

Optimal Inventory Control with Retail Pre-Packs

Long Gao

The A. Gary Anderson School of Management, University of California, Riverside, CA 92521, long.gao@ucr.edu

Douglas J. Thomas

Smeal College of Business, The Pennsylvania State University, University Park, PA 16802, USA, dthomas@psu.edu

Michael B. Freimer

SignalDemand, San Francisco, CA 94111, USA, michael.freimer@signaldemand.com

A *pre-pack* is a collection of items used in retail distribution. By grouping multiple units of one or more stock keeping units (SKU), distribution and handling costs can be reduced; however, ordering flexibility at the retail outlet is limited. This paper studies an inventory system at a retail level where both pre-packs and individual items (at additional handling cost) can be ordered. For the single-SKU, single-period problem, we show that the optimal policy is to order into a “band” with as few individual units as possible. For the multi-period problem with modular demand, the band policy is still optimal, and the steady-state distribution of the target inventory position possesses a semi-uniform structure, which greatly facilitates the computation of optimal policies and approximations under general demand. For the multi-SKU case, the optimal policy has a generalized band structure. Our numerical results show that pre-pack use is beneficial when facing stable and complementary demands, and substantial handling savings at the distribution center. The cost premium of using simple policies, such as strict base-stock and batch-ordering (pre-packs only), can be substantial for medium parameter ranges.

Key words: periodic review policies, pre-pack, batch ordering, steady-state distribution, demand correlation, Markov Decision Process.

History: Received: April 2011; Accepted: November 2013 by Panos Kouvelis, after 3 revisions.

1. Introduction

In production and distribution systems, materials often flow in fixed batch sizes or *pre-packs*, see, e.g., Litchfield and Narasimhan (2000), Smunt and Meredith (2000), and Blackburn and Scudder (2009). The use of pre-packs has the advantage of smoothing production (Chao et al. 2005), increasing the productivity in warehouses, reducing the number of order lines for retail stores, and improving the operating efficiency of supply chains (van der Vlist 2007). On the other hand, goods that are produced and shipped in large quantities or pre-packs are often disaggregated into individual items (called *loose* hereafter) at a break-bulk point (i.e., a distribution center) before continuing on to the point of final consumption. While this break-bulk operation incurs substantial facilities and handling costs, it might be worthwhile since it permits retail locations to receive small, frequent replenishment.

This work was motivated by discussions with a U.S.-based apparel retailer with essentially two aforementioned options for the distribution of goods. The first option is to have overseas contract manufacturers

produce, pack and ship a large quantity of a single SKU which is then unpacked at a domestic distribution center (Gao et al. 2013b). As stores place replenishment orders, individual items are selected and shipped from the distribution center. Option two is to have overseas manufacturers package either multiple units of a single SKU or multiple units of *similar* SKUs (such as different sizes of the same style and color shirt) into smaller collections called *pre-packs*. These pre-packs are also received at the distribution center where they are subsequently picked and shipped to retail stores. These options are illustrated in Figure 1.

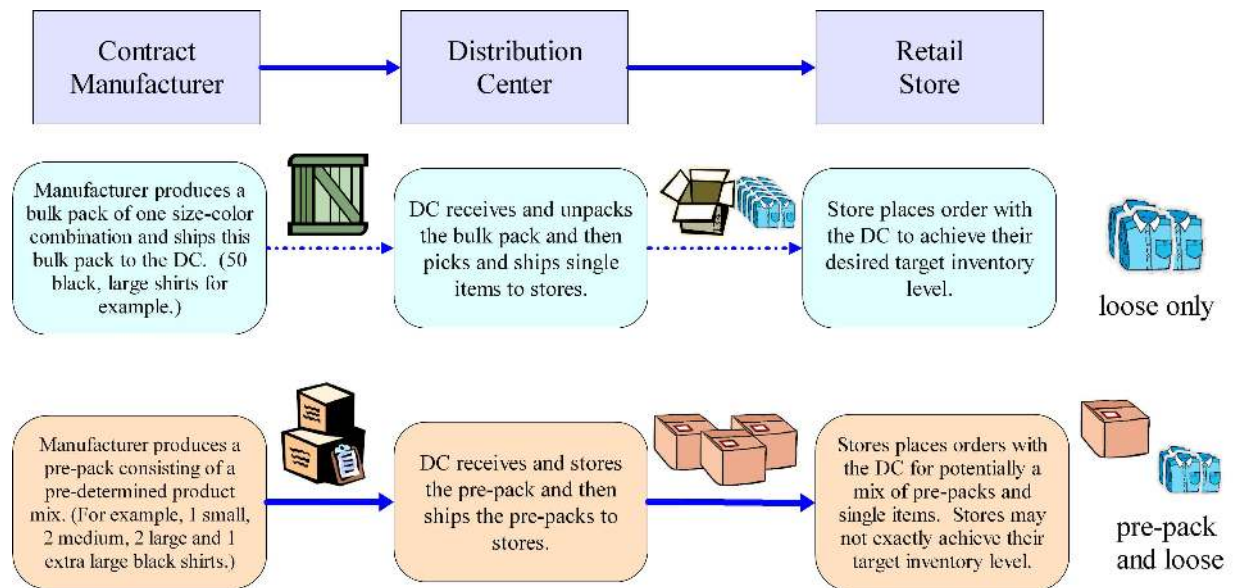


Figure 1 Typical flow of goods for both pre-pack and non-pre-pack distribution.

The advantage of the pre-pack option in this case is that handling costs at the distribution center can be substantially reduced, but ordering flexibility at the store is also limited. For the single-SKU setting, if there is only one pre-pack configuration, and the store can *only* order pre-packs, the optimal inventory policy for minimizing linear holding and shortage costs at the store is some version of the (R, nQ) policy—order an integer number of pre-packs of size Q to bring the inventory above R once the inventory position is below the reorder point R . In practice, the retail store may be able to order pre-packs as well as individual items, and pre-packs may include multiple SKUs. For such settings, it is no longer clear what the optimal policy will be.

This paper addresses the following inventory control questions in the presence of pre-pack and loose ordering options: (1) What is the optimal policy if both pre-packs and loose can be ordered? (2) What are the effects of the cost structure and the demand profile on the performance of the optimal policy? (3) What is the cost of employing some commonly used simple policies? When are they good approximations to the optimal policy?

To answer these questions, we consider a periodic-review inventory system with one distribution center and one retail store. The random consumer demand occurs at the retail store only. Besides ordering an integer number of pre-packs, the store can also order loose items at unit incremental handling penalty. Unfulfilled demand is fully backlogged. The inventory policy seeks to minimize the sum of additional handling cost of ordering loose units, inventory holding and backlogging costs.

Analytically, we characterize the structure of the optimal policy for a variety of scenarios, ranging from traditional base-stock policy to the well-studied pre-pack-only (R, nQ) policy. In particular, we show that for the single-SKU, single period case, the optimal policy under certain conditions possesses a simple band structure where one orders the minimum number of loose items to get into this band. For the multi-period case with Q -modular demand, the optimal policy still has the band structure due to the Q -periodicity of the value function; additionally the target inventory position is semi-uniformly distributed in steady state. Based on the band structure and the steady-state distribution of the inventory position, we develop an efficient procedure to compute the optimal policy for modular demand as well as an effective approximation for arbitrary discrete demand with sufficiently large volume. The band structure also extends to the multi-SKU case, albeit in a more complicated form. Intuitively, when the additional cost of ordering individual units is prohibitively large, the pre-pack-only policy is optimal for the single-SKU setting. Interestingly, under full backlogging, this is not the case for the multi-SKU setting for anything other than perfectly positively correlated demands, because the pre-pack-only policy cannot control the growing demand disparities between distinct SKUs over time. We explore this issue further in §4.

An important question related to the inventory control policy is the *design* of the pre-pack. In general, accumulating more items into a pre-pack leads to greater handling savings but less ordering flexibility at the retail outlet. In addition to this trade-off, firms must contend with different product mixes across retail locations. For one particular product category, the apparel retailer mentioned above used two different kinds of pre-packs: (1) a collection of different *sizes* of the same style-color combination and (2) a collection of different *colors* of the same size-style (known as a “rainbow-pack.”) Consider the size-mix issue and the data summarized in Table 1. This average size distribution shown in this table suggests that a pre-pack consisting of 1 small, 2 medium, 2 large and 1 extra large shirts might work reasonably well for most stores, and in fact, this 1-2-2-1 size mix was precisely the pre-pack configuration in use. As suggested by the min and max values in Table 1 though, such a pre-pack may work poorly for some stores. Some stores sold no shirts in sizes small and medium while those sizes comprised 70% of sales at some other stores. In addition to the product mix variation across stores, correlation in demand across items grouped together will affect the utility of a particular pre-pack configuration. We first address the inventory control issues for a given pre-pack configuration and then investigate the pre-pack *design* implications by varying pre-pack size and demand correlation across SKUs.

	Small	Medium	Large	Extra Large
Avg	17%	29%	32%	22%
Min	0%	0%	16%	0%
Max	41%	54%	100%	40%

Table 1 Historical percentage of sales by size across over 400 apparel retail outlets in North America

2. Literature Review

This research is related to inventory control with batch ordering and inventory models with two batch sizes. Batch-ordering policies have been well studied under various settings. For a single location, Morse (1958) first studies a batch ordering (R, nQ) policy in a periodic review system. Veinott (1965) shows that the (R, nQ) policy is optimal in both the finite and infinite period settings with zero setup cost, linear inventory holding and shortage penalty costs. Zheng and Chen (1992) derive the long run average cost (holding and backorder cost) associated with an (R, nQ) policy and develop an efficient heuristic algorithm to compute the policy parameters with lower and upper bounds. Chen (2000) generalized Veinott's optimality result to multi-echelon settings where each stage is restricted to a batch-ordering policy. Chao and Zhou (2009) further extend Chen's work to the multi-echelon setting with both batch ordering and fixed replenishment intervals. Cachon (2001) provides exact valuation methods for the periodic review systems with one warehouse and multiple identical retailers under a batch-ordering policy. Axsäter (2000) gives exact results for continuous review systems with compound Poisson demand and nonidentical retailers. Andersson et al. (1998) investigate the (R, nQ) policies in coordinating a decentralized inventory system where the stochastic lead times perceived by the retailers are replaced by their correct averages. Based on a simple approximation, they develop near-optimal reorder points and bounds for the approximation errors. Berling and Marklund (2006) examine the performance of continuous review (R, Q) policies in a one-warehouse multiple-retailer context, and offer a practical means to achieve coordinated control of large size systems. In contrast to the aforementioned work with only batch ordering options, our paper deals with situations with both pre-pack and loose ordering options. Moreover, we establish the suboptimality of the (R, nQ) policy in the general multi-SKU case.

A key notion in analyzing many batch-ordering systems is the uniformity property—the steady state distribution of the inventory position after ordering is uniformly distributed. This property was first shown by Hadley and Whitin (1961) under the assumption of Poisson demands and constant or gamma lead time. Song (2000) extends this uniform equilibrium distribution result to the multiple item setting. Chen (1998) extended the uniformity property to a continuous review, two-echelon inventory system with interdependent demands. We contribute to the literature by establishing a semi-uniform property for the system with both ordering options and modular demand.

The research closest to our study is Henig et al. (1997), and Parkinson and McCormick (2005). Henig et al. (1997) analyze a joint inventory control and contract design problem in which a fixed quantity R is

available at no incremental cost and at a given frequency. Quantities greater than R may be ordered with a per-unit penalty. As with the problem addressed in this paper, the base stock policy is not optimal; instead the policy has two critical levels which determine whether an amount greater than R should be acquired. Parkinson and McCormick (2005) also consider a problem similar to the one discussed in this paper, motivated by a chemical supplier with large and small ocean tankers. In this problem a single product is available via two delivery sizes (pre-packs), one of which is twice as large as the other (our model does not have this restriction). However, neither of them considers multi-SKU pre-packs (and consequently the impact of the demand correlation) and steady-state distribution of the inventory position. Chen et al. (2012) address the problem of designing and ordering pre-packs for pre-season planning. The authors develop an optimization model and heuristic solution approach for pre-season planning and ordering assuming deterministic demand. Our work is complementary to theirs in the sense that we focus on the ongoing management of the inventory system under stochastic demand.

3. Single SKU Pre-pack

The primary benefit of using a pre-pack is a reduction in handling costs through the distribution network. We model this handling efficiency by specifying an incremental per-unit cost penalty δ for units ordered loose. We assume $\delta > 0$ unless specified otherwise. Table 2 summarizes the main notation.

Table 2 Notation

ξ	random demand; $\xi \sim f$; $\mathbb{E}\xi = \mu$, $\text{var}\xi = \sigma^2$
b	unit backorder cost (adjusted)
h	unit holding cost (adjusted)
δ	unit penalty for ordering loose items
β	discount factor; $\beta \in (0, 1)$
q	loose order quantity for the 1-SKU case; $\mathbf{q} \in \mathbb{R}^k$ for the k -SKU case
Q	pre-pack size; $\mathbf{Q} \in \mathbb{R}_+^k$ for the k -SKU case
x	initial inventory with recurrent state space \mathcal{X} for the 1-SKU case; $\mathbf{x} \in \mathbb{R}^k$ for the k -SKU case
y	post pre-pack position; $y = x + nQ$; \mathbf{y} for the k -SKU case
I	target position; $I = y + q$; \mathbf{I} for the k -SKU case
R	indifference level; \mathbf{R} for the k -SKU case
S	base stock level; \mathbf{S} for the k -SKU case
Z	loose stock level; \mathbf{Z} for the k -SKU case
$\underline{y}(x)$	pre-pack function; the highest inventory position below $R + Q$ achievable by ordering pre-packs
$\mathbf{1}_A$	indicator function of event A

We consider a periodic-review, single-product inventory system with both pre-pack and loose ordering options to fulfill stationary random demand $\xi \sim f$. The unit variable costs for ordering pre-packs and loose items are c_0 and c_1 ; hence the unit penalty cost $\delta = c_1 - c_0$. The sequence of events in each period unfolds as follows: First, we observe the *initial inventory* $x \in \mathcal{X}$. Then an order is placed for $n \in \mathbb{Z}_+$ pre-packs of size $Q \in \mathbb{R}_+$ and $q \in \mathbb{R}_+$ loose units to reach the target inventory position $I = x + q + nQ$. The order arrives, and random demand ξ is realized and satisfied to the extent possible. Unfulfilled demand is fully backlogged at unit shortage cost \tilde{b} . The leftover inventory incurs unit holding cost \tilde{h} . The objective is to minimize the

expected total discounted cost $\tilde{V}(x)$ for an infinite horizon. The problem can be formulated as a Markov decision process with the following Bellman equation:

$$\tilde{V}(x) = \min_{y=x+nQ: n \in \mathbb{Z}_+} \min_{I \geq x+nQ} \{ c_0(y-x) + c_1(I-y) + \tilde{h}\mathbb{E}(I-\xi)^+ + \tilde{b}\mathbb{E}(\xi-I)^+ + \beta\mathbb{E}\tilde{V}(I-\xi) \}.$$

This formulation can be simplified by applying the transformation $V(x) = \tilde{V}(x) + c_0x - (1-\beta)^{-1}c_0\mu$ (Song and Zipkin 1993), and the identities $I = \mathbb{E}(I-\xi) + \mu = \mathbb{E}(I-\xi)^+ - \mathbb{E}(\xi-I)^+ + \mu$:

$$V(x) = \min_{n \in \mathbb{Z}_+} \min_{I \geq x+nQ} \{ \delta(I-x-nQ) + h\mathbb{E}(I-\xi)^+ + b\mathbb{E}(\xi-I)^+ + \beta\mathbb{E}V(I-\xi) \}, \quad (1)$$

with the *adjusted* costs $\delta = c_1 - c_0$, $h = \tilde{h} - (1-\beta)c_0$, and $b = \tilde{b} + (1-\beta)c_0$.

In the remainder of the section, we first characterize the structure of the optimal policy for the single-period model in §3.1, then extend it to the multi-period setting and derive the steady state distribution for the target inventory position in §3.2.

3.1. Single Period Analysis

Because the ordering sequence is inconsequential, we can solve problem (1) in two steps: First, solve the pre-pack problem for ordering n pre-packs to the *pre-pack position* $y = x + nQ$; second, solve the loose problem for ordering q loose units to the *loose position* $I = y + q$. Hence the optimal policy (y^*, I^*) prescribes the two positions for each initial inventory x by solving the following program:

$$V(x) = \min_y \{ H(y) : y = x + nQ, n \in \mathbb{Z}_+ \}, \quad \text{pre-pack problem} \quad (2)$$

$$H(y) = \min_I \{ \delta(I-y) + G(I) : I \geq y \}, \quad \text{loose problem} \quad (3)$$

$$G(I) = h\mathbb{E}(I-\xi)^+ + b\mathbb{E}(\xi-I)^+, \quad (4)$$

where $G(I)$ is the inventory cost, including the holding and shortage costs. It is easily verified that $G(I)$ is convex and coercive. $H(y)$ is also convex and coercive for $\delta > 0$ (by the convexity of G , and parts (b), (c), and (d) of the convexity Lemma 1 in Appendix).¹ See Figure 2 for a depiction of these properties: function $G(I)$ is a green, dashed line, while functions $\{ \delta(I-y) + G(I) \}_y$, parameterized by y , are blue, solid lines with the exception of the function from this set with y equal to the indifference target R defined below (red and represented with “—”). This curve has particular significance in characterizing the optimal policy as explained below.

To characterize the optimal policy, we define three anchoring points and a function: (1) pre-pack target S , (2) loose target Z , (3) indifference target R , and (4) the pre-pack function $\underline{y}(x)$.

First, define the *pre-pack target* $S \equiv \arg \min_I G(I)$, i.e.,

$$G'(S) = 0. \quad (5)$$

¹ A function $f : \mathbb{R} \rightarrow \mathbb{R}$ is coercive if $\lim_{\|x\| \rightarrow \infty} f(x) = \infty$.

Lemma 2.(a) below establishes that $S = \arg \min_y H(y)$. Thus, S is the unconstrained optimal inventory position for ordering pre-packs.

Second, define the *loose target* $Z \equiv \arg \min_I [\delta I + G(I)]$, i.e.,

$$Z = \begin{cases} -\infty, & \text{if } \delta + G'(I) > 0, \forall I \in \mathbb{R}, \\ [G']^{-1}(-\delta), & \text{otherwise.} \end{cases} \quad (6)$$

where $[G']^{-1}$ is the inverse function of the derivative of G . The convexity of G ensures that Z is well-defined. Clearly, $I = Z$ minimizes $\delta(I - y) + G(I)$ for each fixed y . Intuitively, Z is the *unconstrained* optimizer for the loose decision I .

Third, define the *indifference target* R by

$$H(R) = H(R + Q). \quad (7)$$

The convexity and coerciveness of H ensure that R is well-defined and that $R \leq S \leq R + Q$. Intuitively, R is the inventory position where the cost of ordering loose items up to Z (if feasible) is the same as the cost of ordering one more pre-pack to $R + Q$. Hence R facilitates the comparison of the pre-pack decisions $y = x + nQ, n \in \mathbb{Z}_+$.

Fourth, define the *pre-pack function* $\underline{y}(x)$ by

$$\underline{y}(x) \equiv \max \{ x + nQ : x + nQ \leq R + Q, n \in \mathbb{Z}_+ \}. \quad (8)$$

By construction, $\underline{y}(x)$ is Q -periodic for $x \leq R + Q$, i.e.,

$$\underline{y}(x) = \underline{y}(x - Q), \quad \forall x \leq R + Q. \quad (9)$$

Thus $\underline{y}(x)$ is the highest position less than or equal to $R + Q$ achieved by ordering only pre-packs given initial inventory $x \leq R + Q$. The purpose of $\underline{y}(x)$ is to partition the state space \mathcal{X} for characterization.

The anchoring points S , Z and R have the following properties. See Figure 2 for an illustration.

LEMMA 2. (a) The pre-pack target $S = \arg \min_y H(y)$, i.e., $H'(S) = 0$.

(b) $R \leq S \leq R + Q$ and $Z \leq S$.

The following theorem characterizes the optimal policy for the single-period, single-SKU problem.

THEOREM 1. Consider the single-period, single-SKU problem in (2).

(a) If $Z \leq R$, the (R, nQ) policy is optimal; i.e., order pre-packs only into the band $(R, R + Q]$. Formally,

$$(y^*, I^*) = (\underline{y}(x), \underline{y}(x)). \quad (10)$$

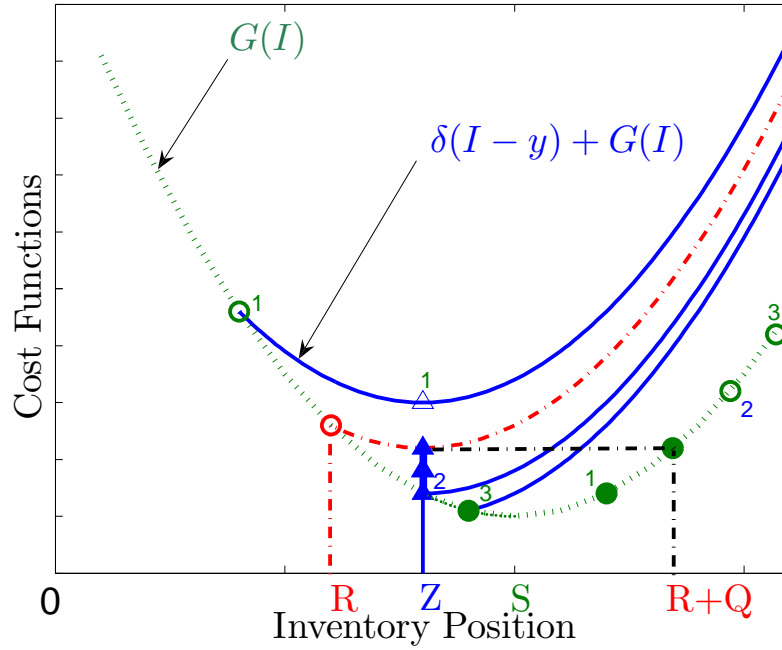


Figure 2 Anchoring points Z , R and S , Functions $G(I)$ and $\delta(I - y) + G(I)$, for the Single-SKU Optimal Policy

(b) If $Z > R$, the band $[Z, R + Q]$ policy is optimal. Specifically, if $\underline{y}(x) \in (R, Z]$, order pre-packs up to $\underline{y}(x)$ and then order $Z - \underline{y}(x)$ loose units; if $\underline{y}(x) \in (Z, R + Q]$, order up to $\underline{y}(x)$ using pre-packs only. Formally,

$$(y^*, I^*) = \begin{cases} (\underline{y}(x), Z), & \text{if } \underline{y}(x) \in (R, Z], \\ (\underline{y}(x), \underline{y}(x)), & \text{if } \underline{y}(x) \in (Z, R + Q]. \end{cases} \quad (11)$$

(c) if the penalty cost $\delta = 0$, then the base-stock S policy is optimal.

Theorem 1 provides a unified framework for characterizing the whole spectrum of inventory policies. Parts (a) and (c) establish, for the two extremes, the optimality of two well-known policies—the (R, nQ) and base-stock policies. Part (b) demonstrates that under certain conditions the optimal policy exhibits a *band structure*. Figure 2 illustrates this structure. (1) If by ordering pre-packs only, we can reach the positions in $(Z, R + Q)$, represented by a solid \bullet , numbered 1 or 3, then it is optimal to order pre-packs to reach one of these points and not order any loose units. Ordering more or fewer pre-packs to levels represented by empty \circ 1 or 3, would not be optimal. (2) If by ordering pre-packs one lands in $[R, Z]$, then it is optimal to reach Z (marked by \triangle) by ordering additional loose units; e.g., position $\triangle 2$ is optimal while $\circ 2$ —reached by ordering one more pre-pack—is not.

In the inventory literature, the band structure also arises in the (s, S) policy (Clark and Scarf 1960). However, the underlying mechanisms for order batching are different. In the (s, S) policy it is driven by the fixed ordering cost; in the $[Z, R + Q]$ policy it arises from the cheaper pre-pack ordering option.

THEOREM 2. *The optimal value function $V(x)$ for the problem (2) is Q -periodic for $x \leq R + Q$.*

Theorem 2 reveals that, unlike a conventional convex ordering cost (Gao and Yang 2012), the cost structure in the presence of both pre-pack and loose ordering options induces periodic behavior in both the optimal policy and the value function. This behavior substantially complicates the multi-period problem. Fig. 3 illustrates the impact of the periodic behavior on $V(x)$. As a result, the conventional backward induction approach for establishing structural results breaks down. Without additional assumptions, the exact characterization for the multi-period case appears unattainable.

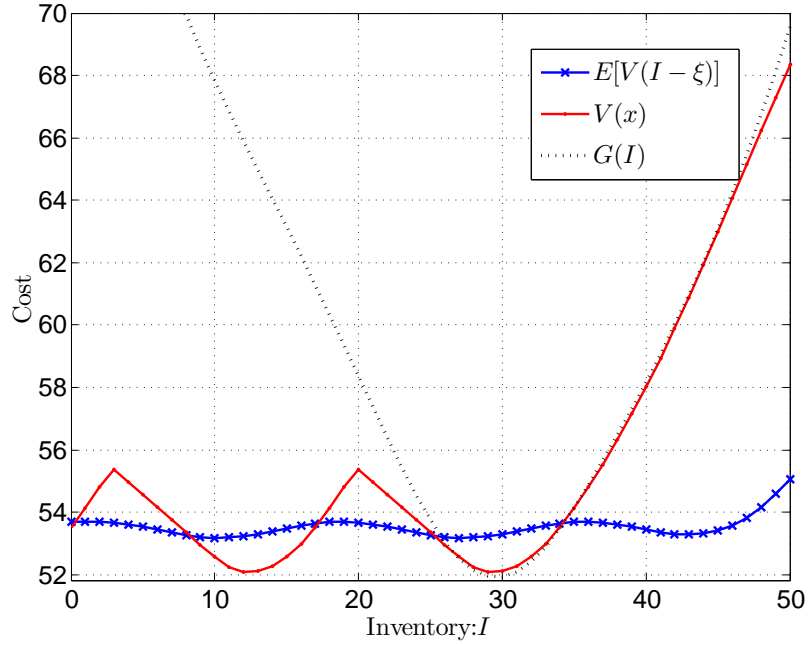


Figure 3 Non-convexity of value function $V(x)$ and $\mathbb{E}[V(I - \xi)]$: $\xi \sim \mathcal{N}(30, 5^2)$, $Q = 17$, $\delta = 0.4$, $h = 1$, $b = 1$

3.2. Multi-Period Analysis

The single-period model in §3.1 extends to the infinite horizon if we redefine $G(I)$ as

$$G(I) \equiv h\mathbb{E}(I - \xi)^+ + b\mathbb{E}(\xi - I)^+ + \beta\mathbb{E}V(I - \xi). \quad (12)$$

To extend the structural results in §3.1 by backward induction, we need to first establish the convexity of $G(I)$, which in turn requires the convexity of $V(x)$ (Li and Gao 2008, Gao et al. 2012, 2013a). Theorem 2, however, demonstrates that the Q -periodicity of $V(x)$ destroys its convexity even in the single-period case. Although the optimal policy can be obtained by solving (1) with standard value- or policy-iteration algorithms, without the guidance of the policy structure, we have to exhaustively search the space $\mathbb{Z}_+ \times \mathbb{R}$ for the optimal decision $(n(x), I(x))$ for each x . The computation becomes even more expensive for multi-SKU cases.

In the rest of this section, we focus on demand with the *modulus property* where sharp characterization is possible.² Let ξ_Q be the residue (remainder) of $\text{mod}(\xi, Q)$, i.e., $\xi = \xi_Q + nQ$, for some $n \in \mathbb{Z}_+$ and $0 \leq \xi_Q < Q$.

ASSUMPTION 1. *The random demand ξ is Q -modular; i.e., $\xi_Q \sim \mathcal{U}[0, Q - 1]$.*

That is, the residue ($\text{mod } Q$) of demand is uniformly distributed. In essence, it requires the demand distribution f satisfies

$$\sum_{n \in \mathbb{Z}_+} f(i + nQ) = \frac{1}{Q}, \quad \forall i = 0, \dots, Q - 1. \quad (13)$$

For example, any uniform distributions $\mathcal{U}[m, m + nQ - 1]$ $m, n \in \mathbb{Z}_+$, and triangular distributions

$$f(x) = \begin{cases} \frac{x}{(nQ)^2}, & x \in [0, nQ], \\ -\frac{x}{(nQ)^2} + \frac{2}{nQ}, & x \in [nQ, 2nQ], \end{cases}$$

are Q -modular. The following lemma is critical for characterization.

LEMMA 3. *If the random variable ξ is Q -modular, and the function $V(I)$ is Q -periodic for $I \leq R + Q$, then $\mathbb{E}V(I - \xi)$ is a constant for $I \leq R + Q$.*

We now characterize the optimal policy in the multi-period setting.

THEOREM 3 (**Optimal Policy**). *Consider the infinite horizon, single-SKU problem in (1) with Q -modular demand.*

- (a) *If $Z \leq R$, the (R, nQ) policy is optimal.*
- (b) *If $Z > R$, the band policy $[Z, R + Q]$ is optimal.*
- (c) *If the penalty cost $\delta = 0$, the base-stock policy with order-up-to level S is optimal.*
- (d) *The optimal value function $V(I)$ is Q -periodic for $I \leq R + Q$.*

For a given policy, the ending inventory I forms a Markov process. We now consider the stationary distribution ϕ of I . For the (R, nQ) policy, Hadley and Whitin (1961) proved the uniformity property of ϕ .

THEOREM 4 (**Hadley and Whitin, 1961**). *Consider a periodic review, stationary, infinite horizon (R, nQ) inventory system with random demand ξ . If the Markov chain is irreducible, then the steady-state inventory position is uniformly distributed over the set $\{R + 1, R + 2, \dots, R + Q\}$.*

The uniformity property is the central notion in the batch ordering literature. It holds for an arbitrary demand distribution. In the context of multiple pre-packs, we have shown that under $Z \leq R$ the optimal policy is also an (R, nQ) policy, and therefore the Hadley and Whitin theorem applies. Unfortunately there is no such general result when $Z > R$. We can, however, establish a similar theorem for Q -modular demand.

² We thank one of the anonymous referees for this key insight.

THEOREM 5 (Semi-uniform distribution). *Under the condition $Z > R$, if demand ξ is Q -modular, and the band policy $[Z, R + Q]$ is implemented, then the ending inventory has a semi-uniform distribution, with equal steady-state probabilities on the points $[Z + 1, R + Q]$ and larger mass at the point Z :*

$$\phi = \left(\frac{Z - R}{Q}, \frac{1}{Q}, \dots, \frac{1}{Q} \right). \quad (14)$$

Theorem 5 makes two theoretical contributions. The first is the simplicity of the distribution form. As such, it generalizes the uniformity property from the pre-pack-only case to the setting where both pre-pack and loose ordering options are available. The second contribution is that computational efficiency can be improved since the semi-uniform distribution enables a direct calculation of the per-period average cost. In view of the inventory balance equation $x = I - \xi$, the distribution $p(x)$ of the initial inventory x is given by $p(x) = \sum_{\xi} \phi(\xi + x) f(\xi)$; the per-period average cost is then obtained via $\mathbb{E}_{\xi}[\delta(I^*(x) - y^*(x)) + L(I^*(x))]$, where $L(I)$ is the inventory holding and backorder costs for policy parameters (Z, R) . As such, expensive dynamic programming algorithms are no longer needed to determine the optimal policy.

The semi-uniformity property and band policy structure also enable efficient computation for the arbitrary demand case. First, many real demands may follow or closely follow the normal or triangular distributions, which are likely to possess the modulus property when demand volume is large relative to the pre-pack size.³ Second, the residual periodic changes of the term $\mathbb{E}V(I - \xi)$ are further “smoothed” out by the expectation operator (e.g., $\mathbb{E}V(I - \xi)$ in Fig. 3), and thus the future value can be reasonably approximated by a constant. Consequently, the band policy serves as a good approximation for arbitrary demand cases. Indeed, our numerical studies confirm this approach as a band policy was optimal for all our problem instances. In summary, our structural results significantly reduce the computational complexity in the multi-period setting.

4. Multi-SKU Pre-pack

A retailer may include multiple SKUs in the same pre-pack. For example, as described in Figure 1, pre-packs could accumulate different sizes of the same style-color. In this section, we extend the single-period, single-SKU model in §3.1 to the k -SKU case. The random demand $\boldsymbol{\xi} = (\xi_1, \dots, \xi_k) \in \mathbb{R}_+^k$ has joint distribution function f . The store can order pre-packs containing $\mathbf{Q} \in \mathbb{R}_+^k$ units as well as loose for each SKU. If loose units $\mathbf{q} \in \mathbb{R}_+^k$ are ordered, the store pays a total penalty cost $\boldsymbol{\delta} \cdot \mathbf{q} = \sum_{i=1}^k \delta_i q_i$, where δ_i is the unit penalty cost for the i^{th} SKU, and q_i is the number of loose units of SKU i ordered. The rationale underlying the multi-SKU scenario is the same as in the single-SKU setting: the extra cost of ordering loose items should be justified by the savings in shortage cost. We follow the convention that vectors are in bold font, that all equalities and inequalities are in the point-wise sense, that the cube $[\mathbf{a}, \mathbf{b}] = \times_{i=1}^k [a_i, b_i]$, that $(a, b) = \emptyset$ if $a > b$.

³ For the normal distribution, this is based on the distribution symmetry and the three-sigma-rule. An example is provided in §5.1.

After observing the initial inventory $\mathbf{x} \in \mathcal{X} = \mathbb{R}^k$, we first order $n \in \mathbb{Z}_+$ pre-packs to the *pre-pack position* $\mathbf{y} = \mathbf{x} + n\mathbf{Q}$ and then order $\mathbf{q} \in \mathbb{R}_+^k$ loose units to the *target position* $\mathbf{I} = \mathbf{y} + \mathbf{q}$. We solve the following problem for the optimal policy $(\mathbf{y}^*, \mathbf{I}^*)$.

$$V(\mathbf{x}) = \min_{\mathbf{y}} \{ H(\mathbf{y}) : \mathbf{y} = \mathbf{x} + n\mathbf{Q}, n \in \mathbb{Z}_+ \}, \quad \text{pre-pack problem} \quad (15)$$

$$H(\mathbf{y}) = \min_{\mathbf{I}} \{ \boldsymbol{\delta} \cdot (\mathbf{I} - \mathbf{y}) + G(\mathbf{I}) : \mathbf{I} \geq \mathbf{y} \}, \quad \text{loose problem} \quad (16)$$

where $G(\mathbf{I}) \equiv \sum_{i=1}^k G_i(I_i)$, and $G_i(I_i) = h_i \mathbb{E}(I_i - \xi_i)^+ + b_i \mathbb{E}(\xi_i - I_i)^+$. It is easily verified that G_i , G , and $\boldsymbol{\delta} \cdot (\mathbf{I} - \mathbf{y}) + G(\mathbf{I})$ are convex and coercive for $\delta > 0$. $H(\mathbf{y})$ is also convex and coercive (by the convexity of G , the separability of $H(\mathbf{y})$ in each dimension, and parts (b), (c), and (d) of the convexity Lemma 1 in the Appendix).

Now we introduce a critical notion for the k -SKU case: direction line. Let $\ell_{\mathbf{x}} \equiv \{ \mathbf{x} + t\mathbf{Q} : t \in \mathbb{R} \}$ be the line passing through point \mathbf{x} with direction \mathbf{Q} . Let $h_{\mathbf{x}}(t) \equiv H(\mathbf{x} + t\mathbf{Q})$. Then the pre-pack problem (15), $V(\mathbf{x}) = \min_{n \in \mathbb{Z}_+} h_{\mathbf{x}}(n)$, is essentially a single dimensional problem of choosing $n \in \mathbb{Z}_+$ along the line $\ell_{\mathbf{x}}$. Viewed along the line $\ell_{\mathbf{x}}$, the function $h_{\mathbf{x}}(t)$ inherits convexity and coerciveness from $H(\mathbf{y})$.

We specify the three anchoring points \mathbf{Z} , $\mathbf{R}_{\ell_{\mathbf{x}}}$, and $\mathbf{S}_{\ell_{\mathbf{x}}}$ along with the pre-pack function $\underline{\mathbf{y}}(\mathbf{x})$ as follows. Define the *loose target* $\mathbf{Z} = \arg \min_{\mathbf{I}} [\boldsymbol{\delta} \cdot \mathbf{I} + G(\mathbf{I})]$, i.e., for SKU $i = 1, \dots, k$,

$$Z_i = \begin{cases} -\infty, & \text{if } \delta_i + G'_i(I_i) > 0, \forall I_i \in \mathbb{R}, \\ [G'_i]^{-1}(-\delta_i), & \text{otherwise.} \end{cases} \quad (17)$$

We now specify the $\ell_{\mathbf{x}}$ -specific points $\mathbf{R}_{\ell_{\mathbf{x}}}$ and $\mathbf{S}_{\ell_{\mathbf{x}}}$. Define the k -SKU *indifference target* $\mathbf{R}_{\ell_{\mathbf{x}}} \in \ell_{\mathbf{x}}$ by

$$H(\mathbf{R}_{\ell_{\mathbf{x}}}) = H(\mathbf{R}_{\ell_{\mathbf{x}}} + \mathbf{Q}), \quad (18)$$

i.e., we are indifferent between staying at $\mathbf{R}_{\ell_{\mathbf{x}}}$ and ordering one more pre-pack to reach $\mathbf{R}_{\ell_{\mathbf{x}}} + \mathbf{Q}$ on line $\ell_{\mathbf{x}}$. For each line $\ell_{\mathbf{x}}$, define the *pre-pack target* $\mathbf{S}_{\ell_{\mathbf{x}}} \equiv \arg \min_{\mathbf{y} \in \ell_{\mathbf{x}}} H(\mathbf{y})$, the unconstrained optimizer of $H(\mathbf{y})$ on $\ell_{\mathbf{x}}$. Define the *pre-pack function* $\underline{\mathbf{y}}(\mathbf{x})$ by

$$\underline{\mathbf{y}}(\mathbf{x}) \equiv \max_{n \in \mathbb{Z}_+} \{ \mathbf{x} + n\mathbf{Q} : \mathbf{x} + n\mathbf{Q} \leq \mathbf{R}_{\ell_{\mathbf{x}}} + \mathbf{Q} \}, \quad (19)$$

as the highest position below $\mathbf{R}_{\ell_{\mathbf{x}}} + \mathbf{Q}$ achievable by ordering only pre-packs given initial inventory $\mathbf{x} \leq \mathbf{R}_{\ell_{\mathbf{x}}} + \mathbf{Q}$. $\underline{\mathbf{y}}(\mathbf{x})$ is \mathbf{Q} -periodic; i.e., $\underline{\mathbf{y}}(\mathbf{x}) = \underline{\mathbf{y}}(\mathbf{x} - \mathbf{Q})$, for $\mathbf{x} \leq \mathbf{R}_{\ell_{\mathbf{x}}} + \mathbf{Q}$. Moreover, for each $\ell_{\mathbf{x}}$, $\mathbf{R}_{\ell_{\mathbf{x}}} \leq \underline{\mathbf{y}}(\mathbf{x}) \leq \mathbf{R}_{\ell_{\mathbf{x}}} + \mathbf{Q}$.

THEOREM 6 (Optimal Policy for k -SKU). Consider the problem (15)-(16).

(a) For initial inventory $\mathbf{x} \in \mathcal{X}$, the optimal policy $(\mathbf{y}^*, \mathbf{I}^*) = (y_i^*, I_i^*)_{i=1}^k$ is given by

$$(y_i^*, I_i^*) = \begin{cases} (\underline{y}_i(\mathbf{x}), Z_i), & \text{if } \underline{y}_i(\mathbf{x}) \in (R_{i, \ell_{\mathbf{x}}}, Z_i], \\ (\underline{y}_i(\mathbf{x}), \underline{y}_i(\mathbf{x})), & \text{if } \underline{y}_i(\mathbf{x}) \in (Z_i, R_{i, \ell_{\mathbf{x}}} + Q_i]. \end{cases} \quad (20)$$

(b) The optimal value function $V(\mathbf{x})$ is \mathbf{Q} -periodic for $\mathbf{x} \leq \mathbf{R}_{\ell_{\mathbf{x}}} + \mathbf{Q}$.

Theorem 6 generalizes the band structure from the 1-SKU to the k -SKU case. Under the optimal policy, the recurrent initial states $\mathbf{x} \in \mathcal{X}$ form a Markov process. Let $\mathcal{L}_{\mathcal{X}} \equiv \{\ell_{\mathbf{x}} : \mathbf{x} \in \mathcal{X}\}$ be the set of all parallel lines along direction \mathbf{Q} generated by \mathcal{X} . For each line $\ell \in \mathcal{L}_{\mathcal{X}}$, the ending inventory forms the line segment $[\mathbf{Z}, \mathbf{R}_{\ell} + \mathbf{Q}] \cap \ell$, where \mathbf{Z} and \mathbf{R}_{ℓ} are the loose target in (17) and line- ℓ specific indifference point in (18). By (18) and the convexity of H , \mathbf{R}_{ℓ} is continuous with respect to $\ell \in \mathcal{L}_{\mathcal{X}}$, and so is line segment $[\mathbf{Z}, \mathbf{R}_{\ell} + \mathbf{Q}] \cap \ell$. Consequently, the set $\mathcal{B} = \bigcup_{\ell \in \mathcal{L}_{\mathcal{X}}} [\mathbf{Z}, \mathbf{R}_{\ell} + \mathbf{Q}] \cap \ell$ of the ending inventories \mathbf{I} exhibits the *generalized band structure*; see Fig. 4 in §5 for an illustration.

Theorem 6 has important implications for practice. First, because the pre-pack decision $\underline{\mathbf{y}}(\mathbf{x})$ is decided jointly via (19), unlike the 1-SKU case, certain $y_i(\mathbf{x})$ may deviate multiple Q_i from the targeted Z_i . This calls for more frequent adjustment by ordering loose, which enhances the value of the loose ordering option in the k -SKU case. Second, the loose decision can be managed individually via (20), because I_i^* only depends on $y_i(\mathbf{x})$. Therefore, three classes of replenishment policies may work: (1) for SKU i with $Z_i \leq R_{i,\ell_{\mathbf{x}}}$, no loose is ordered and the optimal policy is $(R_{i,\ell_{\mathbf{x}}}, nQ_i)$; (2) for SKU with $Z_i \in (R_{i,\ell_{\mathbf{x}}}, R_{i,\ell_{\mathbf{x}}} + Q_i]$, the band $[Z_i, R_{i,\ell_{\mathbf{x}}} + Q_i]$ policy is optimal; (3) for SKU with $Z_i > R_{i,\ell_{\mathbf{x}}} + Q_i$, always order up to Z_i . While the first two classes of policies mirror Parts (a) and (b) of Theorem 1 for the single SKU case, interestingly, the third class for the k -SKU case never arises in the 1-SKU case. This is because space \mathbb{R}^1 is completely ordered and $Z \leq S \leq R + Q$ by Lemma 2.(b). In contrast, in the partially ordered space \mathbb{R}^k , \mathbf{Z} no longer shares the same line $\ell_{\mathbf{x}}$ with $\mathbf{R}_{\ell_{\mathbf{x}}}$, $\underline{\mathbf{y}}(\mathbf{x})$ and $\mathbf{S}_{\ell_{\mathbf{x}}}$ (except for the special case of the line $\ell_{\mathbf{Z}}$). This allows $\ell_{\mathbf{x}}$ -specific $R_{i,\ell_{\mathbf{x}}}$, S_i , and $y_i(\mathbf{x})$ to depart multiple Q_i away from Z_i for some SKU- i , which leads to this new class of control policies based on ordering additional loose.

We now examine the effects of demand correlation on the control policy and the pre-pack design.⁴

THEOREM 7. (a) *The pre-pack-only policy is suboptimal for the multi-SKU setting as long as demands across SKUs are not perfectly positively correlated (\exists SKU $i, j \leq k$, s.t. $\rho_{ij} < 1$).*

(b) *Consider choosing one of two SKUs i and j , to pack with a selected SKU s . Other things being equal, if correlation coefficients $\rho_{is} > \rho_{js}$, then SKU- i should be chosen.*

Part (a) reveals the key role of demand correlation across items in managing multi-SKU pre-packs. As demand correlation drives the distribution of \mathbf{x}_t via the dynamics $\mathbf{x}_t = \mathbf{I}(\mathbf{x}_{t-1}) - \boldsymbol{\xi}_{t-1}$, it shapes both the control policy and pre-pack design. When the demands (ξ_1, \dots, ξ_k) are *imperfectly correlated*, the pre-pack-only policy—the central notion of the batch ordering literature—can never be optimal for the k -SKU setting. This is because the cumulative demand disparity, and thus the difference in inventory position between SKUs, follows a random walk and can become arbitrarily large; thus, a loose ordering adjustment will be needed to avoid unbounded costs.

⁴ We thank the senior editor and the review team for this insight.

Part (b) provides guidance on pre-pack design. A main challenge of multi-SKU pre-pack design is to limit the random walk effect of demand disparities, thus reducing the need for expensive loose ordering to adjust inventory positions. Because the random walk effect is less pronounced for more positively correlated demands, firms should choose such SKUs when designing pre-packs. Indeed, when the demands are *perfectly positively correlated* ($\rho_{ij} = 1, \forall i, j = 1, \dots, k$), the problem simplifies to the single-SKU setting from §3, where either the pre-pack-only (R, nQ) or band policy is optimal.

5. Computational Results and Managerial Insights

This section quantifies the sensitivity of the optimal policy (§5.1) and the cost of employing suboptimal policies (§5.2).

5.1. Impact of δ, Q, cv and ρ on the Optimal Policy

We first investigate the impact of four key parameters: the penalty cost δ , demand variability cv , demand correlation ρ (across multiple SKUs), and pre-pack size Q . They affect the target inventory position region, the total cost TC , and the long-run fraction PP of goods obtained via pre-pack.

We design the tests on a two-SKU pre-pack setting as follows. We consider normal demand with mean $\mu_1 = \mu_2 = \mu = 5$, four levels of coefficients of variation cv (by changing σ), and three levels of demand correlation, $\rho \in \{-0.6, 0, 0.6\}$. Demand is first discretized over the grid $\{0, 1, \dots, \mu + 3\sigma\} \times \{0, 1, \dots, \mu + 3\sigma\}$ and then normalized to 1. The discretization and truncation effects result in actual demand $cv \in \{0.12, 0.38, 0.49, 0.60\}$. We set pre-pack size $Q_1 = Q_2 = Q \in \{3, 4, 5, 6\}$, and unit penalty $\delta \in \{0.01, 0.5, 1, 5\}$. The baseline case is specified as $\mu = 5, cv = 0.38, \rho = 0, Q = 4, \delta = 0.5, h = 1, b = 10, \beta = 0.99$.

The rationale for our experimental choices is as follows. We choose the normal distribution, because it fits empirical demand patterns well, permits different levels of demand correlation, and is likely to have the modulus property. For example, for the baseline case, the residual (mod Q) of the marginal demand has the distribution $(0.2504, 0.2543, 0.2504, 0.24500)$, approximately Q -modular. To ensure the nonnegativity of demands of high variability, we truncate and normalize the corresponding distributions. For accuracy, we compute the optimal solution using value iteration, a standard algorithm in dynamic programming (p.158, Puterman 1994).

Figure 4 depicts the optimal region of the post-ordering inventory position for varying levels of δ . Darker shading indicates greater frequency of visit to that inventory position—the steady-state distribution ϕ of I .⁵ We observe that higher penalty cost leads to higher total cost TC , more frequent deployment of pre-packs, and a wider region of the optimal inventory position. Indeed, for small penalty cost ($\delta = 0.01$), the base stock policy is optimal. When the penalty cost is substantial ($\delta = 0.5, 1, 5$), the optimal inventory region

⁵ Given the optimal policy, we calculate the stationary distribution, setting the largest value in a particular graph to black, zero value equal to white, with all other values on a linear grayscale in between these two extremes.

expands, indicating that the store will be more willing to tolerate the deviation from the optimal base stock level in order to avoid excess handling costs associated with loose ordering.

(μ_1, μ_2)	(5,5)	(5,6)	(5,4)	(5,7)
TC	7.62	10.42	10.17	12.37
PP	94.2%	89.1%	89.3%	87%
$1 - PP$	5.8%	10.9%	10.7%	13%

Table 3 The Impact of Demand Disparity on Loose Usage

Interestingly, even for high penalty cost $\delta = 5$, 5.8% of demand is still met via loose ordering. If this were the single-SKU case,⁶ no loose items would have been ordered, as later shown in Figure 8. This distinction highlights the second role of the loose option—controlling demand disparity across SKUs—the main insight of Theorem 7. Indeed, when the pre-pack composition does not perfectly match the average demand ratios, some loose items must be ordered to control this demand disparity. As shown in Table 3, while 5.8% loose ordering in case (5, 5) is purely driven by randomness, the increasing loose usage in the other three cases is driven by both random fluctuation and the mismatch between the pre-pack configuration and the demand profile. The result is also supported by size distribution data from our apparel retailer. As suggested by Table 1, although the configuration of size mix 1-2-2-1 works reasonably well for many stores, stores with atypically large or small size customer bases would require substantially more loose ordering.

Figure 5 records the optimal inventory position region for varying levels of pre-pack size Q . As Q increases, the optimal inventory position region expands and total cost increases, since a larger Q results in less ordering flexibility. With the same δ , ordering a larger pre-pack is less attractive to the store, so the fraction of goods obtained via pre-pack decreases as Q increases. Note that in these cases, for the sake of simplicity, we vary one parameter at a time. In practice, however, we would expect greater handling efficiencies at the DC with a larger pre-pack, resulting in a greater per-unit handling penalty cost δ associated with ordering loose items.

Figure 6 depicts the optimal inventory position region for varying levels of coefficient of variation cv . Higher relative demand variation results in a lower fraction of product obtained via pre-pack and higher total cost. Comparing the cases of $cv = 0.49, 0.60$, we see the distribution of inventory position under $cv = 0.60$ spreading out more evenly. This is because greater demand variability requires more frequent inventory adjustment via loose items.

Figure 7 depicts the optimal inventory position region for varying levels of demand correlation ρ . As shown in the figure, total cost TC decreases and the fraction PP of goods obtained via pre-pack increases as ρ increases. Recall that in this multi-SKU case, ordering via pre-packs means ordering along the diagonal

⁶ Or equivalently, if demands for the two SKUs are perfectly positively correlated.

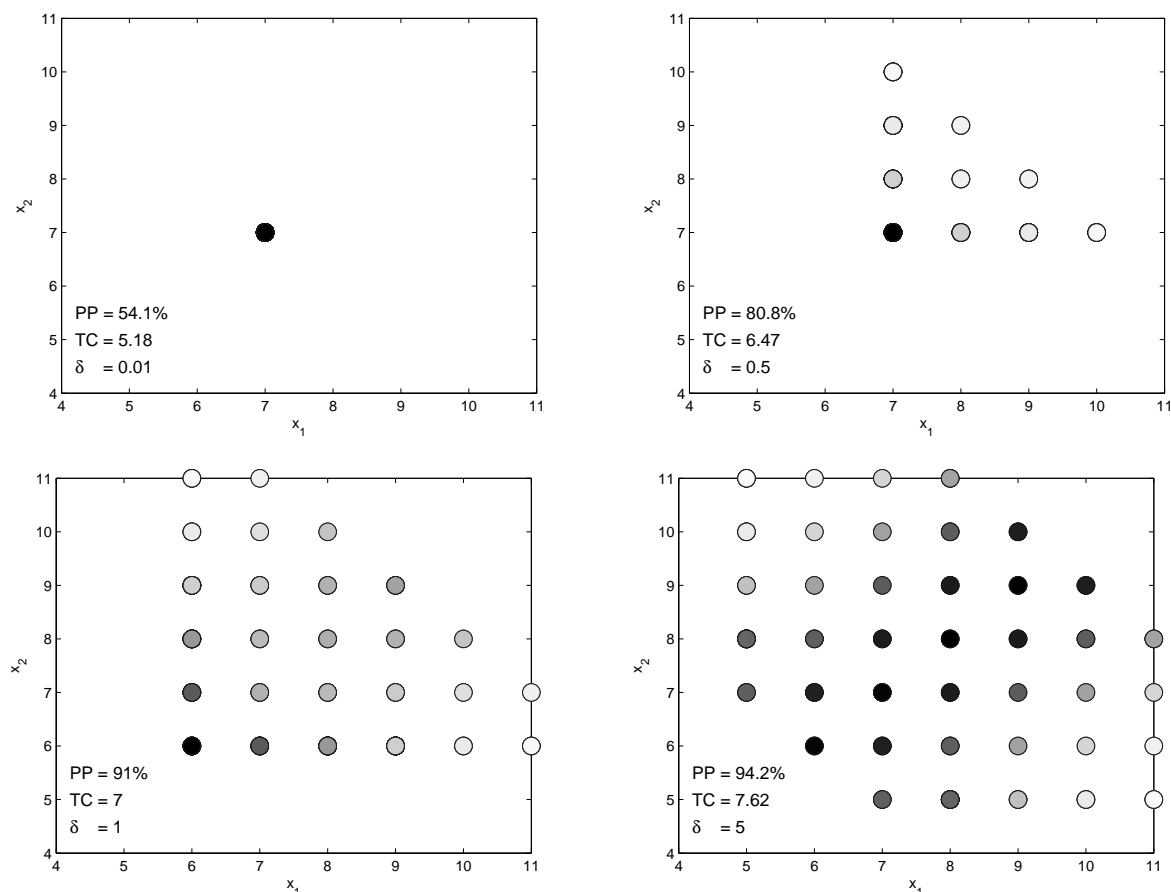


Figure 4 Optimal inventory position region (darker shading indicates greater frequency of occurrence in steady-state) with varying unit handling penalty cost $\delta \in \{0.01, 0.5, 1, 5\}$ with $Q = (4, 4)$, $cv = 0.38$, $\rho = 0$. Total expected system cost TC and long-run fraction of demand satisfied via pre-pack ordering PP are also shown.

determined by the pre-pack configuration. When demands are positively correlated, realized demand values are more likely to occur along a similar diagonal, allowing the store to increase its use of pre-packs. In this case, the optimal inventory region is unchanged for $\rho \in \{-0.6, 0, 0.6\}$, but one can see differences in the stationary distributions, with greater concentration on the $x_1 = x_2$ diagonal for $\rho = 0.6$. These observations suggest that, given a choice, a retailer would prefer to bundle positively correlated items in multi-SKU pre-packs. In our apparel retail example, the configuration of the 1-2-2-1 size mix is designed based in part on this principle that demands for different sizes of the same style-color tend to be positively correlated.

In summary, our observations in this section suggest that, from an operational perspective, a higher fraction of pre-packs should be used when demand is stable and the handling savings are substantial. From a design perspective, the right pre-pack size should properly balance the loss of ordering flexibility and the handling savings. Regardless of the cost ratios, pre-packs are more effective when they bundle items with positively correlated demands.

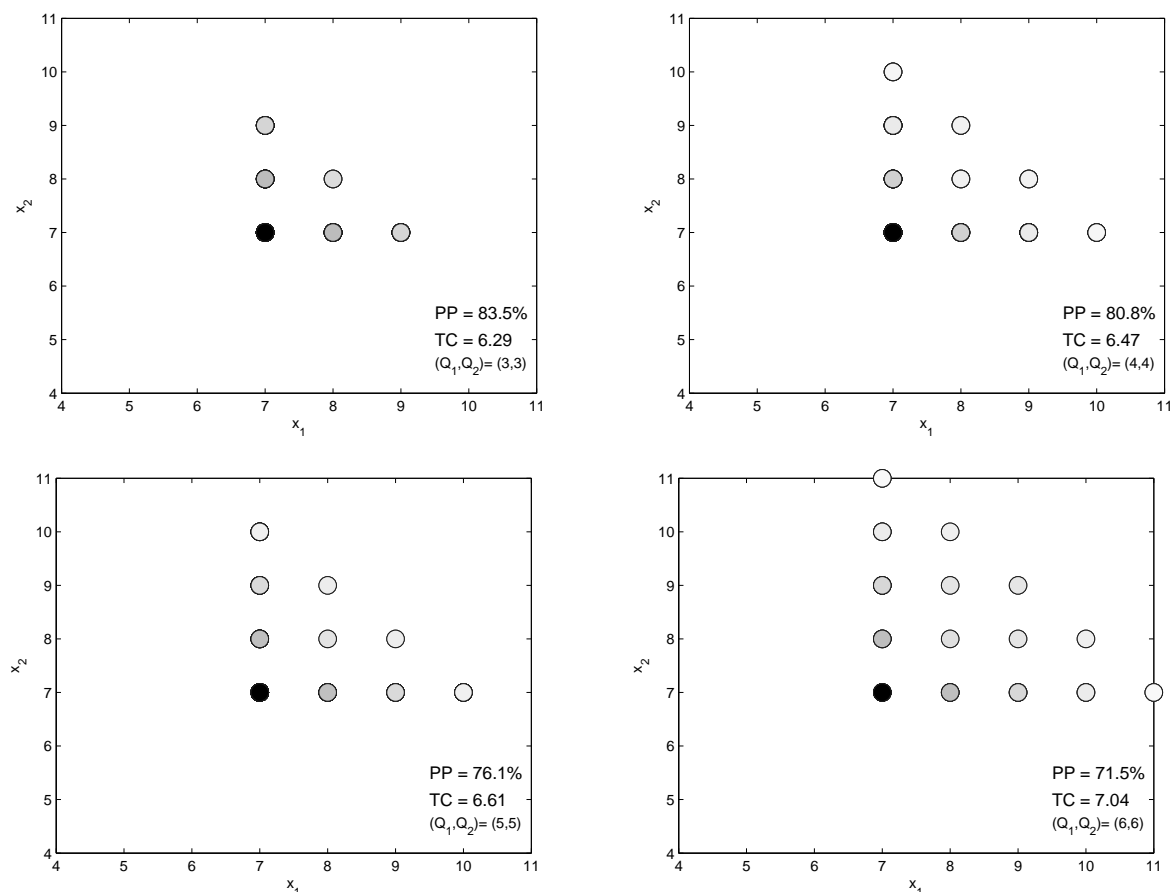


Figure 5 Optimal inventory position region (darker shading indicates greater frequency of occurrence in steady-state) with varying Pre-Pack Size $Q_1 = Q_2 \in \{3, 4, 5, 6\}$ with $\delta = 0.5, cv = 0.38, \rho = 0$. Total expected system cost TC and long-run fraction of demand satisfied via pre-pack ordering PP are also shown.

5.2. Comparison of Policies for Single-SKU

As the optimal policy may be difficult to compute or implement in practice, a store may choose an easy-to-implement alternative policy. In this section, we investigate the cost of adopting two such simple, suboptimal policies in a single-SKU setting: (1) strict base-stock policy, where the store uses a base-stock level that is always achieved exactly, by ordering loose items if necessary, and (2) Pre-pack-only policy: the store never orders loose. For the single-SKU case examined in this section, the pre-pack-only policy is equivalent to the (R, nQ) policy.

Using the single-SKU version of the baseline problem from the previous section (normal demand with $\mu = 5, cv = 0.38, h = 1, b = 10, Q = 4$) we vary one parameter at a time, computing the increase in total cost incurred by using a simpler policy in place of the optimal policy. Figures 8 and 9 summarize the results of this experimentation. The graphs are organized in pairs, showing the total cost for all three policies (optimal, strict base-stock and pre-pack only), as well as the percentage increase from optimal cost for the

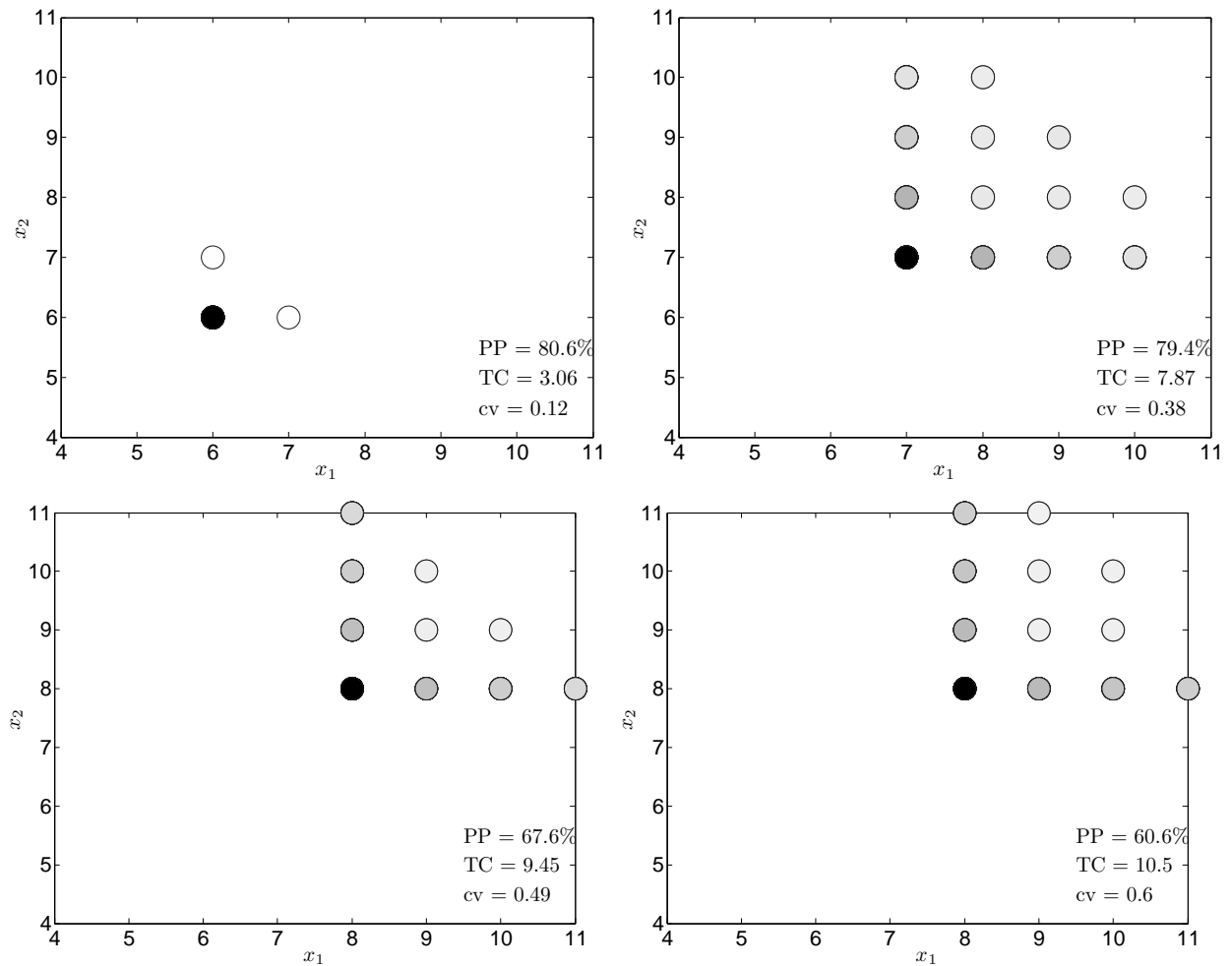


Figure 6 Optimal inventory position region (darker shading indicates greater frequency of occurrence in steady-state) with varying Coefficient of Variation $cv \in \{0.12, 0.38, 0.49, 0.60\}$ with $\delta = 0.5, Q = (4, 4), \rho = 0$. Total expected system cost TC and long-run fraction of demand satisfied via pre-pack ordering PP are also shown.

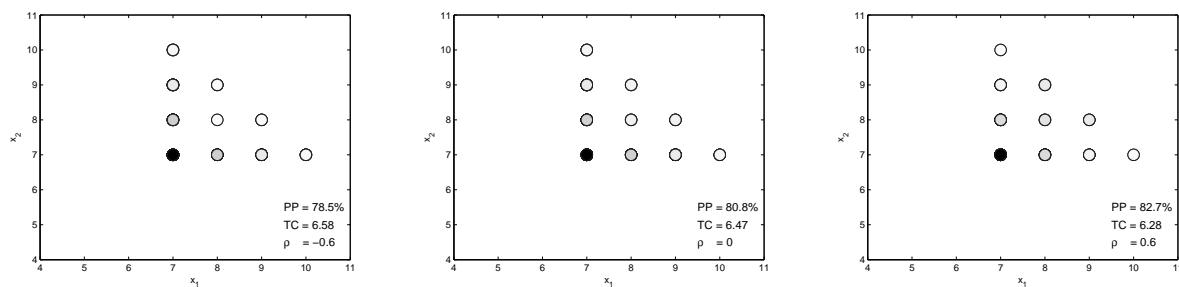


Figure 7 Optimal inventory position region (darker shading indicates greater frequency of occurrence in steady-state) with varying Correlation Coefficient $\rho \in \{-0.6, 0, 0.6\}$ with $\delta = 0.5, Q = (4, 4), cv = 0.3$. Total expected system cost TC and long-run fraction of demand satisfied via pre-pack ordering PP are also shown.

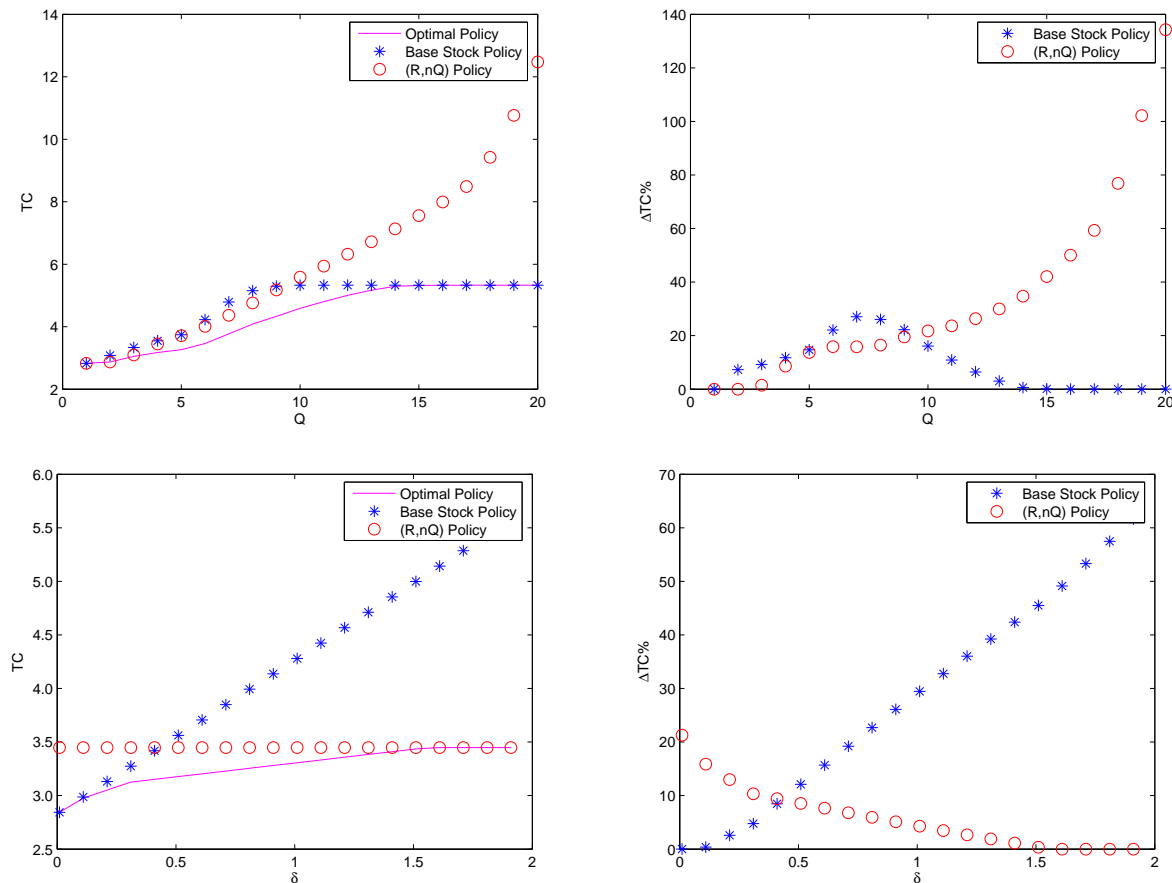


Figure 8 Total cost impact of alternative ordering policies with varying pre-pack size Q and varying unit handling penalty cost δ . ($Q = 4, h = 1, b = 10, \delta = 0.5, cv = 0.38$ unless otherwise specified.)

two suboptimal policies as Q and δ (Figure 8) and b and cv (Figure 9) are varied.

As demonstrated in Figure 8, the cost of all policies must increase as the pre-pack size Q increases. When the pre-pack size is equal to one, all policies are equivalent. As Q increases, the performance of the pre-pack-only policy deteriorates. While the strict base-stock policy is suboptimal for some “medium” range of Q , at certain point Q becomes so large that the optimal policy chooses not to use the pre-pack. Therefore, for small pre-pack size ($Q < 4$), the pre-pack-only policy serves as a good approximation to the optimal policy; for large pre-pack size ($Q > 10$), the strict base-stock policy is optimal; however, for medium pre-pack size ($5 \leq Q \leq 10$ in our experiments here), the optimal policy can provide substantial improvement. In our experiments here, savings are at least 15% better than simpler policies for middle parameter values.

As with the pre-pack size, any increase in the penalty cost δ will increase total cost for optimal and strict base-stock policies. The pre-pack-only policy cost is obviously not affected by δ . Observe that at $\delta = 0$, the optimality gap of the pre-pack-only policy, $\Delta TC = 21\%$, quantifies the *value of ordering flexibility*. For very small δ , it is always worth ordering loose items to get to the ideal inventory position, so the optimality

gap for the base-stock policy is zero. For very large δ , the cost of ordering loose items becomes prohibitive; thus, for some value of δ , e.g., $\delta = 0.4$ in this case, the performance of the pre-pack-only policy and the base-stock policy are roughly equal, with base-stock being optimal for very small δ and pre-pack only being optimal for very large δ .

Figure 9 depicts the implications of varying shortage cost b and coefficient of variation cv . With integer demand, as we have here, the total cost for all three policies is piecewise-linear and increasing in b . The kinks in these total cost curves occur when the policy changes. The total cost curves show similar behavior as the coefficient of variation in demand varies. More variability leads to higher cost, and the percentage increase from optimal cost has no clear pattern since the integer nature of demand causes “jumps” in the rate of change in total cost when the policy shifts.

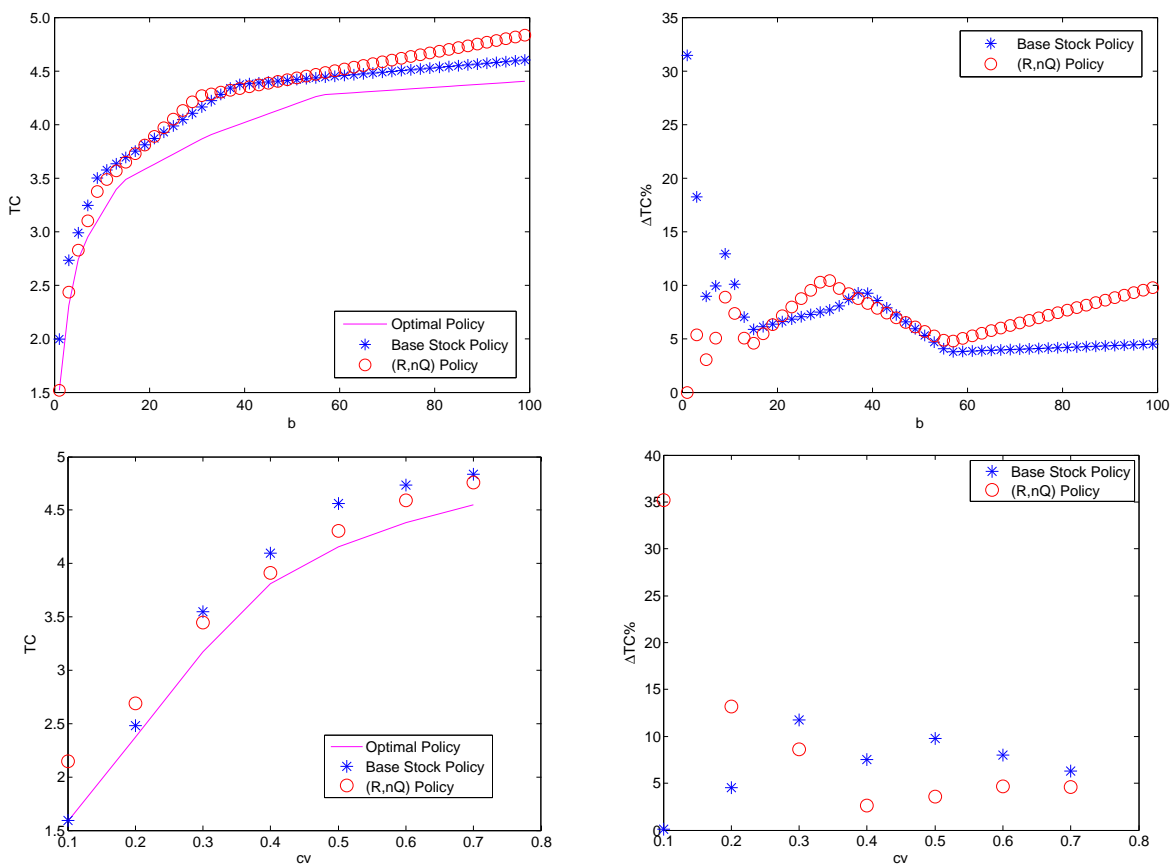


Figure 9 Total cost impact of alternative ordering policies with varying shortage cost b , and varying coefficient of variation cv . ($Q = 4, h = 1, b = 10, \delta = 0.5, cv = 0.38$ unless otherwise specified.)

6. Conclusions

In this research, we study an inventory control problem where a retail store has the option of ordering both pre-packs and individual units. For the single-SKU case, we show that, under certain conditions, the

optimal policy for a single period is to order into a band, ordering as few individual units as possible. The policy is either (R, nQ) or base-stock policy for extreme cases. For the multi-period case with Q -modular demand, we show that the band structure still holds by virtue of the Q -periodicity, and that the steady-state distribution of the target inventory position possesses a semi-uniform structure. These structural properties not only significantly reduce the computational efforts for the optimal policy, but also greatly facilitate the development of effective approximations. For the multi-SKU case, the optimal policy, though more complicated, is structurally similar to the single-SKU band policy. We further characterize the impact of demand correlation on inventory control and pre-pack design. In contrast to the single-SKU case, the pre-pack-only policy is no longer optimal because of its inability to control the rising demand disparity across multiple SKUs. To control such disparity and total cost, the firms should pack together more positively correlated items.

Our numerical experiments on the policy sensitivity and performance comparison lend insight to effective inventory management with pre-packs. For the tests on the 2-SKU case, we find that pre-packs can be cost effective for managing stable and positively correlated demand streams, especially when handling savings are significant. Since the optimal policy may be difficult to implement, we compare the performance of the optimal policy to two simple policies—strict base-stock and pre-pack only—in the single-SKU setting. For extreme values of pre-pack size and incremental per-unit handling cost, one of these two policies performs well; for moderate values, the cost increase associated with using these simpler policies can be substantial.

This paper has addressed how a firm could control inventory given a pre-pack. Although our numerical study sheds some lights on the effects of pre-pack size and demand correlation, we have not explicitly addressed the problem of how the pre-packs should be designed in general. Such a design problem would require the specific relationship between handling savings and the size of the pre-pack; e.g., a larger pre-pack would equal less inventory flexibility but presumably greater handling savings. Pre-pack design also depends on the demand profile of the products involved, potentially across multiple stores. For example, dynamic demand substitution creates positive correlation in demand when a stockout in one item transfers the demand to another related item; umbrella branding is also likely to generate complementary demands where the popularity in one product increases the demand in another product under the same brand.

Acknowledgments

The authors wish to express our sincere gratitude to the department editor, senior editor and two anonymous referees. The feedback from the editorial team was extremely helpful both in strengthening the results and improving the exposition of the paper.

Appendix

Convexity Lemma

The following lemma gathers the relevant properties of convex functions. Its proof can be found in Boyd and Vandenberghe (2004) and Topkis (1998).

LEMMA 1 (Convexity). (a) Define $h \circ g(x) = h(g_1(x), \dots, g_m(x))$, with $h: \mathbb{R}^m \rightarrow \mathbb{R}$, $g_i: \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, \dots, m$. Then $h \circ g(x)$ is convex if h is convex and nondecreasing in each argument, and g_i is convex for each i .

(b) If $h: \mathbb{R}^m \rightarrow \mathbb{R}$ is a convex function, then $h(Ax + b)$ is also a convex function of x , where $A \in \mathbb{R}^m \times \mathbb{R}^n$, $x \in \mathbb{R}^n$, and $b \in \mathbb{R}^m$.

(c) Assume that for any $x \in \mathbb{R}^n$, there is an associated convex set $C(x) \subset \mathbb{R}^m$ and $\{(x, y) : y \in C(x), x \in \mathbb{R}^n\}$ is a convex set. If $h(x, y)$ is convex and the function $g(x) \equiv \inf_{y \in C(x)} h(x, y)$ is well defined, then $g(x)$ is convex over \mathbb{R}^n .

(d) If $f(x)$ and $g(x)$ are convex on X and $\alpha, \beta > 0$, then $\alpha f(x) + \beta g(x)$ is convex on X .

(e) Assume that $F(\xi)$ is a distribution function on Y . Assume also that $f(x, \xi)$ is convex in x on a lattice X for each $\xi \in Y$, and integrable with respect to $F(\xi)$ for each $x \in X$. Then $g(x) \equiv \int_Y f(x, \xi) dF(\xi)$ is convex in x on X .

Proof of Lemma 2

(a). By the definition of Z and the convexity of $[\delta(I - y) + G(I)]$ in I , we have, for $y < Z$,

$$H(y) = \min_{I \geq y} [\delta(I - y) + G(I)] = \delta(Z - y) + G(I),$$

and, for $y \geq Z$,

$$H(y) = \delta(y - y) + G(y) = G(y).$$

Hence

$$H(y) = \begin{cases} \delta(Z - y) + G(Z), & \text{if } y < Z, \\ G(y), & \text{if } y \geq Z. \end{cases} \quad (21)$$

Taking the derivative at $y = S$ yields $H'(S) = G'(S) = 0$.⁷ This along with the convexity of $H(y)$ shows that $S = \arg \min_y H(y)$.

(b). The inequalities $R \leq S \leq R + Q$ follow from the convexity and coerciveness of H and $H'(S) = 0$. The inequality $Z \leq S$ holds because $G'(\cdot)$ is increasing, and $H'(Z) = G'(Z) = -\delta < 0 = G'(S)$. ■

Proof of Theorem 1

(a). First, we argue that the optimal pre-pack decision is $y^* = \underline{y}(x) \in (R, R + Q]$. Indeed, by the convexity of H , $H(y) \geq H(\underline{y}(x))$ for any pre-pack position $y \in (-\infty, R] \cup [R + Q, \infty)$. Second, having arrived at $y = \underline{y}(x)$, it is not optimal to order any loose if $R \geq Z$, because $\frac{\partial}{\partial I} [\delta(I - \underline{y}(x)) + G(I)] > 0$ (by the convexity of the function $[\delta(I - \underline{y}(x)) + G(I)]$ in I and $\underline{y}(x) > R \geq Z$). Thus $I^* = \underline{y}(x)$.

(b). Suppose $Z > R$. By Lemma 2.(b) we have $R < Z \leq R + Q$. Depending on the post-prepack-ordering position $\underline{y}(x)$, we have two cases for loose decision I^* . If $\underline{y}(x) \in (R, Z]$, then $H(\underline{y}(x)) = \delta(Z - \underline{y}(x)) + G(Z)$ by (21); i.e., $I^* = Z$. If $\underline{y}(x) \in (Z, R + Q]$, then $H(\underline{y}(x)) = G(\underline{y}(x))$ by (21); i.e., $I^* = \underline{y}(x)$. This completes the proof of part (b).

(c). If $\delta = 0$, by (5) and (6), loose target Z and pre-pack target S coincide. By (5), base-stock policy S is optimal for $V(x) = \min_{(y, I)} \{G(I) : y = x + nQ, n \in \mathbb{Z}_+, I \geq y\} = \min_{I \geq x} G(I)$. ■

⁷ For a convex function f we use its maximal subgradient as its derivative at each x .

Proof of Theorem 2

It suffices to show that $V(x) = V(x - Q)$, $\forall x \leq R + Q$ for three cases.

Case 1: $Z \leq R$ and $\delta > 0$. For $x \leq R + Q$, we have

$$\begin{aligned} V(x) &= \begin{cases} G(\underline{y}(x) + Q), & \underline{y}(x) \leq R \\ G(\underline{y}(x)), & \underline{y}(x) > R \end{cases} && \text{by Theorem 1.(a)} \\ &= \begin{cases} G(\underline{y}(x - Q) + Q), & \underline{y}(x - Q) \leq R \\ G(\underline{y}(x - Q)), & \underline{y}(x - Q) > R \end{cases} && \text{by } Q\text{-periodicity of } \underline{y}(x) \text{ in (9)} \\ &= V(x - Q). && \text{by Theorem 1.(a)} \end{aligned}$$

Case 2: $Z > R$ and $\delta > 0$. This can be similarly established by using Theorem 1.(b) and the Q -periodicity of $\underline{y}(x)$.

Case 3: $\delta = 0$. The base-stock policy is optimal and we have $V(x) = G(S) = V(x - Q)$ for $x \leq R + Q$. ■

Proof of Lemma 3

Since $\xi_Q \sim \mathcal{U}[0, Q - 1]$, for $I \leq R + Q$, we have

$$\begin{aligned} \mathbb{E}V(I - \xi) &= \sum_{\xi} f(\xi)V(I - \xi) \\ &= \sum_{i=0}^{Q-1} \left[\sum_{\{\xi: \xi_Q = i\}} f(\xi)V(I - \xi) \right] \\ &= \sum_{i=0}^{Q-1} \left[\sum_{n \in \mathbb{Z}_+} f(i + nQ)V(I - (i + nQ)) \right] && \text{by } \xi = \xi_Q + nQ \text{ and } \xi_Q = i \\ &= \sum_{i=0}^{Q-1} \left[\sum_{n \in \mathbb{Z}_+} f(i + nQ) \right] \cdot V(I - i) && \text{by } Q\text{-periodicity of } V \\ &= \frac{1}{Q} \sum_{i=0}^{Q-1} V(I - i) && \sum_{n \in \mathbb{Z}_+} f(i + nQ) = \frac{1}{Q} \text{ by (13)} \\ &= \frac{1}{Q} \sum_{i=1}^Q V(i), && \text{by } Q\text{-periodicity of } V \end{aligned}$$

where the last term $\frac{1}{Q} \sum_{i=1}^Q V(i)$ is a constant. ■

Proof of Theorem 3

(a)-(c) Observe that the analysis in §3.1 relies on the convexity of $G(I)$. Consequently, if we can show the convexity of $G(I) = h\mathbb{E}(I - \xi)^+ + b\mathbb{E}(\xi - I)^+ + \beta\mathbb{E}V(I - \xi)$, then all the analysis in §3.1 applies. Since $h\mathbb{E}(I - \xi)^+ + b\mathbb{E}(\xi - I)^+$ is convex in I , it suffices to show the convexity of $\mathbb{E}V(I - \xi)$. However, Lemma 3 along with the hypothesis of Q -periodicity from part (d) imply that $\mathbb{E}V(I - \xi)$ is a constant, and therefore convex for $I \leq B + Q$.

(d) The proof of this part parallels that of part (d) in Theorem 1. ■

Proof of Theorem 5

Let $E \equiv \{Z, Z + 1, \dots, R + Q\}$ be the set of the target inventory positions. Assume $Z > R$. For demand ξ with probability function f , we first define the set E_{in} for $i \in E$:

$$E_{in} \equiv i - E + nQ = \{i - j + nQ : j \in E\} = [i - R - Q + nQ, i - Z + nQ] \cap \mathbb{N}. \quad (22)$$

Thus, for each given inventory i , demand values can be classified into two sets: $\cup_n E_{in}$ is the set of demand values such that it requires only pre-packs to get into E after fulfilment of ξ , while $(\cup_n E_{in})^c$ is the set of demands that requires ordering both pre-packs and loose. Then the transition probability p_{ij} can be computed as:

$$p_{ij} = P\{\xi \in \cup_n E_{in} : \xi = nQ + i - j\} = \sum_{n=0}^{\infty} f(nQ + i - j), \quad \text{if } j \neq Z \quad (23)$$

$$p_{ij} = P\{\xi \in (\cup_n E_{in})^c\} + P\{\xi = i - Z + nQ\} = 1 - \sum_{l=Z+1}^Q \sum_{n=0}^{\infty} f(nQ + i - l), \quad \text{if } j = Z \quad (24)$$

For a given policy π , $\mathcal{P} = [p_{ij}]$ is its associated Markov matrix. Denote $a_{\Delta} = \sum_{n=0}^{\infty} f(nQ + \Delta)$, $d_i = 1 - \sum_{j=Z+1}^Q a_{i-j}$. Therefore,

$$p_{ij} = \begin{cases} a_{i-j}, & \text{if } j \neq Z \\ d_i, & \text{if } j = Z. \end{cases} \quad (25)$$

Transition matrix \mathcal{P} becomes

$$\mathcal{P} = \begin{pmatrix} d_Z & a_{-1} & a_{-2} & \cdots & a_{Z-(R+Q)} \\ d_{Z+1} & a_0 & a_{-1} & \cdots & a_{Z+1-(R+Q)} \\ d_{Z+2} & a_1 & a_0 & \cdots & a_{Z+2-(R+Q)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{R+Q} & a_{R+Q-Z-1} & a_{R+Q-Z-2} & \cdots & a_0 \end{pmatrix}. \quad (26)$$

The distribution ϕ of I is the solution to the balance equation (27) and the normalization equation (28) (Xu et al. 2007, Hwang et al. 2010):

$$\phi = \phi \cdot \mathcal{P} \quad (27)$$

$$\sum_I \phi(I) = 1. \quad (28)$$

Since $\xi_Q \sim U[0, Q - 1]$, the following statements hold:

$$a_i = a_j = a = 1/Q, \quad d_i = d_j = d = 1 - (R + Q - Z)a = (Z - R)/Q, \quad \forall i, j \in E. \quad (29)$$

Then the transition matrix \mathcal{P} can be simplified as:

$$\mathcal{P} = \begin{pmatrix} d & a & \cdots & a \\ d & a & \cdots & a \\ \vdots & \vdots & \ddots & \vdots \\ d & a & \cdots & a \end{pmatrix} \quad (30)$$

$\phi = \left(\frac{Z-R}{Q}, \frac{1}{Q}, \dots, \frac{1}{Q}\right)$ is a solution to (27) and (28) with matrix form (30). Furthermore, for irreducible Markov Chains, the solution is unique. Hence, ϕ follows a semi-uniform distribution with mass $\frac{Z-R}{Q}$ at point Z , and mass $\frac{1}{Q}$ at other points. ■

Proof of Theorem 6

(a). We first show the optimal pre-pack decision is $\mathbf{y}^* = \underline{\mathbf{y}}(\mathbf{x}) \in (\mathbf{R}_{\ell_x}, \mathbf{R}_{\ell_x} + \mathbf{Q}]$. Indeed, for any $\mathbf{y} = \mathbf{x} + n'\mathbf{Q} \in (-\infty, \mathbf{R}_{\ell_x}] \cup (\mathbf{R}_{\ell_x} + \mathbf{Q}, \infty)$, we have $H(\mathbf{y}) = H(\mathbf{x} + n'\mathbf{Q}) \geq H(\mathbf{R}_{\ell_x}) = H(\mathbf{R}_{\ell_x} + \mathbf{Q})$, where the inequality follows from the convexity of $h_{\mathbf{x}}(t) = H(\mathbf{x} + t\mathbf{Q})$ on line ℓ_x . Thus $\mathbf{y}^* = \underline{\mathbf{y}}(\mathbf{x})$.

Now we show the optimality of the loose decision I^* in (20). Given $\mathbf{y}^* = \underline{\mathbf{y}}(\mathbf{x})$, the optimal loose decision I_i^* for SKU i only depends on $\underline{y}_i(\mathbf{x})$, because the optimization is separable in the following sense:

$$\max_{\mathbf{I} \geq \underline{\mathbf{y}}(\mathbf{x})} [\boldsymbol{\delta} \cdot (\mathbf{I} - \underline{\mathbf{y}}(\mathbf{x})) + G(\mathbf{I})] = \sum_{i=1}^k \max_{I_i \geq \underline{y}_i(\mathbf{x})} [\delta_i(I_i - \underline{y}_i(\mathbf{x})) + G_i(I_i)]. \quad (31)$$

The same argument in the proof of Theorem 1.(b) establishes the optimality of I_i^* in (20).

(b). The \mathbf{Q} -periodicity of $V(\mathbf{x})$ comes from the facts the optimal policy is prescribed by $\underline{\mathbf{y}}(\mathbf{x})$ and that $\underline{\mathbf{y}}(\mathbf{x})$ is \mathbf{Q} -periodic. ■

Proof of Theorem 7

(a). We prove the 2-SKU case in detail. The general k -SKU case follows the same line of argument by picking any two distinct SKUs.

Let $\mathbb{E}\xi_i = \mu$, $\text{var}(\xi_i) = \sigma^2$, $i = 1, 2$, and $\rho < 1$. Let $D_i(\tau)$ be the cumulative demand for SKU i over the initial τ periods, and $S(\tau)$ the cumulative supply over the same τ periods. For sufficiently large τ , by the Central Limit Theorem, $D_i(\tau) \sim \mathcal{N}(\tau\mu, \tau\sigma^2)$, $\Delta_\tau \equiv D_1(\tau) - D_2(\tau) \sim \mathcal{N}(0, 2\tau(1-\rho)\sigma^2)$. Let $m \equiv \min\{h_1, h_2, b_1, b_2\}$. Under the pre-pack-only policy, the period τ cost is

$$\begin{aligned} G_\tau(\mathbf{I}_\tau) &= \sum_{i=1}^2 [h_i \mathbb{E}(I_{i,\tau} - \xi_{i,\tau})^+ + b_i \mathbb{E}(\xi_{i,\tau} - I_{i,\tau})^+] && I_{i,\tau} \text{ and } \xi_{i,\tau} \text{ are for SKU } i \text{ in period } \tau \\ &\geq m \mathbb{E}|D_1(\tau) - S(\tau)| + m \mathbb{E}|D_2(\tau) - S(\tau)| && \text{by the law of conservation and } |x| = x^+ + x^- \\ &\geq m \mathbb{E}|(D_1(\tau) - S(\tau)) - (D_2(\tau) - S(\tau))| && \text{the triangle inequality and the linearity of } \mathbb{E}[\cdot] \\ &= m \mathbb{E}|\Delta_\tau| \\ &\geq m \mathbb{E}[|\Delta_\tau| \mathbf{1}\{|\Delta_\tau| > 2\sigma_\tau\}] && \text{since } \mathbb{E}[|X|] = \mathbb{E}[|X| \mathbf{1}_A] + \mathbb{E}[|X| \mathbf{1}_{A^c}] \geq \mathbb{E}[|X| \mathbf{1}_A] \\ &\geq m \cdot (2\sigma_\tau) \cdot \Pr\{|\Delta_\tau| > 2\sigma_\tau\} && \text{since r.v. } |\Delta_\tau| > 2\sigma_\tau \text{ on set } \{|\Delta_\tau| > 2\sigma_\tau\} \\ &= 0.091m \sqrt{2\tau(1-\rho)\sigma^2}, && \text{since } \Delta_\tau \sim \mathcal{N}(0, 2\tau(1-\rho)\sigma^2) \text{ and } \Pr(|Z| > 2) = 0.0455 \end{aligned}$$

which implies the per-period cost $G_\tau(\mathbf{I}_\tau) \rightarrow +\infty$ as $\tau \rightarrow +\infty$. Hence the pre-pack-only policy cannot be optimal.

(b). By part (a) and $\rho_{is} > \rho_{js}$, the pre-pack (i, s) has stochastically smaller demand disparity than the pre-pack (j, s) , since $2\tau(1-\rho_{js})\sigma^2 < 2\tau(1-\rho_{is})\sigma^2$. For any given policy, this implies smaller total cost for pre-pack (i, s) . ■

References

- Andersson, J., S. Axsäter, J. Marklund. 1998. Decentralized multiechelon inventory control. *Production and Operations Management* **7** 370–386.
- Axsäter, S. 2000. Exact analysis of continuous review (R, Q) policies in two-echelon inventory systems with compound Poisson demand. *Operations Research* **48** 686–696.
- Berling, P., J. Marklund. 2006. Heuristic coordination of decentralized inventory systems using induced backorder costs. *Production and Operations Management* **15** 294–310.
- Blackburn, J., G. Scudder. 2009. Supply chain strategies for perishable products: the case of fresh produce. *Production and Operations Management* **18** 129–137.
- Boyd, S., L. Vandenberghe. 2004. *Convex optimization*. Cambridge university press.

- Cachon, G.P. 2001. Exact Evaluation of Batch-Ordering Inventory Policies in Two-Echelon Supply Chains with Periodic Review. *Operations Research* **49** 79–98.
- Chao, J., M. Chen, A. Deng, H. Miao, A. Newman, S. Tseng, C.A. Yano. 2005. Safeway designs mixed-product pallets to support just-in-time deliveries. *Interfaces* **35** 294–307.
- Chao, X., S.X. Zhou. 2009. Optimal policy for a multiechelon inventory system with batch ordering and fixed replenishment intervals. *Operations Research* **57** 377–390.
- Chen, F. 1998. Echelon reorder points, installation reorder points, and the value of centralized demand information. *Management Science* **44** S221–S234.
- Chen, F. 2000. Optimal policies for multi-echelon inventory problems with batch ordering. *Operations Research* 376–389.
- Chen, S., J. Geunes, A. Mishra. 2012. Algorithms for multi-item procurement planning with case packs. *IIE Transactions* **44** 181–198.
- Clark, A.J., H. Scarf. 1960. Optimal policies for a multi-echelon inventory problem. *Management science* **6** 475–490.
- Gao, Long, Michael Gorman, Ting Luo. 2013a. Managing load acceptance in intermodal transportation with heterogeneous leadtime and dynamic forecast. *UC Riverside working paper* .
- Gao, Long, Zhaolin Li, Biying Shou. 2013b. Information acquisition and voluntary disclosure in an export-processing system. *Production and Operations Management* .
- Gao, Long, Susan H Xu, Michael O Ball. 2012. Managing an available-to-promise assembly system with dynamic short-term pseudo-order forecast. *Management Science* **58** 770–790.
- Gao, Long, Nan Yang. 2012. Dynamic supply risk management with procurement diversification, discretionary selling and signal-based forecast. *UC Riverside working paper* .
- Hadley, G., TM Whitin. 1961. A family of inventory models. *Management Science* **7** 351–371.
- Henig, M., Y. Gerchak, R. Ernst, D. F. Pyke. 1997. An inventory model embedded in designing a supply contract. *Management Science* **43** 184–189.
- Hwang, Johye, Long Gao, Wooseung Jang. 2010. Joint demand and capacity management in a restaurant system. *European Journal of Operational Research* **207** 465–472.
- Li, Zhaolin, Long Gao. 2008. The effects of sharing upstream information on product rollover. *Production and Operations Management* **17** 522–531.
- Litchfield, J., R. Narasimhan. 2000. Improving job shop performance through process queue management under transfer batching. *Production and Operations Management* **9** 336–348.
- Morse, P.M. 1958. *Queues, Inventory and Maintenance*. John Wiley & Sons.
- Parkinson, A., S. T. McCormick. 2005. Optimal replenishment with two delivery sizes. Manufacturing and Service Operations Management Conference Presentation, Evanston, IL.

- Puterman, Martin L. 1994. Markov decision processes: Discrete stochastic dynamic programming.
- Smunt, T.L., J. Meredith. 2000. A comparison of direct cost savings between flexible automation and labor with learning. *Production and Operations Management* **9** 158–170.
- Song, J. 2000. A note on assemble-to-order systems with batch ordering. *Management Science* **46** 739–743.
- Song, J.S., P. Zipkin. 1993. Inventory control in a fluctuating demand environment. *Operations Research* **41** 351–370.
- Topkis, D.M. 1998. *Supermodularity and complementarity*. Princeton Univ. Pr.
- van der Vlist, P. 2007. *Synchronizing the retail supply chain*. Erasmus Research Institute of Management.
- Veinott, A. 1965. The optimal inventory policy for batch ordering. *Operations Research* **13** 424–432.
- Xu, Susan H, Long Gao, Jihong Ou. 2007. Service performance analysis and improvement for a ticket queue with balking customers. *Management science* **53** 971–990.
- Zheng, Y., F. Chen. 1992. Inventory policies with quantized ordering. *Naval Research Logistics* **39** 285–305.