A MULTI-ECHELON INVENTORY MODEL WITH FIXED REORDER INTERVALS

by
Stephen C. Graves

WP #3045-89-MS June 1989

A MULTI-ECHELON INVENTORY MODEL
WITH FIXED REORDER INTERVALS

Stephen C. Graves
A. P. Sloan School of Management
Massachusetts Institute of Technology
Cambridge MA 02139

Draft 2.0, June 1989

This paper develops a new model for studying multi-echelon inventory systems with stochastic demand. For the model we assume that each site in the system orders at preset times according to an order-up-to policy, that delivery times are deterministic, and that the demand processes are stochastic with independent increments. We introduce a new scheme for allocating stock in short supply, whicl we call virtual allocation and which permits significant tractability. We exercise the model on a set of test problems for two-echelon systems to get insight into the structure of good policies. We find that the least-inventory policy puts the safety stock at the retail sites (lower echelon) with the central warehouse (upper echelon) primarily serving as a central ordering agent. Nevertheless, the test problems show some benefit from holding stock at the central warehouse, even though it will stock out with high probability. Furthermore we are able to show that the virtual allocation rule is near optinal for the set of test problems.

1. Introduction

Multi-echelon inventory systems are of great practical interest and significance. Most consumer and industrial finished goods are distributed through multi-echelon inventory systems of one sort or another. Spare parts for office equipment, computers, automobiles, and military hardware are commonly provided through multi-echelon systems. Any enterprise with geographically-dispersed demand, economies of scale in production and/or transportation, and market-driven service requirements must typically rely on a multi-echelon inventory system to remain competitive. Multi-echelon inventory systems are also common in production contexts, particularly in multi-plant operations where the inventories may act to decouple one facility from another.

Over the past thirty years there has been much progress in developing an inventory theory for these multi-echelon systems. In particular, for deterministic demand there are very effective procedures for setting reorder intervals for a wide range of systems (Roundy 1985, Maxwell and Muckstadt 1985). For serial systems with stochastic demand, there are approaches for finding good or optimal order policies for both the periodic review case (Clark and Scarf 1960, Federgruen and Zipkin 1984c) and continuous review case

(De Bodt and Graves 1985). For one-for-one systems with stochastic demand, there is a rich literature on models and algorithms for finding stockage policies for multi-echelon systems, e.g.,Sherbrooke (1968), Simon (1971), Muckstadt (1973), Graves (1985), Svoronos and Zipkin (1988b), Axsater (1988).

The more general problem with stochastic demand, a distribution or general network, and batch or periodic ordering, seems to be much harder, and progress here has been slower. Most of the work considers a two-echelon distribution system with identical retail sites and Poisson demand, and then develops an approximate model of system cost or performance as a function of stockage levels; a simulation is used to evaluate the approximate model. Noteworthy examples are the papers by Deuermeyer and Schwarz (1981) and Svoronos and Zipkin (1988a) for the continuous review case, by Jackson (1988) for a periodic review case, and by Eppen and Schrage (1981) and Federgruen and Zipkin (1984b, 1984c) for a periodic review case in which the central depot holds no stock. Jackson (1988) and Schwarz (1989) provide excellent reviews of this literature with several additional citations; rather than duplicate these papers, I refer the reader to their discussions of the literature.

In this paper I present a new model for multi-echelon inventory systems. For this model I require two key assumptions: a fixed schedule for replenishments for all sites in the system, and a seemingly simplistic allocation rule in which stock at an upper echelon is virtually committed as demand occurs at a lower echelon in the system. All sites follow a base stock (or order-up-to) policy, which is the same policy as considered by Jackson. The assumptions for the model are stated in the next section, Section 2, and a discussion of the assumptions follows in Section 3. The model is developed for the case of Poisson demand, deterministic transhipment (or lead) times, and a distribution network topology. In Section 4, I provide an exact characterization of the inventory at any time at any site in the system. In Section 5, I use an example to show how to exercise the model in a general context for a given stockage policy. We can simplify the computational requirements of the model for a two-echelon system. I show this in Section 6, and introduce a two-moment approximation for the inventory distribution. A set of test scenarios are analyzed in Section 7; these tests provide insight into the structure of the optimal stockage policy and, in particular, show how inventory should be allocated across a two-echelon system for these scenarios. A lower bound for the allocation policy is described in the next section; the evaluation of this lower bound on the test problems shows that the assumed policy, namely virtual allocation, results in near-optimal inventories. Finally, in Section 9, I discuss how the model and analyses might extend to permit more general demand processes, stochastic lead times, and general network topologies.

2. Assumptions

Network Topology: We consider a multiechelon distribution system consisting of M inventory sites, i =1, 2, ... M. Each site j has a single supplier $i = \rho(j)$, except for site 1 which is replenished from an exogenous source; we assume that the sites are numbered such that $\rho(j) < j$.

We term sites with no successors as *retail* sites. The remaining sites act as storage and consolidation facilities, and are called *transhipment* sites. The unique sequence or path of sites from site 1, to a retail site is the *supply chain* for the retail site. We call site 1 the *central warehouse* (CW).

Demand: The distribution system supplies a sirgle good or commodity from a single source (external supplier for site 1) to a population of customers. We assume that customer demand occurs only at retail sites. (The model can easily permit customer demand at the intermediate sites by splitting the site into two sites: a pure transhipment site and a retail site.) Customer demand that cannot be met by inventory on hand is backordered until sufficient inventory is available. The demand at each retail site is given by an independent Poisson process. For each retail site j, we define $D_j(s, t)$ as the demand over the time interval (s, t]; for each transhipment site we define $D_j(s, t)$ as

$$D_j(s, t) = \sum D_i(s, t)$$

where the summation is over the immediate successors to site j, that is, i such that $j = \rho(i)$. For each site i, either retail or transhipment, λ_i is the expected demand rate per unit time.

Fixed Scheduling: We assume that for each site we are given a schedule of preset times at which each site places its replenishment orders on its supplier; that is, for each site j we are given $p_j(m)$ for m = 1, 2, ... where $p_j(m)$ is the time at which site j places its mth order. We also assume that the times at which the corresponding shipments arrive at the site are preset and known. In particular, $r_j(m)$ is the time at which site j receives its mth shipment, where the mth shipment follows the mth order. However, the quantity received in the mth shipment may not exactly match the quantity ordered in the mth order. The supplier may ship less than ordered when there is an inventory shortage, and will make up the shortfall on subsequent shipments. We define $p_j(m)$ to be less than $p_j(m+1)$, and require that $r_j(m)$ is less than $r_j(m+1)$, and $p_j(m)$ is less than or equal to $r_j(m)$. The difference $r_j(m) - p_j(m)$ is the lead time for the m^{th} order by site j, and may vary by order. The requirement that $r_j(m)$ is less than $r_j(m+1)$ means that there is no order crossing.

Ordering Policy: At each order occasion, a site follows an order-up-to policy based on echelon stock. Since we assume that all customer demand is met, this policy translates into an order quantity which replenishes all demand since the last order occasion. Namely, at its m^{th} order occasion $p_j(m)$, site j places an order equal to $D_j[\ p_j(m-1), p_j(m)\ ]$.

To initiate this policy we assume that each site has at time zero an initial inventory (which we term the base stock or order-up-to point), given by B_j for site j. We assume that the first order is placed at $p_j(1) > 0$; we require that $B_j \geq 0$ for all j.

Order Filling Policy: Suppose site i is the supplier to site j, i.e. $i=\rho(j)$. Then site i ships to site j only when there is an order occasion ʿʸ site j. The amount shipped depends on the inventory availability at the supplier and on the allocation policy by the supplier. When site i does not completely fill the order by site j, then the unfilled portion of the order is treated as a backorder on the supplier. This backorder remains open at least until the next order occasion by site j; only then will the supplier try again to fill the outstanding order.

We define the random variable $T_j(m)$ to represent the coverage provided by the supplier on the occasion of the m^{th} order by site j. In particular, the amount shipped by the supplier at time $p_j(m)$ [to arrive at time $r_j(m)$] is given by $D_j[\ T_j(m-1), T_j(m)\ ]$. Since the supplier cannot cover demand that has not yet occurred, we must have that $T_j(m) \leq p_j(m)$. If $T_j(m) < p_j(m)$, then the supplier did not completely fill the order by site j, and $D_j[\ T_j(m), p_j(m)\ ]$ remains on backorder. If $T_j(m) = p_j(m)$, then site i has filled the m^{th} order by site j.

The primary thrust of the analysis in the next section is to characterize the random variable $T_j(m)$. To do this we require an assumption about how the supplier allocates inventory when faced with a possible shortage.

The CW (site 1) is an exception to this discussion since its supplier is external to the system. We assume that this external supplier is completely reliable and fills every order exactly as scheduled [i.e., $T_1(m) = p_1(m)$].

Virtual Allocation: We assume that each transhipment site observes in real time the demand processes at all of its downstream retail sites. Since each site follows an order-up-to policy, each demand event will eventually trigger a sequence of replenishment requests from the retail site up the supply chain to the CW: each site on the supply chain increases its next order by one due to a particular demand occurrence. In anticipation of these replenishment requests, we assume that at the time of a demand event each transhipment site on the supply chain acts as if the replenishment order were placed concurrently. Each site on the supply chain commits a unit of its inventory, if available, to replenish the downstream site; however, the actual shipment of this unit to the downstream site does not occur until the next order occasion. In effect, a site will take a unit from its uncommitted inventory and load it into a waiting truck that is destined for the next site in the supply chain. However, this truck does not depart until the next scheduled departure time. Once a unit has been committed for shipment to a downstream site (e.g., loaded onto the truck), it cannot be uncommitted and will wait until the next order occasion when the actual shipment occurs.

If a site on the supply chain does not have an uncommitted unit to commit at the time of the demand event, the site creates a backorder and adds this to the current list of outstanding orders. When inventory becomes available to service these outstanding orders, they are filled in the order in which they were created. Again, though, the filling of the outstanding order results only in committing the inventory to the next scheduled shipment.

3. Discussion of Assumptions

Network Topology: We present the model and analysis for a pure distribution system. We sketch how to extend the model to permit more general networks with assembly sites in the last section of the paper; however, it is not clear that the analysis required by the model will remain computationally feasible for this case. Hence, the primary presentation is for distribution systems.

Demand: The analysis requires that the demand process at each site has independent increments and has unit demands; for convenience and tractability, we present the model for the most common instance of this type of process, a Poisson process. Of these two requirements, the requirement of independent increments seems to be the more crucial. As will be seen, the analysis focuses on the time when the uncommitted stock at a site is depleted. As long as the demand process has independent increments, we can find the time at which the uncommitted inventory first reaches or crosses zero. For the case of unit demands, at the depletion time there is exactly zero uncommitted inventory and no shortages. For non-unit demands (e.g., a compound Poisson process), there may be shortages at the depletion time, the existence of which complicates the analysis. In the last section, we describe how to extend the analysis to this case.

Fixed Scheduling: There are several comments that need to be made about this assumption. First, the motivation for the assumption is multi-item distribution systems in which there are regularly-scheduled shipments between sites. Regularly-scheduled shipments are common in practice in order to achieve an efficient utilization of available transportation resources and/or to procure transportation services at least cost. In a multi-item inventory system, where the replenishment of each item occupies a small portion of a truckload, transportation economies will dictate a replenishment schedule that fosters consolidation of item shipments. As such, we may interpret the times for order placement, $p_j(m)$, as the times for a dispatch of a truck or rail car from the supplier destined for site j.

Second, the assumption of fixed scheduling separates the problem of determining the reorder intervals from the problem of setting safety stocks. There has been extensive study and significant progress on determining reorder intervals for distribution systems (e.g., Roundy 1985, Graves and Schwarz 1977) and for more general networks (Maxwell and Muckstadt 1985). By assuming fixed scheduling, we effectively presume a hierarchical approach in which the reorder intervals are determined first, followed by setting the safety stocks. Alternately, one might envision a procedure that iterates between setting the reorder intervals and finding the required safety stocks.

Third, we have assumed that the lead time for each order, $r_j(m)$ - $p_j(m)$, is known and deterministic. In the last section we discuss how to extend the model to permit stochastic lead times.

Fourth, the model and analysis do not require any assumptions about the pattern of the order schedule, other than it is fixed and known. Nevertheless, we would expect that the fixed schedule would exhibit some regularity. For instance, the lead time to deliver an order is likely to be constant for a site from order to order, and the time between placing successive orders could be constant for a site. Also, we might expect that the order schedule is nested: whenever a site receives an order, each immediate successor places an order. Although the analysis does not require it, the order policy implicitly assumes some regularity in the order schedule, as discussed next.

Ordering Policy: Given the assumption of fixed scheduling, then an order-up-to policy with base stocks is most reasonable, provided that there is a regular pattern of replenishments and constant lead times. The order-up-to (base stock) level is set, roughly, to cover demand between the time of placing an order, $p_j(m)$, and the time of the receipt of the next order, $r_j(m+1)$; that is, the inventory needs to cover demand over the review period, $p_j(m+1)$ - $p_j(m)$, plus the lead time, $r_j(m+1)$ - $p_j(m+1)$. If for site j this time interval $r_j(m+1)$ - $p_j(m)$ is highly variable, then site j will need to adjust its order-up-to (base stock) level to accommodate the variability in the replenishment intervals. We can build this accommodation into the model, but ignore it here to simplify the presentation.

In order to implement this policy (as well as virtual allocation), we need to assume that there is an information system which permits a site to observe demand as it occurs at the successors to the site. That is, each site will know its echelon inventory in real or near-real time. This is technically feasible and exists in some systems for high value or high volume goods.

Order Filling Policy: This policy is consistent with the notion of a fixed schedule, particularly a fixed transportation schedule. A truck goes on a regular schedule from supplier to site; on each shipment occasion, the supplier tries to fill as much of the outstanding order from the site as possible. There are no emergency or unscheduled shipments.

Virtual Allocation: One motivation for this assumption is to obtain tractability in the analysis. The assumption states that each transhipment site allocates its inventory on a virtual basis as demand occurs. Inventory at a site is committed in a FIFO (first-in-first-out) fashion where the timing of the demand events sets the order of allocation. On the one hand, this scheme can be viewed as an equitable allocation rule in which a site's inventory is always applied to the oldest outstanding orders. On the other hand, the scheme is not optimal in that it does not account for the relative need of downstream sites for inventory replenishment. For instance, it may be desirable to uncommit an inventory unit, which had been destined for one site with ample safety stock, and redirect it to another site with a more critical need for replenishments. In this respect, the virtual allocation rule will not perform as well as a dynamic allocation rule that takes into account more information.

To gain insight into the non-optimality of virtual allocation, we compare it versus an idealized allocation policy. On a set of examples we show that the difference in inventory requirements between the best inventory policy, assuming virtual allocation, and the best inventory policy for the idealized allocation, is quite small.

4. Model Analysis: Characterization of Inventory

The first step in the development of the model is to specify the inventory at each site. Let $I_j(t)$ denote the echelon inventory at site j at time t: that is, $I_j(t)$ is the inventory at site j, plus all of the inventory at or in transit to the successors to site j, minus any backorders at the retail sites served by j. Let $\mathbb{B}_j$ denote the echelon base stock for site j; that is, $\mathbb{B}_j = B_j + \Sigma\, \mathbb{B}_i$ where the summation is over the immediate successors to j. Then we argue that

$$I_j(t) = \mathbb{B}_j - D_j[\, T_j(m), t\,] \qquad (1)$$

for $r_j(m) \leq t < r_j(m+1)$ and where $T_j(m)$ represents the coverage provided by the supplier on the m^{th} shipment to site j. We will characterize $T_j(m)$ below.

The argument for (1) is immediate. By assumption, $I_j(0) = \mathbb{B}_j$. The demand on the echelon stock at site j up to time t is $D_j[\, 0, t\,]$. The total replenishments to the echelon stock as of time t are $D_j[\, 0, T_j(m)\,]$ by the definition of $T_j(m)$. The current inventory is the initial inventory at time 0 minus the demand over the time interval [0,t] plus the replenishments over this interval [0,t]. A crucial requirement for this result is that there be no lost sales.

We can use a similar argument to characterize the available inventory at each site, denoted by $A_j(t)$: $A_j(t)$ is the inventory at site j at time t that has not been committed for shipment to another site. A negative value for $A_j(t)$ corresponds to outstanding orders or backorders at site j at time t. By noting that $A_j(0) = B_j$ and repeating the argument for (1), we have

$$A_j(t) = B_j - D_j[\, T_j(m), t\,] \quad , \qquad (2)$$

where $r_j(m) \leq t < r_j(m+1)$. Implicit in this development is the fact that the demand process that depletes the available inventory is the same as that for the echelon inventory; similarly the replenishment process is the same for the available inventory as for the echelon inventory.

To use (1) or (2) we need to specify $T_j(m)$. At time $p_j(m)$, site j places an order for $D_j[\, p_j(m-1), p_j(m)\,]$ on its supplier, site i where $i = \rho(j)$; site i ships $D_j[\, T_j(m-1), T_j(m)\,]$ where $T_j(m)$ either equals $p_j(m)$ if sufficient stock is available, or equals the time at which the supplier ran out of available inventory to allocate to site j. Suppose n is such that $r_i(n) \leq p_j(m) < r_i(n+1)$; that is, at time $p_j(m)$, site i has received its n^{th} shipment, but has not yet received its $n+1^{st}$ shipment. We define $S_i(n)$ as the depletion or runout time for the n^{th} shipment to site i; that is, based on the receipt of the n^{th} shipment, site i is able to cover or replenish the demand processes of its successor sites up through time $S_i(n)$. Hence, if $S_i(n)$ occurs after $p_j(m)$, $T_j(m) = p_j(m)$; if $S_i(n)$ occurs before $p_j(m)$, $T_j(m) = S_i(n)$. Thus, we have

$$T_j(m) = \min [\, p_j(m), S_i(n) \,] \quad . \qquad (3)$$

To use (3) we need to characterize $S_i(n)$. Although we term $S_i(n)$ as the runout time for the n^{th} shipment to site i, this time may occur before the actual receipt of the n^{th} shipment, as will be seen. We will show that

$$S_i(n) = \min \{\, s\text{:}\ D_i[\, T_i(n), s \,] = B_i \,\} \quad . \qquad (4)$$

To demonstrate (4), we consider two cases. First, suppose from (4) that $S_i(n) \geq r_i(n)$; then from (2) we see that $S_i(n)$ corresponds to the first instant $s \geq r_i(n)$, when the available inventory, $A_i(s)$, reaches zero, provided we ignore all subsequent shipments to site i. Thus, the n^{th} shipment runs out at time $S_i(n)$. The second case is when $S_i(n) < r_i(n)$. Here the receipt of the n^{th} shipment does not cover all of the outstanding orders and leaves the available inventory negative, i.e. $A_i(s) < 0$ at $s=r_i(n)$. In particular, from (2) we observe that at $s=r_i(n)$, $A_i(s) = -D_i[\, S_i(n), s \,]$. Thus, in this case the n^{th} shipment covers the demand processes through but not beyond time $S_i(n)$, and we say that the shipment ran out at time $S_i(n)$.

We note from (4) the relation between the runout time $S_i(n)$ and the coverage $T_i(n)$ for the n^{th} shipment to site i. The difference between these two times is the buffer time provided by the base stock B_i at site i. When the base stock is zero, the runout time is the same as the coverage time. When the base stock is positive, the difference between the two times is a random variable with a gamma distribution with parameters (λ_i, B_i); that is, $S_i(n) - T_i(n)$ is a gamma random variable with mean and variance given by B_i/λ_i and $B_i/(\lambda_i)^2$. Furthermore, the random variable $S_i(n) - T_i(n)$ is independent of $T_i(n)$, due to the assumption that the demand process has independent increments.

This completes the general characterization of the inventory process. In summary, to characterize the echelon or available inventory at a site j at time t, we need to identify the relevant shipment m for time t and then characterize $T_j(m)$, the coverage provided by the m^{th} shipment to site j. To characterize $T_j(m)$, we need to identify the relevant shipment n to supplier i for the m^{th} order placed by site j; then we have to characterize $S_i(n)$, the runout time for the n^{th} shipment to site i. To characterize the runout time $S_i(n)$, we need to characterize $T_i(n)$, the coverage provided by the n^{th} shipment to site i. We can continue in this fashion up the supply chain until we reach the CW. Here we can stop since the external supplier is reliable and hence, $T_1(m) = p_1(m)$ for all orders by the CW (site 1).

Before closing this section, we might try to relate this model to more traditional inventory concepts and approaches. From (1) or (2) we see that the inventory at time t is a base stock level minus the demand over an interval of length t - $T_j(m)$, which we can write as

$$t - T_j(m) = [t - r_j(m)\,] + [\, r_j(m) - p_j(m)\,] + [\, p_j(m) - T_j(m)\,]\,.$$

The first component is the time since the receipt of the last shipment, the second component is the lead time for the m^{th} order, and the third component is the shortfall in resupply due to a runout by the supplier, site i. Thus, we see here the role of the replenishment lead times and supplier runouts in determining the inventory at a site at a random time t.

We also note that the model just characterizes the echelon or available inventory at a site at a time instant; thus, from the model we can obtain the probability density function for the inventory level at any time. For determining the best inventory policy, though, we might want to know the average inventory for a given policy (i.e., specification of base stocks). To determine this, one might first find the average inventory at each time instant and then integrate this over an appropriate time interval; however, care must be taken with the choice of time interval to ensure a proper accounting. Ideally, there would be a cyclic pattern to the ordering schedule and then the time interval for integration would cover an entire order cycle for the system. For instance, if we have a single-cycle ordering policy (described in Section 7), then one could integrate over the time interval between the receipt of two successive shipments to the CW, say from $r_1(m)$ to $r_1(m+1)$. When the ordering schedule is not cyclic, it is not clear that average inventory is well defined, other than for a specified time interval.

5. Example

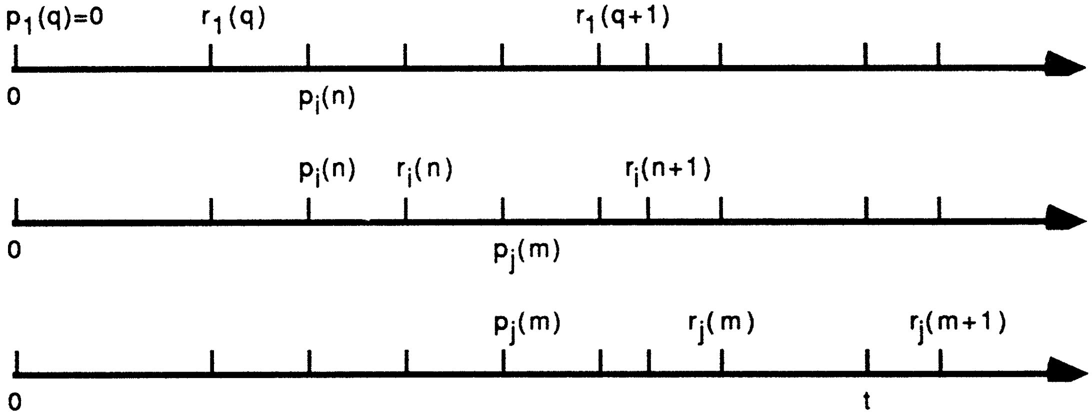
As an example consider a supply chain consisting of three sites, namely sites 1, i and j where $1 = \rho(i)$ and $i = \rho(j)$. Suppose we want to characterize the echelon or available inventory at site j at time t, i.e., either $I_j(t)$ or $A_j(t)$. To do this, we first need to identify the relevant orders for the supply chain at time t. Suppose integers m, n and q are such that

$$r_j(m) \leq t < r_j(m+1),$$
$$r_i(n) \leq p_j(m) < r_i(n+1),$$
and $$r_1(q) \leq p_i(n) < r_1(q+1).$$

Thus, at time t, site j has received its m^{th} shipment; at the time of the placement of site j's m^{th} order, site i had received its n^{th} shipment; and, at the time of the placement of site i's n^{th} order, the CW (site 1) had received its q^{th} shipment.

We depict in Figure 1 the time diagram for the relevant orders, where we arbitrarily set the time for the placement of the q^{th} order by the CW to be time zero: $p_1(q) = 0$. (Equivalently, we restate the time scale by subtracting $p_1(q)$ from the original times.)

Figure 1: Time Diagram for Relevant Orders



For the remainder of this section, we will drop the order indices for the relevant orders for ease of presentation. It should be understood, though, that the timing of events is as depicted in Figure 1. Whenever we refer to the relevant order by site 1, i or j, we are referring to the q^{th}, n^{th} or m^{th} order, respectively.

Now to evaluate (1) and (2) for site j at time t, we start at the top of the supply chain and sequentially characterize the random variables for the coverage and runout times for the relevant shipments. Since the supplier to the CW is completely reliable, the coverage provided by the relevant (q^{th}) shipment, which is received at r_1, is given by

$$T_1 = p_1 = 0 .$$

From (4), we can now evaluate the runout time for this shipment by

$$S_1 - T_1 = G(\lambda_1, B_1) ,$$

where $G(\lambda, B)$ is a gamma random variable with parameters (λ, B). Since $T_1 = 0$, we have that $S_1 = G(\lambda_1, B_1)$.

We now repeat these steps for site i, first using equation (3) to evaluate T_i and then using equation (4) to evaluate S_i. The coverage from the relevant (n^{th}) shipment to site i is given by

$$T_1 = \min [p_i, S_1] ,$$

where p_i is a known constant, and S_1 is a gamma random variable. Hence the distribution of T_i is a truncated gamma distribution with a mass point at $T_i = p_i$.

The runout time for the relevant (n^{th}) shipment to site i is found from (4):

$$S_i = T_i + G(\lambda_i, B_i) .$$

Thus, we obtain the distribution for the runout time by convolving the coverage random variable (a truncated gamma) with an independent gamma random variable with parameters (λ_i, B_i).

For site j and subsequent sites in the supply chain, we repeat these steps. The coverage provided by the relevant (m^{th}) shipment to site j is obtained by truncating at p_j the runout time from the upstream site:

$$T_j = \min [p_j, S_i] \quad .$$

The characterization of the coverage random variable gives the distribution of the inventory, either from (1) or (2), at site j at time t. Given the distribution of T_j, we first obtain the distribution of $D_j[T_j, t]$, namely the uncovered demand from the last shipment until time t.

$D_j[\ T_j,\ t\ ]$ is a mixture of Poisson random variables with means $\lambda_j(t - T_j)$, where the mixture is over the distribution of T_j. Knowledge of the distribution of $D_j[\ T_j,\ t\ ]$ gives immediately the distribution of the echelon inventory or available inventory from (1) or (2), respectively.

The runout time for the relevant shipment to site j is found by convolving the distribution for the coverage random variable with that for a gamma random variable:

$$S_j = T_j + G(\lambda_j, B_j)\ .$$

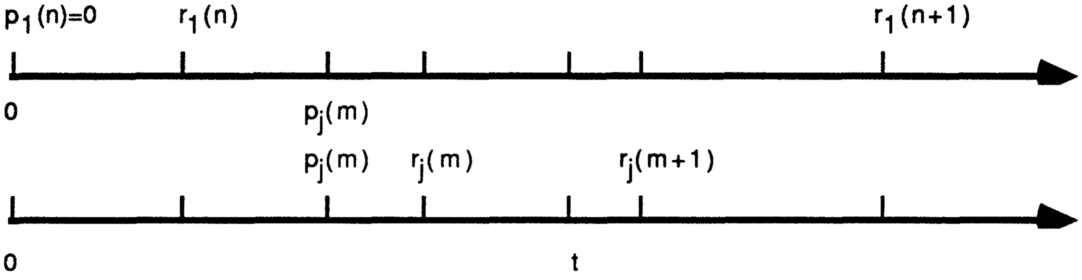
If site j is a transhipment site, then the runout time is needed to characterize the inventories at the successors to site j.

From this description we can find the distribution of the inventory *at any site at any time* for the given model assumptions. From a computational standpoint the most difficult steps are the determination both of the runout time, which requires the convolution of two random variables, and of the uncovered demand $D_j[\ T_j,\ t\ ]$, which requires the evaluation of a mixture of Poissons. In the next section we show how these calculations simplify for the case of a two-echelon system.

6. Two-Echelon System

Suppose that we restrict attention to a two-echelon system. As an example consider a supply chain consisting of sites 1 and j where $1 = \rho(j)$ and site j is a retail site; note that the CW (site 1) may also supply several other retail sites. Again we want to characterize the echelon or available inventory at site j at time t, i.e., either $I_j(t)$ or $A_j(t)$. The relevant orders for the supply chain at time t are m for site j and n for the CW ; that is, $r_j(m) \leq t < r_j(m+1)$, and $r_1(n) \leq p_j(m) < r_1(n+1)$. We define time such that $p_1(n) = 0$, and depict the timing of these events in Figure 2 below.

Figure 2: Time Diagram for Two Echelon System



Thus, at time t, site j has received its m^{th} shipment; the inventory at site j at time t depends on the coverage from the m^{th} shipment, which depends on the time of the placement of site j's m^{th} order, and on the runout time of the relevant shipment to the CW, namely its n^{th} shipment. Again, for convenience we drop the indices for the orders.

To characterize the inventory at site j at time t, we need to evaluate the coverage provided by the most recent shipment. The coverage from this shipment to site j is given by

$$T_j = \min [p_j, S_1] , \qquad (5)$$

where p_j is a constant, and S_1 is a gamma random variable with parameters λ_1 and B_1.

The first two moments for the coverage time T_j are found directly from (5) to be

$$E[T] = E[S] + p \Pr\{ S > p \} - \int_p^\infty x f_S(x)\, dx \qquad (6),$$

and

$$E[T^2] = E[S^2] + p^2 \Pr\{ S > p \} - \int_p^\infty x^2 f_S(x)\, dx \qquad (7),$$

where $T = T_j$, $p = p_j$ and $S = S_1$ and $f_S(x)$ is the probability density function for S_1.

We can simplify equations (6) and (7) by using the fact that S_1 has a gamma distribution $G(\lambda, B)$ with parameters $\lambda = \lambda_1$ and $B = B_1$. In particular, we can express the partial expectations in (6) and (7) in terms of the cumulative probability distribution for gamma random variables.

$$\begin{aligned} \int_p^\infty f_S\, dx &= \Pr\{ G(\lambda, B) > p \} \\ \int_p^\infty x f_S\, dx &= \frac{B}{\lambda} \Pr\{ G(\lambda, B+1) > p \} \qquad (8) \\ \int_p^\infty x^2 f_S\, dx &= \frac{B(B+1)}{\lambda^2} \Pr\{ G(\lambda, B+2) > p \} \end{aligned}$$

Furthermore, for B an integer, we note that

$$\Pr\{ G(\lambda, B) > p \} = \Pr\{ \Pi(\lambda p) < B \},$$

where $\Pi(\lambda p)$ denotes a Poisson random variable with mean λp. Hence, the three partial expectations in (8) can be found from the cumulative distribution function for a single Poisson random variable, namely $\Pi(\lambda p)$.

Given the first two moments of the coverage time T_j, we can use this to characterize the inventory at site j at time t. To do this, we need to determine the uncovered demand $D_j[T_j, t]$, which appears in equations (1) and (2). For a Poisson demand process with parameter λ_j, we can show that

$$E\{ D_j[T_j, t] \} = \lambda_j (t - E[T_j]) \qquad (9)$$

and

$$\mathrm{Var}\{ D_j[T_j, t] \} = \lambda_j (t - E[T_j]) + (\lambda_j)^2 \mathrm{Var}[T_j] \qquad (10).$$

I have found that the negative binomial distribution with the same first two moments is a very accurate approximation to the distribution of $D_j[T_j, t]$. [An exception occurs when $\mathrm{Var}[T_j(m)] = 0$, in which case $D_j[T_j(m), t]$ has a Poisson distribution.] As an illustration see Table 1 where I compare the actual to the approximate (negative binomial) distribution for a case with $\lambda_1 = 36$, $\lambda_j = 2$, $p_j = 2$, $t = 4$, and three possible values for B_1, representing low, medium and high stockage levels at the CW. Graves (1985) and Lee and Moinzadeh (1987) have previously shown the effectiveness of a negative binomial approximation for multi-echelon systems with one-for-one ordering and batch ordering, respectively.

As we increase the mean of $D_j[T_j, t]$ (e.g., by increasing λ_j, increasing the difference $t - p_j$, or decreasing B_1), the normal distribution becomes an increasingly good approximation for the negative binomial distribution, and thus can be used as an alternative approximation for the distribution of $D_j[T_j, t]$.

We can use the approximate distribution for $D_j[T_j, t]$ in (2) to obtain an approximate distribution for the available inventory at time t at site j as a function of the base stock level B_j. We would then set B_j so as to achieve some desired service level.

The determination of the moments for the uncovered demand at site j at time t requires knowledge of the timing of the relevant orders at the CW and at site j, the demand rates for site j and for the entire system, and the base stock at the CW . The base stock at site j (which equals the echelon base stock since j is a retail site) is only needed in the characterization of the inventory at time t, from either (1) or (2). This suggests an iterative optimization approach for finding the base stock levels for all sites, which achieve a desired service level with a minimum inventory investment. In particular, one would first set the base stock at the CW , and then determine the minimum base stocks needed at each retail site to provide the desired service level (e.g., 0.97 probability of stockout); then, by searching over possible settings for the base stock at the CW , one could find the overall best setting for the stockage levels. To help in this search, we note the changes in (6) and (7) as we increase the base stock at the CW from B_1 to B_1+1. In particular, if we let $\Delta E[T_j]$ and $\Delta E[T_j^2]$ denote the change, then we can show that

$$\Delta E[T_j] \;=\; \Pr\{ \Pi(\lambda_1 p_j) > B_1 \}/\lambda_1$$

and $\Delta E[T_j^2] = 2\,(B_1+1)\,\Pr\{\,\Pi(\lambda_1 p_j) > B_1+1\,\}/(\lambda_1^2)$.

We can use these results to see immediately the change on the moments for the uncovered demand when increasing by one the base stock at the CW . To get started we note that when $B_1 = 0$, the runout time S_1 is deterministic and equal to $p_1(n)$, which is 0 by assumption; hence, from (5) and the fact that $p_j(m) \geq r_1(n) \geq p_1(n) = 0$, we see that T_j is deterministic and equal to $S_1 = 0$ when $B_1 = 0$.

When the supply chain contains more than two sites, the runout time at an intermediate site, e.g., S_i in the previous example, will not have a nice distribution; hence, the calculation of the moments of the coverage time, T_j, is not as straightforward. However, the moments of S_i are readily available from the coverage time provided by the relevant shipment from the supplier of site i. In particular, the runout time for the relevant shipment to site i is

$$S_i = T_i + G(\lambda_i, B_i),$$

where T_i and $G(\lambda_i, B_i)$ are independent random variables. Thus, the mean and variance for S_i are found to be

$$E[\,S_i\,] = E[\,T_i\,] + B_i/\lambda_i$$

and

$$\mathrm{Var}[\,S_i\,] = \mathrm{Var}[\,T_i\,] + B_i/(\lambda_i)^2.$$

This suggests that we might suppose that S_i has a gamma distribution with parameters consistent with the mean and variance given above. With this approximation, we can reapply (6) and (7) to evaluate the first two moments of T_j, the coverage time for a site downstream of site i.

7. Computational Study

We have performed some computational experiments to understand better how to set the order-up-to levels in a two-echelon system. In particular, we desire some general insight into how to locate inventory across the echelons to provide the best service. We use the approximation outlined above for this computational investigation.

We consider 16 test scenarios. For each scenario the system demand rate equals 36, i.e., $\lambda_1 = 36$. We then assume we have a set of N identical retail sites, where we let N = 2, 3, 6, or 18. Thus, the demand rate at the retail sites has four possibilities, $\lambda_j = 18, 12, 6$ or 2.

For each demand possibility, we need to specify the ordering schedule and the replenishment lead times. We assume here that the order placement schedule is periodic with the CW ordering less frequently than the retail sites. For each retail site we specify a review period, call it rp_j, such that $p_j(m+1) = p_j(m) + rp_j$ for all m. We assume that the review period for each retail site is one time unit, i.e., $rp_j = 1$. The review period for the CW, rp_1, is a multiple of the review period for the retail sites, and for our experiments, we set $rp_1 = 2$ or $rp_1 = 5$. The replenishment lead time for the CW or a retail site, call it rt_1 or rt_j, is such that $r_1(m) = p_1(m) + rt_1$ and $r_j(m) = p_j(m) + rt_j$ for all m. Furthermore, we assume a single-cycle, nested ordering policy [Graves and Schwarz 1977] such that each retail site places an order upon the receipt of an order by the CW ; for instance, when $rp_1 = 2$, then every second order by a retail site coincides with the receipt of an order by the CW , i.e., $p_j(2m - 1) = r_1(m)$ for m = 1, 2, ...

We are now ready to specify the 16 scenarios for our computational study. For each of the four demand possibilities we consider four order policies as follows:

a) $rt_1 = 1, \ rp_1 = 2, \ rt_j = 1;$
b) $rt_1 = 1, \ rp_1 = 2, \ rt_j = 5;$
c) $rt_1 = 4, \ rp_1 = 5, \ rt_j = 1;$
d) $rt_1 = 4, \ rp_1 = 5, \ rt_j = 5.$

For each demand case and order policy, we want to find the best inventory policy: the order-up-to levels, i.e., B_1 and B_j, which give a desired service level with the minimum amount of inventory. We specify the service level as the probability that a retail site stocks out during the order cycle for the CW, where an order cycle runs from the receipt of one shipment to the receipt of the next shipment. To measure this service level, we need to find that time t within an order cycle of the CW for which the probability of stockout at the retail site is greatest. We claim that for site j this time t is just prior to $r_j(m+1)$, where $r_j(m)$ is the last order to be replenished from an order by the CW (e.g., if $rp_1 = 5$, then this is the fifth order by the retail site in the order cycle.). It should be clear that within an order cycle at the retail site the highest probability of no inventory occurs at the end of the cycle, i.e., just prior to the receipt of the next shipment. Furthermore, within an order cycle for the CW, the last shipment from the CW to a retail site will never provide relatively more coverage than earlier shipments in the order cycle.

If we consider an order cycle for the CW in which the order is placed by the CW at time zero, then for the four order policies we are interested in the available inventory where

a) $p_j(m) = 2, t = 4;$
b) $p_j(m) = 2, t = 8;$
c) $p_j(m) = 8, t = 10;$
d) $p_j(m) = 8, t = 14.$

To understand the problem setting, consider case (c). The CW places its order at $p_1 = 0$, which we term its first order, and receives the corresponding shipment at time $r_1 = p_1 + rt_1 = 4$; it will not order again until $p_1 = p_1 + rp_1 = 5$, the shipment for which is received at time 9. Retail site j places an order in every period ($rp_j = 1$) and receives the corresponding shipment in the next period ($rt_j = 1$). The order placed at time 4 is the first order by the retail site that is served by the first order placed by the CW. The order placed at time 8 is the last (and fifth) order by the retail site that is served by the first order placed by the CW. The order placed at time 8 by retail site j arrives at time 9, and the next receipt at site j will be the first shipment served by the second CW order and will arrive at time 10. Hence, the inventory at site j will be lowest just prior to time 10 and the relevant problem parameters are $p_j = 8$ and $t = 10$.

The four order policies are distinguished by the time between the order placement by the CW ($p_1 = 0$) and by the retail site ($p_j = 2$ or 8), and the time between the placement of an order and the receipt of the next order ($t - p_j = 2$ or 6). Presumably, the inventory at the CW would be sensitive to the first dimension, whereas the inventory at the retail site would be more sensitive to the second dimension.

We consider four service levels, namely $\alpha = .80, .90, .95$ and $.975$, and thus, have 64 test problems: 16 scenarios, each with four service levels. The test problem is to find the inventory policy (B_1, B_j) with the least amount of inventory for which the probability that $I_j(t) \geq 0$ is at least α. That is, we set the inventory policy so that for each retailer the probability of stockout at the end of the order cycle (time t) is $1 - \alpha$. We solve each test problem by searching over a range of values for the base stock at the CW ; for each value of B_1 we find the moments for $D_j[\, T_j, t\,]$ and then use the negative binomial approximation for the distribution of $D_j[\, T_j, t\,]$ to set the base stock at the retail site, B_j, to satisfy the service criterion.

We use $\mathbb{B}_1 = B_1 + NB_j$ as an approximate measure of the total inventory for a given policy. The average inventory level for the two-echelon system equals $\mathbb{B}_1$ minus the expected inventory on order to the CW plus the expected backorders at the retail sites. The expected inventory on order to the CW is a constant (namely $\lambda_1 rt_1$) and does not depend on the inventory policy (B_1, B_j). The expected backorders at the retail sites does depend on the inventory policy, but is extremely small and relatively insensitive to the inventory policy for a fixed (and high) service level. Hence, we ignore the expected backorder component and just use the minimization of $\mathbb{B}_1$ as our objective in the test problems.

The results for the 16 scenarios are shown in Table 2 and Figures 3-6. Table 2 gives the minimum echelon base stock $\mathbb{B}_1$ for each of the 64 test problems, and the smallest value for B_1 which gives the minimum-inventory policy. Each figure corresponds to an order policy, and plots the echelon base stock $\mathbb{B}_1$ for each demand case as a function of the base stock B_1 at the CW for service level $\alpha = .95$. The general shape of these functions is the same for the other service levels. Actually, these figures give the lower envelope of the function; due to the requirement of integer base stocks, the actual function is not smooth but jagged. As we increase B_1 by one unit, either we can reduce the base stock at each retail site by one unit for a total reduction in $\mathbb{B}_1$ of N-1 units of inventory, or there is no change in the base stocks at the retail sites for a net addition to $\mathbb{B}_1$ of one unit. For clarity I have smoothed the functions to illuminate the general form of the functions.

Three observations are very apparent from the table and these figures. First, the optimal choice for the base stock at site 1 is less than $\lambda_1 p_j$ in each case. The quantity $\lambda_1 p_j$ is of interest because it equals the expected system demand from the time that the CW orders (p_1=0) to the time (p_j) that the retail sites place their last order in the order cycle for the CW. The fact that the base stock at site 1 is less than $\lambda_1 p_j$ signifies that the CW expects to run out of uncommitted stock prior to serving the last order in its order cycle. Indeed, for the minimum-inventory stocking policy, the CW will stock out during an order cycle with a very high probability. For instance, when $p_j = 2$, the probability of stockout at the CW is 0.98 for $B_1 = 55$, 0.92 for $B_1 = 60$, and 0.78 for $B_1 = 65$. When $p_j = 8$, the probability of stockout at the CW is 0.95 for $B_1 = 260$, 0.85 for $B_1 = 270$, and 0.67 for $B_1 = 280$.

Second, from the shape of the functions in the figures, we see that the total safety stock in the system is less sensitive to understocking the CW than overstocking. In particular one sees that if one were to set the order-up-to level at the CW to achieve a conventional service level (e.g., $B_1 = \lambda_1 p_j + k \sqrt{\lambda_1 p_j}$ for some k, $1 \le k \le 3$), there would be a substantial overinvestment in inventory in these examples.

Third, we note that in each figure the best choice of B_1 seems insensitive to the number of retail sites. This suggests to me that the case of retailers with non-identical demand rates would show similar behavior. That is, suppose we have three retailers with demand rates λ_j = 6, 12, and 18 for j = 2, 3 and 4; I would expect that system inventory as a function of B_1 would be a combination of the functions for the identical demand cases given in the figures and would be minimized at $B_1 < \lambda_1 p_j$.

8. Lower Bound

For the computational study we assumed a two-echelon system with a single-cycle, nested ordering policy. With the assumption of a virtual allocation policy, we found the inventory needed to provide a preset level of service. In this section, we examine the assumption of virtual allocation and give a lower bound on the inventory by relaxing this assumption.

We again consider a two-echelon system with identical retail sites and a single-cycle, nested ordering policy. We focus on the last retail order within an order cycle, which is placed at time p_j by all retail sites. (For notational convenience, let p_j be the time of placement of the last order for all retail sites and let T_j be the coverage provided by the corresponding shipment to the CW , applicable to all retail sites; we drop the order indices unless needed for a clarification of the event timing.) $D_1[T_j, p_j]$ is the total uncovered demand that cannot be shipped by the CW to the retail sites at time p_j. The virtual allocation policy spreads this shortfall, the uncovered demand, over the retail sites according to the demand experienced by the sites over the interval $(T_j, p_j]$; that is, $D_j[T_j, p_j]$ is not covered by the shipment made by the CW to site j at time p_j.

As an alternate policy, suppose that we can spread the total uncovered demand evenly over all of the retail sites, ignoring integrality restrictions. That is, we assume that the CW makes shipments such that for each retail site the uncovered demand is $D_1[T_j, p_j]/N$. In effect, when the CW ships to the retail sites, it tries to equalize the inventories at the retail sites. This is equivalent to the allocation assumption made by Eppen and Schrage (1981). As discussed there, this allocation scheme is not always feasible and requires an assumption of balanced inventories. We contend that the analysis of this policy gives a lower bound on system inventory over all feasible allocation policies.

For this assumption we can restate (2) as

$$A_j(t) = B_j - D_1[T_j, p_j]/N - D_j[\, p_j, t\,] \ , \qquad (11)$$

for p_j and t such that $p_j(m) \leq r_j(m) \leq t < r_j(m+1)$. T_j remains the same as given before by (3) and (4). From (11), we see that for this policy, which we term the equal-inventory allocation, the uncovered demand at time t for a retail site j is given by

$$D_1[T_j, p_j]/N + D_j[\, p_j, t\,] \qquad (12).$$

The expectation and variance of (12) are found to be

$\lambda_j(t - E[T_j])$ and

$\lambda_j (t - p_j) + \lambda_j (p_j - E[T_j])/N + (\lambda_j)^2 \text{Var}[T_j]$,

respectively. In comparison, the uncovered demand at time t for site j for the virtual allocation policy is $D_j[T_j, p_j]$ with expectation and variance given by (9) and (10).

Thus, we see that the uncovered demand from the equal-inventory allocation has the same expectation as from the virtual allocation, but has less variance by an amount

$$\lambda_j (N-1)(p_j - E[T_j])/N \quad .$$

Hence, the equal-inventory allocation policy should require less inventory at the retail site for any base stock B_1 at the CW , than is required by virtual allocation. Furthermore, it should be clear that no feasible allocation policy can do better (in terms of a lower variance for the uncovered demand) than the equal-inventory allocation.

For the equal-inventory allocation, we can show formally that the best choice for the base stock at the CW is $B_1 = 0$ for the case of identical retailers, and a single-cycle, nested ordering policy, provided we allow non-integer B_j. Rather than give a formal proof, we just present the intuition for the result. The intuitive demonstration is to argue that for equal-inventory allocation, there is no benefit from maintaining an inventory of any sort at the CW . This is because when one assumes that it is always feasible to equalize inventories at an order occasion, then in effect one is assuming that when the retail sites order the system can do as much transhipment as is necessary to equalize the inventories. That is, assuming the feasibility of equal-inventory allocation is the same as allowing constraint-free transhipment between retail sites at any order occasion. When transhipment is permitted, there is no need to hold back inventory at a central facility, i.e., at site 1, and thus the best choice for B_1 is 0. This argument extends to the case of non-identical retailers and to a more general ordering policy in which we just require that the retail sites order simultaneously. When we restrict the base stock at the retail sites to be integer, zero base stock at the CW is not necessarily the minimum-inventory policy; however, the best choice will typically be quite close to zero, since the only reason to hold stock at the CW is due to the indivisibility of a stock unit. One reason for presenting this result is to suggest that in the context under consideration the best inventory policy under the assumption of equal-inventory allocation is a highly idealized (and infeasible) policy, which would require an extensive amount of transhipment to realize.

In Table 2 we compare the echelon inventory $\mathbb{B}_1$ required by the equal-inventory allocation versus that for virtual allocation for the 64 test problems. [For this exercise, we approximate the distribution of uncovered demand for the equal-inventory allocation, i.e., (12), by a negative binomial, Poisson or binomial distribution depending on whether its mean is greater than, equal to, or less than its variance; this is the same approximation scheme as developed by Lee and Moinzadeh (1987), who have shown its effectiveness for a multi-echelon system with a batch ordering policy.] The non-optimality or cost in terms of additional inventory for the virtual allocation policy is very small for these examples. Even though the virtual allocation policy may seem simplistic, the inventory required is not significantly more than that required by the best possible allocation policy. Furthermore, virtual allocation is always feasible, whereas we require an assumption of balanced inventories (or costless transhipment) in order for the equal-inventory allocation to be feasible. We expect, though, that for demand processes with larger coefficients of variation, the non-optimality of virtual allocation will increase. This is because the policy allocates inventory at time p_j without consideration of the demand over the interval $(T_j, p_j]$; with more demand variation, more information is lost by ignoring the demand over this interval $(T_j, p_j]$. Nevertheless, this exercise provides additional evidence and justification for the assumption of virtual allocation for the context of fixed scheduling.

9. Conclusions and Extensions

The model developed in this paper permits the examination of a range of inventory policies for a multi-echelon system. From the computational studies for a two-echelon system, as reported in the previous two sections, we can make several observations about the role of the CW in these systems. The multi-echelon literature (e.g., Schwarz 1989) identifies two reasons for the CW, one to pool risk over the replenishment time for the outside supplier and the other to pool risk over the retail sites by rebalancing periodically the retail inventories. Eppen and Schrage (1981) call the first the joint ordering effect, and the second the depot effect. The joint ordering effect does not require the CW to hold stock, whereas the depot effect does.

We can use the computational studies in this paper to comment on these effects. First, we note that $B_1 = 0$ represents a system in which the CW holds no stock and effectively is just a central ordering agency. Hence, in Figures 3-6, the y-intercept shows the inventory for the system with the joint ordering effect, but no central warehouse. From the figures we see the incremental value of the depot effect, namely the reduction in system inventory from going from the y-intercept to the function minimum. The benefit seems fairly small except when there are many retailers (N=18 and λ_j=2). We note though that the assumption of virtual allocation may understate the benefits from the depot effect; however, by comparison with a lower bound (see Table 2), we found that the additional inventory reduction from an optimal allocation policy is very small, especially when the number of retailers is 6 or less. Finally, we observe that although there are benefits via the depot effect from having the CW hold stock, the policy with the least system inventory still results in the CW stocking out with high probability. Thus, in terms of safety stock, the policy with the least inventory puts all of the safety stock at the retail sites.

In the remainder of this section we discuss how the model and analysis might extend along three dimensions. In particular, we relax three assumptions: the Poisson demand process, an arborescent system, and deterministic replenishment lead times. In each case we suggest the nature of the extension and, as appropriate, identify unresolved issues. We treat each case separately and do not consider the simultaneous relaxation of more than one assumption.

Demand The presentation and implementation of the model are easiest and cleanest for a Poisson demand process. However, we can envision using the model for other demand processes, possibly as an approximation. This extension is important, since in many contexts the demand process has a variance greater than the mean demand, and is not well modeled by a Poisson process. We describe two cases here.

First, suppose the demand process is compound Poisson. Hence, the process has independent increments, but possibly non-unit demands. The difficulty in applying the model is that the inventory process at a supply site (e.g., site 1) is not skip-free, and hence, can skip over the zero state at which the site runs out. The analysis has assumed that the supply site has exactly zero inventory at its runout time (see (4)). For the case of compound Poisson demand, we need to model the "overshoot" when the supply at a site runs out. We redefine the runout time given in (4) by

$$S_i(n) = \min \{ s: \; D_i[\, T_i(n), s \,] \geq B_i \} \qquad ,$$

and define the overshoot as

$$O_i(n) = D_i[\, T_i(n), S_i(n) \,] - B_i \qquad . \qquad (13)$$

Suppose j (i = ρ(j)) is the site whose demand caused the runout at site i at time $S_i(n)$. Then, if the runout time occurs before the order occasion $p_j(m)$ (i.e., $T_j(m) = S_i(n)$), then $O_i(n)$ needs to be added to the uncovered demand $D_j[\, T_j(m), t]$ as part of (1) and (2). The crux of this extension of the model is to characterize the overshoot, given by (13), and then to combine it with the uncovered demand in (1) and (2) to obtain the inventory process at site j. The details will depend upon the actual demand distribution. A possible approximation would be to estimate the moments of the overshoot and then to assume that each successor has a likelihood of causing the runout proportional to its demand rate.

The second case is when we model the demand over an interval (s,t) as coming from a normal distribution with mean μ(t-s) and variance σ^2(t-s). Furthermore, we assume the demand process has independent increments. Note that these assumptions permit both non-integer demand and the possibility of negative demand over an interval. Nevertheless, this is a common model of the demand process for inventory systems, and the inventory model given by equations (1) - (4) applies directly for this demand process. The difference between the runout time and the coverage for a given order, $S_i(n) - T_i(n)$, is no longer a gamma random variable. Instead, $S_i(n) - T_i(n)$ is the first passage time between 0 and B_i for a Brownian motion process with parameters μ and σ^2. The distribution of this first passage time is known (e.g., Karlin and Taylor 1975, pg 363); the cumulative distribution function for this first passage time and its partial expectations can be written in terms of the normal distribution, e.g., Hadley and Whitin, 1963, pp 143-

148. Hence, it may be possible to develop simplifications analogous to (8) for evaluating (6)-(7) for the case of normal demand.

Network Topology We have developed the model for an arborescent system in which each site has a unique supplier. Suppose we relax this assumption and consider site j as an assembly operation, which requires supplies from a set of sites, say i = 1, ... j-1. That is, to create a unit of inventory at site j requires one unit from each supply site. We again assume a fixed order schedule and virtual allocation of available stock at every site. For site i being a supplier to site j, define $T_{ij}(t)$ as the coverage provided by the last shipment from i to j that will have arrived by time t. We can write $T_{ij}(t)$ in terms of the time that j placed the last order and the runout time at site i in an expression analogous to (3). We define $T_j(t)$ as the coverage provided by the set of suppliers to site j as of time t: $T_j(t) = \min \{T_{ij}(t)\}$ where the minimization is over the supply sites i = 1, ... j-1. Then the available inventory at j at time t is

$$A_j(t) = B_j - D_j[\, T_j(t), t\,] \quad .$$

The challenge for the analysis of this model is the characterization of $T_j(t)$; $T_j(t)$ is not only the minimum of a set of random variables, but these random variables, $\{T_{ij}(t)\}$, are not independent. The dependence across the $\{T_{ij}(t)\}$ is due to the fact that the supply sites service the same demand process whenever they supply a common assembly, e.g., site j. Also, there is an open question of how good or appropriate is the assumption of virtual allocation in this context.

Stochastic Lead Times Suppose that the order occasions $p_j(m)$ are fixed and certain, and that for site j, its supplier only ships at time $p_j(m)$. However, the times at which a shipment is received, namely $r_j(m)$, are uncertain. Thus, the lead time for an order, $r_j(m) - p_j(m)$, is stochastic. We again assume that the orders do not cross, $r_j(m-1) \leq r_j(m)$, and that the lead times are always non-negative. We assume that the realization of $r_j(m)$ does not depend on the demand process or the amount ordered. Note that the orders placed by a site remain the same as for the case with certain lead times; that is, at time $p_j(m)$ site j orders an amount $D_j[\, p_j(m-1), p_j(m)]$.

Consider time t and define $\pi_j(m, t)$ as the probability that site j has received its m^{th} shipment by time t. These probabilities can be derived from the specification of the stochastic lead times; Zipkin (1986) gives one model for stochastic lead times, which is consistent with the above assumptions and applicable to this analysis. Then the available inventory at time t equals

$$A_j(t) = B_j - D_j[\, T_j(m), t\,]$$

with probability $\pi_j(m, t)$, where $T_j(m)$ is given by

$$T_j(m) = \min [\, p_j(m), S_i(n)\,]$$

with probability $\pi_i(n, p_j(m))$. $S_i(n)$ is unchanged and given by (4). Hence, the characterizations of $A_j(t)$ and $T_j(m)$ entail the evaluation of a mixture over the possible realizations for the last order received as of time t or time $p_j(m)$, respectively. Depending upon the nature of the stochasticity, this may not be an easy task. Nevertheless, the structure of the model remains valid.

Acknowledgement: The author wishes to thank Professors Anant Balakrishnan, Hau Lee, Don Rosenfield, Lee Schwarz, Larry Wein and Paul Zipkin for their helpful comments on an earlier draft of the paper.

References

Axsater, S., "Simple Solution Procedures for a Class of Two-Echelon Inventory Problems," Working Paper, Lulea University, Lulea Sweden (1988).

Clark, A. J. and H. Scarf, "Optimal Policies for a Multi-Echelon Inventory Problem," Management Science, 6 (1960), 465-490.

De Bodt, M. A. and S. C. Graves, "Continuous-Review Policies for a Multi-Echelon Inventory Problem with Stochastic Demand," Management Science, 31 (1985), 1286-1299.

Deuermeyer, B. L. and L. B. Schwarz, "A Model for the Analysis of System Service Level in Warehouse-Retailer Distribution System," in Multi-Level Production/Inventory Control Systems: Theory and Practice, Schwarz, L. B. (ed.) TIMS Studies in the Management Sciences, Vol. 16, Amsterdam, North-Holland, (1981) 163-193.

Eppen, G. and L. Schrage, "Centralized Ordering Policies in a Multi-Warehouse System with Lead Times and Random Demand," in Multi-Level Production/Inventory Control Systems: Theory and Practice, Schwarz, L. B. (ed.) TIMS Studies in the Management Sciences, Vol. 16, Amsterdam, North-Holland, (1981) 51-67.

Federgruen, A. and P. Zipkin, "Approximations of Dynamic Multilocation Production and Inventory Problems," Management Science, 30 (1984a), 69-84.

Federgruen, A. and P. Zipkin, "Allocation Policies and Cost Approximations for Multilocation Inventory Systems," Naval Research Logistics Quarterly, 31 (1984b), 97-129.

Federgruen, A. and P. Zipkin, "Cɔmputational Issues in an Infinite-Horizon, Multiechelon Inventory Model," Operations Research, 32 (1984c), 818-836.

Graves, S. C., "A Multiechelon Inventory Model for a Repairable Item with One-for-One Replenishment," Management Science, 31 (1985), 1247-1256.

Graves, S. C. and L. B. Schwarz, "Single Cycle Continuous Review Policies for Arborescent Production/Inventory Systems," Management Science, 23 (1977), 529-540.

Hadley, G. and T. M. Whitin, Analysis of Inventory Systems, Prentice Hall, Englewood Cliffs, NJ, 1963.

Jackson, P. L., "Stock Allocation in a Two-Echelon Distribution System or 'What to Do Until Your Ship Comes in'," Management Science, 34 (1988), 880-895.

Karlin, S. and H. M. Taylor, A First Course in Stochastic Processes, Second Edition, Academic Press, New York, 1975.

Lee, H. L. and K. Moinzadeh, "Two-Parameter Approximations for Multi-Echelon Repairable Inventory Models with Batch Ordering Policy," IIE Transactions, 19 (1987), 140-149.

Maxwell, W. L. and J. A. Muckstadt, "Establishing Consistent and Realistic Reorder Intervals in Production-Distribution Systems," Operations Research, 33 (1985), 1316-1341.

Muckstadt, J. A., "A Model for a Multi-Item, Multi-Echelon, Multi-Indenture Inventory System," Management Science, 20 (1973), 472-481.

Roundy, R., "98%-Effective Integer-Ratio Lot-Sizing for One-Warehouse Multi-Retailer Systems," Management Science, 31 (1985), 1416-1430.

Schwarz, L. B., "A Model for Assessing the Value of Warehouse Risk-Pooling: Risk-Pooling over Outside-Supplier Leadtimes," Management Science, in press, 1989.

Sherbrooke, C. C., "METRIC: A Multi-Echelon Technique for Recoverable Item Control," Operations Research,16 (1968), 122-141.

Simon, R. M., "Stationary Properties of a Two Echelon Inventory Model for Low Demand Items," Operations Research, 19 (1971), 761-777.

Svoronos, A. P. and P. Zipkin, "Estimating the Performance of Multi-Level Inventory Systems," Operations Research, 36 (1988a), 57-72.

Svoronos, A. P. and P. Zipkin, "Evaluation of One-for-One Replenishment Policies for Multiechelon Inventory Systems," Working Paper, Columbia University, New York NY (1988b).

Zipkin, P., "Stochastic Leadtimes in Continuous-Time Inventory Models," Naval Research Logistics Quarterly, 33 (1986), 763-774.

Figure 3: System Inventory for Order Policy with p = 2, t = 4, Service Rate = 0.95

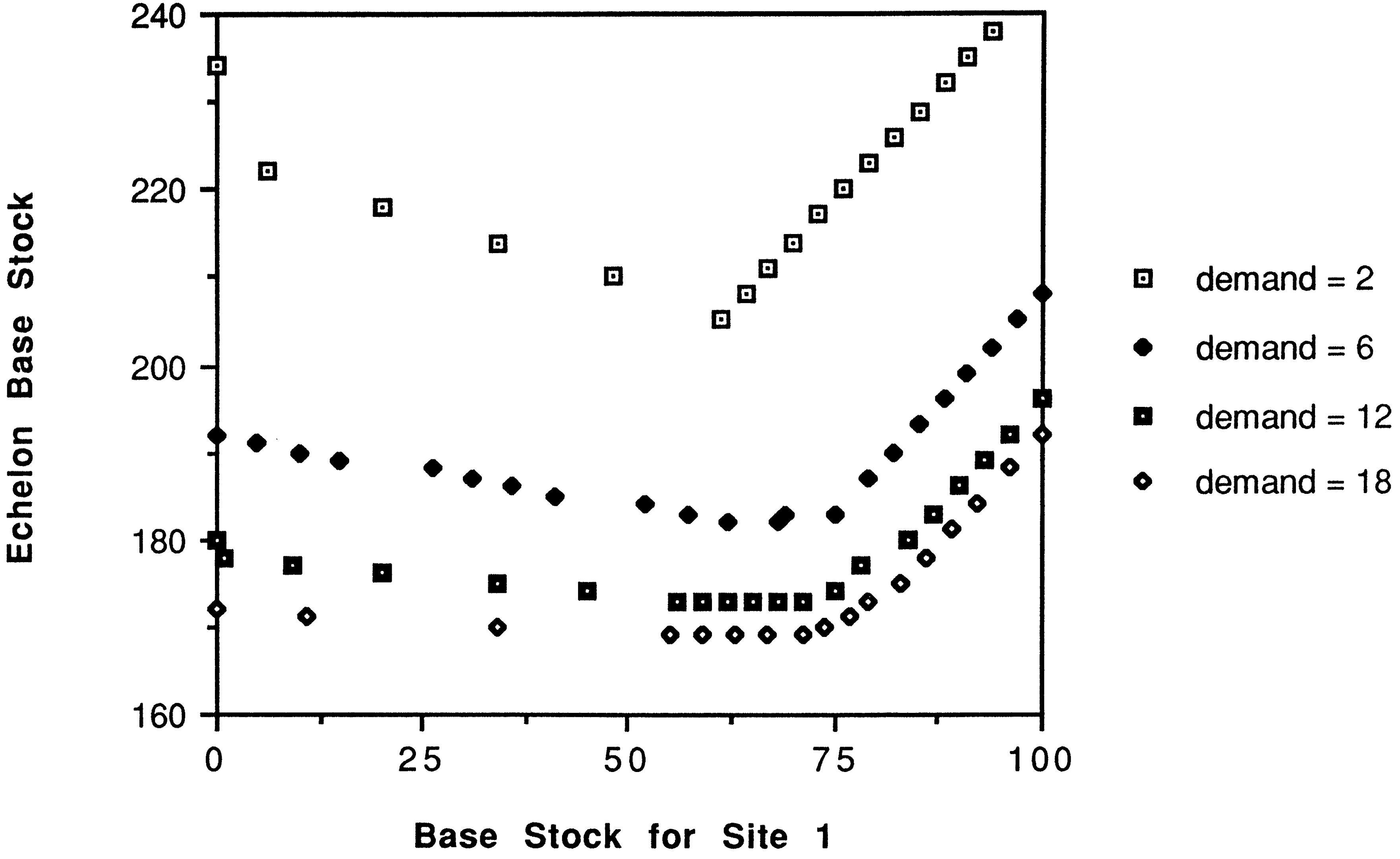


Figure 4: System Inventory for Order Policy with p = 2, t = 8 Service Rate = 0.95

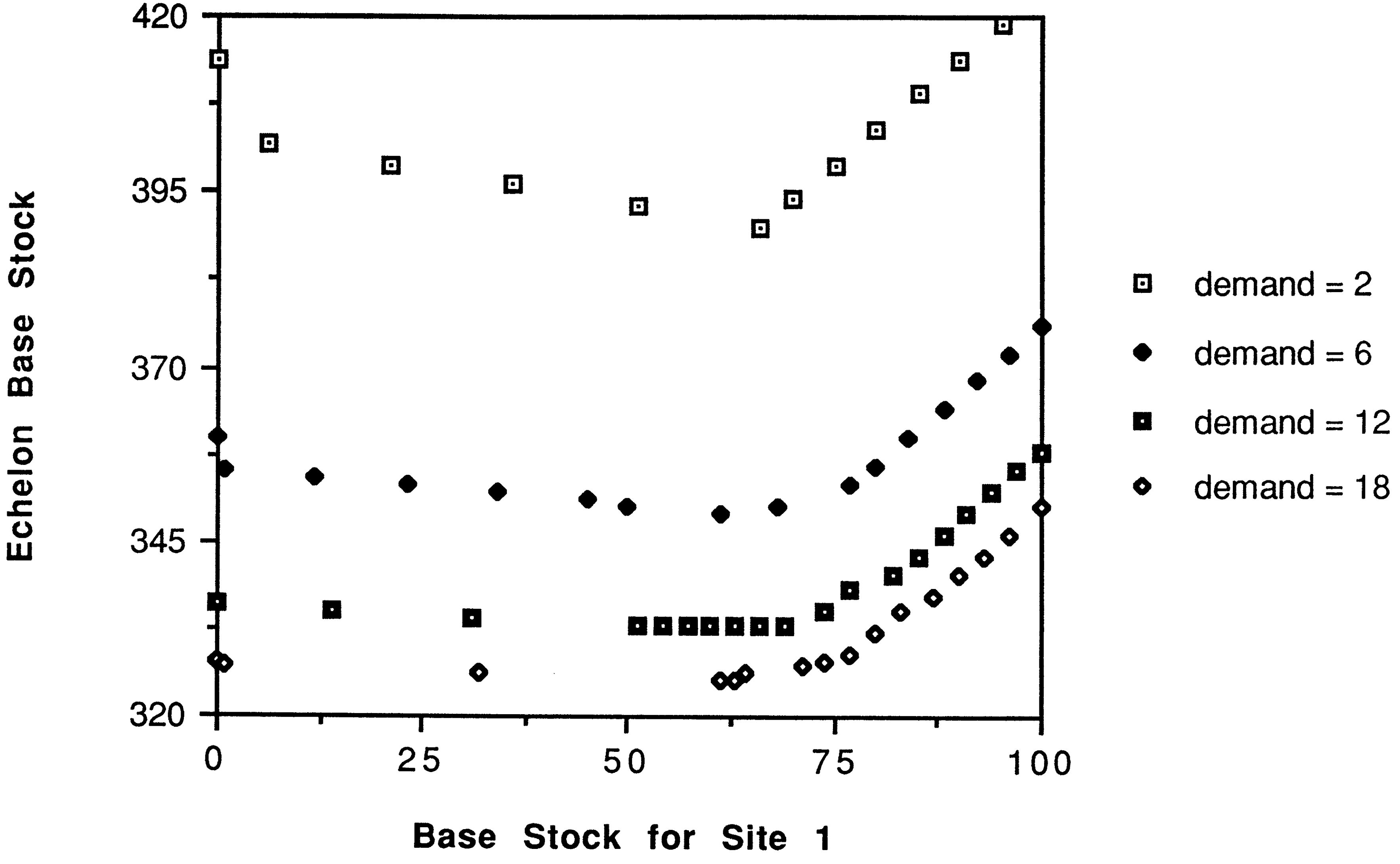


Figure 5: System Inventory for Order Policy with p = 8, t = 10, Service Rate = 0.95

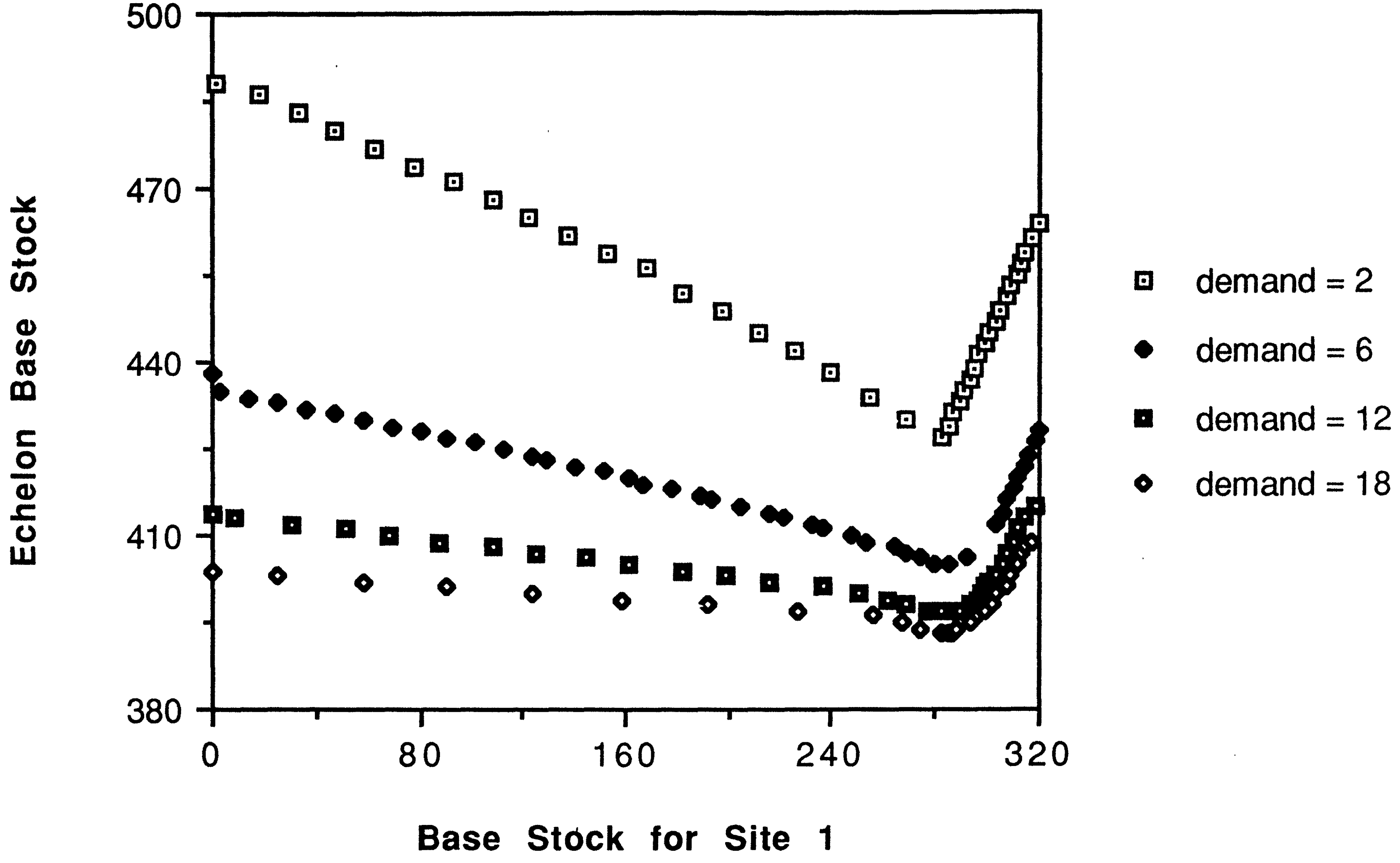


Figure 6: System Inventory for Order Policy with p = 8, t = 14, Service Rate = 0.95

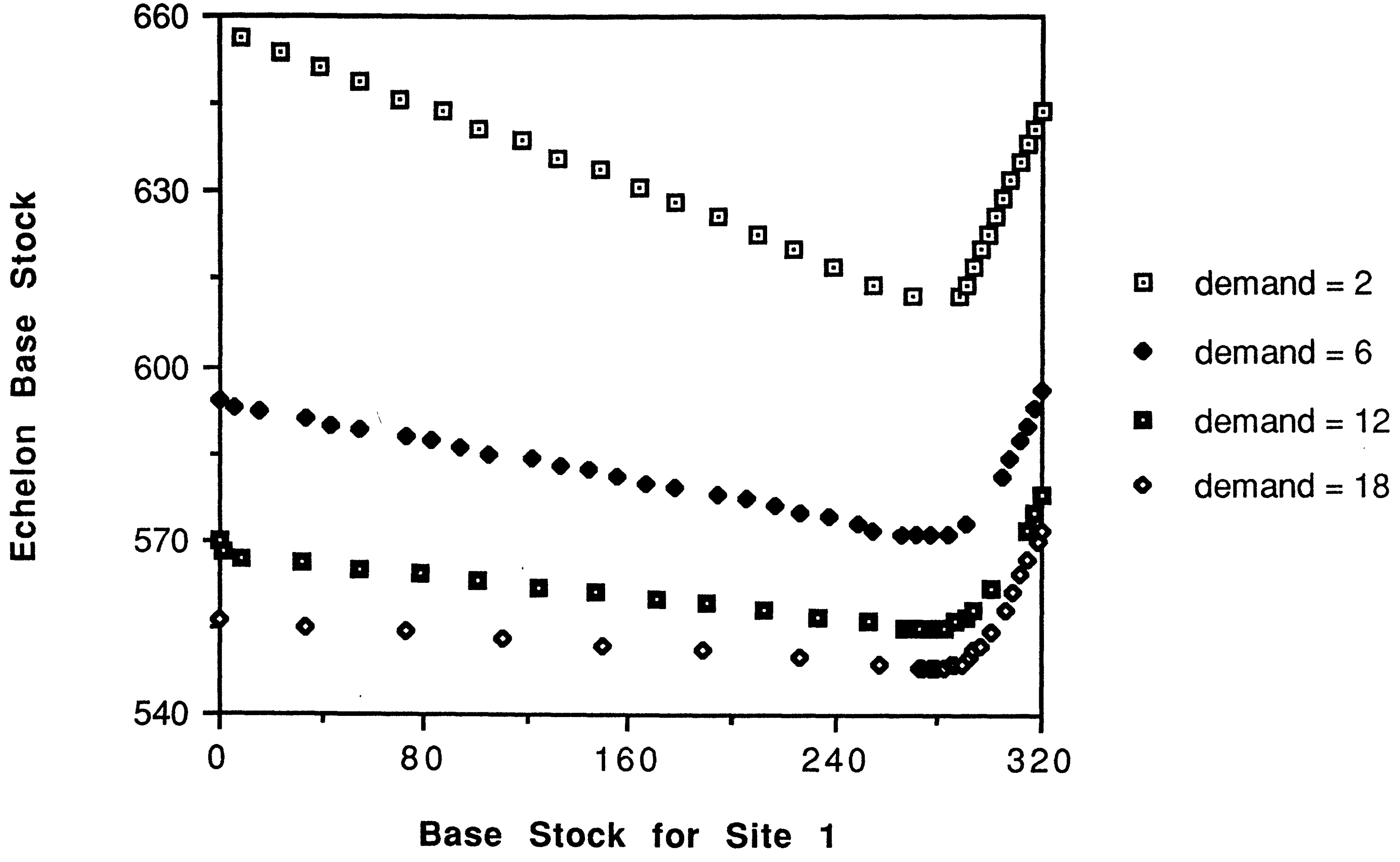


Table 1: Comparison of Actual to Approximate Distribution for Uncovered Demand

Pr (D=i)	$B_1 = 50$		$B_1 = 65$		$B_1 = 80$	
	Actual	Approx.	Actual	Approx.	Actual	Approx.
i = 0	5.05E-6	8.94E-6	5.30E-6	6.06E-6	5.59E-6	5.74E-6
1	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
2	0.0004	0.0005	0.0004	0.0004	0.0004	0.0004
3	0.0015	0.0019	0.0016	0.0016	0.0016	0.0017
4	0.0044	0.0053	0.0047	0.0049	0.0049	0.0050
5	0.0106	0.0119	0.0113	0.0116	0.0119	0.0120
6	0.0214	0.0227	0.0229	0.0232	0.0240	0.0240
7	0.0370	0.0377	0.0396	0.0397	0.0414	0.0414
8	0.0561	0.0556	0.0600	0.0599	0.0626	0.0626
9	0.0758	0.0739	0.0811	0.0806	0.0842	0.0841
10	0.0925	0.0895	0.0987	0.0980	0.1020	0.1018
11	0.1031	0.0998	0.1095	0.1088	0.1124	0.1122
12	0.1060	0.1034	0.1117	0.1111	0.1136	0.1135
13	0.1014	0.1002	0.1054	0.1052	0.1061	0.1061
14	0.0909	0.0913	0.0926	0.0929	0.0921	0.0921
15	0.0769	0.0786	0.0763	0.0768	0.0747	0.0748
16	0.0619	0.0643	0.0592	0.0598	0.0569	0.0570
17	0.0477	0.0501	0.0435	0.0440	0.0408	0.0409
18	0.0354	0.0373	0.0303	0.0307	0.0277	0.0278

Table 1 (continued)

Pr (D=i)	$B_1 = 50$		$B_1 = 65$		$B_1 = 80$	
	Actual	Approx.	Actual	Approx.	Actual	Approx.
19	0.0254	0.0266	0.0202	0.0204	0.0179	0.0179
20	0.0178	0.0183	0.0128	0.0129	0.0109	0.0109
21	0.0121	0.0121	0.0078	0.0078	0.0064	0.0064
22	0.0080	0.0077	0.0046	0.0045	0.0036	0.0036
23	0.0052	0.0048	0.0026	0.0025	0.0019	0.0019
24	0.0033	0.0029	0.0014	0.0013	0.0010	0.0010
25	0.0021	0.0027	0.0008	0.0007	0.0005	0.0005
26	0.0013	0.0009	0.0004	0.0003	0.0002	0.0002
27	0.0007	0.0005	0.0002	0.0002	0.0001	0.0001
28	0.0004	0.0003	0.0001	0.0001	4.93E-5	4.60E-5
29	0.0002	0.0001	4.90E-5	3.50E-5		
30	0.0001	0.0001				
31	7.21E-5	3.70E-5				

Table 2: Optimal Stockage Levels for Test Problems

Problem Parameters			Service Level			
p_j	t	λ_j	$\alpha = 0.80$	$\alpha = 0.90$	$\alpha = 0.95$	$\alpha = 0.975$
2	4	2	60, 168, 166*	62, 188, 183	61, 205, 197	77, 221, 208
		6	59, 161, 160	64, 172, 170	62, 182, 179	64, 190, 186
		12	44, 158, 157	55, 166, 165	56, 173, 171	62, 179, 177
		18	44, 156, 156	53, 163, 162	55, 169, 168	60, 174, 173
2	8	2	63, 333, 332	58, 364, 361	66, 390, 386	72, 414, 407
		6	53, 317, 316	53, 335, 333	61, 349, 347	62, 362, 360
		12	37, 310, 309	58, 322, 322	51, 333, 332	54, 342, 341
		18	7, 307, 306	37, 317, 316	61, 325, 325	53, 333, 332

* x, y, z: x is the optimal choice for the base stock B_1 at site 1.

y is the minimum echelon stock $\mathbb{B}_1$ for virtual allocation policy.

z is the minimum echelon stock $\mathbb{B}_1$ for equal-inventory allocation policy.

Table 2: Optimal Stockage Levels for Test Problems (continued)

Problem Parameters			Service Level			
p_j	t	λ_j	$\alpha = 0.80$	$\alpha = 0.90$	$\alpha = 0.95$	$\alpha = 0.975$
8	10	2	262, 388, 384*	283, 409, 401	283, 427, 414	281, 443, 424
		6	261, 381, 380	274, 394, 392	279, 405, 401	283, 415, 409
		12	258, 378, 377	266, 389, 388	277, 397, 396	281, 404, 403
		18	245, 377, 376	272, 386, 386	283, 393, 393	285, 399, 399
8	14	2	264, 552, 549	278, 584, 578	270, 612, 601	275, 635, 621
		6	254, 536, 535	261, 555, 553	265, 571, 567	278, 584, 580
		12	250, 529, 529	267, 543, 542	267, 555, 554	271, 565, 563
		18	248, 526, 526	268, 538, 538	272, 548, 548	271, 557, 556

* x, y, z: x is the optimal choice for the base stock B_1 at site 1.

y is the minimum echelon stock B_1 for virtual allocation policy.

z is the minimum echelon stock B_1 for equal-inventory allocation policy.