit costs. Resupply times are actual values. We also vary one emergency resupply time to illustrate the effect of changing the relative speed of emergency resupply. The unit costs have been multiplied by a constant to hide their true values.

Tradeoff Curves for Different Approaches

Are multi-echelon models worth using? We, of course, can not answer this question in general. However, we can see if they are worth using for the case we examined. To do this we will compare the performance–cost tradeoffs achieved using two different methods for calculating stock levels. The first method is the Level Decomposition method, and the second is the Item Decomposition Method, both of which are described in §3. Table I gives cost and demand information for a few representative items from the data set. In addition to examining the decisions for these items, we will also give the overall performance achieved for various levels of investment when using the two approaches.

<table>
<thead>
<tr>
<th>Item</th>
<th>Warehouse Daily Demand Rate</th>
<th>Unit Cost (all locations)</th>
<th>Normal Warehouse Resupply (from DC to Warehouse)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0753</td>
<td>$234.60</td>
<td>12 days</td>
</tr>
<tr>
<td>2</td>
<td>0.0004</td>
<td>89.67</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>0.2242</td>
<td>6.72</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>0.0020</td>
<td>280.83</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>0.0144</td>
<td>192.36</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>0.0006</td>
<td>91.47</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>0.0046</td>
<td>358.83</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>0.0087</td>
<td>218.67</td>
<td>12</td>
</tr>
</tbody>
</table>

Note: These items were selected to represent high and low demand and cost, and to give the same total time-weighted backorders in the Table II below.

The normal resupply time from plant to DC was 15 days. The normal warehouse resupply time was 12 days, as shown above. Emergency resupply times used were faster, and $ET_2$, the emergency resupply from the plant to warehouses was 5 days. Two different $ET_1$ values were used, where $ET_1/ET_2 = 0.1$ and $ET_1/ET_2 = 0.3$, to study the effect of this ratio on the difference in stock levels and performance produced by the competing methods. (Recall that $ET_1$ is the emergency resupply time from the DC to a warehouse, if the DC has the product on hand.)

The comparisons were made by first computing stock levels using the two approaches. These stock levels were then evaluated using (2) to determine the expected time a customer must wait before a demand is satisfied. The Level Decomposition
approach was implemented by computing stock levels for various fill rates at each echelon. The expected emergency backorder days per customer was then computed for each combination of warehouse and DC service levels. For example, when $ET_1 / ET_2 = 0.1$, one combination that leads to a solution that is not dominated by another pair in the expected emergency backorder—days versus total investment tradeoff is a fill rate of 0.95 at the DC and a fill rate of 0.75 at the warehouse. The stock levels for the eight items from Table I computed using the Level Decomposition method for the 0.95-0.75 fill rates and the stock levels obtained using the Item Decomposition approach to achieve approximately the same level of performance are given in Table II. This table allows us to compare qualitatively the decisions made using the two approaches before quantitatively examining the overall performance.

<table>
<thead>
<tr>
<th>Item</th>
<th>Warehouse</th>
<th>DC</th>
<th>Expected Emergency Backorder-Days/yr</th>
<th>Stock Levels</th>
<th>DC</th>
<th>Expected Emergency Backorder-Days/yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>25</td>
<td>48.58</td>
<td>1</td>
<td>23</td>
<td>21.12</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0.33</td>
<td>0</td>
<td>0</td>
<td>2.00</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>57</td>
<td>1.00</td>
<td>6</td>
<td>69</td>
<td>0.54</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>4.51</td>
<td>0</td>
<td>0</td>
<td>11.00</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>7</td>
<td>10.39</td>
<td>0</td>
<td>6</td>
<td>14.08</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0.69</td>
<td>0</td>
<td>0</td>
<td>3.33</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>2</td>
<td>8.10</td>
<td>0</td>
<td>1</td>
<td>16.38</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>4</td>
<td>9.47</td>
<td>0</td>
<td>3</td>
<td>16.39</td>
</tr>
</tbody>
</table>

83.07 84.84

To achieve a similar level of expected emergency backorder-days per year, the Level Decomposition approach uses $12,152.28 investment for this sample of eight items, as compared to $10,253.67 for the Item Decomposition Method. The Item Decomposition does not place as much stock at the DC as Level Decomposition does in cases where the warehouse has a large stock. For example, item 3 has a 0.004 probability of stockout at the warehouse. The 12 additional units of stock at the DC increase the probability of not having a warehouse stockout by less than 0.001. For items with a low warehouse service level, Item Decomposition provides superior DC service. For example, the one unit carried at the DC for item 4 means that the warehouse will have that unit available with a probability of 0.66 when it is needed. For item 1, Level Decomposition performs better, but it does so by carrying one unit of item one (at a cost of $234.60 each) at each of the 15 warehouses. Item Decomposition allocates those dollars more effectively across the other items.

The overall performance can best be seen in the two tradeoff curves shown below. In Figure 2 we plot expected emergency backorder days per demand versus total investment for both methods, using $ET_1 = 0.5$ and $ET_2 = 5.0$, giving $ET_1 / ET_2 = 0.1$. The 4 curves for Level Decomposition (the single-echelon approach) represent warehouse fill rates of 0.60, 0.75, 0.85, and 0.95. The 5 dots on each line represent 0.50, 0.60, 0.75, 0.85, and 0.95 DC fill rates respectively. Figure 3 displays the same levels,
assuming $ET_1 = 1.5$ and $ET_2 = 5.0$, giving $ET_1/ET_2 = 0.3$. (The values on both axes have been multiplied by a constant to disguise the data.)

As shown in Figure 2, the difference in investment for the same level of performance is substantial. For example, to achieve a level of 0.40 expected emergency backorder days per demand, the best Level Decomposition approach uses a high fill rate at the DC coupled with a 0.50 fill rate at the warehouse. It requires over twice as much inventory ($31,000$ for the sample items) as the Item Decomposition ($15,000$ for the sample items). Since multi-echelon systems often involve over $100$ million total investment, a total inventory reduction of over $50$ million may be possible (after the new system settles down).

In the evaluation the two sets of stock levels are given equal access to emergency shipments. The decisions made are evaluated using the time-weighted objective (2). This objective is used since it was the objective stated by management. We also observed that if the fill rate constraint in the Level Decomposition method is replaced by a constraint on the total expected time-weighted backorders at each level, the difference in performance between the Level Decomposition and Item Decomposition models is virtually the same as we have reported when the service level constraint is measured in terms of fill rate at each location. This happened because for a similar total investment the stock levels for most items were the same whether the fill rate or the expected time weighted backorder measure was used to constrain performance.
Although the stock levels were normally the same, the range of items having a positive stock level increased slightly using the expected time weighted backorder constraint with a corresponding modest reduction in depth of stock of some items. The overall result was that the difference in performance was negligible when using these two different constraints. Thus the large differences we have noted appear to be attributable to the Item Decomposition approach and not the form of the constraint used in the Level Decomposition approach.

In Figure 2 you can imagine an "efficient frontier" or "envelope curve" for the Level Decomposition approach, consisting of the best performance that can be obtained for any level of investment. It is drawn tangent to points that are not dominated by a point with less investment and lower average backorder days per demand. That curve will use a low level of warehouse service over most of the performance range in the example portrayed in Figure 2. The efficient frontier is the set of undominated points from which a manager can choose an inventory policy; it gives the tradeoff curve that can be obtained using the single-echelon approach. A possible graph of the frontier is shown as a dashed line in Figures 2 and 3. To obtain these graphs more precisely, more points would have to be calculated. In particular, in Figure 2 it appears that with a warehouse fill rate of 85% a DC fill rate of over 95%

![Figure 3. Tradeoff Curves—ET_1/ET_2 = 0.3.](image)

The supposition that warehouse investment is high costs will not materialize. The approach (time-extended high costs for all methods in the system) will not materialize in the time-extended low-decision system.

We follow with a demonstration on a single-backorder system.
may be appropriate. The exact shape of the efficient frontier is not known, but in the 0.25 to 0.50 backorder-days per-demand range, the form shown seems to be correct. Throughout most of the performance range, the Level Decomposition approach requires close to twice as much investment as used by the Item Decomposition approach to achieve the same level of performance.

Figure 3 illustrates the effect of a longer DC to warehouse emergency resupply time. Each curve shifts up to the right, as is expected since emergency resupply time is longer. \( ET_2 \) remains constant. More inventory is needed to give the same level of performance since the DC loses some of its leverage. The lowest warehouse fill rate curve now crosses the curve for the next level. That is, 0.60 warehouse fill rate does not dominate the 0.75 warehouse service level. Rather than building a high-service inventory at the DC, it is preferable to push some inventory out to the warehouse, due to the longer emergency resupply time.

Item Decomposition has a smaller percent improvement (the Level Decomposition investment is about 70% higher, rather than over 100% higher as in Figure 2) when the DC to warehouse emergency resupply time lengths. When \( ET_1 / ET_2 \) approaches 1, that is, the emergency resupply from the plant is as fast as from the DC, there would be less and less inventory carried at the DC. Of course, determining the values of \( ET_1 \) and \( ET_2 \) is a management decision that is based on the cost of providing different speeds of service. The benefit of faster service in reducing inventories can be an input to those decisions.

5. Concluding Remarks

The role of DC stock is different from the role of warehouse stock. The DC stock is supposed to "fill the pipeline" and support the warehouse. It does not best support the warehouse by carrying inventory to give a high fill rate in the same items carried extensively at the warehouse level. It is efficient to centralize safety stock for relatively high cost, low demand items. Single-echelon methods such as "Level Decomposition" will not use the DC to provide additional support in these items; multi-echelon methods such as Item Decomposition will.

The larger the number of low demand items, the more important a multi-echelon approach will be. (This makes a repair-parts inventory system particularly appropriate.) The tighter the inventory budget, the more important it will be to use multi-echelon methods and thereby avoid duplicating significant amounts of safety stocks at both the DC and warehouses. However, multi-echelon methods perform much better for all levels of inventory budgets. Since almost all large multi-echelon inventory systems involve thousands of very low demand items and a constrained budget, methods that systematically take advantage of the system structure should be appropriate in all such systems. The amount of the savings will depend on the factors mentioned above. In systems such as the spare parts system we studied, with almost all low-demand items, the potential saving is large enough to be of strategic importance.

We have performed some additional analyses not reported here but that verify the following comments. First, in a large military inventory system the superiority of a time-weighted objective function over an incident oriented objective function was demonstrated by examining two multi-echelon models, one with a time-weighted backorder objective and one with a backorder occurrences objective. In the military system and the industrial system described throughout this paper, only a time-weighted
objective makes sense. In both cases the “customers” cannot go elsewhere and the demand is backordered. The result of any stockout is a machine that will not work until the part arrives. Thus speed of response is crucial, and a model that uses a time-weighted objective is necessary. In other situations, of course, the appropriate objective to be used is a management decision, and the model must be built around the appropriate objective.

Second, the “Days of Supply” method of controlling inventories, which is widely used in practice, was also examined. It performed so badly in our example that it was not presented in the comparisons. This is worth saying because, even though the weaknesses of the Days-of-Supply method are well known, it is still in widespread use.

Multi-echelon methods do require centralized information and control. However, many organizations have such information systems to facilitate control. The computation cost of the multi-echelon approach, Item Decomposition, is similar to that required by single-echelon methods such as Level Decomposition. Thus, as our example has demonstrated, it very well may be worth the effort to implement multi-echelon inventory models.

1 This research was supported in part by the Office of Naval Research under Grant Number N00014-75-C-1172 TASK NR042-335.

References