DETERMINING AND ALLOCATING
CAPACITY-DRIVEN SAFETY STOCK
IN MULTI-ECHELON SYSTEMS

by

Edward W. Chan, John A. Muckstadt,
and James A. Rappold
Determining and Allocating Capacity-Driven Safety Stock in Multi-Echelon Systems

Edward W. Chan*
John A. Muckstadt*
James A. Rappold†

May 1998
December 1998
February 1999

Abstract
In this paper, we study a multi-item, multi-echelon capacitated production and inventory system under periodic review. The approach is designed to plan and allocate production of items for which it is possible to determine probability distributions describing the demand process at each location. We begin by presenting a stochastic dynamic program for decision making in this environment. Due to its computational complexity, this formulation is of little practical value. We develop an alternative approximation approach for making production and allocation decisions which is computationally efficient. It uses a combination of the inventory shortfall random variable to describe the state of the system and a new capacity allocation mechanism that mitigates the impact of inventory imbalances among items and locations. The approach is suitable for systems with tens of thousands of items across many locations. We conclude by reporting the results of a large simulation study. This study demonstrates the effectiveness of the approach under a wide variety of scenarios. When the proposed methods are used in systems with short lead times, both the existence of imbalance and its impact on costs are negligible.

Key Words: Multi-echelon inventory, Finite capacity, Stochastic dynamic program, Approximations.

*School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853
†School of Business, University of Wisconsin, Madison, WI 53706
1 Introduction

The profitable management of large-scale production and distribution systems requires careful coordination and allocation of resources, namely capacity and materials, to balance operating costs with customer service. Through our observations in industry, production capacity allocation decisions and inventory stocking decisions in multiechelon systems are normally made myopically without explicit capacity considerations. This leads to an over commitment of inventory to slow moving items, for which it is often difficult to estimate future demand, and frequent stockouts of fast moving items. As a result, materials managers too often find themselves in the unfortunate position of having a large amount of stock and a substantial quantity of backorders. As discussed in Muckstadt (1997), this occurs even in reengineered systems with low production flow times.

In this paper, we extend the single item capacitated multi-echelon planning model developed in Rappold and Muckstadt (1998) to the multiple item case. Establishing base stock levels in multi-location, multi-item systems requires careful planning since mis-allocations of capacity and inventory can easily result in inventory imbalances among locations and among items. Due to limited production capacity, these imbalances may be difficult to correct and, as a result, may persist for many time periods. At the highest echelon, the system under study consists of a capacitated production facility in which $P$ items are fabricated, processed, or assembled, and a co-located central warehouse where these items may be stored. The central warehouse subsequently supplies the items to $M$ regional distribution centers, each of which may experience highly variable customer demand for each item. A diagram of the system is shown in Figure 1.

We assume the system operates as follows. Decisions are made and transactions occur only at specific points in time, or periodically. Each time period begins with all distribution centers first receiving inventories which were shipped from the central warehouse at the end of the previous period, and then receiving a set of random customer orders. These orders are then satisfied or backordered. The production facility simultaneously receives information about the demand that has occurred at each distribution center, then allocates capacity for
production among items, then produces the items, and finally allocates and ships some or all of this production to the regional distribution centers.

The contributions of this paper are two-fold. First, we develop a computationally tractable method for allocating production capacity among items and for allocating production among the regional distribution centers. This analytic framework has implications for the design and use of information systems used to support these activities. Second, we test and verify the quality and appropriateness of the approach via an extensive simulation study.

The production and shipping allocation model we propose and solve is an approximation. This approximation is obtained by assuming that all system inventory can be reallocated both among items and locations each period. Clearly, this approximation is useful only if period ending inventories are not often significantly mis-allocated among items and locations. We say the inventory is in a state of capacity imbalance when there is too much inventory of one item in the system and location imbalance when there is more inventory of a particular item at one regional distribution center than is desirable. These imbalances occur as a consequence of (1) the degree of capacity limitation (or capacity utilization), (2) the degree of variability found in the demand processes for each item at each regional distribution center, and (3) the production and shipping allocation rules that are employed. We show that our approximation method is effective and that the existence and impact of inventory imbalances on system performance is slight. We show this by establishing lower bounds on
systems costs and then by demonstrating through a set of simulation experiments that the
difference between the lower bound and the observed performance is small. We also compare
our allocation scheme to that of a more traditional equal fractile allocation.

Our model and solution approach are based on several key assumptions. We first assume
that the system operates in the manner we described. Production and shipping lead times
are assumed to be less than a period and one period in length, respectively. We assume
that the demand for each item at each distribution center is independent and identically
distributed from period-to-period over an infinite horizon. Unsatisfied demand is backlogged.
The demand distribution and parameter values may differ by item and distribution center.
Furthermore, for each item, we assume that demand among distribution centers is either
independent or positively correlated. The relevant costs are backorder and holding costs for
each item at each location, and there are no fixed costs in the system that affect production
and distribution decisions. For ease of exposition, we also assume that the time required
to produce each item is identical. A similar assumption is made in Glasserman (1996).
This permits us to state the production capacity in terms of the number of units that can be
produced per period. Alternatively, we could assume that demand random variables measure
the amount of production time required to produce the number of units ordered.

The rest of the paper is organized as follows. In section 2 we review the relevant literature.
In section 3 we introduce notation used throughout the paper and review a single item
model for the capacitated production system presented in Rappold and Muckstadt (1998).
In section 4 we begin by constructing a dynamic programming formulation of the decision
problem. This is an exact model. Unfortunately, due to the computational burden posed by
this formulation, it is impractical for solving the problem except for trivial situations. The
approximation procedure we propose for solving the problem is presented next. We show how
the results of the single item analysis provides the basis for the multi-item approximation
procedure. Since this approximation method yields decisions that may not be optimal, in
the sense of minimizing expected per period operating costs, we structure a comprehensive
test to measure its effectiveness. The test consists of a set of simulation experiments. In
section 5 we describe this test and demonstrate that the proposed approximation is highly effective. To measure the effectiveness, we show how to obtain a lower bound on the expected per period cost. We then compare the calculated lower bound costs with the corresponding costs calculated through the simulation experiments. Concluding remarks are given in the final section.

2 Literature Review

As stated earlier, this paper is a multi-item extension of the single item model developed in Rappold and Muckstadt (1998). We refer the reader to Rappold and Muckstadt (1998) for a more thorough review of the relevant literature pertaining to this problem.

The primary difficulty in solving this problem is in understanding the detrimental effects of inventory imbalances. Zipkin (1984) shows that for equal and low coefficient of variation demand processes at each location, a myopic allocation rule stochastically minimizes imbalance in future periods. This however is not true with unequal coefficients of variation. It is also shown that inventory imbalances can be significant when using a myopic allocation rule when demand variability is high.

Federgruen and Zipkin (1986a,b) establish the optimality of the modified base-stock policy used to control production in a discrete-time single item, single location finite capacity production system. More recently, DeCroix and Arreola-Risa (1998) prove the optimality of the modified base stock policy for a multi-item, single location capacitated facility. Computation of the optimal policy is demonstrated for the case when all items have identical demand processes. For the non-identical item case, heuristics are developed.

Similar results have been established for the single location, multi-item non-stationary demand case. We refer the reader to Kapuscinski and Tayur (1996) and Aviv and Federgruen (1997) for a detailed discussion of this work. Glasserman (1996) structures a multi-item, single location capacity allocation problem under continuous review in which total capacity
is divided and permanently dedicated to each item. See Tayur (1996) for a recent literature survey in this area.

The most significant difference between previous work and ours is the simultaneous consideration of capacity allocation decisions with inventory allocation decisions in a capacitated multi-item, multi-location system, in which demand is stochastic.

3 A Review of the Single-Item Problem

In this section, we review the single-item model described in Rappold and Muckstadt (1998). As we will see, this work provides a foundation for the multi-item analysis that we present subsequently. They assume a single-item is produced in a capacity-constrained facility and then allocated to regional distribution centers. We first introduce the following notation.

Throughout this section and the remainder of the paper, the subscript $i = 1, \ldots, M$, refers to regional distribution center (DC) $i$ and $i = 0$ refers to the central warehouse. We let the subscript $j = 1, \ldots, P$, refer to item $j$. Let $\mathcal{I} = \{0, 1, \ldots, M\}$ denote the set of all stocking locations, including the central warehouse, and $\tilde{\mathcal{I}} = \mathcal{I} \setminus \{0\} = \{1, 2, \ldots, M\}$ denote the set of all regional distribution centers. Let $\mathcal{J} = \{1, 2, \ldots, P\}$ be the set of all items.

**System Parameters**

- $M$ Number of distribution centers;
- $P$ Number of items;
- $C$ Available production capacity per period (measured in units of items);

**Costs**

- $h_{ij}$ Installation holding cost at location $i \in \mathcal{I}$ for item $j \in \mathcal{J}$ per unit per period;
- $b$ Backorder cost per unit per period at each DC for each item (there is no backorder cost at the central warehouse). For simplicity, each item at each distribution center is assumed to have an identical backorder cost. However, for
implementation, all of the models and algorithms presented may be extended
to consider varying backorder costs for each distribution center;

\[ h_{ij}^e = (h_{ij} - h_{uj} > 0) \] The incremental echelon holding cost at DC \( i \in \bar{I} \) for item \( j \in \mathcal{J} \) per unit per period;

**Decision Variables**

\( T \) The target total inventory level for the system (measured in units of items, or production capacity);

\( T_j \) The target system inventory level for item \( j \in \mathcal{J} \);

\( T_{ij} \) The target system inventory level for item \( j \in \mathcal{J} \) at location \( i \in \mathcal{I} \);

\( \hat{s}_{ij} \) The optimal order-up-to level at DC \( i \in \bar{I} \) for item \( j \in \mathcal{J} \) when using echelon holding costs, \( h_{ij}^e \);

\( \hat{S}_j \) \( (= \sum_{i \in \mathcal{I}} \hat{s}_{ij}) \) The sum of the \( \hat{s}_{ij} \) order-up-to levels for item \( j \in \mathcal{J} \) across all DCs when using echelon holding costs;

\( \hat{S} \) \( (= \sum_{j \in \mathcal{J}} \hat{S}_j) \) The sum of the order-up-to levels across all items and DCs when using echelon holding costs;

\( s_{ij} \) The optimal order-up-to level at DC \( i \in \bar{I} \) for item \( j \in \mathcal{J} \) when using installation holding costs, \( h_{ij} \);

\( S_j \) \( (= \sum_{i \in \mathcal{I}} s_{ij}) \) The sum of the order-up-to levels \( s_{ij} \) for item \( j \in \mathcal{J} \) across all DCs when using installation holding costs;

\( S \) \( (= \sum_{j \in \mathcal{J}} S_j) \) The sum of the order-up-to levels across all items and DCs when using installation holding costs;

\( x_{ij} \) The inventory on-hand (or backordered) of item \( j \in \mathcal{J} \) at DC \( i \in \bar{I} \) after demand has occurred, but before production and allocation decisions are made;
\( y_{ij} \) The stock level at DC \( i \in \mathcal{I} \) of item \( j \in \mathcal{J} \) after production, allocation, and shipment from the central warehouse (which may be less than desired depending on the availability of stock);

\( x_{0j} \) The inventory on-hand of item \( j \in \mathcal{J} \) at the central warehouse before production and allocation \( (x_{0j} \geq 0) \);

\( y_{0j} \) The stock level at the central warehouse of item \( j \in \mathcal{J} \) after production, allocation, and shipment to the DCs \( (y_{0j} \geq 0) \);

Random Variables and Probability Distribution Functions

\( D_{ij} \) A non-negative random variable representing the one period demand at DC \( i \in \mathcal{I} \) for item \( j \in \mathcal{J} \) where \( \Pr\{D_{ij} = 0\} < 1 \);

\( p_{ij}(k) \) \( (= \Pr\{D_{ij} = k\} \ \forall k \geq 0) \) The discrete nonnegative probability mass function with finite first and second moments representing the demand at distribution center \( i \in \mathcal{I} \) for item \( j \in \mathcal{J} \);

\( D_j \) \( (= \sum_{i \in \mathcal{I}} D_{ij}) \) A random variable representing the one period total system demand for item \( j \in \mathcal{J} \);

\( \bar{D} \) \( (= \sum_{j \in \mathcal{J}} \sum_{i \in \mathcal{I}} D_{ij}) \) A random variable representing the one period total system demand;

\( V \) A non-negative random variable representing the stationary total system inventory shortfall in a period where the shortfall measures the difference between the system target inventory level, \( T \), and the achievable inventory level;

\( \pi(k) \) \( (= \Pr\{V = k\} \ \forall k \geq 0) \) The stationary probability distribution of \( V \), if it exists;

\( A_j \) A random variable representing the total amount of item \( j \in \mathcal{J} \) in the system after production of item \( j \), which is available for allocation to the DCs;
\( A \) \((= T - V)\) A random variable representing the stationary net inventory in the system after production which is available for allocation to items and to distribution centers. Note that \( \sum_{j \in \mathcal{J}} A_j = A \).

Although we examine only the single-item case in this section, we will use the \( j \) subscript on variables since it will be useful in later sections. Let \( A_j \) denote the total amount of item \( j \in \mathcal{J} \) in the system at the end of a period available for allocation to the \( M \) distribution centers and central warehouse. The optimal allocation of \( A_j \) to locations depends on incremental echelon holding costs, backorder costs, and expected future demand.

Recall that \( x_{ij} \) denotes the net inventory at location \( i \in \mathcal{I} \) for item \( j \in \mathcal{J} \) after demand occurs but before production and allocation occur, and that \( y_{ij} \) is the inventory position at location \( i \in \mathcal{I} \) for item \( j \in \mathcal{J} \) after production and allocation occur, where \( \sum_{i \in \mathcal{I}} y_{ij} = A_j \). For convenience, we consider \( x_{ij} \) and \( y_{ij} \) to be integer. Define the single period, single item newsvendor function for \( i \in \mathcal{I} \) and \( j \in \mathcal{J} \) as:

\[
G_{ij}(y_{ij}) = \begin{cases} 
  h_{ij} \sum_{k=0}^{y_{ij}} (y_{ij} - k)p_{ij}(k) + b \sum_{k=y_{ij}}^{\infty} (k - y_{ij})p_{ij}(k), & \text{for } y_{ij} \geq 0, \\
  b[E(D_{ij}) - y_{ij}], & \text{for } y_{ij} < 0.
\end{cases}
\] (3.1)

For item \( j \in \mathcal{J} \), define the aggregated expected cost function across distribution centers as:

\[
G_j(y_{ij}, \ldots, y_{Mj}) = \sum_{i \in \mathcal{I}} G_{ij}(y_{ij}).
\] (3.2)

It is easily shown that \( G_{ij}(\cdot) \) and \( G_j(\cdot) \) are convex functions and that the integer-valued base stock levels \( s_{ij} \) which minimize \( G_{ij}(\cdot) \) are given by:

\[
s_{ij} = \min \left\{ s \in \mathbb{Z} : \Pr\{D_{ij} \leq s\} \geq \frac{b}{h_{ij} + b} \right\},
\] (3.3)

where \( \mathbb{Z} \) is the set of all integers. For item \( j \in \mathcal{J} \), let \( S_j = \sum_{i \in \mathcal{I}} s_{ij} \) be the sum of the base-stock levels which minimize \( G_{ij}(\cdot) \).

Our analysis is divided into two cases. In the first case, \( A_j \geq 0 \), and in the second case, \( A_j < 0 \). Let \( \bar{D}_j = \sum_{i \in \mathcal{I}} D_{ij} \) be a random variable representing the total system demand per
period for item \( j \in \mathcal{J} \). We define the following system cost function \( F_j(\cdot) \) for item \( j \in \mathcal{J} \) as:

\[
F_j(A_j; y_{1j}, \ldots, y_{Mj}) = \begin{cases} 
(1) & h_{0j}(A_j - \sum_{i \in \mathcal{I}} y_{ij}) + G_j(y_{1j}, \ldots, y_{Mj}), & \text{for } A_j \geq 0, \\
(2) & b[E(\bar{D}_j) - A_j], & \text{for } A_j < 0.
\end{cases} \tag{3.4}
\]

The central warehouse inventory allocation for item \( j \in \mathcal{J} \) is \( y_{0j} = A_j - \sum_{i \in \mathcal{I}} y_{ij} \) which is held to protect against future shortages attributable to the limitations of production capacity and to balance expected backorder costs with incremental echelon holding costs. That is, if production capacity were unlimited, then \( y_{0j} = 0 \ \forall j \in \mathcal{J} \) and \( y_{ij} = s_{ij} \ \forall i \in \mathcal{I}, \forall j \in \mathcal{J} \).

Suppose \( A_j \geq 0 \) after production in a given period, but before allocation to the \( M \) distribution centers. We allocate item \( j \in \mathcal{J} \) by solving the following allocation problem:

\[
(P_{\text{Single Item}}) \quad \min_{y_{1j}, \ldots, y_{Mj}} \quad F_j(A_j; y_{1j}, \ldots, y_{Mj}) \tag{3.5}
\]

\[\text{s.t.} \quad \sum_{i \in \mathcal{I}} y_{ij} \leq A_j \]

\[y_{ij} \geq 0, \text{ integer } i \in \mathcal{I}. \tag{3.6}\]

This allocation problem ignores any potential inventory location imbalances that may exist due to inventory that was on hand at a distribution center at the end of the period. That is, we require \( y_{ij} \geq 0 \) rather than \( y_{ij} \geq x_{ij} \) in (3.6). Clearly, imbalances could and will occur periodically. On occasion, the optimal \( y_{ij} \) could be less than \( x_{ij} \). We assume that this does not occur in the optimization model. By ignoring the constraint \( y_{ij} \geq x_{ij} \), the allocation model yields a lower bound on the true cost for \( M > 1 \), and the exact cost for \( M = 1 \). See Rappold and Muckstadt (1998) for a complete discussion. Note that \( F_j(\cdot) \) is convex in \( A_j \).

The unconstrained solution to \((P_{\text{Single Item}})\) yields the integer-valued base stock levels \( \hat{s}_{ij} \) for \( i \in \mathcal{I}, \ j \in \mathcal{J} \) based on incremental echelon holding costs for each location as:

\[
\hat{s}_{ij} = \min \left\{ s \in \mathcal{Z} : \Pr\{ D_{ij} \leq s \} \geq \frac{h_{0j} + b}{h_{ij} + b} = \frac{h_{0j} + b}{h_{ij}^c + h_{0j} + b} \right\}. \tag{3.7}
\]

We assume \( h_{ij}^c > 0 \) so that the inventory allocation problem is not trivial. For item \( j \in \mathcal{J} \), let \( \hat{S}_j = \sum_{i \in \mathcal{I}} \hat{s}_{ij} \) be the sum of the base stock levels \( \hat{s}_{ij} \) that minimize \( F_j(\cdot) \). Note that for
every $j \in \mathcal{J}$, $S_j \leq \hat{S}_j$ since it is assumed that $b_{ij} \geq 0$. Also, observe that for any level of inventory $A_j$, location $i \in \tilde{I}$ will never stock more than $\hat{s}_{ij}$ units of item $j$. If $A_j > \hat{S}_j$, then it is best to hold the remaining inventory $A_j - \hat{S}_j$ at the central warehouse. There is no economic benefit to carrying more units of item $j$ at any distribution center.

Now, suppose $A_j < 0$ so that the system is in a backorder state for item $j \in \mathcal{J}$. Operationally, this case is of little practical significance since changes to the system capacity would very likely be made to avoid system-wide backorders, such as by adding capacity through overtime or by outsourcing production. However, when determining the optimal system order-up-to level, $T$, it is necessary to assign a cost to the event ($A_j < 0$). We will approximate this expected system cost, and temporarily depart from our requirement that $y_{ij}$ be integer for all $i \in \tilde{I}, j \in \mathcal{J}$. For $A_j < 0$, we define:

$$F_j(A_j; y_{1j}, \ldots, y_{Mj}) = b[E(\bar{D}_j) - A_j]$$

(3.8)

where

$$y_{ij} := A_j \cdot \frac{E(D_{ij})}{E(\bar{D}_j)}, \quad i \in \tilde{I},$$

$$y_{0j} = 0.$$

We denote by $L_j(A_j)$ the minimum one period expected cost of allocating the available system inventory for item $j \in \mathcal{J}, A_j$, across the $M$ distribution centers and the central
warehouse. $L_j(A_j)$ may be stated as:

\[
L_j(A_j) = \begin{cases} 
(1) & h_{0j}(A_j - \hat{S}_j) + G_j(\hat{s}_{1j}, \ldots, \hat{s}_{M_j}) \quad \text{for } A_j \geq \hat{S}_j, \\
 & \text{where } y_{ij} = \hat{s}_{ij}, \quad i \in \tilde{I}, \\
 & y_{0j} = A_j - \hat{S}_j, \\
(2) & \min \sum_{i \in \tilde{I}} G_j(y_{ij}, \ldots, y_{M_j}) \quad \text{for } 0 \leq A_j < \hat{S}_j, \\
 & \text{s.t. } \sum_{i \in \tilde{I}} y_{ij} = A_j, \\
 & \quad y_{ij} \geq 0, \text{ integer, } i \in \tilde{I}, \\
 & \quad y_{0j} = 0, \\
(3) & b \cdot \left[ E(D_j) - A_j \right] \quad \text{for } A_j < 0, \\
 & \text{where } y_{ij} = A_j \cdot \left[ \frac{E(D_{ij})}{E(D_j)} \right], \quad i \in \tilde{I}, \\
 & y_{0j} = 0.
\end{cases}
\]

(3.9)

It is easily shown that $L_j(A_j)$ is continuous and convex in $A_j$ [Rappold and Muckstadt (1998)].

4 The Multi-Item Models

4.1 A Dynamic Programming Formulation

The multi-echelon production and inventory allocation problem we have described may be formulated as a dynamic program as follows. Let $X = [x_{ij}]$ be a $(M + 1) \times P$ matrix representing the net inventory levels of item $j \in J$ at location $i \in I$ after demand occurs, but before production and allocation decisions have been made. Similarly, let $Y = [y_{ij}]$ be a $(M + 1) \times P$ matrix of net inventory levels after production and allocation decisions. Let $D = [d_{ij}]$ be a $(M + 1) \times P$ matrix of random variables corresponding to the one period demand for item $j \in J$ at location $i \in I$, with $Pr\{d_{0j} = 0\} = 1, \forall j \in J$ since the demand only occurs at the distribution centers. Let $C$ be the total available production capacity per
period. We define the system expected one period cost function to be the sum of the one period expected newsvendor cost functions at each distribution center plus the holding cost incurred at the central warehouse for each item. It is given as,

$$G(Y) = \sum_{j \in J} \left( \sum_{i \in I} G_{ij}(y_{ij}) + h_{0j}y_{0j} \right),$$

(4.10)

where \(\lim_{|y_{ij}| \to \infty} G(Y) = \infty\) for any \(y_{ij}, j \in J, i \in I\). We define the set of feasible decisions \(Y\) given \(X\) as,

$$R(X) = \left\{ Y : \sum_{j \in J} \sum_{i \in I} (y_{ij} - x_{ij}) \leq C \text{ and } y_{ij} \geq x_{ij}, \forall j \in J, \forall i \in I \right\}.$$  

(4.11)

A dynamic programming formulation of this problem may be stated as,

$$f(X) = \min_{Y \in R(X)} [G(Y) + E_D(f(Y - D))].$$

(4.12)

For there to exist a solution to (4.12), the condition \(\sum_{j \in J} \sum_{i \in I} E(d_{ij}) < C\) must hold (Federgruen and Zipkin (1986a)).

Since it is our objective to determine inventory stocking levels for situations in which there are several thousand items stocked at tens of locations, solving (4.12) directly is impractical for the infinite horizon case. Instead, we will develop an approximation to this dynamic program.

### 4.2 An Approximation to the Dynamic Program

The approach we now describe for solving the multi-item production and allocation problem is an approximation of the dynamic program formulation given in (4.12). The recursion given by (4.12) measures the immediate expected cost, \(G(Y)\), and the impact of making the decisions \(Y\) on future expected costs, measured by \(E_D(f(Y - D))\), given \(X\). As before, \(X\) represents the state of the system at the beginning of a period.

The approximate model is based on the structural results developed in Federgruen and Zipkin (1986a,b) and DeCroix and Arreola-Risa (1998). They show that a modified base stock policy is optimal in capacitated systems when there are no fixed costs. Based on their
results, the approximation model assumes that the following single critical number policy is followed. In each period, the policy is to produce enough to raise the total system inventory up to this critical number, $T$, or to produce an amount equal to the period’s production capacity, whichever is smaller.

### 4.3 Allocating Inventory to Items and Locations

Since production is controlled via a modified base stock policy characterized by $T$, $A$ is the amount of inventory in the system following production, where

$$A = T \wedge (\sum_{ij} x_{ij} + C). \quad (4.13)$$

We define $\tilde{R}(X, A)$ to be the set of feasible decisions $Y$ given $X$ and $A$ as,

$$\tilde{R}(X, A) = \left\{ Y : \sum_{j \in J} \sum_{i \in I} y_{ij} = A, \quad y_{0j} \geq 0 \ \forall j \in J, \text{ and } y_{ij} \geq x_{ij}, \forall j \in J, i \in I \right\}. \quad (4.14)$$

Note that $y_{0j}$ need not be greater than or equal to $x_{0j}$ since this inventory may be shipped to the distribution centers. As shown in the next lemma, it is clear from (4.13) that any decision in $\tilde{R}$ satisfies the capacity restriction.

**Lemma 1**

$$\sum_{ij} y_{ij} = T \wedge (\sum_{ij} x_{ij} + C) \implies \sum_{ij} (y_{ij} - x_{ij}) \leq C.$$

**Proof:** Suppose $\Sigma_{ij} y_{ij} = T$. Then $\Sigma_{ij} y_{ij} = T \leq \Sigma_{ij} x_{ij} + C$, which implies that $\Sigma_{ij} (y_{ij} - x_{ij}) \leq C$. Suppose $\Sigma_{ij} y_{ij} = \Sigma_{ij} x_{ij} + C$. Then $\Sigma_{ij} (y_{ij} - x_{ij}) = C$. \(\square\)

The method we propose for finding $Y$ requires $Y \in \tilde{R}(X, A)$, as does the dynamic programming recursion. Rather than using $G(Y) + E_D(f(Y - D))$ to evaluate the expected cost of selecting $Y$, we employ an approximation. We again use $G(Y)$ to measure the expected immediate cost consequence of choosing $Y$. However, we measure the effect of selecting $Y$ on future expected costs in a different way.

Recall that we employ a policy that attempts to raise the system inventory level to $T$ each period. If capacity were unlimited, then $T = S$ and each distribution center $i$ would
stock precisely $s_{ij}$ units of item $j$, with no inventory held at the central warehouse. In this case, there is no impact of current decisions on future ones since imbalances cannot occur. However, due to capacity restrictions, $T$ may exceed $S$. We will call this excess inventory $(T - S)$ the capacity-driven safety stock of the system.

When $T \leq \hat{S}$, then all of the capacity-driven safety stock will be shipped and held at the distribution centers. This is shown in equation (4.10). Consequently, in our approximation, no inventory will be held centrally in this case. This occurs because the consequences of imbalances are ignored in these calculations.

If $T > \hat{S}$, then $(T - \hat{S})$ units will be held centrally and each distribution center $i \in \tilde{I}$ will hold $\hat{s}_{ij}$ units of item $j \in J$. Then the remaining question of interest is how to allocate the $(T - \hat{S})$ units of centrally-held capacity-driven safety stock among the $P$ items. By properly allocating this safety stock, we can reduce the possibility of capacity imbalance.

Suppose $T > \hat{S}$. In the approximation, distribution center $i$ will be allocated $\hat{s}_{ij}$ units of item $j$ for two reasons. First, $s_{ij}$ units protect against demand uncertainty over the one period transportation lead time. Second, an additional $(\hat{s}_{ij} - s_{ij})$ units are held due to the marginal benefit of holding the available stock at the distribution center rather than the central warehouse. While this additional stock adds more protection against demand variation, it may be detrimental to system costs should inventories become imbalanced across locations or items. Hence, this allocation may not be optimal. Note that as $h_{ij}^c \to 0$, $\hat{s}_{ij} \to \infty$.

When deciding how to allocate the $(T - \hat{S})$ units amongst the items, we would like to ensure that these units are allocated such that this inventory is unlikely to wait for long periods of time before being demanded and shipped to the distribution centers. In this sense, we would like to allocate this stock to those items in which turnover is fastest. The classical newsvendor-like allocations made when finding $s_{ij}$ and $\hat{s}_{ij}$ trade off holding plus backorder costs based on right tail demand probabilities; they do not consider how the presence of finite capacity limits possible future allocations, nor do they consider how capacity is utilized over time.
We propose an allocation function which is designed to assign production capacity to those items which are most likely to be demanded in the near future. Suppose for the time being that holding and backorder costs are identical among items and locations. In this case, we would expect the dynamic program to suggest storing central inventory in those items for which the probability of being demanded in the very near future is higher than for other items. Equivalently, we would expect the dynamic program to minimize the total expected number of periods these inventories are held. The stock held at the distribution centers exists to protect against demand variation in the next period; when allocating centrally held stock, the dynamic program trades off expected future holding costs among the items that will occur beyond the next period. This observation leads to an alternative to computing $\mathbb{E}_P(f(Y - D))$ in (4.12) that is a computationally efficient approximation model for capacity allocation decisions.

Let $D_{ij}^{(t)}$ be a random variable representing the cumulative demand of item $j \in \mathcal{J}$ at location $i \in \mathcal{I}$ over $t$ time periods, where $t$ is a positive integer. Furthermore, let $\bar{D}_j^{(t)} = \sum_{i \in \mathcal{I}} D_{ij}^{(t)}$. Also, let $\phi_j^{(t)}(x) = \Pr\{\bar{D}_j^{(t)} = x\}$ and $\Phi_j^{(t)}(x) = \Pr\{\bar{D}_j^{(t)} \leq x\}$. We assume that $\Phi_j^{(t)}(0) < 1$ for all $t > 0$ and for all $j \in \mathcal{J}$. We define a function for item $j \in \mathcal{J}$ denoted by $Q_j(y_{0j})$, representing the expected number of inventory-periods associated with having $y_{0j}$ units of item $j$ at the central warehouse:

$$Q_j(y_{0j}) = \sum_{t=2}^{\infty} \mathbb{E}[y_{0j} - \bar{D}_j^{(t)}]^+, \quad (4.15)$$

where $y_{0j} \geq 0 \ \forall j \in \mathcal{J}$. Since $y_{0j} = [A_j - \hat{S}_j]^+$, $Q_j(\cdot)$ may be expressed alternatively as a function of $A_j$ as

$$Q_j(A_j) = \begin{cases} 0 & \text{for } A_j < \hat{S}_j \\ \sum_{t=2}^{\infty} \mathbb{E}[A_j - \hat{S}_j - \bar{D}_j^{(t)}]^+ & \text{for } A_j \geq \hat{S}_j. \end{cases} \quad (4.16)$$

$Q_j(y_{0j})$ represents the expected central on-hand inventory through time period $t$ starting initially with $y_{0j}$ units and assuming there is no further production. This is the minimum total expected number of inventory-periods beyond the first period that $y_{0j}$ units are held.
Observe that once a unit of item \( j \in J \) is produced, the minimum holding cost associated with carrying this unit in the system is \( h_{0j} \). If the unit were held subsequently at a regional distribution center at the end of a period, its holding cost would be higher since \( h_{ij} > h_{0j} \). On the other hand, if the unit of item \( j \in J \) held at the central warehouse were shipped to a distribution center only when needed to satisfy a known customer demand, then no incremental holding costs would be incurred at a regional distribution center, and the unit would incur only a charge of \( h_{0j} \). Thus, \( h_{0j} Q_j(y_{0j}) \) results in a lower bound on the cost associated with having \( y_{0j} \) units of item \( j \in J \) in the system.

Therefore, \( \sum_{j \in J} h_{0j} Q_j(y_{0j}) \) underestimates but approximates the expected future holding costs associated with an allocation of \( y_{0j} \) units to item \( j \) at the central warehouse. When production exceeds the requirements of all the distribution centers, that is, when \( A > \hat{S} \), we will allocate \( A - \hat{S} \) units using this function. Stated differently, the approximation model will force production in items that are more likely to be required by customers in the immediate future. Before presenting the approximation model that includes the use of this expected holding cost function, we will explore some of its mathematical properties.

An alternate way of writing \( Q_j(\cdot) \) is based on a result from renewal theory. Let \( \bar{D}^{(t)}_j(\cdot) \) be the cumulative demand over \( t \) periods for item \( j \) and let \( \Pr\{\bar{D}^{(t)}_j \leq x\} = \Phi^{(t)}_j(x) \) be the \( t \)-fold convolution of \( D \), as before. Thus, \( \Phi^{(t)}_j(x) \) is the probability that demand over \( t \) periods is less than or equal to \( x \) units. Similarly, we may define another random variable \( N_j(x) \) representing the number of periods required to consume \( x \) units of item \( j \). From renewal theory, we know that,

\[
\Pr\{\bar{D}^{(t)}_j \leq x\} = \Pr\{N_j(x) \geq t\} = \Phi^{(t)}_j(x). \tag{4.17}
\]

See Ross (1970) for details. From a simple summation by parts, we may restate the allocation function \( Q(\cdot) \) as,

\[
Q_j(y_{0j}) = \sum_{t=2}^{\infty} \mathbb{E}[y_{0j} - \bar{D}^{(t)}_j]_+ = \sum_{t=2}^{\infty} \sum_{k=0}^{y_{0j}-1} \Phi^{(t)}_j(k). \tag{4.18}
\]

It is also easy to show,
Proposition 2

(i) \( Q(y) \) is convex in \( y \);

(ii) \( 0 \leq Q(y) < \infty \).

In summary, we propose to make the actual capacity and inventory allocation decisions in each period by solving

\[
(P_{\text{Approximate}}) \quad \tilde{f}(X, A) = \min_{Y \in \tilde{R}(X, A)} G(Y) + \sum_{j \in J} h_{0j} Q_j(y_{0j}),
\]

in a rolling horizon manner rather than by solving the exact dynamic program (4.12). Thus \( \sum_{j \in J} h_{0j} Q_j(y_{0j}) \) is used in the approximation model instead of the term \( E_D(f(Y - D)) \) in (4.12) to find \( Y \). Note that (4.19) is a convex program. Let \( Y^*_R \) be the optimal solution to (4.19). The objective function in (4.19) is not the expected incurred cost in a period, but rather a mechanism used to allocate production capacity to items and inventory to locations in the current period recognizing the consequences of decisions on future periods. The expected incurred cost in the current period is given by \( G(Y^*_R) \).

Furthermore, in the simulation study reported in section 5, we use (4.19) to make the production and allocation decisions in each period in combination with the system order-up-to level, \( T \). In the next section, we show how to compute an appropriate value for \( T \).

4.4 Determining the Order-Up-To-Level, \( T^* \)

Finally, we must determine the value of \( T \). To do so, we define and solve a relaxation of (4.19). Let \( U(A) \) be the set of feasible allocation decisions in a given period in which there is a total of \( A \) units of items in the system:

\[
U(A) = \left\{ Y : \sum_{j \in J} \sum_{i \in I} y_{ij} = A, y_{0j} \geq 0 \quad \forall j \in J \right\}.
\]

(4.20)

For \( A < 0, y_{0j} = 0, \forall j \in J \), since backorders do not occur at the central warehouse. To approximate the allocation decisions given \( A \) units in the system, we solve,

\[
(P_{\text{Relaxed}}) \quad \min_{Y \in U(A)} G(Y) + \sum_{j \in J} h_{0j} Q_j(y_{0j}).
\]

(4.21)
The objective function in (4.21) is clearly convex. Observe that (4.21) ignores any possible imbalances which may exist between items and locations, since only the total amount of system inventory, \( A \), is considered when making the allocation decision. Let \( Y^*_U \) be the optimal solution to (4.21). Then the expected incurred cost is given by,

\[
\mathcal{J}(A) := G(Y^*_U).
\]  

(4.22)

To find the system order-up-to-level, \( T^* \), we use the value of \( \mathcal{J}(A) \) for all \( A \). Recall that the inventory shortfall, \( V \), is the difference between the order up to level, \( T \), and the amount of system inventory available for allocation, \( A \). Equivalently, \( A = T - V \). As described in Rappold and Muckstadt (1998), the evolution of the shortfall random variable may be described as a Markov chain. Necessary and sufficient conditions for a stationary probability distribution to exist is \( E(D) < C \) [Prabhu (1980)]. Let \( \pi(k), k = 0, 1, 2, \ldots \) be the stationary probability distribution of \( V \), if it exists. From the strong law of large numbers for Markov chains, we know that the long run average cost per period converges almost surely to its expectation [Resnick (1992)]. Thus, we find \( T^* \) by solving,

\[
Z(T^*) = \min_T E_V(\mathcal{J}(T - V)) = \sum_{k \geq 0} \mathcal{J}(T - k) \pi(k).
\]  

(4.23)

Since \( \mathcal{J}(A) \) is convex in \( A \), \( T^* \) can be found quickly and efficiently using either a marginal analysis algorithm or a simple search procedure.

There are two reasons why this relaxed problem is an appealing approximation. First, due to the relative flatness of the newsvendor functions in the region of optimality, imbalance constraints are unlikely to cause large cost increases. Second, because of the way in which inventory is allocated in the system on a periodic basis, imbalance situations can only exist when inventory shortfall is positive \( (V > 0) \) or, equivalently, when the amount of available inventory in the system is less than \( T \). This establishes an upper bound on the long-run fraction of periods in which imbalance may exist. Namely, it is \( \Pr\{A < T\} = \Pr\{V > 0\} \).
5 Validation of the Proposed Approximation

Ideally, to validate the approximation model, we would have preferred to compare the average per period cost obtained using the approximation model with that obtained by solving the dynamic program (4.12). Unfortunately, due to the size of the state space, this precise computation is not possible. Our strategy is therefore to trap the value of the optimal expected cost between an upper and lower bound for a wide variety of system scenarios and to understand the circumstances in which this gap is large or small.

Let $z^*$ represent the optimal expected per period cost computed from the dynamic program (4.12).

Let $z_*$ denote the average per period cost obtained by simulating a system in which the approximation model (4.19) was used to make operating decisions. We know that $z_*$ is an upper bound on the optimal cost since it is derived from using a lower bound on the system inventory level $T$ while adhering to imbalance restrictions. The approximation model requires decisions to be made at the beginning of each period given the current state, $X$. The corresponding production and allocation decisions, $Y$, generated by this model need not correspond to the ones that would have resulted from solving (4.12). Hence, when implementing the production and allocation decisions, along with the choice of $T$, we will incur costs that may exceed those that would have been obtained if the solution to the dynamic program had been implemented.

Let $z_{lb}$ represent the lower bound on the expected per period cost which is obtained from solving (4.23). While many lower bounds exist, the one we have chosen is the objective function value obtained when solving (4.23). We know that $z_{lb}$ is a lower bound since it ignores any possible imbalances that may exist. Therefore,

$$z_{lb} \leq z^* \leq z_*,$$

and we will examine the relative error $(z_* - z_{lb}) / z_{lb}$.

To verify the quality of the solution provided by the approximation model, we conducted a large simulation study and estimated this relative cost error for a wide set of scenarios.
In this section, we will describe the experimental design in detail, report statistics obtained, and summarize the results based on the data.

5.1 Experimental Design

We have two objectives in conducting these experiments. First, we wish to understand how well the models we developed perform. Second, we want to gain further insight into the dynamic behavior of the system under different environmental and economic conditions. In these experiments, we are managing a system in which there are 5 locations and 10 items.

The system attributes, or experimental factors, of interest are the following:

(1) the variability of the demand process for each item at each location;
(2) the distribution of demand for capacity across items and locations;
(3) the capacity utilization rate of the production facility;
(4) the cost of backordering unsatisfied customer demand;
(5) the holding costs at the central warehouse relative to the holding costs at the distribution centers.

We now describe each of these experimental factors in greater detail.

(1) The demand occurring each period for each item at each location is assumed to be a negative binomially distributed random variable. These random variables are assumed to be independent across items, location, and periods. Additionally, we assume identical Variance-to-Mean ratios (VTMR) for each item at each location. As a consequence, the random variable describing the total system demand each period is also negative binomially distributed. The mean aggregate demand over all items and locations is 500 units per period in all of the experiments conducted. However, the manner in which these 500 units are allocated among items and locations vary in the experiments. Thus, we are able to exam cases with different demand rates by item and by location. Therefore, observe that the individual coefficients of variation among items and locations may vary substantially.
Three values of the Variance-to-Mean ratio are examined: 1.01, 5, and 10. Note that when the VTMR is 1.01 for each item at each location, the demand is approximately Poisson distributed.

(2) While the total system demand per period is fixed at 500 units, the expected demand per item and location varied. We consider two cases. In the first, the expected demand per period is identical for each item at each location. That is, the expected demand for item item at each location is 10 units per period. In the second case, the expected demands by item and location are given in Figure 2. The purpose of examining these two cases is to see how imbalance and its effect might be related to the manner in which demand is distributed across items and locations with varying coefficients of variation.

<table>
<thead>
<tr>
<th>DC</th>
<th>Product</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>21.9</td>
<td>16.2</td>
<td>14.6</td>
<td>13.6</td>
<td>13.0</td>
<td>12.4</td>
<td>12.0</td>
<td>11.7</td>
<td>11.4</td>
<td>11.2</td>
<td>136.0</td>
<td>27.6%</td>
</tr>
<tr>
<td>2</td>
<td>16.2</td>
<td>12.0</td>
<td>10.8</td>
<td>10.1</td>
<td>9.6</td>
<td>9.2</td>
<td>8.9</td>
<td>8.7</td>
<td>8.5</td>
<td>8.3</td>
<td>102.3</td>
<td>20.5%</td>
</tr>
<tr>
<td>3</td>
<td>14.6</td>
<td>10.8</td>
<td>9.7</td>
<td>9.1</td>
<td>8.6</td>
<td>8.3</td>
<td>8.0</td>
<td>7.8</td>
<td>7.6</td>
<td>7.4</td>
<td>92.0</td>
<td>18.4%</td>
</tr>
<tr>
<td>4</td>
<td>13.6</td>
<td>10.1</td>
<td>9.1</td>
<td>9.5</td>
<td>9.1</td>
<td>7.8</td>
<td>7.5</td>
<td>7.3</td>
<td>7.1</td>
<td>7.0</td>
<td>96.0</td>
<td>17.2%</td>
</tr>
<tr>
<td>5</td>
<td>13.0</td>
<td>9.6</td>
<td>8.6</td>
<td>8.1</td>
<td>7.7</td>
<td>7.4</td>
<td>7.1</td>
<td>6.9</td>
<td>6.8</td>
<td>6.6</td>
<td>81.7</td>
<td>16.3%</td>
</tr>
<tr>
<td>Total</td>
<td>79.2</td>
<td>56.7</td>
<td>52.9</td>
<td>49.4</td>
<td>46.9</td>
<td>45.1</td>
<td>43.6</td>
<td>42.4</td>
<td>41.3</td>
<td>40.4</td>
<td>500.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15.6%</td>
<td>11.7%</td>
<td>10.6%</td>
<td>9.9%</td>
<td>9.4%</td>
<td>9.0%</td>
<td>8.7%</td>
<td>8.5%</td>
<td>8.3%</td>
<td>8.1%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: Distribution of demand across items and locations

(3) As one might believe, the presence of finite capacity does have an impact on the validity of the approximation model we have proposed. Thus it is important that the experiment consider various utilization rates. Specifically, we examine three utilization rates: 80%, 90%, and 95%. In order to maintain identical simulated demand streams, capacity per period was adjusted to 625, 556, and 526 units per period, respectively.

(4) The two types of costs present in the models are holding costs and backorder costs for each period. We normalized the holding costs so that the installation holding cost for each item at each distribution center is $1 per unit per period. Backorder costs per period are only incurred the distribution centers and are assumed to be identical for all items at all distribution centers. Two values of backorder costs are considered: $5 and $10 per unit.
per period. As the ratio of backorder costs to holding costs \( b/h \) increases, safety stock also increases.

(5) Holding costs at the central warehouse \( (h_0) \) are set to be a fraction of the installation holding costs at the distribution centers \( (h) \). In the experiment, the central warehouse holding costs were chosen to be identical for each item. According to the allocation problem (3.4), as the ratio of these holding costs \( (h_0/h) \) decreases, the incentive to hold stock centrally increases as well as the incentive to increase the amount of inventory in the system. Specifically, two values for the ratio \( (h_0/h) \) are considered: 0.1 and 0.9.

The experimental factors and their respective factor levels are summarized in Table 1.

5.2 The Simulation

When running the simulation experiments, certain system attributes were exploited. First, in each period when the system inventory level attains the target level \( T \), a system renewal occurs. Let a *cycle* be the time between successive renewals. Clearly, the average cycle length increases as the capacity utilization and the demand process variation increases. In our comparisons, each scenario that was simulated consisted of 5,000 cycles. Observe that the number of periods simulated varied substantially between scenarios. Ten replications of each scenario was run, with each using different random seeds. All random seeds used exhibited anti-cycling characteristics. Furthermore, to reduce the experimental bias, each replication was repeated using its antithetic random variate for each scenario. Thus, a total of 20 replications were performed for each system scenario for total of 100,000 cycles.

5.3 Performance Measures

To assess the effectiveness of the model, the following four performance measures were calculated for each scenario:

1. Average actual total system cost per period;
2. Average number of units in capacity imbalance per period;
3. Average number of units in location imbalance per period;
(4) Average total number of units in imbalance per period.

Let us define the performance measures more precisely.

(1) The average total system cost per period measures the average actual total holding and backorder costs incurred for each item at each location.

(2) The average number of units in capacity imbalance compares the actual number of units of an item \( j \) in the system after production decisions are made with the \textit{ideal} number of units of \( j \) assuming no capacity imbalance exists. Note the capacity imbalance is indifferent to the precise location of the units of item \( j \).

(3) The average number of units in inventory imbalance compares the actual number of units of inventory at location \( i \) after allocation decisions are made with the ideal number of units of inventory at location \( i \) assuming no inventory imbalance exists. Note that location imbalance does not consider the mix of items at location \( i \), but rather how much production’s worth.

(4) The average total imbalance measures the difference between the actual units of item \( j \) at location \( i \) and the ideal number of units of item \( j \) at location \( i \) for a given amount of available system inventory, \( A \).

It follows directly that both the capacity imbalance measure and the location imbalance measure are bounded above by the total imbalance measure. The formulas for computing these imbalance statistics are given in Table 2.

It is clear that when considering allocation decisions \( \mathbf{Y} \) given \( \mathbf{X} \), for each location-item pair \((i, j), i \in \mathcal{I}, j \in \mathcal{J}\), for which \( y_{ij}^* < y_{ij} \), there exists a pair \((i', j') \neq (i, j)\) such that \( y_{ij}^* > y_{i'j'} \), where \( y_{ij}^* \) is the optimal allocation given \( A \). Thus when one unit is mis-allocated, the imbalance occurs either in two locations or two items. For this reason, to compute the imbalance measures, we multiply the sum of all imbalance occurrences by \(1/2\), so as not to double count the number of units in the state of imbalance.
<table>
<thead>
<tr>
<th>System Attribute</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>VTMR</td>
<td>1.01, 5, 10</td>
</tr>
<tr>
<td>Capacity Utilization</td>
<td>80%, 90%, 95%</td>
</tr>
<tr>
<td>$b/h$</td>
<td>5, 9</td>
</tr>
<tr>
<td>$h_0/h$</td>
<td>0.1, 0.9</td>
</tr>
<tr>
<td>Location Skew</td>
<td>Identical, Skewed</td>
</tr>
<tr>
<td>Item Skew</td>
<td>Identical, Skewed</td>
</tr>
</tbody>
</table>

Table 1: Factors under study

Total Imbalance \[\frac{1}{2} \sum_j \sum_i |Y^*_R(i, j) - Y_A(i, j)|\]
Capacity Imbalance \[\frac{1}{2} \sum_j |\sum_i (Y^*_R(i, j) - Y_A(i, j))|\]
Location Imbalance \[\frac{1}{2} \sum_i |\sum_j (Y^*_R(i, j) - Y_A(i, j))|\]

Table 2: Imbalance Measures

6 Discussion of Results

While both capacity and location imbalance can be of theoretical concern, in the range of system configurations we examined, imbalance had virtually no impact on average system costs and occurred minimally as a function of the total amount of inventory in the system. With adequate production capacity, imbalance situations can be quickly resolved. Consequently, the relative cost errors of the approximation model were very small.

6.1 Inventory Levels

As a function of system parameters, inventory levels varied dramatically. In high demand variance and high capacity utilization cases, inventory levels increased substantially. In response to increasing incremental echelon holding costs ($h_0/h$), inventory levels decreased since the minimal cost associated with having an additional unit of inventory in the system ($h_0$) increased. These inventory levels are summarized in Figure 3.
<table>
<thead>
<tr>
<th>Utilization</th>
<th>VTMR</th>
<th>b/h</th>
<th>h0/h</th>
<th>0.1</th>
<th>0.1</th>
<th>0.1</th>
<th>0.1</th>
<th>0.9</th>
<th>0.9</th>
<th>0.9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>80%</td>
<td>1.01</td>
<td>646</td>
<td>1.3</td>
<td>646</td>
<td>1.3</td>
<td>706</td>
<td>1.4</td>
<td>706</td>
<td>1.4</td>
<td>676</td>
<td>(1.4)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>809</td>
<td>1.6</td>
<td>809</td>
<td>1.6</td>
<td>968</td>
<td>1.9</td>
<td>968</td>
<td>1.9</td>
<td>889</td>
<td>(1.8)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>899</td>
<td>1.8</td>
<td>899</td>
<td>1.8</td>
<td>1,151</td>
<td>2.3</td>
<td>1,151</td>
<td>2.3</td>
<td>1,025</td>
<td>(2.1)</td>
</tr>
<tr>
<td>90%</td>
<td>1.01</td>
<td>646</td>
<td>1.3</td>
<td>646</td>
<td>1.3</td>
<td>706</td>
<td>1.4</td>
<td>706</td>
<td>1.4</td>
<td>676</td>
<td>(1.4)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>818</td>
<td>1.6</td>
<td>818</td>
<td>1.6</td>
<td>976</td>
<td>2.0</td>
<td>976</td>
<td>2.0</td>
<td>897</td>
<td>(1.8)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>917</td>
<td>1.8</td>
<td>917</td>
<td>1.8</td>
<td>1,167</td>
<td>2.3</td>
<td>1,167</td>
<td>2.3</td>
<td>1,042</td>
<td>(2.1)</td>
</tr>
<tr>
<td>95%</td>
<td>1.01</td>
<td>651</td>
<td>1.3</td>
<td>651</td>
<td>1.3</td>
<td>711</td>
<td>1.4</td>
<td>711</td>
<td>1.4</td>
<td>681</td>
<td>(1.4)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>840</td>
<td>1.7</td>
<td>840</td>
<td>1.7</td>
<td>997</td>
<td>2.0</td>
<td>997</td>
<td>2.0</td>
<td>919</td>
<td>(1.8)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>991</td>
<td>2.0</td>
<td>970</td>
<td>1.9</td>
<td>1,241</td>
<td>2.5</td>
<td>1,222</td>
<td>2.4</td>
<td>1,106</td>
<td>(2.2)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>802</td>
<td>1.6</td>
<td>800</td>
<td>1.6</td>
<td>968</td>
<td>1.9</td>
<td>956</td>
<td>1.9</td>
<td>879</td>
<td>(1.8)</td>
</tr>
</tbody>
</table>

Figure 3: System Order-up-to Levels (Days of Supply)

6.2 Relative Cost Errors

The relative cost errors and their standard deviations are reported in Figure 4. In all cases, the cost errors were very small and did not differ significantly from zero. Across all system configurations tested, the distribution of relative cost errors are shown in Figure 5. From this, we observe that 50% of the systems tested resulted in cost errors less than 0.27%, and 95% of the systems tested resulted in cost errors less that 1.0%. In particular, for the approximate Poisson demand case, the relative cost error was virtually zero. Consequently, the approximation model is very effective in determining appropriate inventory levels.

<table>
<thead>
<tr>
<th>Utilization</th>
<th>VTMR</th>
<th>b/h</th>
<th>h0/h</th>
<th>Average</th>
<th>StdDev</th>
<th>Average</th>
<th>StdDev</th>
<th>Average</th>
<th>StdDev</th>
<th>Average</th>
<th>StdDev</th>
<th>Average</th>
<th>StdDev</th>
<th>Average</th>
<th>StdDev</th>
<th>Average</th>
<th>StdDev</th>
<th>Average</th>
<th>StdDev</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>80%</td>
<td>1.01</td>
<td>646</td>
<td>1.3</td>
<td>0.04%</td>
<td>0.38%</td>
<td>0.04%</td>
<td>0.38%</td>
<td>0.01%</td>
<td>0.47%</td>
<td>0.01%</td>
<td>0.47%</td>
<td>0.01%</td>
<td>0.47%</td>
<td>0.01%</td>
<td>0.47%</td>
<td>0.03%</td>
<td>0.43%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>809</td>
<td>1.6</td>
<td>0.01%</td>
<td>0.51%</td>
<td>0.01%</td>
<td>0.51%</td>
<td>0.03%</td>
<td>0.62%</td>
<td>0.03%</td>
<td>0.62%</td>
<td>0.01%</td>
<td>0.56%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>899</td>
<td>1.8</td>
<td>0.00%</td>
<td>0.52%</td>
<td>0.00%</td>
<td>0.52%</td>
<td>0.06%</td>
<td>0.63%</td>
<td>0.06%</td>
<td>0.63%</td>
<td>0.02%</td>
<td>0.67%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>1.01</td>
<td>646</td>
<td>1.3</td>
<td>0.07%</td>
<td>0.40%</td>
<td>0.07%</td>
<td>0.40%</td>
<td>0.16%</td>
<td>0.48%</td>
<td>0.16%</td>
<td>0.48%</td>
<td>0.11%</td>
<td>0.44%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>818</td>
<td>1.6</td>
<td>0.06%</td>
<td>0.40%</td>
<td>0.06%</td>
<td>0.40%</td>
<td>0.08%</td>
<td>0.47%</td>
<td>0.08%</td>
<td>0.47%</td>
<td>0.07%</td>
<td>0.43%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>917</td>
<td>1.8</td>
<td>0.06%</td>
<td>0.40%</td>
<td>0.06%</td>
<td>0.40%</td>
<td>0.09%</td>
<td>0.47%</td>
<td>0.09%</td>
<td>0.47%</td>
<td>0.06%</td>
<td>0.44%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95%</td>
<td>1.01</td>
<td>651</td>
<td>1.3</td>
<td>0.25%</td>
<td>0.33%</td>
<td>0.25%</td>
<td>0.33%</td>
<td>0.36%</td>
<td>0.39%</td>
<td>0.36%</td>
<td>0.39%</td>
<td>0.31%</td>
<td>0.36%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>840</td>
<td>1.7</td>
<td>0.11%</td>
<td>0.37%</td>
<td>0.11%</td>
<td>0.37%</td>
<td>0.13%</td>
<td>0.41%</td>
<td>0.13%</td>
<td>0.41%</td>
<td>0.12%</td>
<td>0.39%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>991</td>
<td>2.0</td>
<td>0.06%</td>
<td>0.53%</td>
<td>0.03%</td>
<td>0.66%</td>
<td>0.00%</td>
<td>0.56%</td>
<td>0.06%</td>
<td>0.57%</td>
<td>0.01%</td>
<td>0.65%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>802</td>
<td>1.6</td>
<td>0.05%</td>
<td>0.44%</td>
<td>0.06%</td>
<td>0.44%</td>
<td>0.10%</td>
<td>0.51%</td>
<td>0.11%</td>
<td>0.51%</td>
<td>0.08%</td>
<td>0.43%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4: Average Relative Cost Errors and Standard Deviations
Figure 5: Distribution of Cost Errors across System Scenarios

6.3 Imbalance

The number of units in imbalance varied as a function of the particular system parameters tested. The average number of units in imbalance as a fraction of total system inventory is reported in Figure 6. In most situations, imbalance was negligible due to the short transportation lead times which allow a quick correction of an imbalance situation, provided there is adequate stock in the system. However, in high demand variance and high capacity utilization cases, the system is often unable to correct an imbalance situation quickly. Despite the presence of imbalance, the number of units in imbalance remains very small as a percentage of total system inventory and consequently has little impact on costs. In particular, even when the average percentage imbalance was largest, the average cost difference between the upper and lower bounds was only 0.06%. 
<table>
<thead>
<tr>
<th>Utilization</th>
<th>b/h</th>
<th>h0/h</th>
<th>5</th>
<th>9</th>
<th>5 Total</th>
<th>9 Total</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1</td>
<td>0.9</td>
<td>0.1</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>80%</td>
<td>1.01</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.01%</td>
<td>0.01%</td>
<td>0.01%</td>
<td>0.01%</td>
<td>0.01%</td>
<td>0.01%</td>
</tr>
<tr>
<td>80% Total</td>
<td>1.01</td>
<td>0.01%</td>
<td>0.01%</td>
<td>0.01%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.01%</td>
</tr>
<tr>
<td>90%</td>
<td>1.01</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.01%</td>
<td>0.01%</td>
<td>0.01%</td>
<td>0.01%</td>
<td>0.01%</td>
<td>0.01%</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.09%</td>
<td>0.09%</td>
<td>0.09%</td>
<td>0.07%</td>
<td>0.07%</td>
<td>0.06%</td>
</tr>
<tr>
<td>90% Total</td>
<td>1.01</td>
<td>0.04%</td>
<td>0.04%</td>
<td>0.04%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.04%</td>
</tr>
<tr>
<td>95%</td>
<td>1.01</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.13%</td>
<td>0.02%</td>
<td>0.05%</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>2.50%</td>
<td>0.18%</td>
<td>1.40%</td>
<td>2.05%</td>
<td>0.15%</td>
<td>1.11%</td>
</tr>
<tr>
<td>95% Total</td>
<td>1.05</td>
<td>0.08%</td>
<td>0.57%</td>
<td>0.91%</td>
<td>0.07%</td>
<td>0.49%</td>
<td>0.52%</td>
</tr>
<tr>
<td>Total</td>
<td>0.37</td>
<td>0.04%</td>
<td>0.21%</td>
<td>0.32%</td>
<td>0.04%</td>
<td>0.18%</td>
<td>0.19%</td>
</tr>
</tbody>
</table>

Figure 6: Average Number of Units in Imbalance as a Percentage of Total System Inventory

7 Conclusion

In this paper, we have constructed a multiple location supply chain model which differs from previous work by the consideration of finite supply and multiple items. We formulate the exact problem as a stochastic dynamic program whose solution is impractical for large systems in which several thousand items are managed at tens of locations. To address this difficulty, we constructed an approximation model in which the solution can be computed very quickly on a common personal computer. The computational requirements increase at a rate of $\log_2((M + 1)P)$. Through a set of computational experiments, we demonstrated that the resulting solutions, when implemented, are very close to optimal.

This paper forms the basis for continued research in the area of capacitated multiechelon systems in which production and distribution decisions must be coordinated to achieve a reliable and efficient supply chain.
References


