Capacitated Multi-Item Inventory Systems with Random and Seasonally Fluctuating Demands: Implications for Postponement Strategies

Yossi Aviv • Awi Federgruen
Olin School of Business, Washington University, St. Louis, Missouri 63130
Columbia Graduate School of Business, New York, New York 10027

We address multi-item inventory systems with random and seasonally fluctuating, and possibly correlated, demands. The items are produced in two stages, each with its own lead-time; in the first stage a common intermediate product is manufactured. The production volumes in the first stage are bounded by given capacity limits. We develop an accurate lower bound and close-to-optimal heuristic strategies of simple structure. The gap between them, evaluated in an extensive numerical study, is on average only 0.45%. We use the model to investigate the benefits of various delayed product differentiation (postponement) strategies, as well as other strategic questions, including (i) the benefits of flexible versus dedicated production facilities; (ii) the trade-off between capacity and inventory investments; and (iii) the trade-off between capacity investments and service levels.

(*Multi-Item; Multi-Echelon; Inventory Model; Markov Decision Process; Dynamic Programming; Design for Postponement; Capacity*)

Introduction and Summary

Multi-item production/inventory systems continue to present major challenges that result from the compounding effects of several complicating factors. First, demands for the items fluctuate, partially because of predictable and periodically varying causes and partially because of intrinsic uncertainties that cannot be attributed to identifiable or predictable exogenous factors. Determining appropriate safety stocks as a hedge against the random demand components is especially difficult under nonstationary parameters and in the presence of capacity limits requiring inventory buildups to meet peak demands.

The above factors represent a challenge even for a single item. In addition, it is usually necessary to address a complete family of items via an integrated model, e.g., because items share critical bottleneck resources or have correlated demands. Finally, the products in the family are often differentiated by a limited set of features. This can be exploited by a *design for postponement* strategy in which a common intermediate product is manufactured in a first phase, with the differentiating options and features postponed until a second phase.

Take, for example, one of the cases most widely used in operations management classes to demonstrate the benefits of postponement, i.e., the redesign of the European DeskJet Printer line of Hewlett Packard; see Kopczak and Lee (1993), reprinted, e.g., in Flaherty (1996). In the early 1990s, HP Europe sold six different printers in this line in as many distinct markets, differentiated in terms of the power supply module, power cord terminator, and manual.

0025-1909/01/4704/0512$5.00
1526-5501 electronic ISSN

Traditionally, localization took place in the Vancouver plant (*factory localization*). The case demonstrates the benefits of postponing localization until the printers arrive at the European distribution center (*DC localization*). The analysis of the case (see, e.g., the teaching note by Flaherty et al. 1996) assumes that (i) the localization process is instantaneous; (ii) the capacity in the Vancouver plant is well in excess of demands, even in the peak months; (iii) the monthly fluctuations are i.i.d. across time, ignoring seasonal patterns; and (iv) sales of the different items (in the different European markets) are independent of each other.

In reality, each of the above assumptions (i)–(iv) may be violated. The localization process in the DC requires approximately one week. It is therefore more appropriate to view the system as a two-stage process. Second, as the sales volumes continued to rise, the Vancouver plant capacity turned into a possible bottleneck, especially in peak months. Third, it is well known that the industry experiences several seasonal peaks, in particular at the end of the summer (the back-to-school period) and at the end of the year (because of holiday shopping and year-end budget considerations). These peaks are clearly reflected in the case's sample data. As a consequence, a significant part of the month-to-month sales variations is due to seasonal patterns as opposed to representing noise around a constant year-round mean. Finally, sales in neighboring European markets are correlated as well.

The HP case and the surrounding articles have stirred much interest in postponement among researchers and practitioners. This paper looks deeper at the benefits of postponement, seeking to understand the environments in which postponement will yield major cost reductions. We focus on factors (i)–(iv) listed above. The specifics are illustrated with the HP case and, because no existing tractable model appears to address the combination of the four factors (i)–(iv), an analytical model is built and analyzed to provide a tool for estimating the savings in other settings. The analysis also results in easily implementable and close-to-optimal strategies for two-stage production-distribution processes of the above type.

The first major consideration is the extent to which seasonality exists and can be forecasted. The data in the HP case are limited to a single year and are therefore insufficient to arrive at a proper statistical separation between seasonal fluctuations and intrinsic randomness around a predictable seasonal mean demand curve. The purpose of our numerical investigations is to illustrate that the magnitude of the benefits of postponement can be quite different depending on what part of the monthly sales fluctuations is due to the seasonality pattern and what part is intrinsic randomness. For example, we estimate that if there is no seasonality, then postponement can reduce total costs per month by about 16%. If very strong seasonality exists (and can be correctly forecasted), then the basic costs without postponement can be reduced by about 32%. An additional 19% can be saved (on the optimal basic costs, assuming seasonality) through postponement. (These are compounded savings, so the total savings from recognizing the seasonality and implementing postponement is about 45%.) If postponement is implemented but very strong seasonality is not recognized, so that a stationary policy is used, then postponement saves only about 13%. This example shows that the *relative* benefits due to postponement (i.e., the percentage cost savings) can be severely underestimated (by as much as 50%) when failing to detect and account for seasonality patterns. Moreover, the cost performance under postponement is 37% lower when the proper seasonality pattern can be forecasted and accounted for, as compared to when such seasonality patterns are ignored.

The analysis so far assumes that there is sufficient factory production capacity at all times. As capacity becomes more limited, the benefits of postponement are reduced. For example, if mean weekly demand equals 90% of factory capacity, postponement reduces costs by 7% when there is no seasonality and about 10% under very strong seasonality. There is less to be saved when capacity is limited because the factory has fewer options and must produce nearly at capacity most of or all of the time, regardless of the demand stream.

Next, the analysis up to this point assumes that demands for the final products are independent of each other. The presence of correlations between the demand processes has a significant impact on the magnitude of the benefits of delayed product differentiation. Assume, for example, that the correlations between all pairs of items are identical. If mean weekly demand equals 80% of factory capacity, and in the presence of moderate seasonality, postponement reduces the basic costs by (i) 16% when product demands are *independent*, (ii) 6% if the pairwise correlation is $\rho = 0.5$, and (iii) 23% and 33% when $\rho = -0.25$ and -0.5, respectively. In fact, the cost improvement *decreases* monotonically with the correlation ρ. This is explained by the fact that the benefits of postponement arise from the ability to pool the risks associated with individual patterns during the first procurement stage preceding the point of differentiation; the larger the correlation between the demands, the smaller the impact of risk pooling.

Finally, in many (re)design processes of production/distribution systems, various options may prevail to select the points of differentiation instead of focusing on two *extreme* designs, immediate or *factory-localization* and postponement to the very end of the process as in *distribution center* (*DC*)-*localization*. It is important to understand how the benefits of postponement grow as the point of differentiation is varied. For example, if the mean weekly demand equals 80% of factory capacity, under uncorrelated demands and moderate seasonality, differentiating the products after the 1st (2nd, 3rd, 4th) week of a five-week total lead time reduces costs by 4% (9%, 14%, 20%) when compared to the case of no product differentiation.

Most of the literature on stochastic inventory systems with capacity constraints deals with a single final product; see Ciarallo et al. (1994) and, for models with periodically varying parameters, Aviv and Federgruen (1997a) and Kapuscinski and Tayur (1998) and the references therein. Song and Zipkin (1993), Cheng and Sethi (1995), and Beyer and Sethi (1997) deal with the uncapacitated single-item model, in which the parameters fluctuate as a function of an underlying Markov chain. (The periodic structure treated here can be modeled within this framework, with the period type as the state of the Markov chain.)

Evans (1967) is the first to address a version of the multi-item model. He characterizes the structure of an optimal policy in the case of two items, lost sales, and stationary data. Metters (1998) presents several heuristics for the special case of the model examined here, in which both manufacturing stages are instantaneous and without characterization or measurement of the associated optimality gaps (other than for two-item models). His two basic heuristics determine a base-stock level for each period as the period's newsboy solution, i.e., disregarding capacity limits and periodic fluctuations. Actual production volumes are determined by a single-stage linear program in which demands are replaced by their means. Our general approach in developing lower bounds and heuristic strategies bears similarity to that employed in Eppen and Schrage (1981), Federgruen and Zipkin (1984a, 1984b), and Chen and Zheng (1994) for uncapacitated stationary models (but possibly more complex production costs).

In addition, our paper contributes to the literature on *delayed product differentiation* strategies. Lee and Tang (1997) recently provided a classification of possible design changes in the production and distribution processes that result in delayed product differentiation, along with a few analytical models to evaluate some of these design changes. One alternative mechanism to enhance benefits of delayed product differentiation is to resequence various differentiating manufacturing operations. In the apparel industry, for example, families of garments are often differentiated by color and style/size combinations; a knitting operation specifies the style/size combination, while a dyeing operation determines the color. Lee and Tang (1998) initiated an analytical model to determine which sequence of operations results in optimal operational performance; see Kapuscinski and Tayur (1999) and Federgruen (1999) for refinements of the analysis and Aviv and Federgruen (1998) and Garg and Lee (1998) for recent survey chapters.

In §1 we specify the model and the required notation. As is the case for almost all multi-item inventory models, an exact analysis is intractable. We first (§2) develop a lower-bound approximation and heuristic strategies for the case of a single-stage production process, i.e., settings where products need to be differentiated from the onset. In §3 we extend our bounds and heuristic strategies for the general case where product differentiation can be postponed and production occurs in two stages. In §4 we report on a numerical study of synthetic problem instances carefully designed to gauge the accuracy of the lower bounds and the optimality gap of the heuristic strategies, as well as to investigate the impact of several system parameters on various performance measures. The numerical study is also used to provide insights into a number of fundamental design questions. In addition to (i) the benefits of a design for postponement strategy, these include, characterization of (ii) the benefits of flexible versus dedicated production facilities and (iii) the trade-offs between capacity, inventory investments, and service level. In §5 we revisit the HP DeskJet Printer supply-chain case, addressing all of the abovementioned complications. In §6, we develop extensions of our basic model to settings where the intermediate (undifferentiated) product can be kept in stock. This section also contains a brief discussion of settings with unknown parameters for the demand distributions and intertemporal correlations.

1. A Basic Model

A company produces and sells J products. Inventories are monitored periodically. Demands in each period follow a given multivariate distribution with arbitrary correlations between items. Initially we assume, as in virtually all inventory models, that demands in different periods are independent and their distributions are perfectly known.

Production occurs in up to two phases. In the first phase of $L \geq 0$ periods, a common intermediate product (or *blank*) is manufactured. Product differentiation occurs in a second phase of l_j periods for product $j = 1, \ldots, J$. The special case where $L = 0$ corresponds with settings where final products need to be differentiated from the onset, i.e., where they are manufactured in a *single* phase, without the intermediary of a common blank. The size of any period's order for blanks is limited by a capacity constraint. No capacity limit is imposed on the second production phase. Indeed, in most practical localization or assembly operations, such as Hewlett-Packard's, capacity can be increased rather easily, as needed.

We initially assume that, upon release of the batch of blanks, the batch is allocated among the final products, i.e., no inventories of blanks are maintained. Sometimes it is physically impossible or highly expensive to store the intermediate product (perhaps because it is highly perishable, or dangerous, as in smelting processes). In other settings, intermediate inventories are avoided as a company policy to reduce the lead time and minimize material-handling costs. Moreover, even if intermediate inventories *can* be maintained, this option only exists if the production process is designed in two phases, i.e., under delayed differentiation. Assessing the benefits of delayed differentiation under the restriction of zero intermediate stock thus results in a *lower* bound on the full benefits achievable. Gallego and Zipkin (1997) recently showed, albeit for stationary and uncapacitated systems, that the benefits of intermediate stock are very minor. Nevertheless, §5.1 extends our results to allow for intermediate stock.

Unsatisfied demand is backlogged. Production costs in both phases are proportional with the production volumes. (In particular, there are no fixed-cost components.) All other cost components incurred for a product (in particular, holding and shortage costs) are a function of the product's inventory position = inventory on hand + blanks being transformed into units of the final product − backlogs. The capacity limit, cost parameters and functions, and the demand distributions vary periodically, with periodicity K. The objective is to minimize expected discounted costs over a finite or infinite horizon or their long-run average value.

We now specify the exact dynamics of the decision process and introduce the basic notation. At the beginning of each period t, a decision is made whether to order a batch of blanks, and if so, of what size. Any batch ordered in period $t - L$ is completed at the beginning of period t and needs to be allocated to the J final products. Demands in the kth period of any cycle of K periods are identically distributed as the vector $d^k = (d_1^k, \ldots, d_J^k)$. Let,

x_j = the inventory position of item j at the beginning of a period, *before* allocation of this period's production batch of blanks;

y_j = the inventory position of item j at the beginning of a period, *after* allocation of this period's production batch of blanks; and

b^k = the capacity in periods of type k ($k = 1, \ldots, K$).

To simplify the notation, we assume that all cost parameters are stationary. (The extension where these parameters depend on the period type k, is straightforward.)

γ = the variable first-stage production cost rate for blanks; and

c_j = the variable second-stage production cost rate for item j.

$\alpha \le 1$ is the discount factor. We assume that the expected value of all other cost components for item j ($j = 1, \ldots, J$) that are charged to a period of type k ($k = 1, \ldots, K$) can be given as a function $G_j^k(y_j)$. Assume, for example, that the carrying (backlogging) cost incurred for an end-of-the-period inventory (backlog) of x_j^+ (x_j^-) units is given by a function $h_j(x_j^+)$ ($p_j(x_j^-)$). With a standard accounting device, we charge to each period the expected discounted holding and backlogging costs incurred a lead time later, i.e., $\overline{G}_j^k(y_j)$ where

$$\overline{G}_j^k(y_j) = \alpha^l \mathsf{E}\{h_j([y_j - d_j^k - d_j^{k+1} \cdots - d_j^{k+l_j}]^+) + p_j([d_j^k + d_j^{k+1} \cdots + d_j^{k+l_j} - y_j]^+)\}, \quad (1)$$

and where all superscripts are taken mod K.

We make the following assumptions regarding the functions $\overline{G}_j^k$ and the finiteness of moments of the demand distributions: We write $\phi(x) = O(\psi(x))$ for any pair of functions $\phi(\cdot)$, $\psi(\cdot)$ if a constant C exists such that $\phi(x) \le C[\psi(x)+1]$ for all x.

ASSUMPTION 1. *$\overline{G}_j^k$ is convex and* $\lim_{|y|\to\infty} \overline{G}_j^k(y) = \lim_{|y|\to\infty}[c_j^k y + \overline{G}_j^k(y)] = \infty$ *for all* $j = 1, \ldots, J$; $k = 1, \ldots, K$.

ASSUMPTION 2. $\overline{G}_j^k(y) = O(|y|^r)$ *for some positive integer* r ($j = 1, \ldots, J$; $k = 1, \ldots, K$).

ASSUMPTION 3. $\mathsf{E}[(d_j^k)^r] < \infty$ ($j = 1, \ldots, J$; $k = 1, \ldots, K$).

Convexity of the one-step expected-cost functions $\{\overline{G}_j^k : j = 1, \ldots, J; k = 1, \ldots, K\}$ is satisfied under most commonly used cost structures; e.g., in (1) it holds when the functions $\{h_j, p_j\}$ are linear or convex. The second part of Assumption 1 is satisfied whenever the asymptotic marginal backlogging cost is in excess of the period's variable production cost rate; it precludes the trivial and unrealistic case where it is *never* beneficial to carry stock in anticipation of demands. Assumption 2 is similarly general; if the $\overline{G}_j^k$-functions are of the form given by (1), it is satisfied with $r = 1$ when $h_j(\cdot)$ and $p_j(\cdot)$ are linear or piecewise linear, and with $r \ge 2$ when these functions are bounded by polynomials. Assumption 3 is necessary to guarantee that the expected cost over a single or multiperiod horizon remains finite; it is often required to ensure that the functions $\overline{G}_j^k(\cdot)$ themselves are finite; see Equation (1). Finally, we write $k^+ = (k \bmod K)+1$. These three assumptions are required in the single-item case as well; see Aviv and Federgruen (1997a).

2. The Single-Stage Production Model: Lower-Bound Approximations and Proposed Strategies

In this section we specify the model for the case where $L = 0$, i.e., where production occurs in a single stage. The model can be formulated as a Markov Decision Process (MDP) with countable state space $S = \{(x, k) : x \text{ is integer}, k = 1, \ldots, K\}$ and (finite) action sets $A(x, k) = \{y : x \le y \text{ and } \sum_{j=1}^J y_j \le \sum_{j=1}^J x_j + b^k\}$. In other words, the state of the system is given by the prevailing vector of inventory positions and the period type k.

Because the state space is of dimension $J+1$, it is impractical to compute an optimal policy. We develop a lower-bound approximation, which corresponds with a *single-item* model, by relaxing the action sets $A(x, k)$ to sets $\widetilde{A}(x, k) = \{y : y \text{ is integer and } \sum_{j=1}^J x_j \le \sum_{j=1}^J y_j \le \sum_{j=1}^J x_j + b^k\}$; in other words, we replace the individual lower bounds $y \ge x$ by the aggregate constraint $\sum_{j=1}^J y_j \ge \sum_{j=1}^J x_j$. The relaxation is equivalent to assuming that if the initial inventory levels of some products are inappropriately high, some of these units can be converted into others. It can be shown (see Aviv and Federgruen 1999, §1) that the value functions in this relaxed model depend on the vector x

only via its aggregate sum $X = \sum_{j=1}^{J} x_j$, and are given by $V_n : \mathbf{Z} \times \{1, \ldots, K\} \to \mathbf{R}$, defined recursively via $V_0 \equiv 0$ and

$$V_n(X, k) = \min_{X \le Y \le X + b^k} \left\{ R^k(Y) + \alpha \mathsf{E} \times \left[V_{n-1}\left(Y - \sum_{j=1}^{J} d_j^k, k^+ \right) \right] \right\}, \quad (2)$$

where

$$R^k(Y) = \min \left\{ \sum_{j=1}^{J} G_j^k(y_j) : \sum_{j=1}^{J} y_j = Y \right\}, \quad (3)$$

and where $G_j^k(y_j) = \overline{G}_j^k(y_j) + (1-\alpha) c_j y_j + \alpha c_j \mathsf{E}[d_j^{k^+}]$. The functions R^k $(k = 1, \ldots, K)$ can be shown to satisfy Assumptions 1 and 2 (see Aviv and Federgruen 1999, §1). We observe that $V_n(X, k)$ represents the minimum expected cost over a horizon of n periods in a *single-item* capacitated model with periodic parameters. It has $D^k = \sum_{j=1}^{J} d_j^k$, $R^k(\cdot)$ and b^k as the demand, one-step expected-cost function and capacity limit in periods of type k. The following theorem is proven in Aviv and Federgruen (1997a) and shows that modified base-stock policies are optimal for the relaxed model for finite and infinite horizons and whether considering total discounted or average costs. A modified base-stock policy initiates in each period a production batch to bring the aggregate inventory position as close as possible to a specific, period-dependent target level. (If the initial aggregate inventory position is above the target level, no new batch is initiated; if the difference between the target level and the prevailing inventory position exceeds the capacity limit, a full-capacity batch is ordered.)

THEOREM 1. (a) (*Finite-horizon model*) *The function* $\widetilde{H}_n(Y, k) \doteq R^k(Y) + \alpha \mathsf{E}[V_{n-1}(Y - D^k, k^+)]$ *has a finite smallest minimizer* $\beta^*_{n,k}$. *The modified base-stock policy with base-stock levels* $\{\beta^*_{n,1}, \ldots, \beta^*_{n,K}\}$ *is optimal for the n-period model.*

(b) (*Infinite-horizon discounted model*: $\alpha < 1$) *Let* $V^*_\alpha(X, k)$ *denote the minimum expected total discounted cost over an infinite planning horizon starting in a period of type k with an inventory position of X units.*

(b-i) $V^*_\alpha = \lim_{n\to\infty} V_n$ (*pointwise*).

(b-ii) V^*_α *is a nonnegative solution of the optimality equation*

$$V(X, k) = \min_{X \le Y \le X + b^k} \{ R^k(Y) + \alpha \mathsf{E}[V(Y - D^k, k^+)] \}. \quad (4)$$

(b-iii) *There exists a policy* f^* *that satisfies the Optimality Equation* (4) *for* $V = V^*_\alpha$, *which is a modified base-stock policy and minimizes the expected infinite-horizon discounted costs.*

(c) (*Infinite-horizon average-cost model*) *Assume* $\sum_k \sum_j \mu_j^k < \sum_k b^k$, *i.e., the expected average* (*aggregate*) *demand per period is less than the average capacity per period.*

(c-i) *There exists a modified base-stock policy that is average-cost optimal, provided that* $\mathsf{E}[(D^k)^{r+1}] < \infty$.

(c-ii) *For an arbitrary parameter* $0 < \tau < 1$, *consider the following value-iteration scheme*: *for all* $X \in \mathbf{Z}, k = 1, \ldots, K$,

$$\widehat{V}_n(X, k) = \min_{X \le Y \le X + b^k} \{ \tau R^k(Y) + (1-\tau) \widehat{V}_{n-1}(X, k) + \tau \mathsf{E}[\widehat{V}_{n-1}(Y - D^k, k^+)] \}. \quad (5)$$

If $\mathsf{E}[(D^k)^{r+2}] < \infty$ $(k = 1, \ldots, K)$, *then the long-run average cost value of the policies generated by* (5), *i.e., the policies achieving the minima in* (5) *converges to the minimum average cost value.*

Thus, for any of the considered criteria, a modified base-stock policy is optimal and can be computed via Scheme (2) or (5); see Aviv and Federgruen (1997a). To execute these recursive schemes one needs only to compute for all $k = 1, \ldots, K$ the convolution D^k of the demand distributions of the individual items $\{d_j^k : j = 1, \ldots, J\}$, as well as the one-step cost functions $R^k(\cdot)$. Because the functions $G_j^k(\cdot)$ are convex, the allocation problem in (3) can be solved by the greedy procedure; see, e.g., Fox (1966). Hence, $R^k(\cdot)$ can be evaluated recursively as follows: Assume $R^k(Y_0)$ has been calculated and $y^*(Y_0)$ achieves the minimum in (3) for $Y = Y_0$. Then,

$$R^k(Y_0 + 1) = R^k(Y_0) + \min_j \{ \overline{G}_j^k(y^*(Y_0)_j + 1) - \overline{G}_j^k(y^*(Y_0)_j) \}, \quad (6)$$

while $y^*(Y_0 + 1)$ is obtained from $y^*(Y_0)$ by incrementing by one any component j that achieves the minimum in (6); see Aviv and Federgruen (1997a).

It is of interest to compare the one-step cost functions $\{R^k\}$ in the relaxed single-item model with those in the original model. The comparison is easiest in the most prevalent case where the functions $G_j^k(\cdot)$ are of the form given in (1) and the demand distributions $\{d_j^k\}$ all belong to the same two-parameter family of distributions. More specifically, assume that $Pr\{d_j^k \leq x\} = F\big((x-\mu_j^k)/\sigma_j^k\big)$ for a common *cdf* $F(\cdot)$ and appropriate pairs of parameters $\{(\mu_j^k, \sigma_j^k) : j = 1, \dots, J;\ k = 1, \dots, K\}$. For example, this is the case if all demand distributions are normal or (translated) gamma with a common shape parameter. Zipkin (1982) shows that when cost rates are identical across items, i.e., $h_j = h$, $p_j = p$ and $c_j = c$,

$$R^k(Y) = \mathsf{E}\{h[Y-\widehat{D}^k]^+ + p[\widehat{D}^k - Y]^+\} + (1-\alpha)cY, \quad (7)$$

where $Pr\{\widehat{D}^k \leq x\} = F\big(x - M^k/\Sigma^k\big)$ with $M^k = \sum_{j=1}^J \sum_{r=0}^{l_j} \mu_j^{k+r}$ and $\Sigma^k = \sum_{j=1}^J \sqrt{\sum_{r=0}^{l_j} (\sigma_j^{k+r})^2}$. Thus, $R^k(\cdot)$ is the one-step cost function in a single-item model with a lead-time demand distribution whose mean M^k is the sum of the items' expected lead-time demands, but whose standard deviation Σ^k is in general *larger* than the standard deviation of the aggregate lead-time demand. When demands are independent, the latter equals $\sqrt{\sum_{j=1}^J \sum_{r=0}^{l_j} (\sigma_j^{k+r})^2}$. The larger standard deviation reflects the penalty paid for demand being spread over J distinct items. This penalty increases as the correlation between the items' demands decreases. Finally, in case h_j, p_j, or c_j are item-dependent, $R^k(\cdot)$ can be approximated *closely* by a function of the Form (7) with h, p, and c appropriate weighted averages of the parameters (see Zipkin 1982).

2.1. Heuristic Strategies

We now describe our proposed strategies. To simplify the exposition, we confine ourselves to the long-run average-cost criterion. The lower-bound approximation suggests heuristic strategies that in each period determine the production quantities in two steps:

Step 1. (*Aggregate Production Quantity*) determination of W, the *aggregate* production quantity across all J items, on the basis of the modified base-stock policy $\beta^* = (\beta^{*1}, \dots, \beta^{*K})$, which is optimal for the lower-bound model. β^{*k} represents the (scalar) base-stock level to be employed in periods of type k ($k = 1, \dots, K$).

Step 2. (*Disaggregation*) disaggregation of W into production quantities for the individual items by solving a specific allocation problem.

W can be specified strictly in accordance with β^*; i.e., in periods of type k ($k = 1, \dots, K$)

$$W^A = \min\{b^k, [\beta^{*k} - X]^+\}, \quad (k = 1, \dots, K). \quad (8)$$

A potential drawback of this specification is that it is made on the basis of the aggregate inventory position X only, even when x, the vector of the items' inventory positions, is highly unbalanced, i.e., when some items are at critically low levels while others are in ample supply. (As we shall show, this situation arises only when the base-stock levels are nonstationary.)

As an alternative to (8) we therefore propose specifying W on the basis of *disaggregate* base-stock levels (and the discrepancies between *individual* inventory positions vis-à-vis them). To this end, we first disaggregate the aggregate base-stock levels $\{\beta^{*k} : k = 1, \dots, K\}$. More specifically, for $k = 1, \dots, K$, let s^{*k} denote the optimal disaggregation of the base-stock level β^{*k} in the (relaxed) allocation Problem (3); in particular, $R^k(\beta^{*k}) = \sum_{j=1}^J G_j^k(s_j^{*k}) = \min\{\sum_{j=1}^J G_j^k(y_j) : \sum y_j = \beta^{*k}\}$. We now specify

$$W^D = \min\left\{b^k, \sum_{j=1}^J [s_j^{*k} - x_j]^+\right\}. \quad (9)$$

Clearly, $W^D \geq W^A$ since $\beta^{*k} - X = \sum_{j=1}^J (s_j^{*k} - x_j) \leq \sum_{j=1}^J [s_j^{*k} - x_j]^+ \geq 0$.

As far as Step 2 is concerned, we disaggregate the total production quantity W, via "myopic allocation," so as to minimize the total expected costs over all J items at the end of the *very first* period in which they become available. For periods of type k ($k = 1, \dots, K$) this disaggregation procedure reduces to solving the allocation problem

$$(P^k): \quad \min\left\{\sum_{j=1}^J G_j^k(y_j) \,\middle|\, \sum_{j=1}^J (y_j - x_j) = W;\ y_j \geq x_j \ (j = 1, \dots, J)\right\}. \quad (10)$$

Myopic allocations are in fact optimal in the lower bound model (Aviv and Federgruen 1999) and exhibit excellent performance in *uncapacitated* multi-item systems (see Federgruen and Zipkin 1984a, 1984c). See Aviv (1998) for an improved allocation mechanism, called "cycle allocations," which improves on "myopic allocations," but in a minor to moderate way only.

Let H^A (H^D) denote the heuristic that selects the aggregate production quantity $W = W^A(W^D)$ and disaggregates by determining a solution to the allocation problems $\{P^k : k = 1, \ldots, K\}$. When these problems have multiple optimal solutions, we specify that under both heuristics the lexicographically smallest solution is selected. We now show that when the system is stationary, i.e., when $K = 1$, H^A and H^D are in fact equivalent in the strong sense that the processes $\{x_n^A\}$ and $\{x_n^D\}$ of inventory positions generated by these heuristics *coincide* after finitely many periods. This implies in particular that the steady-state distribution of inventory positions and the long-run average cost values are identical under both heuristics. On the other hand, our numerical studies demonstrate well that in nonstationary settings, i.e., when $K \geq 2$, the performance of the two heuristics may differ in significant ways.

PROPOSITION 1. *Assume $K = 1$. The processes $\{x_n^A\}$ and $\{x_n^D\}$ coincide after finitely many periods. In particular, the steady-state distribution of inventory positions x and the long-run average cost value are identical under both heuristics H^A and H^D.*

PROOF. Because $K = 1$, we drop the superscript k. We show that both heuristics H^A and H^D choose a vector y^* of inventory positions (after ordering) that satisfies the inequalities

$$x_j \leq y_j^* \leq \max\{s_j^*, x_j\}, \quad j = 1, \ldots, J. \tag{11}$$

This implies that under H^A (H^D), after finitely many periods n^A (n^D), $x \leq s^*$ in *every* subsequent period. Moreover, for $x \leq s^*$, $W^A = W^D = \min\{b, \beta - X\}$, i.e., H^A and H^D prescribe the same aggregate order sizes in every state $x \leq s^*$. Let Y_n^A and Y_n^D denote the aggregate inventory positions after ordering in period n, under heuristics H^A and H^D, respectively. Let $\bar{n} = \max(n^A, n^D)$. Thus, for all $n \geq \bar{n}$, both the $\{Y_n^A\}$ and $\{Y_n^D\}$ processes follow the well-known Lindley equations: $Y_{n+1}^{\cdot} = \min\{\beta, Y_n^{\cdot} - D_n + b\}$ for all $n \geq \bar{n}$. Note also that, since $W^D \geq W^A$ in *every* period,

$$\beta \geq Y_{n+1}^D \geq Y_n^A, \quad \text{for all } n \geq \bar{n}. \tag{12}$$

Since $\mu < b$, a period $n^* \geq \bar{n}$ exists, with probability 1, such that $Y_{n^*}^A = \beta$, and hence, by (12), $Y_{n^*}^D = Y_{n^*}^A = \beta$, while $x_{n^*}^A \leq s^*$ and $x_{n^*}^D \leq s^*$. However, that implies that $y_{n^*}^A = y_{n^*}^D = s^*$, so that for all $n \geq n^*$, the $\{x_n^A\}$ and $\{x_n^D\}$ processes coincide.

To prove (11), assume it does not apply to H^A, i.e., for some item j_1, $y_{j_1}^* > \max\{s_{j_1}^*, x_{j_1}\}$. Also,

$$\begin{aligned}\sum_{j \neq j_1} y_j^* + \max\{s_{j_1}^*, x_{j_1}\} &< \sum_j y_j^* \\ &= \min\{[\beta - X]^+, b\} + X \leq [\beta - X]^+ + X \\ &= \max\{\beta, X\} \leq \sum_j \max\{s_j^*, x_j\} \\ &= \sum_{j \neq j_1} \max\{s_j^*, x_j\} + \max\{s_{j_1}^*, x_{j_1}\}.\end{aligned} \tag{13}$$

Thus, for some $j_2 \neq j_1$: $x_{j_2} \leq y_{j_2}^* < s_{j_2}^*$. Since the functions G_j are convex, we have $[G_{j_1}(y_{j_1}^*) - G_{j_1}(y_{j_1}^* - 1)] + [G_{j_2}(y_{j_2}^*) - G_{j_2}(y_{j_2}^* + 1)] \geq [G_{j_1}(s_{j_1}^* + 1) - G_{j_1}(s_{j_1}^*)] + [G_{j_2}(s_{j_2}^* - 1) - G_{j_2}(s_{j_2}^*)] \geq 0$, because s^* achieves $\min\{\sum_j G_j(y_j) : \sum_j y_j = \beta\}$. But then, by shifting one unit from y_{j1}^* to y_{j2}^*, a lexicographically smaller, *feasible* solution to (P^k) can be obtained, which is at least as good as y^*, contradicting its definition. (The perturbation is feasible since $y_{j_1}^* > x_{j_1}$.)

Similarly, (11) applies to H^D as well, merely replacing (13) by $\sum_{j \neq j_1} y_j^* + \max\{s_{j_1}^*, x_{j_1}\} < \sum_j y_j^* = \min\{\sum_j [s_j^* - x_j]^+, b\} + \sum_j x_j \leq \sum_j [s_j^* - x_j]^+ + \sum_j x_j = \sum_j \max\{s_j^*, x_j\}$. □

3. Two-Stage Production Processes

We now describe how the lower bound and heuristics should be adapted when $L > 0$, i.e., when production occurs in *two* phases. It can be verified (see Aviv and Federgruen 1999) that the single-item, single-stage lower-bound model described by (2) continues to apply, merely replacing $R^k(\cdot)$ by $\widetilde{R}^k(Y) \doteq \mathsf{E}\{R^k[Y - (D^{k+1} + D^{k+1} + \cdots + D^{k+L})]\}$; Y now denotes the systemwide echelon inventory position of blanks, i.e., all blanks being manufactured, transformed into final products, or as part of final products' inventories. $\widetilde{R}^k(\cdot)$ satisfies Assumptions 1 and 2, since the functions $R^k(\cdot)$ do (see Aviv and Federgruen 1999, §2.1). With this modification of the $R^k(\cdot)$ functions, Theorem 1 thus continues to apply.

As to the required modifications of the proposed heuristics, we confine ourselves again to the long-run average-cost criterion. When $L = 0$, it is *useful* to view the choice of the production quantities as occurring in two steps (aggregate production quantity and disaggregation). When $L > 0$, the two steps occur L periods apart and are intrinsically separate.

As to Step 1, we again propose setting the order size W in each period as $W = W^A$ or $W = W^D$, except that in (9) (the expression for W^D), the vectors $\{s^{*k}\}$ are now determined to achieve the minima in the problems $\min\{\sum_{j=1}^{J} G_j^{k+L}(y_j) : \sum_j y_j = \beta^{*k}\}$, for $k = 1, \ldots, K$. It thus suffices to replace $G_j^k(\cdot)$ by $G_j^{k+L}(\cdot)$, allocating the base-stock levels β^{*k} once again with the goal of minimizing total expected inventory holding and backlogging costs a complete production lead time later. See the discussion following (10) for a more detailed rationale.

Whenever a batch of blanks is completed, it is allocated among the final products. This is Step 2 (disaggregation), and we continue to propose the myopic allocations heuristic.

4. Numerical Studies

This section is devoted to a numerical study of a synthetic set of problem instances, conducted to gauge the quality of the lower-bound approximation and proposed heuristics and the associated computational effort, as well as to provide insights into the strategic questions listed in the introduction. We study a total of 123 instances, split into four sets. The first three sets represent single-stage processes, i.e., immediate product differentiation ($L = 0$). All instances have linear holding and backlogging costs and a uniform rate per unit (per period) of $h = 0.05$ and $p = 1$. Also, we fix $\bar{\mu} = \frac{1}{KJ} \sum_{j,k} \mu_j^k$, the average of the mean demands, across all items and period types, at $\bar{\mu} = 40$. All one-period demands are normally distributed.

For each instance, we solve the lower-bound model via the value-iteration method (5) and evaluate the heuristics via simulation over an interval of 450 cycles of K periods each, ignoring the initial 50 cycles. Several hundreds of replicas of these simulations are conducted to obtain sufficiently narrow confidence intervals. Our sets of instances are as follows:

(i) Forty-eight instances with *stationary* ($K = 1$) and *independent* demands (across items). These instances are grouped into 12 sets of four instances each. All instances within a group share the same number of items (J) and the same coefficient of variation (*c.v.*) for the one-period demand distributions. In each group of instances, four values are chosen for the capacity per item, i.e., $b/J = 45, 50, 60$, and 80, hence with utilization rates of 88.89%, 80%, 66.67%, and 50%. Table 1 reports several performance measures. (Recall that for stationary instances, the two heuristics coincide.)

(ii) Thirty instances with *nonstationary* and independent demands. These are grouped into five sets of six instances each, all with $K = 4$, $J = 2$, and a utilization rate of 80%. Each group shares the same demand pattern; i.e., array $[\mu_j^k : j = 1, \ldots, J;\ k = 1, \ldots, K]$ of mean demands. We consider five patterns, see Table 2. Within a group we vary *c.v.* from 0 to 0.5 in increments of 0.1. Patterns A, B, D, and E describe identical items: Pattern A is our benchmark stationary case, while B, D, and E exhibit increasingly severe fluctuations over the course of the cycle. In Pattern C, each of the items experiences significant seasonality, but the patterns dovetail each other perfectly, i.e., the aggregate mean demand per period is stationary. Our results are reported in Table 3.

(iii) Fifteen instances with *correlated* (stationary and nonstationary) demands, partitioned into three groups of five, all with $J = 2$ items, (maximum) periodicity $K = 4$, capacity $b = 100$, $l = 2$, and constant $c.v. = 0.5$. All instances within a group share the same pair of normal, *marginal* demand distributions that, given a fixed value for *c.v.*, are characterized by the array $[\mu_j^k : j = 1, \ldots, J; k = 1, \ldots, K]$ of mean demands. In each period, the same correlation coefficient ρ applies for all pairs of items. The three groups correspond with demand patterns A, B, and C. Within each group we systematically vary ρ from -1 to $+1$ in increments of 0.5.

Table 1 Numerical Results: Stationary Models with Independent Demands ($K = 1$, $L = 0$, $h = 0.05$, $p = 1$)

			$J = 2$				$J = 5$			
			Lower-Bound Model		Heuristic (Simulation)		Lower-Bound Model		Heuristic (Simulation)	
	c.v.	Capacity per item B/J	β^*	*LB*	Average Cost Z^H	Optim. Gap (%)	β^*	*LB*	Average Cost Z^H	Optim. Gap (%)
$l = 0$	0.125	45	97	1.055	1.061 ± 0.000	0.57	242	2.603	2.617 ± 0.000	0.54
		50	97	1.041	1.047 ± 0.000	0.58	242	2.602	2.616 ± 0.000	0.54
		60	97	1.041	1.047 ± 0.000	0.58	242	2.602	2.616 ± 0.000	0.54
		80	97	1.041	1.047 + 0.000	0.58	242	2.602	2.616 + 0.000	0.54
	0.25	45	121	2.377	2.384 ± 0.007	0.29	286	5.263	5.274 ± 0.001	0.21
		50	114	2.110	2.113 ± 0.001	0.14	283	5.203	5.216 ± 0.000	0.25
		60	113	2.081	2.086 + 0.000	0.24	283	5.201	5.214 + 0.000	0.25
		80	113	2.081	2.086 ± 0.000	0.24	283	5.201	5.214 ± 0.000	0.25
	0.5	45	204	6.896	7.001 + 0.127	1.52	397	11.380	11.427 + 0.039	0.41
		50	161	4.753	4.777 ± 0.015	0.50	372	10.521	10.565 ± 0.003	0.42
		60	149	4.219	4.246 ± 0.002	0.64	367	10.399	10.448 ± 0.000	0.47
		80	147	4.162	4.191 ± 0.001	0.70	367	10.396	10.444 ± 0.000	0.46
$l = 2$	0.125	45	269	1.810	1.814 ± 0.000	0.22	672	4.507	4.518 ± 0.000	0.24
		50	269	1.803	1.807 ± 0.000	0.22	672	4.507	4.518 ± 0.000	0.24
		60	269	1.802	1.807 + 0.000	0.28	672	4.507	4.518 + 0.000	0.24
		80	269	1.802	1.807 ± 0.000	0.28	672	4.507	4.518 ± 0.000	0.24
	0.25	45	304	3.752	3.759 + 0.004	0.19	747	9.039	9.055 ± 0.001	0.18
		50	299	3.618	3.624 ± 0.000	0.17	744	9.009	9.023 ± 0.000	0.16
		60	298	3.604	3.608 ± 0.000	0.11	744	9.007	9.022 ± 0.000	0.17
		80	298	3.604	3.608 + 0.000	0.11	744	9.007	9.022 + 0.000	0.17
	0.5	45	399	8.833	8.920 ± 0.102	0.98	912	18.505	18.559 ± 0.022	0.29
		50	367	7.504	7.524 ± 0.008	0.27	893	18.070	18.120 ± 0.001	0.28
		60	357	7.235	7.257 ± 0.001	0.30	889	18.008	18.060 ± 0.000	0.29
		80	356	7.208	7.229 ± 0.000	0.29	888	18.006	18.058 ± 0.000	0.29

± represents 95% confidence interval.

(iv) Thirty instances with $L > 0$, grouped into six groups of five instances. For the sake of brevity, we consider only instances in which all demands are independent and normally distributed with $c.v. = 0.5$. Throughout, we specify the capacity so that the utilization rate is 80%. In each instance, the items share a common second-phase lead time, i.e., $l_j = l$ for all j. Within a group, all instances share the same number of items ($J = 2, 5$) and the same demand pattern (A, C, and E; see Table 2). All thirty instances have a total lead-time $L + l = 4$. Within each group, we systematically consider the five possible decompositions of the total lead time of four periods among L and l.

Table 2 Patterns of Mean Demands

	Period type			
Pattern	$k = 1$	$k = 2$	$k = 3$	$k = K = 4$
A ($\mu_j^k = \mu^k$ for all j)	40	40	40	40
B ($\mu_j^k = \mu^k$ for all j)	25	25	25	25
C $j = 1, 5$	25	35	45	55
$j = 3$	40	40	40	40
$j = 2, 4$	55	45	35	25
D ($\mu_j^k = \mu^k$ for all j)	15	15	25	105
E ($\mu_j^k = \mu^k$ for all j)	0	0	20	140

4.1. Quality of Lower Bound and Heuristics

We begin with the first three sets of instances with $L = 0$. The average gap between the lower bound and the best of the proposed heuristics is 0.4%, indicating that the former is very accurate and the latter

Table 3 Numerical Results: Models with Periodically Varying Parameters and Independent Demands ($J = 2$, $L = 0$, $l = 2$, $h = 0.05$, $p = 1$, $b^k = 100$)

		Lower-Bound Model		Heuristic (Simulation)			
				H^A		H^D	
Pattern	*c.v.*	β^*	*LB*	Average Cost Z^H	Optim. Gap (%)	Average Cost Z^H	Optim. Gap (%)
A	0	240	0.000	0.000 ± 0.000	0.00	0.000 ± 0.000	0.00
	0.1	263	1.442	1.447 ± 0.000	0.35	1.447 ± 0.000	0.35
	0.2	286	2.888	2.891 + 0.000	0.10	2.891 + 0.000	0.10
	0.3	311	4.361	4.366 ± 0.001	0.11	4.366 ± 0.001	0.11
	0.4	338	5.889	5.899 ± 0.003	0.17	5.899 ± 0.003	0.17
	0.5	367	7.504	7.524 ± 0.008	0.27	7.524 ± 0.008	0.27
B	0	220, 270, 250, 230	0.125	0.125 ± 0.000	0.00	0.125 ± 0.000	0.00
	0.1	242, 296, 275, 257	1.637	1.641 ± 0.000	0.24	1.641 ± 0.000	0.24
	0.2	265, 323, 300, 285	3.186	3.190 + 0.001	0.13	3.190 + 0.001	0.13
	0.3	289, 349, 328, 314	4.759	4.767 ± 0.002	0.17	4.767 ± 0.002	0.17
	0.4	314, 376, 357, 344	6.378	6.394 + 0.004	0.25	6.392 + 0.004	0.22
	0.5	341, 402, 390, 377	8.073	8.108 ± 0.010	0.43	8.101 ± 0.010	0.35
C	0	240, 240, 240, 240	0.000	0.000 ± 0.000	0.00	0.000 ± 0.000	0.00
	0.1	264, 264, 264, 264	1.492	1.496 + 0.000	0.27	1.496 + 0.000	0.27
	0.2	288, 288, 289, 289	2.988	2.994 ± 0.000	0.20	2.992 ± 0.000	0.13
	0.3	313, 313, 314, 315	4.515	4.545 ± 0.001	0.66	4.530 ± 0.001	0.33
	0.4	341, 341, 342, 343	6.109	6.183 + 0.004	1.21	6.143 + 0.003	0.56
	0.5	372, 371, 373, 374	7.799	7.935 + 0.009	1.74	7.866 + 0.008	0.86
D	0	220, 290, 290, 330	2.125	2.125 ± 0.000	0.00	2.125 ± 0.000	0.00
	0.1	249, 326, 326, 363	4.166	4.178 + 0.001	0.29	4.178 + 0.001	0.29
	0.2	277, 363, 363, 396	6.225	6.238 ± 0.003	0.21	6.238 ± 0.003	0.21
	0.3	306, 399, 399, 435	8.336	8.366 ± 0.005	0.36	8.352 ± 0.005	0.19
	0.4	334, 435, 438, 478	10.559	10.622 + 0.010	0.60	10.590 + 0.010	0.29
	0.5	363, 471, 482, 527	12.949	13.069 + 0.026	0.93	13.038 + 0.026	0.69
E	0	220, 320, 340, 400	4.000	4.000 + 0.000	0.00	4.000 + 0.000	0.00
	0.1	257, 367, 377, 443	6.456	6.478 + 0.002	0.34	6.478 + 0.002	0.34
	0.2	294, 414, 422, 486	9.093	9.120 ± 0.004	0.30	9.114 ± 0.004	0.23
	0.3	331, 461, 471, 534	11.849	11.918 + 0.009	0.58	11.888 + 0.009	0.33
	0.4	368, 508, 526, 592	14.844	15.005 ± 0.025	1.08	14.962 ± 0.025	0.79
	0.5	405, 554, 591, 661	18.199	18.527 ± 0.071	1.80	18.507 ± 0.068	1.69

is close-to-optimal. Only when $J = 2$ can the optimal strategy be computed in reasonable time, and the exact accuracy and optimality gap assessed. (A single instance requires about four hours of CPU time on a Pentium 266 MHz PC.) In a sample of six two-item instances, the average accuracy gap is 0.03% and the optimality gap is 0.25%.

Each of the two proposed heuristics, as well as the lower bound, require a one-time computation of the optimal base-stock levels $\{\beta^{*1}, \ldots, \beta^{*K}\}$ in the lower-bound model, via (5). The effort to evaluate one of the functions $\widehat{V}_n(\cdot, \cdot)$ is roughly proportional with K, and roughly quadratic in the number of distinct inventory levels X considered. To achieve high precision, we have considered 700 inventory levels in the (truncated) state space. (See Aviv and Federgruen 1997a on how the cost values of inventory levels outside the truncated state space are extrapolated.) For $K = 1$ and 700 inventory levels, an iteration takes on average 4 seconds on a Pentium 266 MHz PC. The number of iterations required to achieve convergence of the base-stock levels depends on the system's utilization rate and the variability of the demands, ranging from a handful of iterations when the *c.v.* is less than 0.15, to 15–25 under high utilization rates and *c.v.*s less than 0.25, and to about 100 under high utilization rates and a *c.v.* equal to 0.5. Once the optimal base-stock levels $\{\beta^{*1}, \ldots, \beta^{*K}\}$ have been computed, a single instance of problem (P^k) requires less than 0.01 seconds.

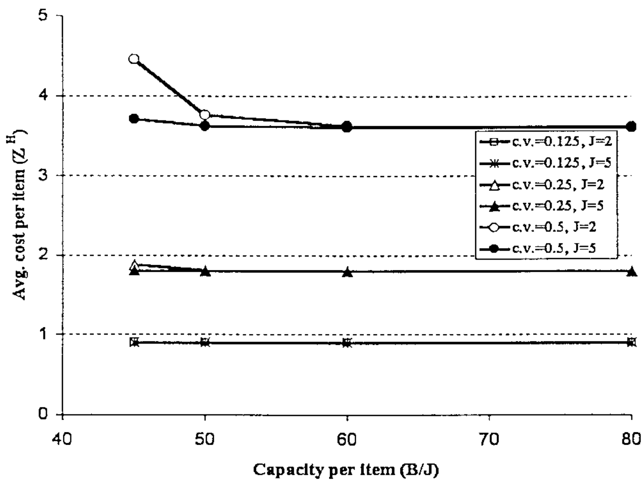
For the first set of instances, we found that the upper bound for the optimality gap $(Z^H - LB)/LB$ is always less than 1.52% and less than 0.70% when the utilization rate is 80% or below. Its average value across all 48 instances equals 0.37% (see Table 1). The optimality gap tends to further decrease as either the utilization rate or the value of *c.v.* decreases or J increases. The monotonicity may be explained by the fact that as J increases the coefficient of variation of *aggregate* demands decreases (by a factor of $\sqrt{J}$ since demands are independent across items). Thus, the lower bound and heuristic can be comfortably used as close-to-optimal when $J \leq 5$, and a fortiori for larger numbers of items.

For the nonstationary instances of our second set of instances, the two proposed heuristics (H^A and H^D) are, in general, distinct. The bounds for the optimality gaps $(Z^H - LB)/LB$ are equally low as for the stationary instances, with an average of 0.51% for H^A and 0.37% for H^D over all instances with $c.v. > 0$. The bounds for the optimality gaps tend to increase with the *c.v.* value and among those in excess of 0.35%, we observe an increase as we move toward pattern E, with large "seasonal" fluctuations. The optimality gap we observed is even lower when J increases. H^D consistently outperforms H^A, in particular under patterns C–E, where large imbalances between the individual items are more likely to occur, see §2.

For our third set of instances with correlated demands, the bound for the gap $(Z^H - LB)/LB$ continues to be equally small under correlated demands. It never exceeds those arising under independent demands by more than 0.93%, and its average value is 0.86% for H^A and 0.52% for H^D. The lower bound, the heuristic's average cost, and the base-stock levels all increase with ρ, and hence the variance of aggregate demands is increased.

Finally, we found that when $L > 0$, the optimality gaps are, if anything, even smaller than in the case $L = 0$. For example, for the instances with $J = 2$, the average upper bound for the optimality gap is 0.84%, and it is never larger than 2.14% (under H^A).

Figure 1 **Long-Run Average Cost (Per Item) as a Function of the Capacity (Per Item),** ($K = 1,\ L = 0,\ l = 2,\ h = 0.05,\ p = 1$)

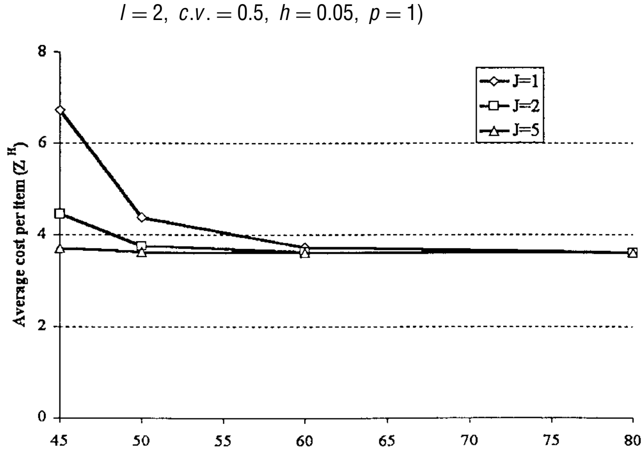


4.2. Managerial Insights

In this section we present several managerial insights we have obtained from our study.

4.2.1. Stationary Models with Independent Demands. Figure 1 exhibits, for the six groups of instances with $l = 2$, the average cost per item under the heuristic strategy as a function of the capacity per item. Figure 2 displays the same relationship for

Figure 2 **Long-Run Average Cost (Per Item) as a Function of the Capacity (Per Item) for** $J = 1, 2,$ **and** 5 **Items** ($K = 1,\ L = 0,\ l = 2,\ c.v. = 0.5,\ h = 0.05,\ p = 1$)



instances with $c.v. = 0.5$ and $J = 1$, $J = 2$, and $J = 5$ identical items.

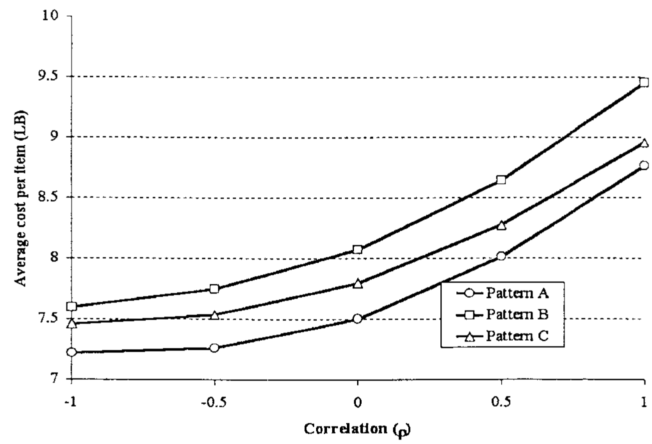
We first note that the costs grow rapidly as the utilization rate or the $c.v.$ value grows. For example, under high utilization and $J = 2$, the cost triples as the $c.v.$ varies from 0.25 to 0.5. An increased utilization rate is of particular impact when $c.v.$ is large or when the items are nonidentical.

Figure 2 quantifies the value of flexible production systems. Clearly, as more and more items are assigned to the same facility and its capacity increased in proportion (i.e., the average capacity per item is kept constant), the average cost per item decreases. This is again explained by the $c.v.$ of aggregate demands decreasing by a factor $\sqrt{J}$, as J increases. In addition, the marginal benefit decreases as J increases. Also, the benefits of flexible capacity are particularly large when the utilization rate is high and almost insignificant when it is less than (say) 80%. A similar trio of curves is obtained for systems with *heterogeneous* items. When the $c.v.$ value is increased, the three curves maintain the same shape and relative position to each other, however, with significantly larger gaps between them, in particular under high utilization rates. In other words, the benefit of flexibility increases with the variability of demands.

4.2.2. Models with Periodically Varying Parameters and Independent Demands. As before, average costs increase rapidly with the $c.v.$ value (see Table 3). Patterns A and C have almost identical costs; this is explained by the fact that aggregate demands have identical means under both patterns, but those under C have somewhat larger variances. Average costs increase rapidly as seasonal fluctuations increase; see Patterns D and E. As in the stationary case, the average cost per item decreases with J. We note again the same benefits of flexibility as in Figure 2. These are particularly large under dovetailing patterns as in C.

4.2.3. Models with Correlated Demands. Figure 3 displays the average cost per item as a function of the correlation between the products' demands for Patterns A, B, and C. The costs grow faster with ρ under Pattern B, in which the mean *aggregate* demand varies considerably over the course of the cycle.

Figure 3 **Long-Run Average Cost (Per Item) as a Function of the Correlation Between the Demands for Different Products** ($J = 2$, $L = 0$, $l = 2$, $b^k = 100$, $c.v. = 0.5$, $h = 0.05$, $p = 1$)



Our models can be used to gauge the value of *flexible* production systems. To this end, we compare the cost values with those of systems with dedicated (item-specific) production equipment. Let Z^D denote the cost of an optimally designed *dedicated* system and $LB(\rho)$ the lower bound for the flexible system as a function of ρ. Z^D is independent of ρ, and since Z^D is the cost of a feasible strategy for the flexible system, $Z^D \geq LB(\rho)$ for all $\rho \leq 1$, i.e.,

$$Z^D \geq \overline{LB} \doteq \max_{-1 \leq \rho \leq +1} LB(\rho). \qquad (14)$$

This inequality and the accuracy of the lower bound LB imply that the benefits of a flexible system can be conservatively estimated by $(\overline{LB} - LB(\rho))$. (As explained in Aviv and Federgruen 1997a, exact computation of Z^D may be tedious. It involves solving a convex program with a nonseparable objective in $J \cdot K$ variables.) With identical items, as in patterns A and B, we have in fact that (14) holds as an equality; i.e.,

$$Z^D = LB(1) = \overline{LB}. \qquad (15)$$

If $\rho = 1$, all items experience identical demands in all periods, so that it is optimal to have identical production quantities for all items in every period. This production rule is implementable in a dedicated system with, for each item, capacity b^k/J in periods of type k. This explains the first equality in (15) and, hence by (14), the second equality as well. We thus

Figure 4 **Long-Run Average Cost (Per Item) as a Function of the Point of Differentiation (L) for Various Demand Patterns** ($L+l=4$, $b^k/J=50$, $c.v.=0.5$, $h=0.05$, $p=1$)

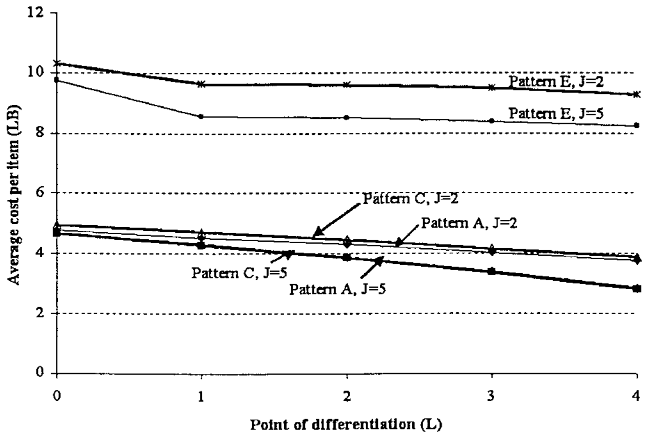
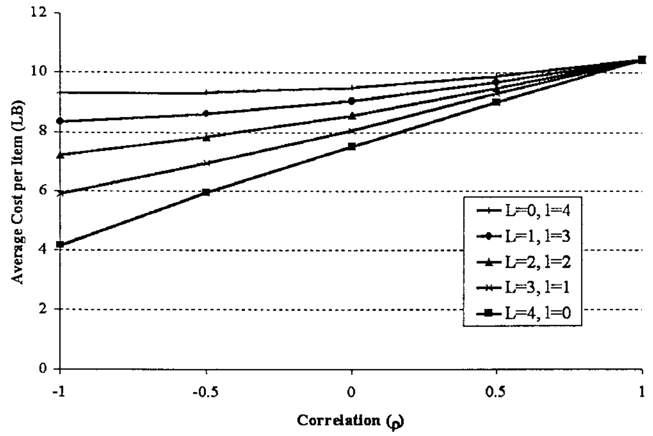


Figure 5 **Long-Run Average Cost (Per Item) as a Function of the Correlation Between the Demands for Different Products** ($K=1$, $J=2$, $L+l=4$, $b^k=100$, $c.v.=0.5$, $h=0.05$, $p=1$)



observe from Figure 3 that the benefits of a flexible system decrease with ρ.

4.2.4. Two-Stage Production Processes. Figure 4 exhibits the benefits of delayed differentiation. Costs decrease significantly within each set of instances as we move from immediate to maximally postponed differentiation. The average decrease is 17.7% for two-item and 31.6% for five-item systems. The large cost savings for five-item systems follow from the benefits of dealing with a single common intermediate product increasing with J. We conclude that postponed product differentiation is particularly beneficial if one wishes to maintain a large degree of product variety.

Finally, to investigate the impact of ρ, the degree of *correlation* between the items' demands, we have, for the instances with $J=2$ and Pattern A, varied ρ between -1 and $+1$ in increments of 0.5; see Figure 5. Benefits from postponed differentiation decrease with ρ, with zero benefits in the limiting case $\rho=1$. As shown in §3, the cost value in this limiting case equals that arising in a dedicated system (since the items are identical). We thus conclude from the graphs in Figure 5 that for a given value of *c.v.*, the benefits of moving from dedicated facilities to a flexible factory often begin to be realized only when this move is accompanied by a redesign of the production process to allow for postponed product differentiation.

5. The Hewlett-Packard Case

We now revisit the HP case, investigating how the postponement benefits vary as the four factors (i)–(iv) listed in the introduction are incorporated into the analysis.

The traditional analysis of the HP case (see e.g., the teaching note by Flaherty et al. 1996) is based on sample sales data for six DeskJet printers over 12 consecutive months. Under assumptions (i)–(iv) listed in the introduction, the system with factory localization behaves like six distinct stationary and uncapacitated single-item models, while it behaves like a single such model facing aggregate demands in the case of DC localization. Solving these models on the basis of the case-specified desired fill rate of 98%, one concludes that safety stocks reduce from 4 weeks to 2.5 weeks when postponing differentiation (similar to the actual reduction from 5.2 weeks to 3.5 weeks achieved by the company; see Lee and Billington 1995). Likewise, supply-chain inventory holding costs decrease by 16%. (Since under assumptions (i)–(iv), the analysis involves single-item models only, these performance measures are exact.) We now revisit the analysis, taking into account that the localization process in the DC requires approximately 1 week. We therefore model the system as a two-stage process whose total lead time of 5 weeks is partitioned as $L=4$ and $l=1$. In addition, we incorporate the impact of capacity limitations, seasonality patterns, and correlated demand processes.

Recall that in the stationary single-item models, the fill-rate constraint is equivalent to minimizing average holding and backlogging costs with a specific corresponding p/h ratio. In our case, the 98% fill rate corresponds with $p/(p+h) = 0.95$, approximately. Maintaining this ratio in the analysis below, we focus on Items AB, AQ, and AU, whose sales volume represents 97% of the total European line. We first compute for each of the 12 months, for the aggregate sales, the ratio of the sales in that month over the monthly average. If all of the monthly variations are due to seasonal fluctuations, these ratios $\{\gamma_k : k = 1, \ldots, 12\}$ represent the seasonality factors. More realistically, only part of the monthly variation is attributable to seasonalities. The data in the HP case are limited to a single year and thus are insufficient to arrive at a proper statistical separation between seasonal fluctuations and intrinsic uncertainties around predictable seasonal means. We therefore consider several seasonality patterns, parameterized by $0 \le \alpha \le 1$, with $\gamma_k(\alpha) = 1 + \alpha(\gamma_k - 1)$, $k = 1, \ldots, 12$; see Figure 6. (The extreme case $\alpha = 1$ has $\gamma_k(\alpha) = \gamma_k$, while $\alpha = 0$ assumes a lack of seasonalities.) Assuming that the *c.v.* value of monthly demands is constant over time, we estimate this value for each of the 3 items from the 12 ratios of the actual sales over their seasonally adjusted means. The sample size adjusted standard deviation of these 12 ratios represents an unbiased estimator of the *c.v.* value. Finally, we represent each month as 4 weeks and assume intertemporal independence between weeks. Table 4 reports the *c.v.* values for the three values of α. (The *c.v.* for weekly demands is thus twice that of monthly demands.)

Table 4 Values of Coefficient of Variation for Various α-Values—The HP Case

	Product/Location		
α	AB	AQ	AU
0	0.355	0.508	0.524
0.5	0.259	0.434	0.525
1	0.195	0.451	0.601

Figure 7 depicts average systemwide costs as a function of the capacity-utilization rate, both under factory and DC localization (FL and DCL, respectively) and for three possible seasonality patterns, with $\alpha = 0, 0.5$, and 1. (These cost measures are based on the analytical lower-bound model, the accuracy of which was demonstrated in §4.) We note that the costs increase significantly as the utilization rate increases and that the costs (and safety stocks) can be significantly reduced if a larger part of the monthly fluctuations can be attributed to seasonal variations, and hence anticipated by adopting a period-dependent modified base-stock policy. In addition, the two dotted line curves depict, for the seasonality patterns with $\alpha = 0.5$ and $\alpha = 1$, the cost incurred if, as in the teaching note, seasonalities are ignored and an optimal corresponding stationary policy is implemented.

Figure 6 Seasonality Patterns for DeskJet Printers Demand in Europe (Based on November 1989 through October 1990 Sales Data)

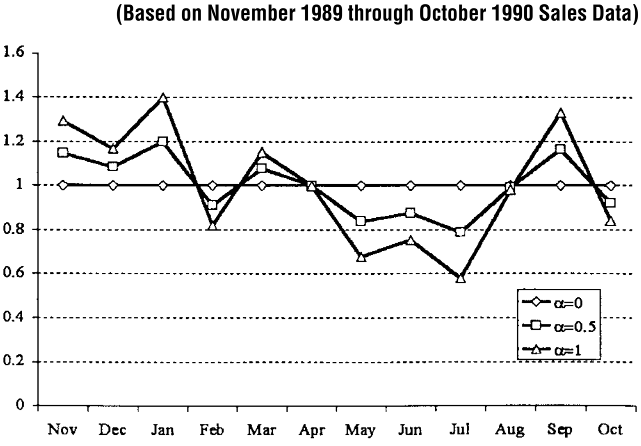
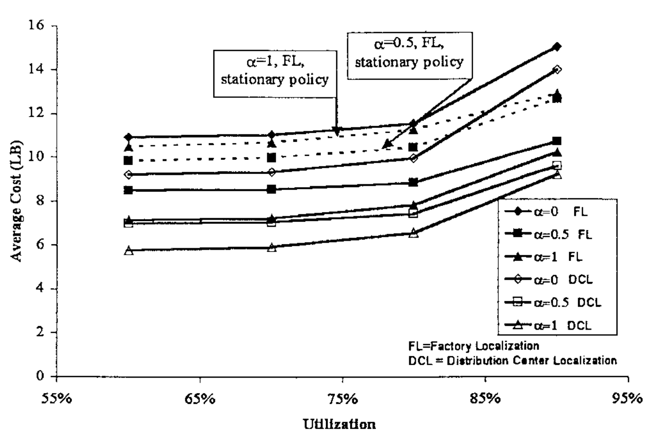


Figure 7 Long-Run Average Cost as a Function of Capacity Utilization (Hewlett Packard Deskjet Printer Supply Chain Case)



The upper (lower) dotted curve should be compared with the ($\alpha = 1$, FL) (($\alpha = 0$, DCL)) curves. Note that the cost increase due to the failure to recognize seasonalities and the adoption of a stationary policy can be as large as 45%! Observe, in addition, that the benefits of postponement decrease as the utilization rate increases. For example, in Figure 7, when $\alpha = 0.5$, the costs may be reduced by 18%, 17%, 16%, and 10% when switching from factory- to DC-localization, and when the capacity utilization equals to 60%, 70%, 80%, and 90%, respectively. As mentioned, less can be saved when capacity is limited because the factory has fewer options and must produce nearly at capacity most of the time, regardless of the demand stream.

Note next that the top three curves in Figure 7 all relate to systems with factory localization (FL), governed by the *stationary* replenishment and allocation policy, which is optimal (in the lower-bound model) in the *absence* of any seasonality pattern. It is noteworthy that the system performs better under this policy when the demand processes exhibit a seasonal pattern, i.e., when $\alpha = 0.5$ or $\alpha = 1$. Perhaps surprisingly, system performance does not necessarily improve as α increases from 0 to 1. This may be explained by the fact that, for some items, the coefficient of variation of monthly demands does not monotonically decrease as α is varied between 0 and 1. Finally, note that the curve corresponding to the case of DC localization and *no* seasonality ($\alpha = 0$, DCL) crosses the dotted curves at large capacity-utilization rates. As explained above, the benefits of postponement decrease significantly as capacity utilization becomes large. For large utilization rates, the benefits may in fact be outweighed by the cost advantages associated with more predictable seasonality patterns.

Figure 8 shows the decrease in systemwide inventories due to postponed differentiation, again as a function of the utilization rate, for $\alpha = 0, 0.5$, and 1. Observe that the relative benefits can be significantly larger under larger seasonal fluctuations. On the other hand, the inventory ratio may fail to be monotone in α, even though both the numerator and denominator are. While the absolute savings increase, the relative benefits decrease significantly as the utilization rate increases for reasons explained above. Also, under more significant seasonal fluctuations, the magnitude of the benefits is considerably less sensitive to fluctuations of the utilization rate, e.g., as one experiences different phases of the business cycle.

Figure 8 Inventory Savings as a Result of Postponement (Hewlett Packard DeskJet Printer Supply Chain Case)

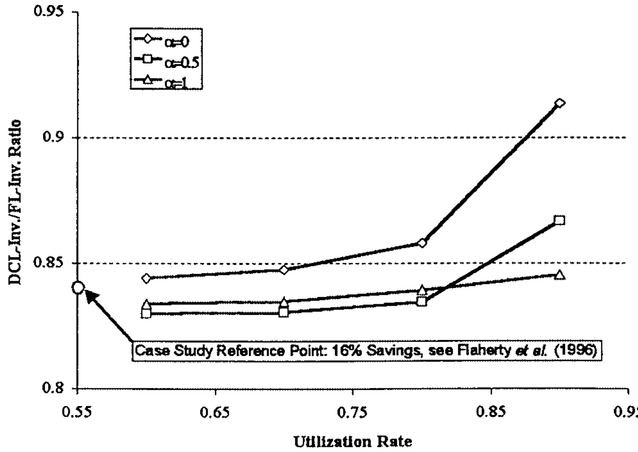


Figure 9 has three graphs showing how the savings due to postponement, measured as a percentage of the base case with *immediate* differentiation ($L = 0$), vary with the point of differentiation. The three graphs correspond with the three seasonality patterns ($\alpha = 0, 0.5$, and 1). All instances have independent demands and a utilization rate of 80%. The savings under seasonality are significantly larger than achieved in the absence of seasonality ($\alpha = 0$), and the

Figure 9 Savings as a Result of Delayed Product Differentiation—The Impact of Seasonality (Hewlett Packard DeskJet Printer Supply Chain Case)

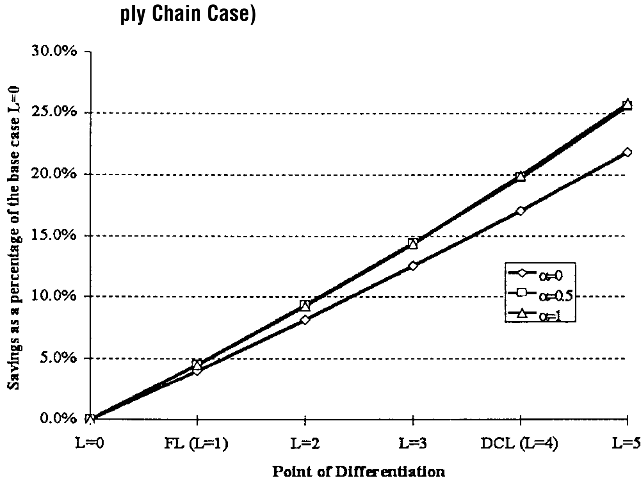
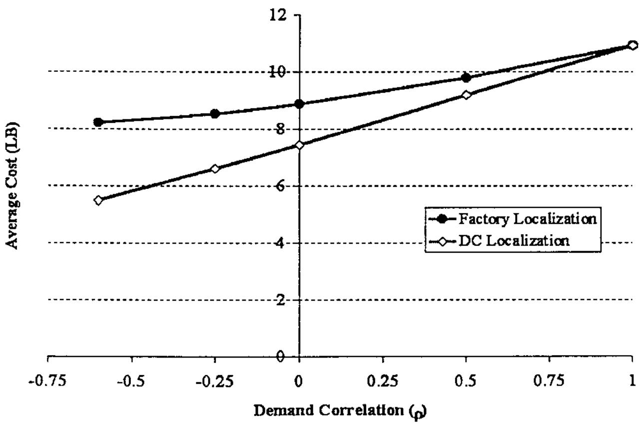


Figure 10 Long-Run Average Cost as a Function of Demand Correlation (Hewlett Packard DeskJet Printer Supply Chain Case)



incremental savings increase as the point of differentiation is postponed. Observe, however, that the savings under $\alpha = 1$ are insignificantly larger than those under $\alpha = 0.5$.

Finally, Figure 10 shows how the costs vary with the demand processes' correlation pattern. To enable a simple parametric representation, we consider instances in which the correlation coefficients of the demands of all pairs of items are identical. All instances have a utilization rate 0.8 and assume a modest seasonality pattern ($\alpha = 0.5$). The upper (lower) graph in Figure 10 corresponds to the case of factory (DC) localization ($L = 1$ and $L = 4$, respectively). Observe that the costs under DC localization are always lower than those achieved under factory localization. The benefits of DC localization increase dramatically, as ρ decreases from 1; they are small under high pairwise correlations (vanishing in the limiting case of $\rho = 1$) since the variance of aggregate demand is close to the sum of the variances of the demands for the individual items.

6. Extensions

In this section we discuss two important extensions to our basic model.

6.1. Intermediate Inventories

In this section, we give a brief outline of how our results can be extended to allow for intermediate inventories. An exact representation of the problem requires a dynamic program with a $(L+J+1)$-dimensional state space. In addition to the vector x of the final items' inventory positions and the period type k, the state description includes

$$v^d = \text{the echelon inventory of blanks before arrival of an order} = \text{inventory of blanks} + X,$$

as well as the vector of outstanding orders for blanks $(w^1, \ldots, w^{L-1})$, where w^r represents the size of the batch of blanks ordered r periods before the beginning of the current period (note that w^L is already taken into account in v^d).

For the sake of brevity, we confine ourselves to the finished goods' holding and backlogging cost structure used in (1). Let h_0 denote the cost of carrying a blank unit in inventory in a given period, and reinterpret h_j as the *incremental* cost of holding a unit of item j in inventory during a given period (beyond h_0). Without loss of generality, $h_j \geq 0$ for all $j = 1, \ldots, J$. We account for all *expected* inventory-carrying costs, by charging the basic rate h_0 for each unit in the echelon inventory and for all $j = 1, \ldots, J$, $\alpha^l h_j$, the incremental holding cost for every *expected* unit of inventory of finished good j, l periods later. The actions at the start of every period consist of (i) $w =$ the number of new blanks to be ordered, (ii) $z =$ the amount to be withdrawn from the inventory of blanks to start the second manufacturing stage, and (iii) the vector $(z_1, \ldots, z_J)$ with z_j the part of z allocated to final item j.

This intractable dynamic program can again be approximated, this time by a *pair* of interdependent dynamic programs for *single*-item inventory systems. The approximation starts, again, with a relaxation of the nonnegativity constraints $z_j \geq 0$. Under this relaxation, the current choice of $(z_1, \ldots, z_J)$ in no way affects the future achievable values of the items' inventory positions or their distributions. We may thus choose the vector $(z_1, \ldots, z_J)$ to minimize immediate costs only, i.e., to solve Problem (3) in periods of type k.

We refer to the pair of single-dimensional dynamic programs as the "finished goods model" and the "blanks model." The former is defined by the following recursion, resembling (2),

$$V_n(X,k) = \min_{z\geq 0}\Big\{R^k(X+z) + \alpha \mathsf{E}\Big[V_{n-1}\Big(X+z-\sum_{j=1}^{J} d_j^k, k^+\Big)\Big]\Big\}. \quad (16)$$

Since $R^k(\cdot)$ is convex and orders in this model are unbounded, Zipkin (1989) shows that a periodic base-stock policy solves the problem and that $V_n(\cdot,k)$ is convex with a finite minimizer $X^*_{n,k}$, for all n and k. The optimal base-stock policy in (16) fails, by itself, to be an optimal allocation policy, simply because its recommended withdrawal quantity may be infeasible given a limited inventory of blanks. However, similar to the Clark and Scarf (1960) model, the functions $V_n(\cdot,k)$ are used to obtain at an optimal strategy for the lower-bound model.

It is shown (Aviv and Federgruen 1999) that if the available echelon inventory Y, at the beginning of period n, is in excess of $X^*_{n,k}$, the optimal withdrawal quantity z^* in (16) is in fact feasible (and optimal). If $Y < X^*_{n,k}$, the withdrawal quantity is limited to $(Y - X_{n,k})$ and the expected finished-goods–related costs are correspondingly higher. We thus define for all n and k, the *induced penalty functions* $P_n(Y,k)$ by

$$P_n(Y,k) = \begin{cases} 0 & \text{if } Y\geq X^*_{n,k} \\ R^k(Y) - R^k(X^*_{n,k}) \\ \quad +\alpha\mathsf{E}\big[V_{n-1}\big(Y-\sum_{j=1}^{J} d_j^k, k^+\big)\big] \\ \quad -\alpha\mathsf{E}\big[V_{n-1}\big(X^*_{n,k}-\sum_{j=1}^{J} d_j^k, k^+\big)\big] & \text{if } Y < X^*_{n,k}, \end{cases}$$

which are convex and nonnegative. These functions are used to specify the *second* dynamic problem, i.e., the "blanks model," which is defined by the recursion

$$W_n\big(w^1,\ldots,w^L; v^d, k\big) = \min_{0\leq w\leq b^k}\Big\{\gamma^k w + h_0^k(v^d+w^L) + P_n(v^d+w^L,k) + \alpha\mathsf{E}\Big[W_{n-1}\Big(w, w^1,\ldots,w^{L-1}; v^d + w^L - \sum_j d_j^k, k^+\Big)\Big]\Big\}. \quad (17)$$

The following can be proven by induction; see Aviv and Federgruen (1999):

(a) Except for certain constant terms, and given the above cost-accounting schemes, the expected total cost with n periods to go, and given that the current period is of type k, is represented by $V_n(X,k) + W_n(w^1,\ldots,w^L; v^d, k)$.

(b) By Aviv and Federgruen (1997a) and the convexity of the $P_n(\cdot,k)$ functions, a modified base-stock policy with period-dependent base-stock levels optimizes (17) and can be found by the simple value-iteration method described there. This policy is an optimal order policy for the lower-bound model. For this lower-bound model, it is further optimal to set the withdrawal quantity as close as feasible to $\{X^*_{n,k}\}$ and to allocate it so as to solve the optimization problem in (3).

The lower-bound model thus provides an easily computable cost estimate. The policy optimizing the model can be transformed into an effective heuristic strategy for the original system; e.g., leave the order and withdrawal policy unaltered and determine allocations via the allocation problems (P^k). Enhancements similar to those suggested in §2.1 can be developed.

6.2. Models with Parameter Learning

Our basic model assumes that all demand distributions are perfectly known from the outset. This assumption is often restrictive: Many products have a short life cycle or are subject to dynamic and competitive market forces. Thus, even the most basic characteristics of the demand distributions (e.g., their means) may fail to be known and parameter estimates can be significantly improved on the basis of observed sales data. Fisher and Raman (1996) show, for certain fashion items, that the accuracy of forecasts can be improved dramatically after observing the first fifth of the sales season. Delayed product differentiation allows one to use observed sales data during the first stage, not just to get updated information about the products' inventory status at the completion of the first stage, but also to base allocations to the individual products on more accurate forecasts of future demand distributions. The same applies when demands are correlated over time, even in the absence of parameter uncertainty.

See Aviv and Federgruen (1997b) for models with parameter learning and intertemporal correlation, analyzed in a Bayesian manner. The approximating analytical models are more complex (resulting in two- or three-dimensional dynamic programs) but remain tractable, in particular since the optimal strategies in these models continue to be of fairly simple structure. One general conclusion in these models is that the benefits of postponement *increase* under parameter learning beyond those in settings where the parameters (e.g., the mean demands) are known in advance. In fact, the larger the degree of initial uncertainty regarding the demand distribution or the larger the degree of intertemporal correlation, the larger the benefits of postponement are. See Aviv and Federgruen (1997b) for a detailed development.

References

Aviv, Y. 1998. Planning models for the design of capacitated multi-stage production and distribution systems. Ph.D. dissertation, Columbia University, New York, UMI #9838877.

——, A. Federgruen. 1997a. Stochastic inventory models with limited production capacity and periodically varying parameters. *Prob. Engrg. Inform. Sci.* **11** 107–135.

——, ——. 1997b. Design for postponement: A comprehensive characterization of its benefits under unknown demand distributions. *Oper. Res.* Forthcoming.

——, ——. 1998. The benefits of design for postponement. S. Tayur, M. Magazine, R. Ganeshan, eds. *Quantitative Modeling for Supply Chain Management*. Kluwer Academic Publishers.

——, ——. 1999. Technical addendum to: Capacitated multi-item inventory systems with random and seasonally fluctuating demands: Implications for postponement strategies. Washington University, St. Louis, MO, and Columbia University, New York, NY.

Beyer, D., S. P. Sethi. 1997. Average cost optimality in inventory models with Markovian demands. *J. Optim. Theor. Appl.* **92**(3) 497–526.

Cheng, F., S. P. Sethi. 1995. Optimality of state-dependent (s, S) policies in Markovian demand inventory models with lost sales. Working paper, Faculty of Management, University of Toronto, Ontario, Canada.

——, Y. S. Zheng. 1994. Lower bounds for multi-echelon stochastic inventory systems. *Management Sci.* **40** 1426–1443.

Ciarallo, F. W., R. Akella, T. E. Morton. 1994. A periodic review, production planning model with uncertain capacity and uncertain demand—optimality of extended myopic policies. *Management Sci.* **40** 320–332.

Clark, A. J., H. Scarf. 1960. Optimal policies for a multi-echelon inventory problem. *Management Sci.* **6** 475–490.

Eppen, G., L. Schrage. 1981. Centralized ordering policies in a multiwarehouse system with leadtimes and random demands. L. Schwarz, ed. *Multi-Level Production/Inventory Control Systems: Theory and Practice.* North-Holland, Amsterdam, The Netherlands, 51–69.

Evans, R. 1967. Inventory control of a multiproduct system with a limited production resource. *Naval Res. Logist. Quart.* **14** 173–184.

Federgruen, A. 1992. Centralized planning models for multi-echelon inventory systems under uncertainty. S. Graves, A. Rinnooy Kan, P. Zipkin, eds. *Handbook in Operations Research and Management Science.* North-Holland, Amsterdam, The Netherlands.

——. 1993. Recent advances in production and distribution management. R. Sarin, ed. *Perspectives in Operations Management*. Kluwer Academic Publishers, Norwell, MA.

——. 1999. Comments on: Variability reduction through operations reversal in supply chain re-engineering. Working paper, Columbia University, New York.

——, P. Zipkin. 1984a. Approximations of dynamic, multilocation production and inventory problems. *Management Sci.* **30** 69–84.

——, ——. 1984b. Allocation policies and cost approximations for multilocation inventory systems. *Naval Res. Logist. Quart.* **31** 97–129.

——, ——. 1984c. Computational issues in an infinite horizon multi-echelon inventory model. *Oper. Res.* **32** 818–832.

Fisher, M. L., A. Raman. 1996. Reducing the cost of demand uncertainty through accurate response to early sales. *Oper. Res.* **44** 87–99.

Flaherty, M. T. 1996. *Global Operations Management*. McGraw-Hill.

——, M. Cohen, L. Kopczak, H. Lee, D. Pyke. 1996. Teaching note for Hewlett Packard: DeskJet printer supply chain (A). The Wharton School, University of Pennsylvania, Philadelphia, PA.

Fox, B. 1966. Discrete optimization via marginal analysis. *Management Sci.* **4** 36–153.

Gallego, G., P. Zipkin. 1997. Bounds, heuristics and approximations for distribution systems. Working paper, Columbia University, New York, NY, and Duke University, Durham, NC.

Garg, A., H. L. Lee. 1998. Managing product variety: An operations perspective. S. Tayur, M. Magazine, R. Ganeshan, eds. *Quantitative Modeling for Supply Chain Management*. Kluwer Academic Publishers.

Gonsalvez, D. 1991. Integrating sales and manufacturing order management processes, applied scheduling: Integration of disciplines. Working paper, Columbia University, New York.

Gross, O. 1956. A class of discrete-type minimization problems. Working paper RM-1644-PR, The RAND Corporation, Santa Monica, CA.

Kapuscinski, R., S. Tayur. 1998. A capacitated production-inventory model with periodic demand. *Oper. Res.* **46**(6) 899–911.

——, ——. 1999. Variance vs. standard deviation—variability reduction through operations reversal. *Management Sci.* **45** 765–767.

Kopczak, L., H. Lee. 1993. Hewlett Packard: DeskJet printer supply chain (A). Case study, Stanford University, Stanford, CA.

Lee, H. 1993. Design for supply chain management: Concepts and examples. R. Sarin, ed. *Perspectives in Operations Management*. Kluwer Academic Publishers, Norwell, MA.

Lee, H. L., C. Billington. 1995. The evolution of supply chain models and practice at Hewlett-Packard. *Interfaces* **25**(5) 42–63.

——, ——, B. Carter. 1993. Hewlett-Packard gains control of inventory and service through design for localization. *Interfaces* **23**(4) 1–11.

——, C. S. Tang. 1997. Modeling the costs and benefits of delayed product differentiation. *Management Sci.* **43**.

——, ——. 1998. Variability reduction through operations reversal. *Management Sci.* **44** 162–173.

Metters, R. 1998. Producing multiple products with stochastic seasonal demand and capacity limits. *J. Oper. Res. Soc.*, **49** 263–272.

Signorelli, S., J. L. Heskett. 1989. Benetton (A). Working paper, Harvard Business School, Boston, MA.

Song, J. S., P. Zipkin. 1993. Inventory control in a fluctuating demand environment. *Oper. Res.* **41** 351–370.

Zipkin, P. 1982. Exact and approximate cost functions for product aggregates. *Management Sci.* **28** 1002–1012.

——. 1989. Critical number policies for inventory model with periodic data. *Management Sci.* **35** 71–80.

Accepted by Luk Van Wassenhove; received November 24, 1997. This paper was with the authors 10 months for 2 revisions.