SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NY 14853-3801

TECHNICAL REPORT

June 2001

Capacitated Production Planning and Inventory Control when Demand is Unpredictable for Most Items: The No B/C Strategy

John A. Muckstadt
Cornell University

David H. Murray
College of William & Mary

James A. Rappold
University of Wisconsin

This research was partially funded by the National Science Foundation (Grant DMI0075627) and by Aspen Technology.

DRAFT

Capacitated Production Planning and Inventory Control when Demand is Unpredictable for Most Items: The No B/C Strategy

John A. Muckstadt
School of Operations Research and Industrial Engineering
Cornell University, Ithaca, NY 14853

David H. Murray
School of Business Administration
The College of William & Mary, Williamsburg, VA 23187

James A. Rappold
School of Business
University of Wisconsin, Madison, WI 53706

June 2001

ABSTRACT

In this paper, we examine a discrete-time, periodic-review production environment that assembles several hundred items and that possesses limited, perhaps random production capacity. The demand for a large subset of these items is highly erratic and extremely difficult, if not impossible, to predict accurately. Consequently, a coordinated production-inventory strategy, such as the *No B/C Strategy* presented in Carr, et al. (1993), is necessary. In such a strategy, inventory is carried only in high demand rate, predictable items and production priority is given to non-stocked items. Production is controlled for stocked items through a modified base stock policy.

A key feature of our approach is that it does not rely on item-level forecasts for each item. Our objective is to develop and test a computationally efficient and accurate procedure for establishing base-stock levels that minimizes the expected holding and backorder costs per period over an infinite horizon.

We consider two alternate operating scenarios and compare them. In the first, we consider an information poor environment and assume that production decisions must be made in advance of observing demand. In the second, we assume an information rich environment and observe demand prior to making production decisions. We quantify the value of this information in terms of operating performance. We demonstrate that as the capacity utilization increases, the value of obtaining advanced demand information, while still valuable, diminishes.

In these environments, there are two types of safety stock needed to minimize expected costs: demand-driven safety stock and capacity-driven safety stock. We demonstrate that for capacity-driven safety stock, traditional equal-fractile allocation policies may lead to substantial inventory imbalances over time and result in lower than expected service levels. To resolve this, we present an alternative means for allocating capacity-driven safety stock that favors storing inventory in items for which the risk of not selling them is lower.

Through a series of numerical experiments, the model is shown to be accurate and on average performs within 0.26% of a lower bound. The information rich case resulted, on average, in 35% lower safety stock and 57% lower expected costs per period than the information poor case. Using our proposed capacity allocation scheme resulted in 23% less units in imbalance than if the newsvendor allocation were used. We conclude with an experiment of the approach in an industrial environment and demonstrate its effectiveness over current management practices.

KEY WORDS: PRODUCTION PLANNING AND CONTROL, MANUFACTURING STRATEGY, NO B/C STRATEGY, VALUE OF INFORMATION, INVENTORY SHORTFALL, COLLABORATIVE PLANNING.

1 Introduction

In this paper, we explore an alternative way of making capacity allocation and inventory stocking decisions that is designed for lean manufacturing environments. These environments have a wide product variety, highly uncertain demand, limited production capacity (perhaps random), drastically shortened flow times, reduced setup times, and small batch sizes. Investments made in new equipment and in employee training have resulted in assembly environments with very short, predictable, and repeatable flow times. Typically, there is an effective maximum production capacity per period. Depending on how the production facility has been designed, this production capacity may be deterministic or random, and may be scalable through the dynamic addition of equipment or labor. The timing of demand information may also be different in different environments. In some cases, production decisions must be made prior to the realization of actual demand. In this case, a forecast is generated for each item over a lead-time and inventory is held due to forecast errors. The lead-time is often used as a deterministic surrogate for limited production capacity. We refer to this inventory as demand-driven safety stock. In other cases, demand information is known prior to making production decisions. If capacity were unlimited, then no finished goods inventory would be needed; however, due to limited capacity, there may or may not be sufficient capacity in a period to satisfy all demand. Consequently, some inventory is needed to protect customer service objectives. We refer to this type of inventory as capacity-driven safety stock. The more volatile and unpredictable the demand is, the more safety stock is needed for a given level of customer service.

Based on our work and observations in both the hydraulic fluid conveyance and electrical controls industries, the existing methodologies used in practice for governing the management, planning, and control of capacity and materials in manufacturing systems are not well suited for these new environments. Specifically, the assumption that demand forecast errors for an item over a short replenishment lead-time can be accurately represented using a probability distribution is flawed, as we shall demonstrate. Thus, production and inventory plans based on resulting inaccurate demand forecasts are themselves flawed and can lead to a significant misallocation of capacity over time. In some cases, this capacity misallocation has resulted in large inventories of low demand rate items and in inventory writedowns.

Examples of the items considered in this research are assembled products, such as computers. In order to receive, process, and fill these orders, firms must balance assembly capacity limitations with finished goods inventory investment and customer service levels. While it is typically not economical to stock a wide breadth of products in anticipation of demand, there are likely to be several popular models for which there is a lower risk of obsolescence. It is in these products that we consider possible stocking in advance of demand in order to protect customer service against insufficient production capacity.

The value of information for production and inventory decision-making depends heavily on the timing of events in the system. If the manufacturing flow-time is greater than the customer order lead-time, then a forecast must be made and inventory must be held. If the manufacturing flow-time is within the customer order lead-time, then the firm can potentially satisfy demand without the need for item-level forecasts for all items. If demand is observed before production decisions are made, this permits a one-day reduction in order fulfillment lead-time for low demand items. On the other hand, if production decisions must be made before the observation of the period's demand, this results in an additional one day of order fulfillment lead-time.

Currently, inventory systems with several hundred or thousand items are managed fundamentally on the notion of ABC inventory, as described in Silver and Peterson (1985). Under this system, items are categorized into either *A*, *B*, or *C* categories. In this categorization, the *A*-type items are fast-selling items, which typically account for 20% of all the items, and which account for 80% of all sales dollars. The *B*-type items are medium-selling items which account for 30% of the items, and which account for 15% of sales dollars. The *C*-type items are the slowest-selling items, which account for the remaining 50% of the items, and which account for the remaining 5% of sales dollars.

This is the classic Pareto analysis. In our experience, however, we rarely see distributions of demand among items that are like this classical one. The one shown in Figure 1 is more representative of the many that we have observed. Furthermore, this categorization does not account for the differences in uncertainty that exists in the demand among items. For some items, demand is uncertain but predictable, while for others the demand is not predictable. That is, the distribution of lead-time forecast errors in one case can be represented accurately with a Poisson, Negative Binomial, Normal or Laplace distributions, for example. In the second case, the coefficient of variation of forecast errors is so large that an accurate prediction of lead-time demand is not possible to obtain, much less the distribution of demand. Examples of time series for some items are given in Figure 2.

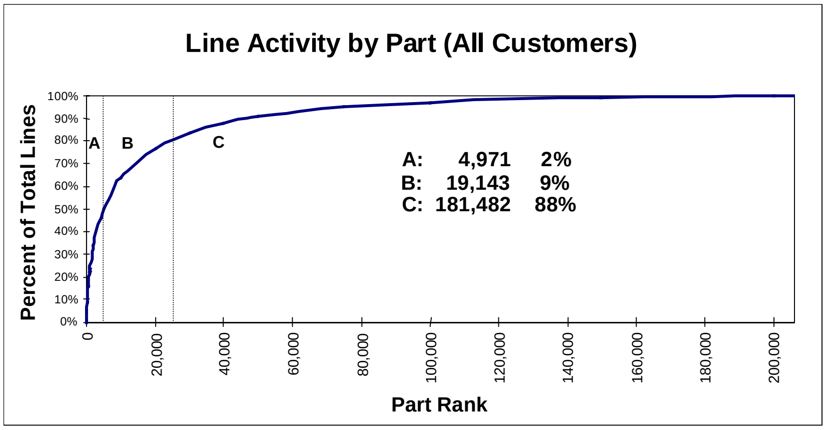

Figure 1. An example part categorization for a processing facility of automotive spares

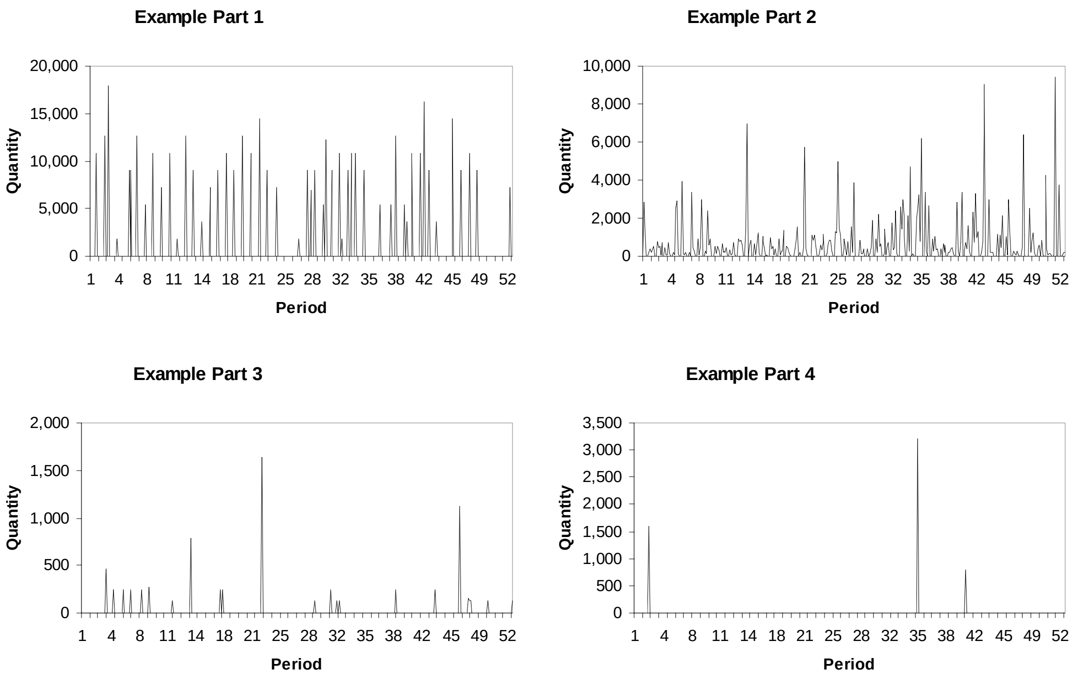

Figure 2. Example Time series for some products.

The research questions of interest are: What is an appropriate production policy for controlling production and inventory? How should capacity be allocated to individual items? How can the base-stock levels be computed efficiently for large-scale systems? What is the value of obtaining advanced demand information?

The approach we present is different from the current paradigm for managing production and inventory in this type of environment, since it considers highly variable demand and limited production capacity. As we have stated, the assumption that the demand process for each item is accurately representable via a probability distribution is inappropriate in many situations. Consequently, we will not assume a forecast is possible for these items. Nor will we assume a fixed lead-time as is common practice.

We propose an approach for managing production capacity and inventory that is called the *No B/C Stocking Strategy*. It was first described in Carr, et al. (1993), and has the following operating characteristics. First, inventory is not held in any *B*- or *C*-type items, but rather only *A*-type items are produced on a make-to-stock basis. Second, *B*- and *C*-type items are given production priority when demand for them occurs. Thus, they are produced on a make-to-order basis. Any excess production capacity remaining in a period after production of the *B*- and *C*-type items may be used to replenish *A*-type item inventory up to a predetermined base stock level. To operate under this policy, *A*-type item inventory is required due to the presence of limited production capacity. Although accurate item-level forecasts may be difficult for many items, it is often easier to accurately characterize the *aggregate demand for capacity* using a probability distribution. One may argue that producing the very low demand items is not appropriate; however, in many instances, the low and unpredictable demand rate items are the most profitable items, since manufacturers of these items are often the "supplier of last resort".

The decision to categorize an item as a make-to-stock item usually involves item-specific attributes such as demand volume, stability, manufacturing lead-time, etc.; however, the categorization decisions and the stock level decisions are not separable in light of capacity limitations. Our model determines the categorization of items into stocked and non-stocked types, depending on the system attributes, such as capacity and demand variability. Items for which a probability distribution for demand is not available or possible to estimate are, by definition, non-stocked items.

Fundamental to our approach is the assumption that the effect of inventory imbalances on cost over time is negligible. *Inventory imbalance* occurs when there is too much of one item and too little of another item in the system for a given level of total inventory. In such situations, we say that the allocated capacity is in a state of imbalance. We make this assumption since it has been shown that in multi-echelon inventory systems with finite production capacity, inventory imbalance does not have a significant impact on system cost. See Rappold and Muckstadt (2000). Under this assumption, we obtain a simple lower bound on the expected system cost per period and demonstrate that this lower bound is very close to the optimal cost.

There are several advantages that we expect from operating the system under such a policy. First and most importantly, the inherent statistical problems surrounding accurate forecasts for item-level demands diminish for *B*- and *C*-type items. Under the No B/C strategy, the aggregate demand for production time is forecasted across all items. Because this aggregation is over a large group of items, the aggregated demand process possesses a much smaller coefficient of variation. This leads to parameter estimates of demand for capacity with a higher level of confidence, and ultimately allows for more accurate forecasts of capacity usage. Second, inventory investment is focused in fewer items, each of which has a high turnover ratio. Third, the management of the reduced inventory items is simplified and many of the indirect costs associated with inventory storage, tracking, and handling are reduced. Fourth, customer service improves due to shorter response times for *B*- and *C*-type items, (since they are produced on a make-to-order basis) and due to high inventory levels of *A*-type items to protect service from insufficient production capacity. Fifth, the system can be managed as a pull system and can be controlled using simple pull signals for the stocked items.

2 Literature Review

There has been much research accomplished in the area of inventory planning for single location production and inventory systems. The research has been in the context of continuous or discrete time models for systems that produce either a single or multiple items and operate with or without setup costs. Our research falls in the category of discrete time, multiple item models with no setup costs. Where our problem differs is wherein past research has assumed the ability to characterize item level demands with a probability distribution, we assume that for the majority of items (*B*- and *C*-type items), no such characterization is possible.

Currently, there are two fundamental approaches for examining this problem. The first approach, by Federgruen and Katalan (1994), regards time as a continuum and applies queuing theory and polling system theory to describe the behavior of the production system. Under this paradigm, *A*-type items are produced according to a cyclic production schedule. The methodology is able to consider both setup times and processing times for each item. The production of *B/C*-type items is triggered once an order for such an item is received. Production priority is given to the *B/C*-type item(s) by either pre-empting the current *A*-type item production or by waiting until the completion of the current *A*-type item. Once the production of the *B/C*-type item is complete, production of the *A*-type items resumes according to the cyclic schedule. Production of *A*-type items is controlled via a base-stock policy in which production of the *A*-type item stops once its inventory level reaches a pre-determined target level, or base-stock level. The framework is computationally efficient and quite general, in that it allows setup times.

As a consequence of this policy, the cycle length (the time between successive production runs of the same *A*-type item) is random and, depending on the volatility of demand for all items, may be highly variable. This cycle time variation increases safety stock requirements and may hinder the ability to provide predictable service to customers when demand is highly uncertain. Because of this variation in lead times, it is difficult to implement in practice. Furthermore, with uncertain lead times, it is difficult to use this approach within a mathematical programming based production and inventory control model.

Our approach likewise assumes that if the individual item demand for *B/C*-type items is unpredictable, then no inventory should be stored in such items. To compensate for this, production priority will be given to *B/C*-type items. Furthermore, to ensure that customer service is protected for *A*-type items, additional *A*-type item inventory will be held. Both approaches are also computationally viable for real-sized production environments. Where they differ is in the fundamental perspective of the system - periodic versus continuous review. Our model has the operational advantage that, because of the repetitive nature of the system, lead times are relatively constant, which is, in all real systems we have examined, a requirement.

More recently, Sox, Thomas, and McClain (1997) also consider a production-inventory strategy in which inventory is held only in high demand items. Their approach is based on a continuous time analysis and uses queueing analysis to model system behavior under a variety of production scheduling disciplines. Their model has the attractive feature that it is simple and as the capacity utilization approaches 100%, relatively more inventory is held in the high demand items. They, like many other researchers, assume that item-level demand processes are Poisson processes. In practice, we have found that the Poisson model is inappropriate because the variation in demand over a short lead-time is much greater than the Poisson model suggests. We are therefore most interested in studying highly uncertain demand environments. Hence, we have elected to model *A*-type item and aggregate *B/C*-type item demand using Negative Binomial distributions.

Much of the past research into this problem has either assumed unlimited supply or allocation rules that are inherently computationally complex to handle imbalanced inventory situations. Basic insights into the problem of inventory imbalances are made in Zipkin (1984). Other work by Aviv and Federgruen (1997) and by Kapuscinski and Tayur (1996) have made great strides in furthering the understanding of these complex systems, but are computationally unattractive when applied to large systems. Approximations such as those developed in Glasserman (1996, 1997) are very insightful in explaining system dynamics as the capacity utilization approaches 100% in single item systems. In our observations of real systems, as the capacity utilization approaches 100%, changes are made to the system in order to complete production on time. Examples of such changes include outsourcing a fraction of the production requirement and using overtime.

The choice of a base stock policy for controlling production and inventory for each item is based on the results developed in Federgruen and Zipkin (1984a) and in DeCroix and Arreola-Risa (1998). Our model differs slightly from theirs since they consider the situation in which production decisions are made prior to the realization of demand. They argue that for systems in which there is finite capacity and no fixed setup costs, a modified base stock policy is optimal. Thus, while a base stock policy may not be the true optimal policy, it is a conceptually easy to follow policy rendering it attractive from a managerial perspective. Ciarallo et al. (1994) examine the case when capacity is finite and random. As we will show subsequently, the case of uncertain capacity can be modeled as increased uncertainty in the demand process.

There are several contributions in this paper. First, we propose the No B/C strategy as a heuristic and easy to understand policy, whose performance we explore and validate. Second, we develop a model to efficiently determine item inventory levels for systems containing a large number of items. Lastly, we describe an industrial implementation of such a strategy.

3 Assumptions and General Notation

In this section we state our modeling assumptions and define the notation that we use throughout the rest of the paper. Model 1 corresponds to the information poor case and Model 2 corresponds to the information rich case. For *x* real, let $[x]^+ := \min\{x, 0\}$.

<u>System Attributes</u>

A	the set of all potentially stocked (*A*-type) items;
B	the set of all non-stocked (*B/C*-type) items;
P	$=\lvert A\rvert$, the total number of *A*-type items;
i	the index for items;
n	the index for time periods;
m	the index for models, *m*=1 or 2, for Model 1 or 2, respectively;
h_i	the holding cost per unit per period for item $i \in A \cup B$;
π_i	the backorder cost per unit per period for item $i \in A \cup B$;
T_m	the system base-stock level;
$\bar{\tau}_m = [\tau_{mi}]$	the vector of base-stock levels for individual items, where $\sum_{i \in A} \tau_{mi} = T_m$.

<u>Random Variables and Probability Distributions</u>

C	the discrete, non-negative random variable (r.v.) representing the per period production capacity;
c_n	the realized production capacity in period *n*;
F_C	the cumulative distribution function of *C*;
A_i	the discrete, non-negative r.v. representing the per period demand for capacity, for item $i \in A$;
F_i	the cumulative distribution function of A_i;
$A_i^{(n)}$	the cumulative demand r.v. over *n* periods for $i \in A$;
$F_i^{(n)}$	the cumulative distribution function of $A_i^{(n)}$;

A the discrete, non-negative r.v. representing the per period aggregate demand for capacity by all A-type items, where $A = \sum_{i \in A} A_i$;

B the discrete, non-negative r.v. representing the aggregate demand for capacity by all *B/C*-type items;

D the r.v. representing the aggregate demand for capacity, where $D=A+B$;

V_n the inventory shortfall in period n for Model 1;

V the stationary inventory shortfall r.v. for Model 1, if it exists.

U_n the inventory shortfall in period n for Model 2;

U the stationary inventory shortfall r.v. for Model 2, if it exists.

<u>Vectors</u>

$\overline{D}$ a $1 \times (P+1)$ vector of all demand r.v.'s, where $\overline{D} = [A_1, A_2, \ldots, A_P, B]$;

$\overline{d}_n$ a $1 \times (P+1)$ vector of realized demands in period n, with d_{ni} the realized demand for item i=1, 2, …, P, in period n, and $d_{n,P+1}$ the realized demand for capacity by all *B/C*-type items in period n;

$\overline{x}_n$ a $1 \times (P+1)$ vector of net inventory levels at the beginning of period n;

$\overline{y}_n$ a $1 \times (P+1)$ vector of net inventory levels after production decisions in period n for Model 1;

$\overline{z}_n$ a $1 \times (P+1)$ vector of net inventory levels after production decisions in period n for Model 2;

We make the following set of assumptions. First, we assume that there are no fixed setup times or costs between items which affect production decisions. We assume that while the production capacity is random each period, it is known in the period before making production decisions. We assume that the system is stable,

$$E(D) < E(C). \tag{1}$$

We assume that we always have adequate capacity to satisfy *B/C*-type item demand in a period,

$$\Pr\{B < C\} = 1. \tag{2}$$

Since *B*/*C*-type items receive production priority over *A*-type items and Pr{B < C}=1, *B*/*C*-type items will never be backordered. This assumption is primarily for convenience and ease of exposition. As we will discuss subsequently, this assumption may be relaxed, as it does not alter the validity of the problem or solution procedure. We assume that customer order lead times for *B/C*-type items are one day. Without loss of generality, we assume that the per unit processing times across the items are identical. Thus, we may express demand for each item in terms of required time when allocating production capacity. We assume that the demand r.v. and the capacity r.v. are stationary and independent between items and periods.

Inventory is pulled from finished goods inventory to satisfy external demand. When stock is inadequate to meet demand, excess demand is backordered. A modified base stock policy is followed to control production and inventory for each *A*-type item. For Model *m*, according to a modified base stock policy, there is a prescribed target inventory level τ_{mi} for each *A*-type item *i*, which is also called an order-up-to level, or base stock level. We denote $T_m = \sum_{i \in A} \tau_{mi}$ as the system target inventory level. The production facility attempts to restore all *A*-type item inventory levels to τ_{mi}, but may not be able to do so due to insufficient capacity. We assume that items produced in a period are available for shipment at the end of that period.

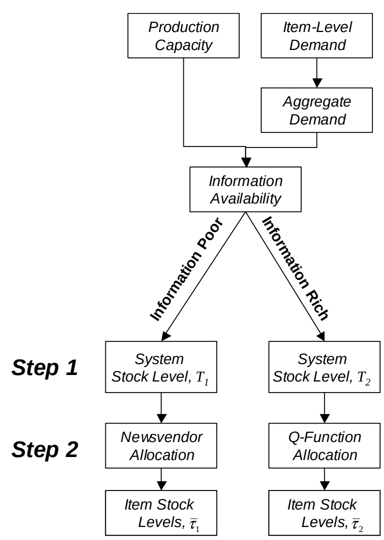

Figure 3. Overview of modeling approach

Our approach for determining these inventory levels is a two-step process shown in Figure 3. First, we determine how much inventory should be held in the system. Second, we allocate this aggregate inventory level to individual *A*-type items using one of two allocation schemes, depending on the information availability.

4 Model 1 – Information Poor Environment

Under this first scenario, production decisions are made prior to observing the period's demand. The sequence of daily events is as follows. At the beginning of period *n*, $\bar{x}_n$ is the vector of net inventory levels (on-hand minus backorders) for the stocked items. Production capacity c_n becomes known. Production decisions are made to satisfy *B/C*-type demand occurring in the previous period and to raise x_{ni} to τ_{1i}, for each $i \in A$, if possible, resulting in post-production net inventory levels of $\bar{y}_n$. The system may be unable to restore the net inventory levels to their targets due to insufficient capacity. We define the inventory shortfall in period n as V_n, measured after production but before demand occurs as described in Tayur (1992). The single period expected cost function is,

$$G(\bar{y}_n) = \sum_{i=1}^{P} h_i E[y_{ni} - A_i]^+ + \pi_i E[A_i - y_{ni}]^+ , \qquad \textbf{(3)}$$

where h_i and π_i are the holding and backorder costs, respectively, for item $i \in A$. This is the familiar newsvendor function. It is the mechanism by which capacity allocation decisions will be made in Model 1. Since *B/C*-type items are given production priority and $\Pr\{B < C\} = 1$, no holding or backorder costs will be incurred for these items. The period's demands $\bar{d}_n$ are then realized and are either satisfied or backordered. Finally, the period's costs are computed. This sequence of events is depicted in Figure 4.

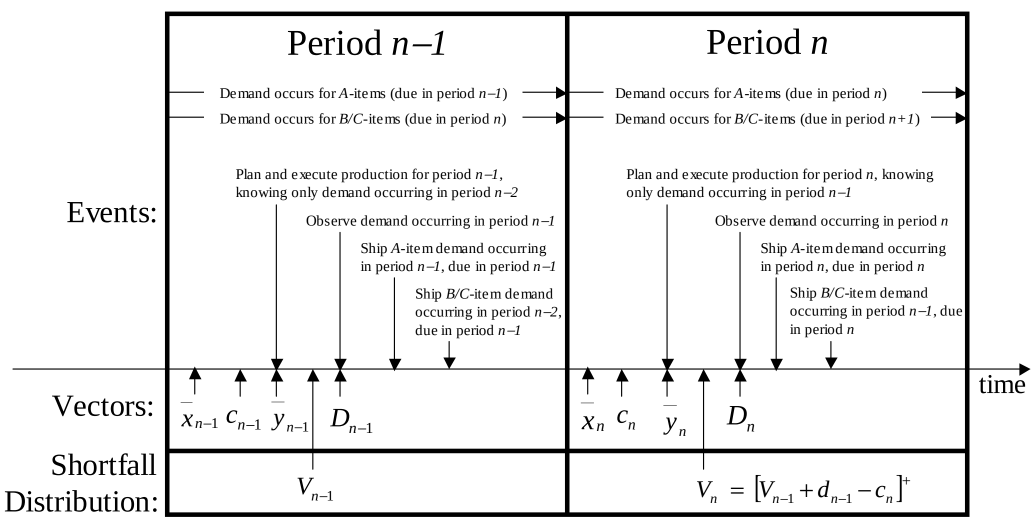

Figure 4. Periodic sequence of events for Model 1 - Production decisions prior to demand observations.

In this case, if the production capacity were infinite, inventories would be held in *A*-type items to minimize the expected holding and backorder costs in the current period.

Our objective is to determine the value of the target aggregate inventory level, T_1, and individual item target levels $\bar{\tau}_1$ that minimizes the expected holding and backorder costs per period over an infinite horizon. For Model 1, the target inventory level, T_1, is defined as the ideal amount of inventory after production but before demand in a period. Any item for which $\tau_{1i} = 0$ is a make-to-order item.

While stating the precise problem is conceptually simple, due to the enormous number of possible states that may exist, solving the associated stochastic dynamic program explicitly is not a practical option. Thus, our strategy will be to construct a very close approximation that is computationally efficient.

4.1 An Exact Formulation

We may state the precise decision problem as,

$$f_n(\bar{x}_n) = E_C\left\{\min_{\bar{y}_n \in R(\bar{x}_n, C)} G(\bar{y}_n) + Ef_{n+1}(\bar{y}_n - \overline{D})\right\}, \tag{4}$$

where

$$R(\bar{x}_n, C) = \{\bar{y} : y_i \geq x_{ni}, \tag{5}$$

$$\sum_{i=1}^{P+1}(y_i - x_{ni}) \leq C\}, \tag{6}$$

is the set of feasible decisions in period *n*. Constraints (5) prevent the intentional removal of inventory, since inventory may be lowered only through demand. We call these the *inventory imbalance* constraints. Constraint (6) limits production to the available capacity in the period.

The results presented in Federgruen and Zipkin (1984a) and in DeCroix and Arreola-Risa (1998) show that a modified base stock policy is optimal when there is capacity constraint and there are no fixed ordering or setup costs. Under this policy, each stocked item has a prescribed target inventory level. Each period, production occurs to raise each item up to its target level, or to get as close as is possible. We will assume that this production policy is used in our approach.

4.2 An Approximation using the Inventory Shortfall Process

Solving (4) each period is not possible due to the enormous number of states and the resulting computational complexity. Instead, we propose using the inventory shortfall process to describe the evolution of the system and solving the following approximation each period. This method is an approximation; it does not necessarily find the optimal production and inventory plan that would be found by solving (4). However, the method is computationally tractable and, as we will demonstrate, is effective in finding good solutions.

4.2.1 The Inventory Shortfall Process

Since capacity is limited, the inventory level may not be restored T_1 after production. We define the deviation from T_1 after production in a period as the inventory shortfall, and denote this quantity as $V_n = T_1 - \sum_{i \in A} y_{ni}$. The inventory shortfall process satisfies the Lindley recursion $V_0 = 0$, $V_n = [V_{n-1} + D - C]^+$. Recall that both D and C are independent, identically distributed random variables and observe that $\{V_n\}_{n>0}$ is a Markov chain. Let V be its stationary r.v., if it exists. Assumption (1) provides necessary and sufficient conditions for the existence of V. From Chan, et al. (1999), the transition matrix for the shortfall process can be described as follows. Let $p_{ij} = \Pr\{V_n = j \mid V_{n-1} = i\}$. Then,

$$p_{ij} = \begin{cases} \sum_{k \geq i} \Pr\{D \leq k-1 \mid C = k\} \cdot \Pr\{C = k\}, & \text{for } j = 0,\ i \geq 0, \\ \sum_{k \geq i-j} \Pr\{D = k + j - i \mid C = k\} \cdot \Pr\{C = k\}, & \text{for } j > 0,\ i \geq 0, \\ 0, & \text{otherwise,} \end{cases} \quad \textbf{(7)}$$

where D represents the random variable for the aggregate demand for capacity over all items in a period, and C is the random variable for capacity. Define $\tilde{C} = C - B$, where B is the aggregate demand for the *B/C*-type items. Since the demand for all items in all periods are independent of each other and that production will first be allocated to *B/C*-type items, we can describe an inventory shortfall process for *A*-type items using the transition probabilities given in (7). Thus,

$\tilde{C}$ is the random capacity available for the production of *A*-type items and $\Pr\{\tilde{C} = k\} = \Pr\{B = C - k\}$. Since $E(D) < E(C)$, we have that $E(A) < E(C) - E(B) = E(\tilde{C})$.

A key fact is that the inventory shortfall process $\{V_n\}_{n>0}$ is independent of the aggregate target inventory level T_1. There are two principle purposes for safety stock in this system. First, safety stock exists to satisfy demand variation in the current period. Since demand is not observed prior to the production decision, a demand forecast is made for each *A*-type item. We refer to this type of safety stock as *demand-driven safety stock*. It exists due to information delays or when customer order lead times are less than manufacturing lead times. Second, safety stock exists to protect the system against insufficient production capacity in previous periods. We refer to this second type of safety stock as *capacity-driven safety stock*.

Instead of solving (4) every period, we use $\{V_n\}_{n>0}$ and propose solving,

$$\min_{\bar{y}_n \in R'(\bar{x}_n, T_1 - V_n)} G(\bar{y}_n) \tag{8}$$

where

$$R'(\bar{x}_n, T_1 - V_n) = \{\bar{y} : y_i \geq x_{ni}, \tag{9}$$

$$\sum_{i \in A} y_i = T_1 - V_n \quad \}. \tag{10}$$

From the inventory imbalance constraints (9), the state of the system, $\bar{x}_n$, at the beginning of period *n* constrains the solution to the production problem. In particular, the set of feasible solutions *R'* is a result of the assumption that a modified base stock policy is followed. The net inventory at the beginning of a period plus the production in a period will not exceed T_1.

4.3 Determining the Target System Inventory Level, T_1

While solving (8) is very effective for making periodic allocation decisions, it cannot be easily used to determine the optimal target inventory level T_1^*, since all possible combinations of beginning inventory levels $\bar{x}_n$ must still be considered. To find T_1^*, we construct an approximation that ignores potential inventory imbalances. We will ignore the allocation of current stock among the items by relaxing the constraints (9). We assume that the inventory is properly allocated among the items so that inventory imbalance is not likely to occur. Imbalance will occur, of course; however, as we will demonstrate, its effect on expected per period cost is not substantial. We assume that state of the system at the beginning of a period is no longer defined by $\bar{x}_n$ and V_n, but rather only by V_n. This results in a drastic simplification in the description of the system evolution. Let,

$$\bar{y}^*(T_1 - V_n) = \arg\min\{G(\bar{y}) : \bar{y} \in R''(T_1 - V_n)\}, \tag{11}$$

where

$$R''(T_1 - V_n) = \left\{ \bar{y} : \sum_{i \in A} y_i = T_1 - V_n \right\}.$$

Equation (11) is simply the vector of optimal production decisions, given a total supply of $T_1 - V_n$ units in period *n*, assuming that inventory imbalances do not occur over time. The resulting expected cost in period *n* is,

$$J(T_1 - V_n) = G\left(\bar{y}^*(T_1 - V_n)\right). \tag{12}$$

This is a lower bound on the true expected cost per period, as any misallocation of inventory among items can only increase this cost. We explore the accuracy of this approximation through a computational experiment in section 6. We appeal to the strong law of large numbers for Markov Chains to establish that the long run average cost per period converges almost surely to its expected value. To determine T_1^* we solve,

$$\min_{T_1} \; E_V\left[J(T_1 - V)\right] = \min_{T_1} \sum_{k \geq 0} J(T_1 - k) \cdot \Pr\{V = k\}, \tag{13}$$

where *V* is the stationary shortfall random variable. As shown in Rappold and Muckstadt (2000), the objective function in (13) is convex and can be solved very quickly.

5 Model 2 – Information Rich Environment

In this section, we alter the periodic sequence of events. We assume that the period's demand is realized prior to production decisions. The sequence of periodic events is shown in Figure 5. At the beginning of a period, $\bar{x}_n$ is the amount of on-hand inventory (or backorders). As before, $x_{ni} = 0, \; \forall i \in B$, since *B/C*-type items receive production priority and $\Pr\{B < C\} = 1$.

Next, demand is realized and a production decision is made. Production occurs first to satisfy *B/C*-type item demand and then to restore *A*-type item inventory levels to their prescribed target inventory levels, $\bar{\tau}_2$.

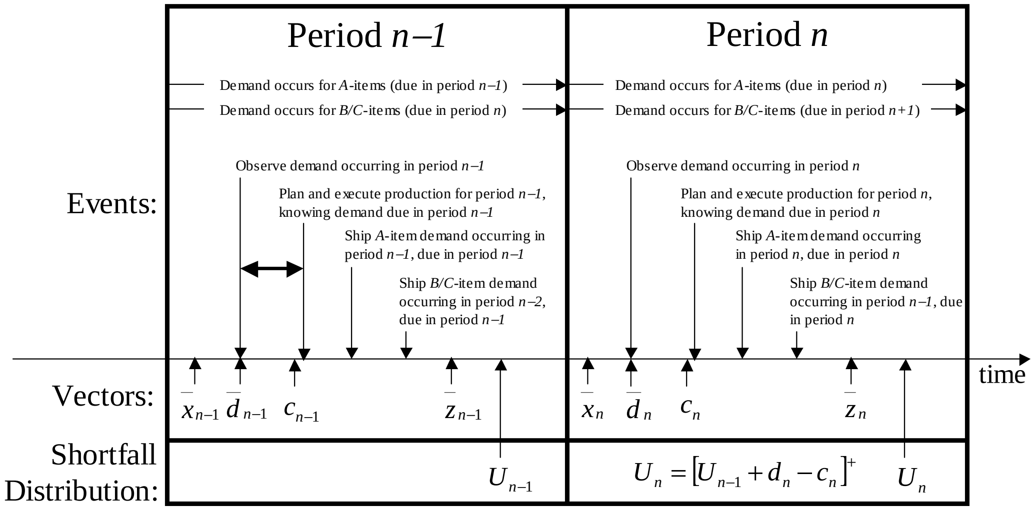

Figure 5. Periodic sequence of events for Model 2 - Demand observed prior to production decisions.

If capacity were unlimited, then the target inventory levels would be zero for all items, since demand could be satisfied in the same period. Equivalently, all items would be make-to-order items, since any item for which $\tau_{2i} = 0$ is a make-to-order item. Therefore, the system target inventory level, T_2, represents solely the *capacity-driven safety stock* in the system. There is no demand-driven safety stock in this case.

We again determine τ_{2i}, $i \in A$, in two steps. First, we determine how much total inventory should be held in the system, across all items. Second, we allocate this aggregate quantity to individual *A*-type items using a novel allocation function.

Let $\bar{x}_n$ and $\bar{z}_n$ be vectors representing net inventory levels (on-hand minus backorders) at the beginning and at the end of period *n*, respectively. Let T_2 denote the aggregate base stock level for Model 2 and τ_{2i} be the base stock level of item *i*, where $\sum_{i \in A} \tau_{2i} = T_2$. We define the one period cost function as,

$$H(\bar{z}_n) = \sum_{i \in A} \left(h_i (z_{ni})^+ + \pi_i (-z_{ni})^+ \right), \tag{14}$$

since the demand for the period is known.

The decision problem is,

$$g_n(\bar{x}_n, \bar{d}_n) = E_C\left\{\min_{\bar{z}_n \in S(\bar{x}_n, \bar{d}_n, C)} H(\bar{z}_n) + Eg_{n+1}(\bar{z}_n, \bar{D})\right\}, \quad \textbf{(15)}$$

where

$$S(\bar{x}_n, \bar{d}_n, C) = \{\bar{z} : z_i \geq x_{ni} - d_{ni}, \quad \textbf{(16)}$$

$$\sum_{i \in A \cup B} (z_i - x_{ni} + d_{ni}) \leq C \;\}. \quad \textbf{(17)}$$

The set of feasible decisions $S(\bar{x}_n, \bar{d}_n, C)$ is a function of the beginning inventory levels, the realized demand in the period, and the production capacity. Constraints (16) are the inventory imbalance constraints. Constraint (17) limits production to the available capacity in the period.

5.1 A New Capacity Allocation Function, $Q(\cdot)$

As before, the state space becomes exceedingly large for reasonable sized problems, rendering (15) unsolvable in practical situations. Rather than solving (15), we will solve an approximation to it. The goal of the dynamic program is to minimize the current period's costs and also produce a mix of items in the proper quantities so that future period's demands can be satisfied while not carrying excessive stock. Thus, an approximation should minimize the sum of the current period's shortage and holding cost plus expected future costs that result from the production decision in this period. When the holding costs among the items are the same, the current production capacity should be allocated to items for which there is likely to be demand that will consume the stock in the very near future. Little, if any, of the current period's capacity should be used to produce items that will not likely be used for many periods in the future.

Let $A_i^{(n)}$ be the random variable representing the cumulative demand for item $i \in A$ over *n* periods and define the following function,

$$Q_i(w) = \sum_{n=1}^{\infty} E\left[w - A_i^{(n)}\right]^+ = \sum_{n=1}^{\infty} \sum_{k=0}^{w-1} F_i^{(n)}(k), \quad \textbf{(18)}$$

as the expected number of inventory-periods associated with stocking *w* units of item *i* in the system. This function was first introduced in Chan, Muckstadt, and Rappold (1999). This function is convex and has a higher value for items that possess a highly variable demand process. When weighted by h_i, the quantity $h_i Q_i(w)$ is a lower bound on all future undiscounted holding costs associated with stocking *w* units of item *i*.

5.2 An Approximation using the Inventory Shortfall Process

Under the definition of the inventory shortfall process described in Figure 5, the shortfall process for Model 1, $\{V_n\}_{n>0}$, is identical to the shortfall process for Model 2, $\{U_n\}_{n>0}$. Thus, the respective stationary distributions are identical. That is $\Pr\{V = k\} = \Pr\{U = k\}$, for $k = 0,1,2,...$

Using $\{U_n\}_{n>0}$, we propose the following approximation to (15),

$$\min_{\bar{z}_n \in S'(\bar{x}_n, \bar{d}_n, T_2 - U_n)} H(\bar{z}_n) + \sum_{i \in A} h_i Q_i(z_{ni}), \tag{19}$$

where the set of feasible solutions is,

$$S'(\bar{x}_n, \bar{d}_n, T_2 - U_n) = \{\bar{z} : z_i \ge x_{ni} - d_{ni}, \quad \sum_{i \in A \cup B} z_i = T_2 - U_n \}. \tag{20}$$

The approximation (19) recognizes the inventory imbalance constraints (20) and makes allocation decisions that minimize the current period's costs and expected future holding costs. It is an approximation since it ignores future expected shortage costs and limits the amount of inventory in the system to T_2, regardless of the precise state of the system. As we shall demonstrate subsequently, this assumption does not have a significant impact on system costs.

5.3 Determining the Target System Inventory Level, T_2

Using (19) to determine a suitable value of T_2 is not possible due to the enormous state space. We again develop an approximation by relaxing the imbalance constraints. We assume the state of the system at the beginning of a period is no longer defined by $\bar{x}_n$ but rather by the value of the inventory shortfall random variable, U_n. We assume that the inventory is properly allocated among the items so that imbalance is not likely to occur.

Let,

$$\bar{z}^*(T_2 - U_n) = \arg\min\left\{H(\bar{z}_n) + \sum_{i\in A} h_i Q_i(z_{ni}) : \bar{z} \in S''(T_2 - U_n)\right\} \quad \textbf{(21)}$$

where,

$$S''(T_2 - U_n) = \left\{\bar{z} : \sum_{i\in A} z_i = T_2 - U_n\right\}.$$

Equation (21) is the vector of optimal production decisions, given a total supply of $T_2 - U_n$ units in period *n*, assuming that inventory imbalances do not occur over time. The resulting expected cost in period *n* is,

$$L(T_2 - U_n) = H\left(\bar{z}^*(T_2 - U_n)\right). \quad \textbf{(22)}$$

To determine a suitable value of T_2 we solve,

$$\min_{T_2} \; E_U\left[L(T_2 - U)\right] = \min_{T_2} \sum_{k\geq 0} L(T_2 - k)\cdot \Pr\{U = k\}, \quad \textbf{(23)}$$

where *U* is the stationary shortfall random variable. The objective function in (23) is convex and can be solved very quickly. Once T_2^* is computed, the individual item target levels may be determined as $\bar{\tau}_2 = \bar{z}^*(T_2^*)$.

In summary, for any level of system inventory $T_2 - U_n$, the vector $\bar{z}^*(T_2 - U_n)$ provides the desired allocation amongst items. Since the term $\sum_{i\in A} h_i Q_i(z_i)$ is not an incurred cost but rather a mechanism for allocating capacity-driven safety stock, the cost associated with having $T_2 - U_n$ units of inventory in the system is given as $L(T_2 - U_n) = H\left(\bar{z}^*(T_2 - U_n)\right)$.

Since T_2^* is only capacity-driven safety stock and the shortfall distributions for both Model 1 and Model 2 are identical, $T_1^* - T_2^*$ is the inventory reduction due to increased information availability. In Model 1, T_1^* is the target inventory prior to demand. The desired ending inventory in a period is $T_1^* - E(D)$. Consequently, $\left(T_1^* - E(D)\right) - T_2^*$ is the safety stock reduction associated with increased information availability.

6 Numerical Study

An extensive simulation study was conducted to measure the accuracy of the proposed approach. Since we assume that inventory imbalance among items can effectively be ignored, one objective in these experiments is to measure the degree to which this assumption caused operating costs to differ from the computed lower bound. Another objective is to provide insight into the value of information in this environment. We also demonstrate the performance of the Q-function over the newsvendor function. In addition to these numerical experiments, we illustrate the effectiveness of the Q-function in an industrial experiment described in section 7.

6.1 The Design of the Experiment

We considered 72 operating environments. In all cases, holding costs were set at $1/unit/period and the backorder costs at $9/unit/period. The operating environments differed in terms of information availability, capacity utilization, demand process variation, number of *A*-type items, and how total *A*-type item demand was distributed among the *A*-type items. The latter two factors contributed no additional insight into the operational dynamics of the system, and so the results that follow do not report these factors explicitly. The critical factors are summarized in Table 1.

The information environments differed in terms of the point in the time period when production planners knew the period's demand. In Model 1, production planners did not have access to the current period's demand and thus had to anticipate demand. We assume that the true demand probability distributions and parameters were known. In Model 2, decision makers had access to demand information before making production decisions that period. In Model 2, we performed each experiment twice, once using the newsvendor allocation function (3) to allocate the available inventories among items, and once using the Q-function (18) for the allocation.

System Factor	**Abbreviation**	**Factor Levels**
Information Availability	*Information*	Model 1, Model 2
Capacity Utilization	*Utilization*	83%, 91%, 95%
Demand Process Variance-to-Mean Ratio	*VTMR*	1.01, 2, 5
Allocation Method (Model 2 only)	*Allocation*	Newsvendor, Q-function

Table 1. Factors under study.

In each of the experiments, we first computed the stationary inventory shortfall distribution by solving the associated Markov chain. Using *V*, we computed T^*, the optimal target system inventory level, and the associated lower bound on the expected cost per period. We then determined the individual item target inventory levels $\overline{\tau}$ by solving (11) and (21).

For each of 10 random seeds, we simulated 2 million cycles of activity, where a cycle is defined to be the time between successive restorations to T^*. Of these cycles, 1 million were generated using the base random number stream, and the remaining 1 million were generated using its antithetic stream. Random demands were generated from negative binomial distributions whose means totaled 100 units per period and variance-to-mean ratios were as indicated in Table 1. The capacity per period was varied to achieve the levels of utilization indicated in Table 1. In periods where there was insufficient capacity to raise the system inventory to T^*, capacity was allocated in accordance with (8) for Model 1, and both (8) and (19) for Model 2. Inventory imbalances arise in the simulation and costs were incurred accordingly. The performance measures of interest are listed in Table 2.

Measure	**Description**
Inventory Levels	The total system inventory level, *T*.
Expected Cost	The expected cost of operating with *T* units of inventory, assuming that no inventory imbalances occur.
Percent Cost Error	The value of $\frac{\lvert ActualCost - ExpectedCost \rvert}{ExpectedCost}$
Average Imbalance	The average number of units in imbalance per period.

Table 2. System performance measures.

The following three subsections describe the performance of our approach in relation to these performance measures.

6.2 Accuracy of the Model

Across all experiments, the actual observed cost differed from the lower bound on expected cost by an average of 0.26%. Indeed, as shown in Figure 6, in all cases the actual operating costs were within 0.53% of the expected cost lower bound. Observe that cost errors tend to increase with VTMR, since higher levels of inventory imbalance are likely to occur as the predictability of item-level demand decreases. In the information-rich environment (Model 2), experimental costs differed from theoretical lower bound costs by 0.26% and in the information-poor environment (Model 1) by 0.14%. This occurred due to the fact that the expected cost per period is significantly lower for Model 2.

Avg Cost Error %		*VTMR*			
Information	*Utilization*	**1.01**	**2**	**5**	**Grand Total**
Model 1	**83.33%**	0.03%	0.04%	0.09%	0.05%
	90.91%	0.02%	0.12%	0.22%	0.12%
	95.24%	0.06%	0.26%	0.44%	0.25%
Model 1 Total		0.03%	0.14%	0.25%	0.14%
Model 2	**83.33%**	0.36%	0.41%	0.39%	0.39%
	90.91%	0.17%	0.42%	0.51%	0.36%
	95.24%	0.19%	0.53%	0.50%	0.41%
Model 2 Total		0.24%	0.46%	0.47%	0.39%
Grand Total		0.14%	0.30%	0.36%	0.26%

Figure 6. Average cost error under different system scenarios.

From these results, we conclude that the inventory imbalance assumption is reasonable and that the approach is quite accurate. We define *inventory imbalance* as the number of units at the end of a period that have been stocked incorrectly. These are units that are either above or below the optimal item-level targets for a given amount of inventory in the system. The intent of the Q-function is to allocate inventory only to those items that have low coefficients of variation, thereby decreasing the number of units in imbalance. Figure 7 shows the average number of units in imbalance in all experiments. Observe that imbalance increases within each model as utilization and demand variation increase. Notice that when imbalance does occur, it is lower for Model 2 than for Model 1 at each level of utilization and demand variation.

Average Units In Inbalance			*VTMR*			
Information	*Allocation*	*Utilization*	**1.01**	**2**	**5**	**Grand Total**
Model 1	**Newsboy**	**83.33%**	0.00	0.01	0.16	0.06
		90.91%	0.00	0.03	0.36	0.13
		95.24%	0.02	0.08	0.61	0.24
Model 1 Total			0.01	0.04	0.38	0.14
Model 2	**Q-Function**	**83.33%**	0.00	0.00	0.02	0.01
		90.91%	0.00	0.00	0.11	0.04
		95.24%	0.00	0.01	0.23	0.08
Model 2 Total			0.00	0.00	0.12	0.04
Grand Total			0.00	0.02	0.25	0.09

Figure 7. Average number of inventory units in imbalance each period.

Avg Cost Error %		VTMR			
Utilization	*Allocation*	**1.01**	**2**	**5**	**Grand Total**
83.33%	*Q-function vs. Newsvendor*	0.00%	0.00%	0.00%	0.00%
90.91%	*Q-function vs. Newsvendor*	0.00%	0.00%	-1.43%	-0.67%
95.24%	*Q-function vs. Newsvendor*	0.00%	0.00%	-2.50%	-1.04%

Figure 8. Change in Avg. Cost Error % when using Q-function instead of Newsvendor allocation for Model 2.

The allocation scheme used in the Model 1 case was the familiar newsvendor method, which, because it considers only a single period, is driven significantly by backorder costs when deciding how much of each item to stock. It is insensitive to the relative amount of variation in the demand patterns of the individual items being stocked over longer periods of time. Model 2 used the Q-function to allocate available inventory, which considers multiple periods into the future. The Q-function stores capacity in the form of inventory that has a higher probability of being consumed in the immediate future. Storing capacity in highly variable items increases the risk of inventory imbalance, and consequently, increases holding and backorder costs. The Q-function therefore creates fewer imbalances than the newsvendor method. For Model 2, Figure 8 shows the change in the cost error when using the Q-function instead of the newsvendor function for allocation decisions. In the industrial experiment discussed in section 7, the demand processes are substantially more variable and the improved performance of the Q-function over the newsvendor function is even more significant.

6.3 Model 1 vs. Model 2: The Value of Information

In this subsection, we examine the performance differences between Models 1 and 2. Figure 9 and Figure 10 contain the values and relative differences, respectively, of the desired period-ending inventory for Models 1 and 2. These levels of safety stock are held to protect the system against uncertainty of two types – demand uncertainty and supply uncertainty. Within each model, inventory increases as VTMR (demand uncertainty) and capacity utilization (supply uncertainty) increase, as one would expect. Between models 1 and 2, the levels of safety stock held are uniformly higher in Model 1 than in Model 2 for a given level of VTMR and capacity utilization, since, in Model 2, demand is known in the current period. In Model 2, there is one period's less demand uncertainty to account for than in Model 1.

Desired Period-Ending Inventory		*VTMR*			
Information	*Utilization*	**1.01**	**2**	**5**	**Grand Total**
Model 1	**83.33%**	11	24	55	30
	90.91%	13	31	74	39
	95.24%	19	45	118	60
Model 1 Total		15	33	82	43
Model 2	**83.33%**	0	0	17	6
	90.91%	5	16	49	23
	95.24%	18	39	107	55
Model 2 Total		8	18	58	28

Figure 9. Desired period-ending inventory levels (safety stock) for different system scenarios.

Relative Difference in Period-Ending Inventory		*VTMR*			
Information	*Utilization*	**1.01**	**2**	**5**	**Grand Total**
Model 1 vs. 2	**83.33%**	-100%	-100%	-69%	-81%
	90.91%	-62%	-48%	-33%	-40%
	95.24%	-5%	-12%	-9%	-9%

Figure 10. Relative change in inventory levels depending on system scenario.

Figure 11 and Figure 12 contain the values and relative differences, respectively, of the expected costs per period for Models 1 and 2. Within each model, the expected cost increases as demand uncertainty and supply uncertainty increase and, between models, the expected costs are uniformly higher in Model 1 than in Model 2 for a given level of VTMR and capacity utilization.

Average of Exp Cost		*VTMR*			
Information	*Utilization*	**1.01**	**2**	**5**	**Grand Total**
Model 1	**83.33%**	$ 43.83	$ 65.94	$ 114.34	$ 74.70
	90.91%	$ 44.47	$ 68.00	$ 123.46	$ 78.64
	95.24%	$ 48.03	$ 77.61	$ 157.34	$ 94.32
Model 1 Total		$ 45.44	$ 70.52	$ 131.71	$ 82.56
Model 2	**83.33%**	$ 1.03	$ 6.81	$ 29.34	$ 12.39
	90.91%	$ 9.83	$ 22.74	$ 61.41	$ 31.33
	95.24%	$ 22.82	$ 46.99	$ 120.45	$ 63.42
Model 2 Total		$ 11.23	$ 25.51	$ 70.40	$ 35.71
Grand Total		$ 28.33	$ 48.01	$ 101.06	$ 59.13

Figure 11. The expected holding and backorder costs per period.

Relative Difference in Expected Costs		*VTMR*			
Information	*Utilization*	**1.01**	**2**	**5**	**Grand Total**
Model 1 vs. 2	**83.33%**	-98%	-90%	-74%	83%
	90.91%	-78%	-67%	-50%	-60%
	95.24%	-52%	-39%	-23%	-33%

Figure 12. The value of information: relative change in the expected cost per period between Models 1 and 2.

Figure 13 and Figure 14 contain the rates of change in expected costs for each model as supply and demand uncertainty increase, respectively. Observe that the rate of change in expected cost as both VTMR and capacity utilization increase is greater for Model 2 than for Model 1. That is, the expected cost per period is more sensitive to VTMR and capacity utilization when operating in the leaner, information rich environment.

Change in Cost as Utilization Increases		*VTMR*		
Information	*Utilization*	**1.01**	**2**	**5**
Model 1	**83.33%**	Base	Base	Base
	90.91%	1%	3%	8%
	95.24%	10%	18%	38%
Model 2	**83.33%**	Base	Base	Base
	90.91%	859%	234%	109%
	95.24%	2126%	590%	311%

Figure 13. Rates of change in expected cost as supply uncertainty increases

Change in Cost as VTMR Increases		*VTMR*		
Information	*Utilization*	**1.01**	**2**	**5**
Model 1	**83.33%**	Base	50%	161%
	90.91%	Base	53%	178%
	95.24%	Base	62%	228%
Model 2	**83.33%**	Base	564%	2761%
	90.91%	Base	131%	525%
	95.24%	Base	106%	428%

Figure 14. Rates of change in expected cost as demand uncertainty increases

This observation has important consequences for supply chain designers attempting to reduce costs by improving the downstream visibility of demand information. Improving the richness of the available information decreases costs, but also increases the system's sensitivity to capacity utilization and aggregate demand uncertainty. This is relevant since the current state-of-practice in production planning and control systems is, at worst, to ignore the effects of capacity constraints entirely and, at best, to capture capacity limits by simple heuristics such as fixed manufacturing lead times. We explore the impact of such practices in the next subsection.

6.4 The Impact of Ignoring Capacity Constraints

Much has been said and written recently about the importance of designing supply chains to provide the kind of information visibility present in Model 2. For example, see Muckstadt, et al. (2001). The results in the previous subsection confirm and quantify these ideas. In this subsection, we reinforce the importance of considering limited capacity in production planning decision-making, and observe that the advantages of increased information availability can deteriorate rapidly if finite capacity is ignored. Figure 15 shows the expected cost per period when ignoring capacity constraints. To construct the *Exp Cost Ignoring Capacity* values, capacity was assumed to be unlimited, thereby ignoring any supply uncertainty. Figure 16 contains the relative differences in expected costs.

			VTMR			
Information	*Utilization*	Data	**1.01**	**2**	**5**	**Grand Total**
Model 1	**83.33%**	Average of Exp Cost	$ 43.83	$ 65.94	$ 114.34	$ 74.70
		Exp Cost Ignoring Capacity	$ 43.83	$ 66.01	$ 115.00	$ 74.95
	90.91%	Average of Exp Cost	$ 44.47	$ 68.00	$ 123.46	$ 78.64
		Exp Cost Ignoring Capacity	$ 44.59	$ 69.29	$ 131.15	$ 81.68
	95.24%	Average of Exp Cost	$ 48.03	$ 77.61	$ 157.34	$ 94.32
		Exp Cost Ignoring Capacity	$ 50.23	$ 88.61	$ 212.85	$ 117.23
Model 1 Average of Exp Cost			$ 45.44	$ 70.52	$ 131.71	$ 82.56
Model 1 Exp Cost Ignoring Capacity			$ 46.22	$ 74.64	$ 153.00	$ 91.29
Model 2	**83.33%**	Average of Exp Cost	$ 1.03	$ 6.81	$ 29.34	$ 12.39
		Exp Cost Ignoring Capacity	$ 1.03	$ 6.81	$ 40.28	$ 16.04
	90.91%	Average of Exp Cost	$ 9.83	$ 22.74	$ 61.41	$ 31.33
		Exp Cost Ignoring Capacity	$ 12.25	$ 38.36	$ 138.13	$ 62.91
	95.24%	Average of Exp Cost	$ 22.82	$ 46.99	$ 120.45	$ 63.42
		Exp Cost Ignoring Capacity	$ 49.57	$ 119.90	$ 354.72	$ 174.73
Model 2 Average of Exp Cost			$ 11.23	$ 25.51	$ 70.40	$ 35.71
Model 2 Exp Cost Ignoring Capacity			$ 20.95	$ 55.02	$ 177.71	$ 84.56
Total Average of Exp Cost			$ 28.33	$ 48.01	$ 101.06	$ 59.13
Total Exp Cost Ignoring Capacity			$ 33.58	$ 64.83	$ 165.36	$ 87.92

Figure 15. Expected costs of considering finite capacity vs. ignoring finite capacity.

			VTMR			
Information	*Utilization*		**1.01**	**2**	**5**	**Grand Total**
Model 1	**83.33%**	*Change in Exp Cost*	0.002%	0.105%	0.574%	0.324%
	90.91%	*Change in Exp Cost*	0.260%	1.897%	6.235%	3.858%
	95.24%	*Change in Exp Cost*	4.599%	14.186%	35.281%	24.289%
		Model 1 Change in Exp Cost	1.706%	5.846%	16.163%	10.573%
Model 2	**83.33%**	*Change in Exp Cost*	0.000%	0.000%	37.305%	29.440%
	90.91%	*Change in Exp Cost*	24.664%	68.697%	124.922%	100.833%
	95.24%	*Change in Exp Cost*	117.202%	155.159%	194.504%	175.514%
		Model 2 Change in Exp Cost	86.624%	115.666%	152.435%	136.784%
Total Change in Exp Cost			18.528%	35.023%	63.628%	48.683%

Figure 16. Relative change in the expected cost per period when ignoring finite capacity.

In all but 2 cases, ignoring the presence of capacity constraints resulted in higher expected costs. As the VTMR and capacity utilization increase, the magnitude of the expected cost differences increases rapidly for both information models. The magnitude of the changes in expected costs are substantially greater for Model 2 than for Model 1. For example, at a utilization level of 95.24% and VTMR of 5, ignoring capacity constraints in the information-poor environment increases expected costs by 35.281%. In the information-rich environment, expected costs increase by 194.504%.

This suggests that as we attempt to extract value from more tightly-coupled information environments, it becomes critically important to consider the effects of capacity constraints explicitly in our decision models. A true understanding of the economics of information can be developed only by simultaneously considering the combined effects of reduced demand uncertainty and increased sensitivity to supply uncertainty.

7 Industrial Implementation

In this section, we test our approach using data from an industrial environment. The purpose of conducting this industrial experiment is to illustrate the performance of our approach compared to (1) current management practices and (2) the newsvendor allocation function, when the demand processes are extremely volatile. Whereas the numerical experiments of section 6 used demand processes with a VTMR of less than 5, the industrial demand data possess an aggregate VTMR of 716. We begin with a short description of the current operational environment, current management practices, and current system performance levels. Then, using the real demand data, we simulate the daily operation of the system under three policies:

1. Current operating policies with current item target level inventories;
2. No B/C Policy using the newsvendor allocation function; and
3. No B/C Policy using the Q-function.

7.1 Background

In this example, we consider a product family of 30 consumable industrial products that are assembled in a manufacturing cell. This cell has gone through extensive reengineering efforts over the past 5 years. Flow times through the cell are short, predictable and are on the order of minutes. The cell has a planned capacity of 904 units per day. The per-unit processing times across the items are approximately the same and setup times between items are very small.

Production and inventory are controlled using common reorder point logic. Each item has a reorder point that is either computed periodically in a system or is established manually by a production planner. Finite production capacity is not taken into account explicitly when computing reorder points. Instead, a deterministic lead-time is used as a surrogate for finite capacity. Reorder points are calculated as the forecasted demand for an item over some fixed replenishment lead-time plus some safety stock. When the inventory position (on-hand plus on-order minus backorders) for an item falls at or beneath its reorder point, a manufacturing order is released to the floor. Each day, a work list is generated by a planning system detailing which items to produce. The work list is created from a mix of backordered items and a list of items that are below their reorder points. The cell team leader decides on a production sequence.

Orders arrive at the cell each day and are known at the start of the shift. In order to meet promised shipment dates, items must either be produced that day or be satisfied from inventory. A Pareto chart of item-level demand is shown in Figure 17. The demand for the Top 7 items constitutes 84% of the total time (capacity) demanded. Observe that each item demands a capacity at a different rate. Each item also has different cost attributes.

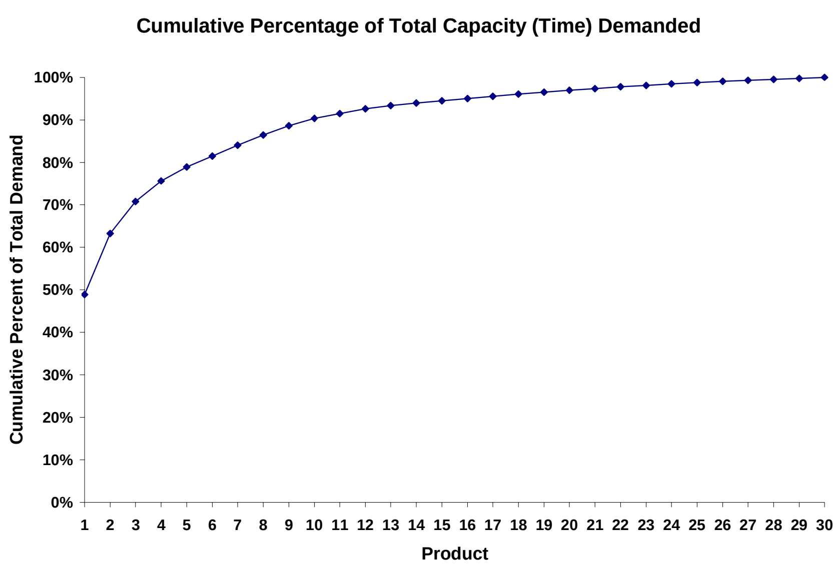

Figure 17. Pareto chart of the top 30 products in the product family

The aggregate demand time series across all items is shown in Figure 18. The average daily demand is 801 units per day, the demand uncertainty, as measured by its standard deviation, is 829 units, and the coefficient of variation is 1.03. The capacity utilization is 89%. Demand frequently exceeds capacity. The system protects customer service by either adding inventory, much of which may never be sold, or by using costly overtime production.

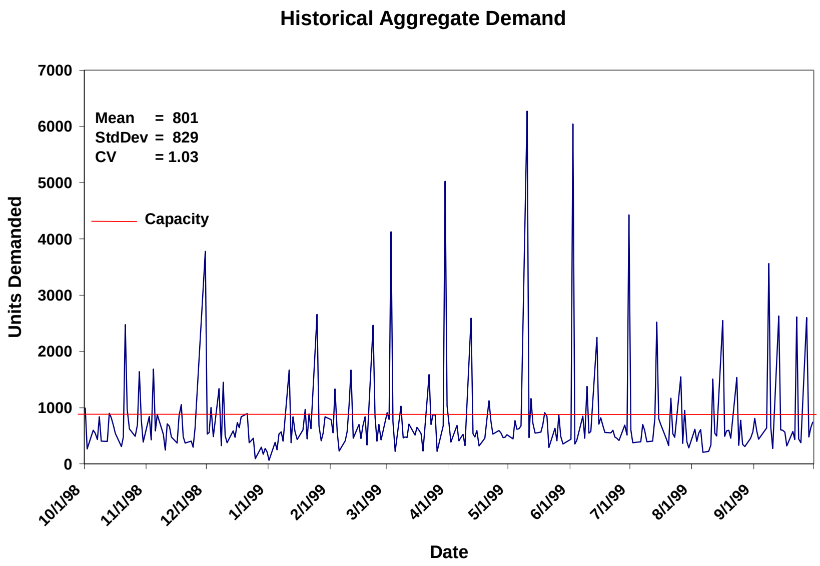

Figure 18. Time series of the total demand for capacity in the facility

The high degree of uncertainty for the aggregate demand process makes accurate forecasting very difficult for the total demand, let alone for individual items. As a result, large inventories are often created and remain in storage for long durations of time. Moreover, when an order does arrive for an item, there is often insufficient inventory to satisfy the order. Consequently, the use of overtime production is frequent.

For example, Figure 19 and Figure 20 contain the demand time series and the current target inventory levels for Products 1 and 14, respectively. For Product 1, the target inventory level was set manually as a result of the inability to generate accurate forecasts. Note that the current target level is set high enough to satisfy the large, periodic spikes in demand. For Product 14, the inventory level is set to 60 units by the planning system. This is approximately 12.8 days of average demand and would have been sufficient to achieve only a 10.1% fill rate. The service level objective for this cell is a 93% fill rate. Under the current operating policy, the firm stocks $182,000 worth of finished goods inventory across 30 items in a central storage facility and reports an 87% average fill rate.

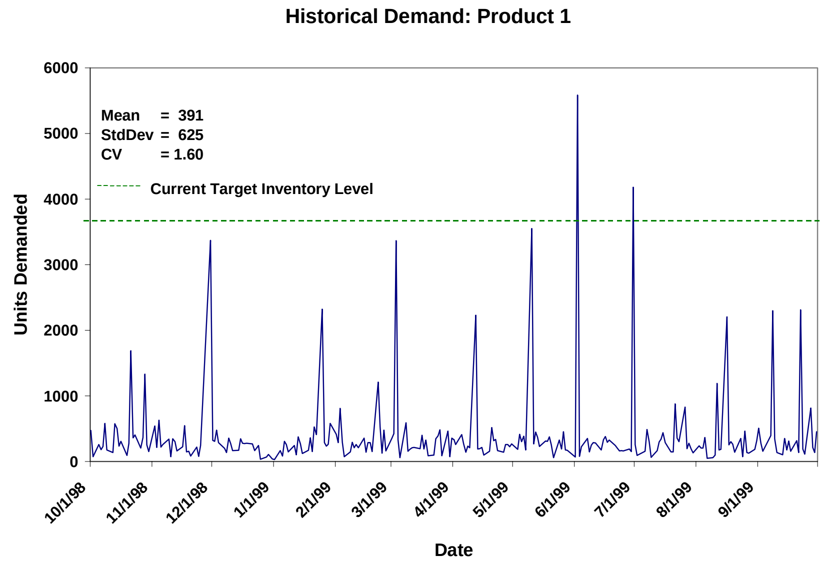

Figure 19. Time series of demand for the highest volume product

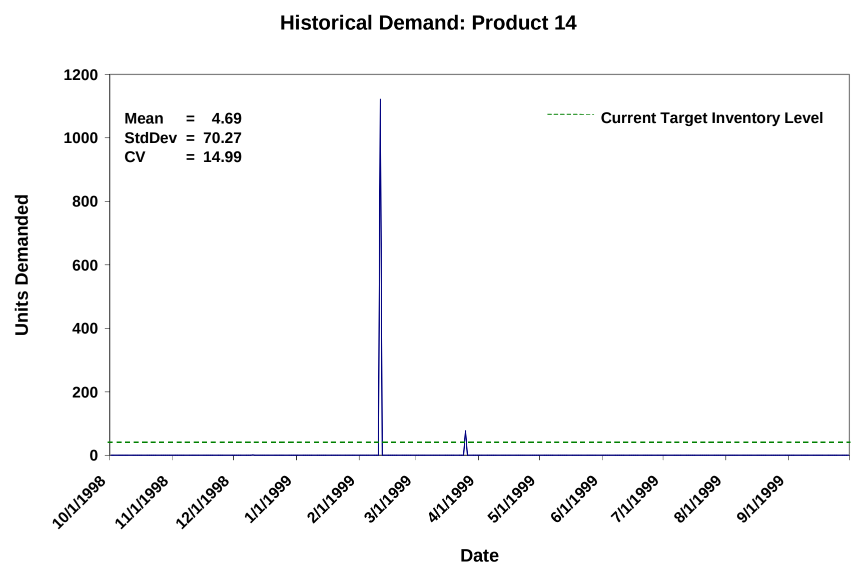

Figure 20. Time series of demand for Product 14

7.2 The Experiment

After examining the demand for each item, we designate the Top 7 products as *A*-type items and the remaining items as *B/C*-type items. The aggregate demand for the *A*-type items is shown in Figure 21 and the aggregate demand for the *B/C*-type items is shown in Figure 22.

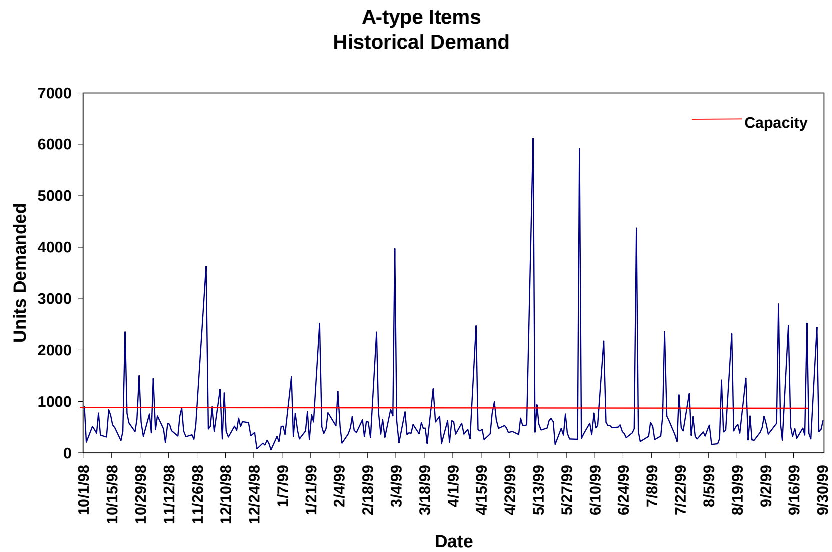

Figure 21. Aggregate demand for Top 7 (A-type) items.

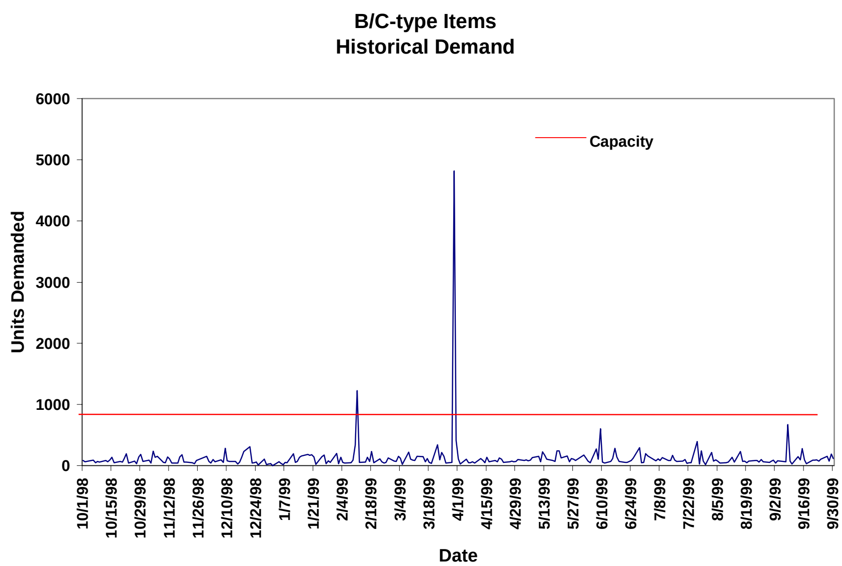

Figure 22. Aggregate demand for B/C-type items.

Observe that while the demand for the *A*-type items often exceeded the daily production capacity, the demand for *B/C*-type items only exceeded the daily production capacity on two days over the course of one year. Usually the production facility has sufficient capacity to produce all of the requirements for the *B/C*-type items, which will receive production priority on a daily basis. Any remaining production capacity will be used to produce the *A*-type items.

The data used in this experiment for the 30 items are summarized in Figure 23. Based on selling margins and discussions with management, backorder costs were estimated at 25 times their respective unit holding costs. We will use the same target system inventory level of 7,039 units. For the three operating policies of interest, item-level target inventory levels are shown.

								Target Inventory Levels		
Product	Holding Cost	Backorder Cost	Demand Mean	Demand Variance	Demand StdDev	VTMR	CV	Current Policy	No B/C - Newsvendor	No B/C - Q-function
1	$ 0.022	$ 0.557	391.4	391,147.1	625.4	999.3	1.60	**3,600**	**2,217**	**4,190**
2	$ 0.030	$ 0.738	115.1	74,614.6	273.2	648.2	2.37	**1,260**	**864**	**713**
3	$ 0.020	$ 0.504	60.3	1,922.8	43.9	31.9	0.73	**480**	**174**	**785**
4	$ 0.029	$ 0.719	38.9	869.9	29.5	22.3	0.76	**252**	**112**	**345**
5	$ 0.008	$ 0.203	26.5	1,794.5	42.4	67.7	1.60	**120**	**3,372**	**863**
6	$ 0.042	$ 1.047	20.3	2,681.5	51.8	132.1	2.55	**120**	**152**	**65**
7	$ 0.039	$ 0.967	20.1	2,456.7	49.6	122.2	2.46	**144**	**148**	**78**
8	$ 0.027	$ 0.681	19.5	88,570.0	297.6	4,535.7	15.24	**156**	-	-
9	$ 0.024	$ 0.594	17.5	1,605.5	40.1	91.5	2.28	**120**	-	-
10	$ 0.062	$ 1.545	14.1	253.8	15.9	18.0	1.13	**84**	-	-
11	$ 0.006	$ 0.147	9.2	107.3	10.4	11.7	1.13	**35**	-	-
12	$ 0.011	$ 0.284	8.6	395.9	19.9	46.0	2.31	**40**	-	-
13	$ 0.072	$ 1.794	6.0	158.8	12.6	26.5	2.10	**42**	-	-
14	$ 0.060	$ 1.493	4.7	4,938.0	70.3	1,053.4	14.99	**60**	-	-
15	$ 0.058	$ 1.440	4.4	62.7	7.9	14.3	1.81	**50**	-	-
16	$ 0.018	$ 0.455	4.4	110.5	10.5	25.2	2.40	**108**	-	-
17	$ 0.056	$ 1.393	4.1	39.8	6.3	9.7	1.53	**36**	-	-
18	$ 0.025	$ 0.619	4.1	201.6	14.2	49.1	3.46	**48**	-	-
19	$ 0.023	$ 0.566	3.9	220.4	14.8	56.8	3.82	**36**	-	-
20	$ 0.037	$ 0.917	3.4	59.5	7.7	17.7	2.30	**34**	-	-
21	$ 0.019	$ 0.487	3.3	1,070.8	32.7	322.5	9.86	**25**	-	-
22	$ 0.026	$ 0.656	3.1	90.1	9.5	28.6	3.01	**42**	-	-
23	$ 0.070	$ 1.741	3.0	101.6	10.1	34.1	3.39	**20**	-	-
24	$ 0.036	$ 0.912	2.5	13.4	3.7	5.3	1.46	**11**	-	-
25	$ 0.032	$ 0.812	2.4	14.3	3.8	5.9	1.56	**20**	-	-
26	$ 0.009	$ 0.228	2.3	183.4	13.5	80.1	5.92	**15**	-	-
27	$ 0.048	$ 1.203	2.1	147.9	12.2	71.7	5.90	**15**	-	-
28	$ 0.022	$ 0.548	2.0	27.5	5.2	14.0	2.67	**12**	-	-
29	$ 0.022	$ 0.555	1.9	7.8	2.8	4.0	1.45	**30**	-	-
30	$ 0.023	$ 0.583	1.9	24.9	5.0	13.3	2.66	**24**	-	-
Total			801.0	573,892.6	757.6	716.4	0.95	**7,039**	**7,039**	**7,039**

Figure 23. Data for the industrial experiment.

We simulated the three operating polices with the various inventory allocations and obtained the results in Figure 24. The simulation of the current policy resulted in a fill rate of 78% and 1,518 units in imbalance on average, or 22% of the total inventory. We attribute the difference in the simulated fill rate of 78% and the actual reported fill rate of 87% to the use of overtime and other changes to the system, such as outsourcing and negotiating with customers. The 78% fill rate serves as a benchmark to which we may compare the other operating policies.

Simulating the No B/C policy using the newsvendor function for inventory allocation decisions resulted in an improved fill rate of 89%. Notice, however, that 15% of the inventory, or roughly 1,061 units each day were imbalanced. This is due to the way in which the newsvendor function determined the item-level allocations, shown in Figure 23. Notice that 3,372 units were held in item 5. The newsvendor function made this decision because item 5 has the lowest unit holding cost.

Simulating the No B/C policy using the Q-function for the item-level allocation decisions resulted in a fill rate of 95% and only 2.2%, or 153 units on average were in imbalance. Instead of holding 3,372 units in item 5, the Q-function selected 4,190 units of item 1 because of its higher probability of being demanded.

Operating Policy	Number of Stocked Items	System Inventory Level (T)	Allocation Used Among Items	Fill Rate Achieved	Average Number of Units in Imbalance	Percent of Total Units in Imbalance
Current Policy	30	7,039 *	Newsvendor	77.96%	1,517.8	21.6%
No B/C Policy using Newsvendor	7	7,039	Newsvendor	89.26%	1,061.4	15.1%
No B/C Policy using Q-function	7	7,039	Q-function	94.73%	152.7	2.2%

* the current item-level target inventory levels were used.

Figure 24. Results of applying the No B/C Policy to an industrial environment.

8 Conclusion

In this paper, we develop a computationally efficient approach for setting base stock levels in a capacitated manufacturing environment in which forecasts for the majority of items are not available. We quantify the value of obtaining advanced demand information for decision-making. The value of this information, while consistently valuable, diminishes as the capacity utilization increases. We illustrate the implications operating the environment without explicit consideration of the finite capacity. We demonstrate the obtaining advanced demand information combined with inadequate capacity planning can result in substantial deviations from optimality.

Bibliography

Carr, S.A., R.A. Güllü, P.L. Jackson, and J.A. Muckstadt, *"Exact Analysis of the No B/C Stock Policy,"* Tech. Report 1051, School of OR&IE, Cornell University, Ithaca, NY, 1993.

Chan, E.W., J.A. Muckstadt, and J.A. Rappold, *"Determining and Allocating Capacity-Driven Safety Stock in Multi-Item, Multi-Echelon Systems,"* School of Business, University of Wisconsin, Madison, WI, 1999.

Ciarallo, F., R. Akella, and T.E. Morton, *"A Periodic Review, Production-Planning Model with Uncertain Capacity,"* Management Science, 40(3), 320-332, 1994.

DeCroix, G.A. and A. Arreola-Risa, *"Optimal Production and Inventory Policy for Multiple Products Under Resource Constraints,"* Management Science, 44(7), 950-961, 1998.

Evans, R., *"Inventory Control of a Multi-Product System with a Limited Production Resource,"* Naval Research Logistics, 14(2), 173-184, 1967.

Federgruen, A. and Z. Katalan, *"The Stochastic Economic Lot Scheduling Problem: Cyclic Base-Stock Policies withy Idle Times,"* Management Science, 1999.*

Federgruen, A. and P. Zipkin, *"An Inventory Model with Limited Production Capacity and Uncertain Demands I: The Average Cost Criterion,"* Mathematics of Operations Research, 11(2), 193-207, 1986.

Glasserman, P., *"Allocating Production Capacity Among Multiple Products,"* Operations Research, 44(5), 724-734, 1996.

Glasserman, P., *"Bounds and Asymptotics for Planning Critical Safety Stocks,"* Operations Research, 45(2), 1997.

Glasserman, P. and S. Tayur, *"The stability of a capacitated, multi-echelon production-inventory system under a base-stock policy,"* Operations Research, 42(5), 913-925, 1994.

Güllü, R. and P.L. Jackson, *"On the continuous time capacitated production/inventory problem with no setup costs,"* Technical Report No. 1054, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY, 1993.

Muckstadt, J. *"A Paradigm Lost"* Technical Report #1180, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY, 1997.

Muckstadt, J.A., D.H. Murray, J.A. Rappold, and D.E. Collins, *"Guidelines for Collaborative Supply Chain System Design and Operation,"* Journal of Information Systems Frontiers, forthcoming, 2001.

Rappold, J. and J. Muckstadt, *"A Computationally Efficient Approach for Determining Inventory Levels in a Capacitated Multiechelon Production-Distribution System"* Naval Research Logistics, 47(5), 377-398, 2000.

Roundy, R.O. and J.A. Muckstadt, *"Heuristic Computation of Periodic-Review Base Stock Policies,"* Management Science, 1999. *

Silver, E.A., D.F. Pyke and R. Peterson, Inventory Management and Production Planning and Scheduling – 3rd ed., Wiley, New York, 1998.

Sox, C., L.J. Thomas, and J.O. McClain, *"Coordinating Production and Inventory to Improve Service,"* Management Science, 43(9), 1189-1197, 1997.

Tayur, S. *"Computing Order-up-to Levels in Capacitated Environments,"* Stochastic Models, 9, 585-598, 1992.

Williams, T., *"Special Products and Uncertainty in Production/Inventory Systems,"* European Journal of Operations Research, 15, 46-54, 1984.

Acknowledgements

This research was partially funded by the National Science Foundation (Grant DMI0075627) and by Aspen Technology.