Newsvendor Bounds and Heuristic for Optimal Policies in Serial Supply Chains

Kevin H. Shang • Jing-Sheng Song
Fuqua School of Business, Duke University, Durham, North Carolina 27708
Graduate School of Management, University of California, Irvine, California 92697
khshang@duke.edu • jssong@uci.edu

We consider the classic N-stage serial supply systems with linear costs and stationary random demands. There are deterministic transportation leadtimes between stages, and unsatisfied demands are backlogged. The optimal inventory policy for this system is known to be an echelon base-stock policy, which can be computed through minimizing N nested convex functions recursively. To identify the key determinants of the optimal policy, we develop a simple and surprisingly good heuristic. This method minimizes $2N$ separate newsvendor-type cost functions, each of which uses the original problem data only. These functions are lower and upper bounds for the echelon cost functions; their minimizers form bounds for the optimal echelon base-stock levels. The heuristic is the simple average of the solution bounds. In extensive numerical experiments, the average relative error of the heuristic is 0.24%, with the maximum error less than 1.5%. The bounds and the heuristic, which can be easily obtained by simple spreadsheet calculations, enhance the accessibility and implementability of the multiechelon inventory theory. More importantly, the closed-form expressions provide an analytical tool for us to gain insights into issues such as system bottlenecks, effects of system parameters, and coordination mechanisms in decentralized systems.
(*Inventory Policies; Stochastic Demand; Serial System; Closed-Form Solutions; Sensitivity Analysis*)

1. Introduction

We consider an N-stage serial supply system with deterministic transportation leadtimes between stages. Stationary random demand occurs at stage 1, which obtains resupply from stage 2, stage 2 obtains resupply from stage 3, and so on. Stage N replenishes its stock from an outside supplier which has ample stock. There are linear ordering and inventory holding costs at all stages. Unsatisfied demands at stage 1 are backlogged and incur a linear backorder cost. This system has been studied extensively since the seminal work by Clark and Scarf (1960), who show that an echelon base-stock policy is optimal for the finite-horizon problem. Federgruen and Zipkin (1984) extend this result to infinite horizon and show that a stationary order-up-to-level policy is optimal.

Chen and Zheng (1994) further streamline and simplify the optimality proof. We refer to Gallego and Zipkin (1999) for a more detailed summary of related work and history of development.

In this paper, we focus on the infinite-horizon problem with the objective of minimizing long-run average cost. It is known that the optimal stationary echelon base-stock policy can be computed through minimizing N nested convex functions recursively. We review this recursion in §2. Despite its deceivingly simple form, however, it is not easy to see the key determinants of the optimal policy and cost from the recursion. It is also not easy to communicate the computational procedure to managers and business-school students who have interests in learning the theory of supply-chain management. For one thing, it

is not easy to implement the algorithm by using simple spreadsheet calculations—a familiar tool for those students and practitioners. These challenges motivated us to look for closed-form approximations that can be easily obtained by using spreadsheets and at the same time can shed light on the effect of system parameters.

This desire echos with the observations of several researchers. For example, according to a survey by Cohen et al. (1994), many companies have failed to implement advanced inventory management methods, and hence there are plenty of opportunities for improvement. Hopp et al. (1997) conjecture that the reason for the failure to implement inventory management methods is the difficulty of using them.

The contribution of this paper is twofold. The first one is computational and implementational. In particular, we develop a simple and surprisingly good heuristic for the optimal echelon base-stock levels, which can be obtained by solving $2N$ separate newsvendor-type problems. More specifically, in §3 we develop an upper and a lower bound on the average total echelon cost function of each stage, provided all downstream stages follow the optimal policy. These cost bounds are the convex cost functions of certain single-stage inventory problems with intuitive physical meanings. The minimizers of the bounding functions form an upper and a lower bound for the optimal echelon base-stock level. The simple average of these bounds in turn forms the heuristic solution for the optimal echelon base-stock level. The end result is a closed-form solution involving the original problem data only. In §4 we perform an extensive numerical study to demonstrate the effectiveness of the heuristic. It is shown that the average relative error of the heuristic is 0.24%, with the maximum error less than 1.5%. It is also shown that the upper bound function for stage-N provides a convenient quick estimate for the optimal system cost.

The second main contribution, which perhaps is more important, is transparency. The simple structures of the bounding functions and the closed-form heuristic solution help open the "multi-echelon black box." They allow us to "know not only *what* the optimal solution is for a given set of input data, but also *why*" (Geoffrion 1976, p. 81), and therefore sharpen

our intuition on how to manage this kind of system. More specifically, using these expressions, we can study the effects of system parameters on the optimal cost and policies analytically. This, in turn, provides guidance on how to allocate critical resources to improve system performance. For example, in §5 it is shown that if resources are limited, then it is a better strategy to shorten the leadtime at stage 1 (the one nearest to the customer) or to reduce the echelon holding cost at stage N (the one nearest to the supplier).

The Clark-Scarf result (for the centralized system) has served as a benchmark for the increasingly active supply-chain research on decentralized systems; see, e.g., Cachon and Zipkin (1999), Chen (1999), Lee and Whang (1999), and Porteus (2000). We hope that the tools developed here will help mitigate the analytical challenge and generate more insights in this line of research. We make an initial attempt in §6; we show that our approximations can simplify the results in Chen (1999) on coordination mechanisms in decentralized supply chains. We also make connection of our work to Porteus (2000).

There have been several other efforts in the literature to construct simple bounds on optimal cost or optimal base-stock levels. Gallego and Zipkin (1999) discuss the issue of stock positioning and construct three heuristics to the optimal system average cost. In the "RD heuristic," they decompose the system into some subsystems and use the shortest-path algorithm to search for upper bounds on the optimal cost. Zipkin (2000) introduces a lower bound on the optimal base-stock levels for a two-stage system by restricting the possibility of holding inventory at the upstream stage. By doing so, the upper echelon cost function reduces to a single-stage cost function. The formulation of our upper-bound cost functions is consistent with this idea in the sense of collapsing an N-stage system into a single-stage system, although the way of construction is different. Using a different approach, Dong and Lee (2001) also develop lower bounds on optimal base-stock levels for systems with convex holding and backorder cost functions, which happen to coincide with our lower bounds when the holding and backorder costs are linear. To our knowledge, there exists no previous effort before ours in

constructing upper bounds on the optimal base-stock levels for the serial system. An exact formula for the optimal base-stock levels, which requires heavy computation, is provided in van Houtum et al. (1996). Gallego (1998) develops closed-form "distribution-free" upper and lower bounds on (Q, r) policies in a single-stage system. Glasserman (1997) establishes bounds and asymptotics for performance measures and base-stock levels in single and serial capacitated systems. One of the main results from his study is that, when the backorder cost is high, the bounds perform well. This is consistent with our findings in this research. Hopp et al. (1997) suggest an easily implementable heuristic control policy for a single warehouse with multiple parts. What they mean by "easily implementable" is a closed-form solution for the control parameters for each part. Our heuristic solution, too, obviously qualifies for this category.

2. Preliminaries

We now provide a brief review of the related existing theory of single- and N-stage inventory systems. Recall that there are linear ordering costs. However, because the long-run average ordering cost is a constant, we ignore this cost in our presentation. Throughout the paper, we focus primarily on the continuous-review, compound-Poisson-demand systems. All the results hold for the periodic-review models with independent and identically distributed (i.i.d.) demands. We refer the reader to Chen and Zheng (1994), Gallego and Zipkin (1999), and Zipkin (2000) for details.

2.1. The Single-Stage System
Consider a single-stage (location), single-item inventory system in which the demand follows a stationary compound Poisson process. There is a constant leadtime L for replenishment orders. There is a linear holding cost for on-hand inventories with unit rate h and a linear backorder cost for backorders with unit rate b. It is known that a base-stock policy with base-stock level s^* is optimal for this system. That is, we monitor the inventory position continuously. Whenever the inventory position is below s^*, we order up to s^*. Otherwise, we do not order.

Denote
$D =$ the leadtime demand.
$F(\cdot) =$ cumulative distribution of D.
$F^{-1}(\theta) = \min\{y \mid F(y) \geq \theta\}$, $0 \leq \theta \leq 1$.
$(x)^+ = \max\{0, x\}$.
$(x)^- = \max\{0, -x\}$.
For any given base-stock policy with base-stock level y, the steady-state on-hand inventory is $I = (y - D)^+$ and the steady-state number of backorders is $B = (y - D)^-$. Thus, the long-run average cost is

$$C(y) = \mathsf{E}[hI + bB] = \mathsf{E}[h(y - D)^+ + b(y - D)^-]. \quad (1)$$

Because s^* is the optimal base-stock level which minimizes (1) over y, we have

$$s^* = F^{-1}\left(\frac{b}{b+h}\right). \quad (2)$$

The cost expression in (1) has exactly the same format as in the single-period newsvendor model with D being the single-period demand, h the overage cost, b the underage cost, and y the order quantity. The solution (2) corresponds to the optimal order quantity. For notational convenience, from now on we refer to the problem with the cost expression in (1) and the corresponding solution (2) as system NV(h, b, D).

2.2. The N-Stage System
Consider a serial inventory system with N stages and a compound-Poisson-demand process $\mathbf{D} = \{D(t), t \geq 0\}$, where $D(t)$ is the cumulative demand in the time interval $(0, t]$. The material flows from stage N to stage 1 where customer demand occurs. An outside supplier with ample stock supplies material to stage N. There are constant transportation leadtimes between stages. Unsatisfied demand is fully backlogged. Let
$j =$ stage index.
$L_j =$ constant transportation leadtime from stage $j + 1$ to stage j.
$D_j =$ leadtime demand for stage $j = D(t + L_j) - D(t)$.
$h'_j =$ installation (local) inventory holding cost rate at stage j.
$h_j =$ echelon inventory holding cost rate at stage $j = h'_j - h'_{j+1}$, ($h'_{N+1} = 0$).
$b =$ backorder cost rate at stage 1.
We denote this system as Series $(N, (h_i, L_i)_{i=1}^N, b, \mathbf{D})$.

It is well known that an echelon base-stock policy is optimal for this system (see, e.g., Chen and Zheng 1994). Define the following state random variables in equilibrium:

B = number of backorders at stage 1.

I'_j = installation inventory at stage j.

T_j = inventory in transit to stage j.

I_j = echelon inventory at stage $j = I'_j + \sum_{i=1}^{j-1}(T_i + I'_i)$.

IN_j = echelon net inventory level at stage $j = I_j - B$.

IO_j = inventory on order at stage j.

IOP_j = echelon inventory-order position at stage $j = IN_j + IO_j$.

IP_j = echelon inventory-transit position at stage $j = IN_j + T_j$.

Because stage N has ample supply from the outside supplier, $IP_N = IOP_N$. An echelon base-stock policy $\mathbf{s} = (s_1, \ldots, s_N)$, where s_j is the echelon base-stock level for stage j, $j = 1, \ldots, N$, works as follows: We monitor the echelon inventory-order position IOP_j for each stage j continuously. Whenever it falls below the target level s_j, we place an order from stage $j+1$ to bring it back to this target. Under any echelon base-stock policy $\mathbf{s}$, the key performance measures can be evaluated as follows:

$$IP_N = s_N, \tag{3}$$

$$IN_j = IP_j - D_j, \quad j = N, \ldots, 1, \tag{4}$$

$$IP_j = IN_{j+1} \wedge s_j, \quad j = N-1, \ldots, 1, \tag{5}$$

where $u \wedge v = \min\{u, v\}$. From these, and assuming $s_0 = 0$, we have

$$I'_j = IN_j - IP_{j-1}, \quad j = 1, \ldots, N, \tag{6}$$

$$B = [IN_1]^-. \tag{7}$$

Let

s_j^* = optimal echelon base-stock level for stage j,
$$j = 1, \ldots, N.$$

Then the optimal echelon base-stock policy $\mathbf{s}^* = (s_1^*, \ldots, s_N^*)$ minimizes the long-run average systemwide cost

$$
\begin{aligned}
C(\mathbf{s}) &= \mathsf{E}\left[\sum_{j=1}^N h'_j I'_j + bB + \sum_{j=2}^N h'_j D_{j-1}\right] \\
&= \mathsf{E}\left[\sum_{j=1}^N h_j IN_j + (b + h'_1)[IN_1]^-\right]
\end{aligned}
\tag{8}
$$

among all $\mathbf{s}$, and can be obtained through the following recursive optimization equations: Set $\underline{C}_0(x) = (b + h'_1)[x]^-$. For $j = 1, 2, \ldots, N$, given C_{j-1}, compute

$$\widehat{C}_j(x) = h_j x + \underline{C}_{j-1}(x), \tag{9}$$

$$C_j(y) = \mathsf{E}[\widehat{C}_j(y - D_j)], \tag{10}$$

$$s_j^* = \arg\min\{C_j(y)\}, \tag{11}$$

$$\underline{C}_j(x) = C_j(s_j^* \wedge x). \tag{12}$$

Here, each $C_j(\cdot)$ is a convex function with a finite minimum point. The optimal systemwide average cost $C^* = C(\mathbf{s}^*) = C_N(s_N^*)$.

Note that we can evaluate any base-stock policy by simply skipping the optimization step in (11).

The relationships between installation and echelon base-stock levels have been established in Axsäter and Rosling (1993). Let $s_j'^*$ be the optimal installation base-stock level at each stage j. Then, $s_j'^* = s_j^* - s_{j-1}^*$ (with $s_0 = 0$) if $s_j^* \geq s_{j-1}^*$ for all j. In general, let $s_j^- = \min_{i \geq j}\{s_i^*\}$, then the policy $\mathbf{s}^*$ is equivalent to $\mathbf{s}^-$, and we can set $s_j'^* = s_j^- - s_{j-1}^-$ with $s_0^- = 0$.

3. The Newsvendor Bounds

From the recursion (9)–(12), we can see that solving s_1^* is the same as solving a newsvendor problem $NV(h_1, b + \sum_{i=2}^N h_i, D_1)$. Thus, from (2),

$$s_1^* = F_1^{-1}\left(\frac{b + \sum_{i=2}^N h_i}{b + \sum_{i=1}^N h_i}\right). \tag{13}$$

However, obtaining s_j^*, $j = 2, \ldots, N$ is not as simple. Minimizing C_j depends on all the previous calculations for stages 1 through $j-1$. Our goal is to bound each C_j by a pair of newsvendor-type functions and use their solutions to construct bounds for s_j^*, $j = 2, \ldots, N$. All proofs in this paper are in the Appendix.

We first make the following important observation based on (9)–(12):

OBSERVATION. For each stage j, the optimal echelon base-stock level s_j^* is completely independent of the decisions at its upstream stages. More precisely, s_j^* is solely determined by b, $h'_1 = \sum_{i=1}^N h_i$, and (s_i^*, D_i, h_i) for $i = 1, \ldots, j$. Thus, s_j^* depends on the upstream stages only through the sum of the echelon holding cost rates at these stages, $\sum_{i=j+1}^N h_i$; it does not depend on $s_{j+1}^*, \ldots, s_N^*$.

This observation motivates the following concepts:

DEFINITION 1. For any fixed $j \geq 1$, we say a policy $(s_1, \ldots, s_j)$ is an *echelon-j base-stock policy* if at each stage i, an echelon base-stock policy is followed with echelon base-stock level s_i, $i = 1, \ldots, j$. An *echelon-j manager* is one who makes the echelon-j base-stock level decisions based on the following *echelon-j information*:

$$\mathscr{I}_j = \left\{ b, (D_i, h_i)_{i=1}^j, \sum_{i=j+1}^N h_i, (s_i^*)_{i=1}^{j-1} \right\}. \quad (14)$$

For any echelon-j base-stock policy $(s_1^*, \ldots, s_{j-1}^*, y)$, let $I_i'(y)$ denote the local on-hand inventory at stage i, $i = 1, \ldots, j$, and $B(y)$ the number of backorders at stage 1. Then these random variables can be obtained recursively according to (3)–(7), with N replaced by j, index j replaced by i, $s_j = y$, and $s_i = s_i^*$, $i < j$. We use the argument y here to emphasize the dependence on y. Using these expressions, we have the following decomposition of C_j, which resembles (8) and plays a pivotal role in the construction of the bounds. Denote $\widetilde{D}_j = \sum_{i=1}^j D_i$.

PROPOSITION 2. *For each $j \geq 2$, $C_j(y)$ is the long-run average cost for echelon-j if the echelon-j base-stock policy $(s_1^*, \ldots, s_{j-1}^*, y)$ is followed, and*

$$C_j(y) = \tau_j + G_j(y), \quad (15)$$

where

$\tau_j = $ *average in-transit holding cost from stage j to stage 1*

$$= \sum_{i=2}^j (h_i + \cdots + h_j) \mathsf{E}[D_{i-1}] = \sum_{i=2}^j h_i \mathsf{E}[\widetilde{D}_{i-1}], \quad (16)$$

$G_j(y) = $ *average holding/backorder cost at stages 1 through j under policy $(s_1^*, \ldots, s_{j-1}^*, y)$*

$$= \mathsf{E}\Big[h_j I_j'(y) + (h_j + h_{j-1}) I_{j-1}'(y) + \cdots$$

$$+ \Big(\sum_{i=1}^j h_i \Big) I_1'(y) + \Big(b + \sum_{i=j+1}^N h_i \Big) B(y) \Big]. \quad (17)$$

Proposition 2 implies that the echelon-j manager is in effect responsible for the installation holding

cost rates $h_i + h_{i+1} + \cdots + h_j$ at stage i, $i = 1, \ldots, j$, and a penalty cost rate $b + \sum_{i=j+1}^N h_i$ at stage 1. Thus, echelon-j has exactly the same structure as

$$\text{Series}\Big(j, (h_i, L_i)_{i=1}^j, b + \sum_{i=j+1}^N h_i, \mathbf{D} \Big).$$

In other words, the echelon-j manager faces a system that is truncated at stage j of the original system: Everything else stays the same as in the original except that the backorder cost rate is increased by $\sum_{i=j+1}^N h_i$, the sum of the echelon holding cost rates of the truncated-off part—stages $j+1$ through N.

The increased backorder cost rate can be interpreted as follows. Although the unit holding cost rate at stage 1 is $h_1' = \sum_{i=1}^N h_i$, the total value added in echelon-j is only $\sum_{i=1}^j h_i$. Hence, for each unit sold, the echelon-j manager perceives a gain of $h_1' - \sum_{i=1}^j h_i = \sum_{i=j+1}^N h_i$ for the system. Consequently, if he cannot satisfy a demand immediately, his perceived backorder cost rate will be the original backorder cost b plus the potential perceived gain $\sum_{i=j+1}^N h_i$.

The decomposition of G_j in Proposition 2 inspires our idea of bounding G_j by two newsvendor-type functions. Essentially, the idea is to keep the backorder cost rate at $b + \sum_{i=j+1}^N h_i$, but to replace the installation holding cost rates at different stages by a single value. Under such a cost structure, there would be no incentive to hold inventory at upper stages ($i = 2, \ldots, j$), so echelon-j would be in effect collapsed into a single-stage system.

More specifically, to form a lower bound cost, we set this single value to be the minimum perceived installation holding cost rate h_j (see (17)). In this way, we obtain a lower bound system $\text{Series}(j, (\underline{h}_i, L_i)_{i=1}^j, b + \sum_{i=j+1}^N h_i, \mathbf{D})$, where $\underline{h}_j = h_j$ and $\underline{h}_i = 0$ for $i < j$. Applying the same algebraic argument that leads to (8), the average cost of any echelon-j base-stock policy equals

$$\mathsf{E}\Big[\sum_{i=1}^j h_j I_i' + \Big(b + \sum_{i=j+1}^N h_i \Big) B + \sum_{i=2}^j h_j D_{i-1} \Big]$$

$$= \mathsf{E}\Big[\sum_{i=1}^j \underline{h}_i IN_i + \Big(\Big(b + \sum_{i=j+1}^N h_i \Big) + \sum_{i=1}^j \underline{h}_i \Big) [IN_1]^- \Big]$$

$$= \mathsf{E}\Big[h_j IN_j + \Big(b + \sum_{i=j+1}^N h_i + h_j \Big) [IN_1]^- \Big]. \quad (18)$$

Clearly, the optimal policy for the lower bound system is to allocate all inventory to stage 1. Conditioning on $IP_j = IOP_j = y$, the optimal policy for the lower bound system is the echelon-j base-stock policy $(y, \ldots, y, y)$. Under this policy, we have $IN_j = y - D_j$ and $IN_1 = y - \widetilde{D}_j$. Then, (18) becomes

$$\mathsf{E}\left[h_j[y - \widetilde{D}_j]^+ + \left(b + \sum_{i=j+1}^{N} h_i\right)[y - \widetilde{D}_j]^- \right] + h_j\mathsf{E}[\widetilde{D}_{j-1}]$$

$$:= G_j^l(y) + \underline{\tau}_j, \tag{19}$$

where $\underline{\tau}_j = h_j\mathsf{E}[\widetilde{D}_{j-1}]$. Because this is the optimal cost for the lower bound system (conditioning on $IP_j = y$), and $C_j(y)$ is bounded below by the cost of a particular policy $(s_1^*, \ldots, s_{j-1}^*, y)$ in this lower bound system, we have

$$C_j(y) = G_j(y) + \tau_j \geq G_j^l(y) + \underline{\tau}_j.$$

Thus, C_j is bounded below by a newsvendor cost function plus a constant.

Symmetrically, by charging the largest installation holding cost rate $\sum_{i=1}^{j} h_i$ at each stage in echelon-j, we obtain an upper bound system. Conditioning on $IP_j = y$, the optimal policy for the upper bound system is again the echelon-j base-stock policy $(y, \ldots, y, y)$, whose long-run average cost is

$$\mathsf{E}\left[\left(\sum_{i=1}^{j} h_i\right)[y - \widetilde{D}_j]^+ + \left(b + \sum_{i=j+1}^{N} h_i\right)[y - \widetilde{D}_j]^- \right]$$

$$+ \left(\sum_{i=1}^{j} h_i\right)\mathsf{E}[\widetilde{D}_{j-1}] := G_j^u(y) + \bar{\tau}_j, \tag{20}$$

where $\bar{\tau}_j = \sum_{i=1}^{j} h_i\mathsf{E}[\widetilde{D}_{j-1}]$. By construction, this cost is an upper bound on the long-run average cost of the same policy in the original system. On the other hand, $C_j(y)$ is a lower bound for the original system, conditioning on $IP_j = y$, as shown in Chen and Zheng (1994). So,

$$C_j(y) = G_j(y) + \tau_j \leq G_j^u(y) + \bar{\tau}_j.$$

That is, C_j is bounded above by a newsvendor cost function plus a constant.

Observe that the constant τ_j (see (16)) is independent of the choice of policies in the original system. So, it is tempting to conjecture that $G_j^l(y) \leq G_j(y) \leq$

$G_j^u(y)$. Because $\underline{\tau}_j \leq \tau_j \leq \bar{\tau}_j$, the conjecture implies tighter bounds. Theorem 3 below shows that this is indeed true.

Obviously, the solutions of these bounding functions are easy to obtain. The next question is, what would be the relationship between these simple solutions and the optimal echelon base-stock level s_j^*? The answer is quite intuitive: Because the upper bound function charges higher inventory cost than the original system, there is less incentive to hold inventory; therefore, its solution is likely to be a lower bound for s_j^*. Symmetrically, the solution of the lower bound function is likely to be an upper bound for s_j^*. In the following, we formalize these results.

First, we set forth notation, some of which has been defined earlier. Let
$\widetilde{D}_j = \sum_{i=1}^{j} D_i =$ total leadtime demand in the subsystem consisting stages 1 through j.
$F_j(\cdot) =$ cumulative distribution function of $\widetilde{D}_j$.

$$G_j^u(y) = \mathsf{E}\left[\left(\sum_{i=1}^{j} h_i\right)(y - \widetilde{D}_j)^+ + \left(b + \sum_{i=j+1}^{N} h_i\right)(y - \widetilde{D}_j)^- \right].$$

$G_j^l(y) = \mathsf{E}[h_j(y - \widetilde{D}_j)^+ + (b + \sum_{i=j+1}^{N} h_i)(y - \widetilde{D}_j)^-].$
$C_j^u(y) = \tau_j + G_j^u(y).$
$C_j^l(y) = \tau_j + G_j^l(y).$

$$s_j^l = \arg\min G_j^u(\cdot) = F_j^{-1}\left(\frac{b + \sum_{i=j+1}^{N} h_i}{b + \sum_{i=1}^{N} h_i} \right).$$

$$s_j^u = \arg\min G_j^l(\cdot) = F_j^{-1}\left(\frac{b + \sum_{i=j+1}^{N} h_i}{b + \sum_{i=j}^{N} h_i} \right).$$

Then, we have:

THEOREM 3. *For any given j and y,*

(a) *$G_j(y)$ is bounded above and below by the cost functions for* $\mathrm{NV}(\sum_{i=1}^{j} h_i, b + \sum_{i=j+1}^{N} h_i, \widetilde{D}_j)$ *and* $\mathrm{NV}(h_j, b + \sum_{i=j+1}^{N} h_i, \widetilde{D}_j)$, *respectively. That is,*

$$G_j^l(y) \leq G_j(y) \leq G_j^u(y).$$

(b) *s_j^* is bounded by the minimizers of G_j^u and G_j^l; i.e.,*

$$s_j^l \leq s_j^* \leq s_j^u.$$

(c) *$C_j^l(y) \leq C_j(y) \leq C_j^u(y)$. In particular, the optimal system cost $C(\mathbf{s}^*)$ satisfies*

$$\tau_N + G_N^l(s_N^u) \leq C(\mathbf{s}^*) \leq \tau_N + G_N^u(s_N^l).$$

(d) *All the inequalities become equalities for $j = 1$.*

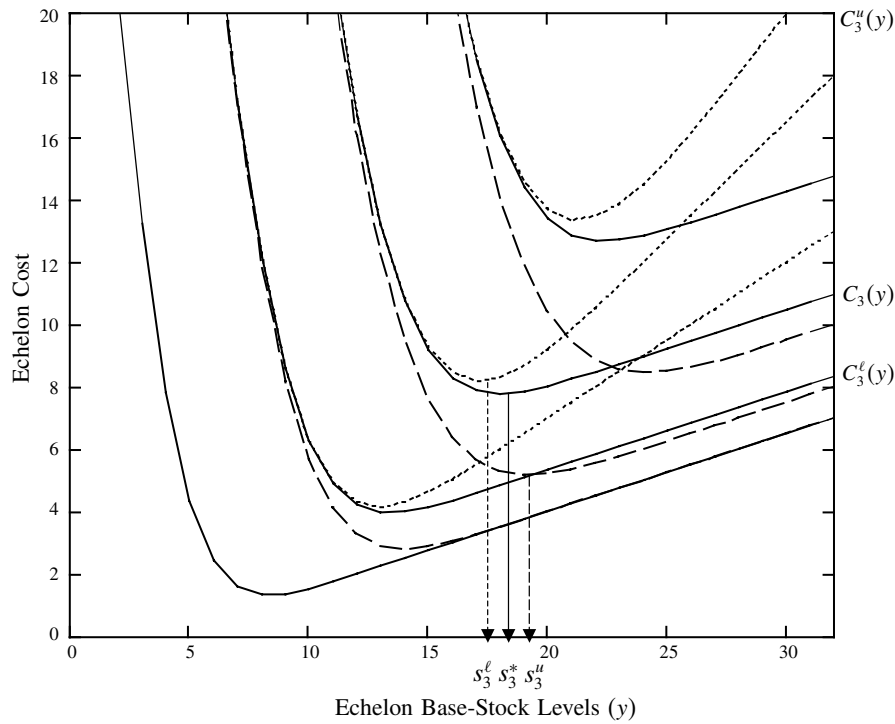**Figure 1   A Typical Graph for Echelon Cost Functions in a Four-Stage System**

Figure 1 is a typical graph of C_j, C_j^u, and C_j^l for a four-stage system. The system parameters are $L_1 = L_2 = L_3 = L_4 = 0.25$, $h_1 = h_2 = h_3 = h_4 = 0.25$, $\lambda = 16$, and $b = 25$. It is clear that $C_1 = C_1^u = C_1^l$ and C_j is bounded by C_j^u and C_j^l for $j = 2, 3$, and 4.

REMARK. Note that when applying this result to periodic-review systems with i.i.d. demands, one needs to be careful about the assumptions on the sequence of events in a period. Most models assume that inbound and outbound shipments occur at the beginning of each period, while inventory-backorder costs are assessed at the end of the period. In this case, we need to consider one more period in calculating $\widetilde{D}_j$. See an example in §6.

It is interesting to note that $C_j^u(y)$ coincides with the upper bound function developed by Dong and Lee (2001), based on a completely different idea. Using the conventional dynamic programming formulation, Dong and Lee inflate the induced penalty cost function to each stage j, $j \geq 2$, by charging a penalty even for sufficient stock. This is equivalent to replacing $(s_j^* \wedge x)$ by x in (12) for all j while keeping the rest of

the optimality recursion as is. Their approach is thus, in effect, identical to that in Zipkin (2000) for the two-stage system, which sets $s_j^* = \infty$ in (12).

4.   The Heuristic and Its Performance

There are several ways of constructing approximations of the optimal base-stock levels by applying Theorem 3. For example:

(i) According to Theorem 3 (b), any convex combination of s_j^l and s_j^u,

$$\alpha s_j^l + (1-\alpha) s_j^u, \quad 0 \leq \alpha \leq 1, \tag{21}$$

can be used to approximate s_j^*. In principle, extensive numerical experiments can be carried out to identify effective values for α.

(ii) Alternatively, instead of working with the solution bounds, we can replace the coefficients of $I_i'(y)$ in (17) by a single convex combination of them to obtain a newsvendor-type system (similar to the bounding systems), and use its solution as the approximate

solution. More specifically, this common coefficient takes the form of

$$\alpha_j h_j + \alpha_{j-1}(h_j + h_{j-1}) + \cdots + \alpha_1 \sum_{i=1}^{j} h_i, \qquad (22)$$

where $0 \le \alpha_k \le 1$ and $\sum_{k=1}^{j} \alpha_k = 1$. The corresponding solution is

$$F_j^{-1}\left(\frac{b + \sum_{i=j+1}^{N} h_i}{b + \sum_{i=j+1}^{N} h_i + \sum_{k=1}^{j} \alpha_k \sum_{i=k}^{j} h_i}\right).$$

Again, in principle, experiments can be carried out to identify good choices of α_k.

In this paper, we focus on one of these possibilities—we use the simple average of the upper and lower bound solutions to approximate the optimal policy. Let s_j^a denote the approximation for s_j^*, Then,

$$s_j^a = \frac{(s_j^l + s_j^u)}{2} \approx s_j^*, \quad j = 2, 3, \ldots, N. \qquad (23)$$

This corresponds to $\alpha = 1/2$ in (21). Expressing in terms of the original problem data,

$$s_j^a = \frac{1}{2}\left[F_j^{-1}\left(\frac{b + \sum_{i=j+1}^{N} h_i}{b + \sum_{i=1}^{N} h_i}\right) + F_j^{-1}\left(\frac{b + \sum_{i=j+1}^{N} h_i}{b + \sum_{i=j}^{N} h_i}\right)\right],$$
$$j = 2, 3, \ldots, N. \qquad (24)$$

If the average is not an integer, we can either round down or round up the value to obtain the nearest integer value for s_j^a. The difference in performance of the two methods is very small in most cases (see some statistics in the discussion of Table 1). In general, when b is small, say smaller than 39—a number observed from our numerical experiments—truncation provides a slightly better approximation. It seems that this rule is independent of holding cost parameters in all examples that we tested in this paper.

Recall that both G_j^u and G_j^l use the same backorder cost rate as the original G_j. So, all three functions are close to each other on the downward side of their bowl-shaped curves. The main differences of the functions are reflected on the upward side of the curves: G_j^u increases the fastest and G_j^l the slowest. However, it is not the *speed* of increase that determines the bottom (the minimum) of the curve. Rather, it is *when* the

curve *turns* from downward to upward that matters. Given that all these curves have similar downward parts, and they all charge some positive inventory costs, it is expected that they all "turn" about the same time. This is why we can expect the approximation to work. Moreover, it is generally understood that these inventory-backorder cost functions are flat around the minimum. So, the increase in system cost by using the approximation is expected to be small.

It turns out, however, that the average of the minimum values of the two bounding functions at stage N, $(C_N^u(s_N^l) + C_N^l(s_N^u))/2$, is not a very accurate estimate for optimal system cost. Instead, the upper bound $C_N^u(s_N^l)$ alone is a much better approximation. This is largely because, for stage N, the lower bound function, which charges the lowest installation holding cost rate for all stages, becomes looser than the upper bound function. Therefore, we propose $C_N^u(s_N^l)$ as a quick estimate for the optimal system cost; i.e.,

$$C_N^u(s_N^l) \approx C(\mathbf{s}^*) = C_N(s_N^*). \qquad (25)$$

We discuss the effectiveness of this cost approximation in §5.

Recall that C_j^u is in effect obtained by ignoring s_{j-1}^* in the optimality recursion. The good performance of $C_N^u(s_N^l)$ indicates that at optimality, the likelihood of having extra inventory in each installation right after the shipments is small (i.e., $IP_{j-1} \approx IN_j$).

We emphasize two aspects in evaluating the effectiveness of the heuristic. First, we show that the difference between s_j^a and s_j^* is very small under different system parameters. Second, we provide the percentage error on the optimal echelon cost to show the effectiveness of the heuristic. The percentage error of the heuristic is defined as

$$\% \text{ error} = \frac{C(\mathbf{s}^a) - C(\mathbf{s}^*)}{C(\mathbf{s}^*)} \times 100\%.$$

Table 1 compares the optimal and the heuristic policies for a four-stage system with $L_1 = L_2 = L_3 = L_4 = 0.25$ and Poisson demand with $\lambda = 16$. We change the cost parameters to demonstrate the effectiveness of our heuristic. We use the round-down rule to obtain integer values for s_j^a in the $b = 9$ cases, and use the round-up method in the $b = 99$ cases. In general, the heuristic appears to perform better for larger b. When $b = 9$, the

Table 1   Optimal and Heuristic Policies: Poisson Demand

| b	h_1	h_2	h_3	h_4	s_1^*	s_2^*/s_2^a	s_3^*/s_3^a	s_4^*/s_4^a	$C(\mathbf{s}^*)/C(\mathbf{s}^a)$	Error (%)
9	0.25	0.25	0.25	0.25	8	13/13	18/18	22/22	12.688/12.688	0.000
	2.5	0.25	0.25	0.25	6	12/12	17/16	21/21	17.947/18.018	0.396
	0.25	2.5	0.25	0.25	9	10/10	16/16	21/21	28.615/28.615	0.000
	0.25	0.25	2.5	0.25	9	14/13	14/14	20/21	39.048/39.085	0.096
	0.25	0.25	0.25	2.5	9	14/13	18/18	18/18	49.387/49.392	0.001
	2.5	2.5	0.25	0.25	6	9/9	16/16	20/20	32.479/32.479	0.000
	2.5	0.25	2.5	0.25	6	13/12	14/14	20/20	43.082/43.084	0.005
	2.5	0.25	0.25	2.5	6	13/12	18/17	18/18	53.424/53.429	0.009
	0.25	2.5	2.5	0.25	9	10/10	13/14	19/20	53.008/53.258	0.472
	0.25	2.5	0.25	2.5	9	10/10	17/17	17/18	63.663/63.667	0.008
	0.25	0.25	2.5	2.5	9	14/13	15/15	17/18	73.107/73.401	0.403
	2.5	2.5	2.5	0.25	6	10/10	13/13	19/20	56.127/56.205	0.139
	2.5	2.5	0.25	2.5	6	10/10	16/17	17/17	66.738/66.739	0.001
	2.5	0.25	2.5	2.5	6	13/12	14/14	17/17	76.470/76.475	0.006
	0.25	2.5	2.5	2.5	9	11/11	14/14	17/17	86.220/86.220	0.000
	2.5	2.5	2.5	2.5	6	10/10	13/14	16/17	88.857/89.347	0.552
99	0.25	0.25	0.25	0.25	11	17/17	22/22	27/27	16.206/16.206	0.000
	2.5	0.25	0.25	0.25	8	16/16	22/21	27/26	27.383/27.518	0.495
	0.25	2.5	0.25	0.25	11	14/14	21/21	27/26	40.155/40.175	0.051
	0.25	0.25	2.5	0.25	11	17/17	19/19	26/26	52.077/52.077	0.000
	0.25	0.25	0.25	2.5	11	17/17	22/22	24/24	63.711/63.711	0.000
	2.5	2.5	0.25	0.25	8	13/14	21/21	26/26	50.523/50.573	0.098
	2.5	0.25	2.5	0.25	8	16/16	19/19	26/26	62.622/62.622	0.000
	2.5	0.25	0.25	2.5	8	16/16	22/21	24/24	74.373/74.387	0.019
	0.25	2.5	2.5	0.25	11	14/14	18/19	26/26	74.564/74.747	0.245
	0.25	2.5	0.25	2.5	11	14/14	21/21	23/24	86.569/86.659	0.104
	0.25	0.25	2.5	2.5	11	17/17	19/19	23/24	97.496/98.039	0.557
	2.5	2.5	2.5	0.25	8	14/14	18/18	25/25	84.263/84.263	0.000
	2.5	2.5	0.25	2.5	8	14/14	21/21	23/23	96.373/96.373	0.000
	2.5	0.25	2.5	2.5	8	16/16	19/19	23/23	107.532/107.532	0.000
	0.25	2.5	2.5	2.5	11	14/14	18/19	23/23	119.151/119.227	0.064
	2.5	2.5	2.5	2.5	8	14/14	18/18	23/23	128.591/128.591	0.000

average percentage error to the optimal cost $C(\mathbf{s}^*)$ is 0.131% with the maximum percentage error 0.552%. In 4 out of 16 cases, the heuristic solution coincides with the optimal solution. When $b = 99$, the average percentage error to the optimal cost is 0.102% with the maximum percentage error 0.557%. In 8 out of 16 cases, the heuristic produces the true optimal solution. (The average and maximum percentage errors are 0.646% and 2.28%, respectively, if rounding up is used for $b = 9$, and 0.214% and 0.957% if rounding down is used for $b = 99$.)

Because the analytical results apply equally to periodic-review systems with i.i.d. demands, we repeated the same experiment for a four-stage system with negative binomial demands. The mean demand

in each period was 16 and its variance 32. In this experiment, the optimal solutions showed the similar pattern to those in Table 1. The heuristic coincided with the optimal in 4 out of 16 cases. The largest percentage error was 1.294% when $b = 9$ and 0.455% when $b = 99$. The average percentage error was 0.23%. Again, the heuristic performed better when $b = 99$.

We next examine the effectiveness of the heuristic under different holding cost structures and different number of stages. We demonstrate the performance of the heuristic for $N = 2, 4, 8, 16, 32,$ and 64 stages. As for the choice of different holding cost forms, we follow those used in Gallego and Zipkin (1999): linear, affine, kink, and jump holding costs. In particular, in the linear holding costs, $h_j = 1/N$. For affine

Figure 2 **Different Holding Cost Forms**
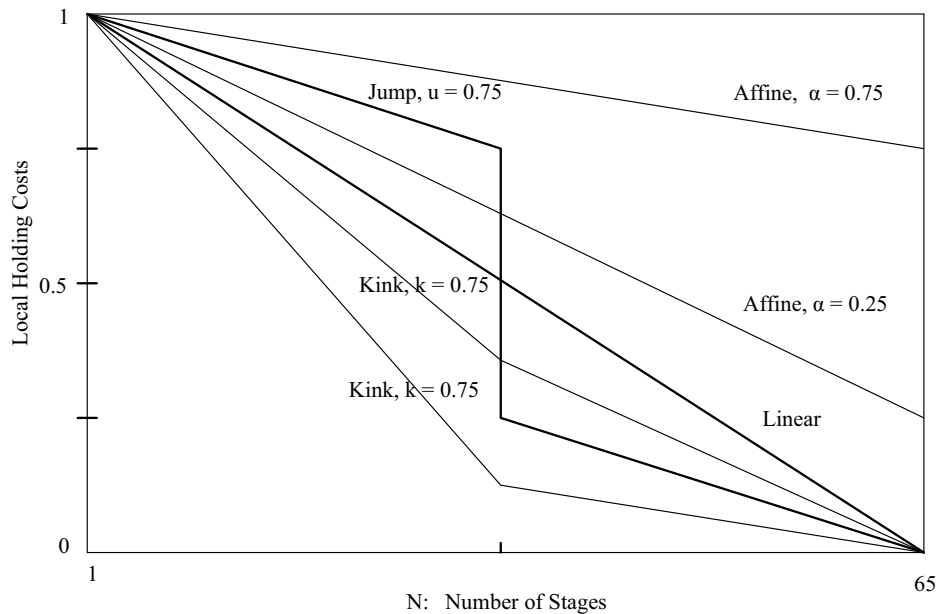
holding costs, $h_N = \alpha + (1-\alpha)/N$ and $h_j = (1-\alpha)/N$, $j = 1, 2, \ldots, N-1$. Here, we compare $\alpha = 0.25$ and 0.75 cases. The kink form is piecewise linear with two pieces. We assume the system changes the holding cost rate in the middle stage. Thus, the general format of the kink holding costs is $h_j = (1-k)/N$, $j \le N/2$ and $h_j = (1+k)/N$, $j > N/2$. Here we also test $k = 0.75$ and 0.25 cases. The last one is the jump holding cost form, where cost is incurred at a constant rate, except for one stage with a large cost. We assume the jump occurs at stage $(N/2)+1$. The general format is $h_j = u + (1-u)/N$, $j = N/2$, and $h_j = (1-u)/N$ otherwise. Figure 2 shows the different holding cost structures. In addition, we assume leadtimes are equally divided among stages and total system leadtime $= 1$. Also, the total system holding cost h_1' is fixed and equal to 1.

The other system parameters are $\lambda = 64$ and $b = 39$. We use the rounding-up method in choosing the integer values for the heuristic policy.

Tables 2–4 report the results. Not surprisingly, in general the percentage error of the heuristic increases as the number of stages N increases. This effect is most obvious when the holding cost has the kink form. However, the performance of the heuristic stays surprisingly good, even for $N = 64$. With all the cost forms and the numbers of stages tested, the worst case of the percentage error is less than 1.3%. The average percentage error is 0.174%.

Figures 3(a) and 3(d) demonstrate the optimal and heuristic policies for a 64-stage system under the linear, jump, affine, and kink holding cost forms. These figures demonstrate that s_j^a is very close to s_j^* in all

Table 2 **Optimal and Heuristic Policies: Linear and Jump Holding Costs**

N	Form	$C(\mathbf{s}^*)$	$C(\mathbf{s}^a)$	Error (%)	Form	$C(\mathbf{s}^*)$	$C(\mathbf{s}^a)$	Error (%)
64	Linear	47.590	47.859	0.565	Jump	46.075	46.107	0.068
32		47.151	47.289	0.292	$u = 0.75$	45.217	45.231	0.031
16		46.265	46.335	0.151		43.500	43.518	0.042
8		44.529	44.555	0.059		40.069	40.080	0.028
4		41.015	41.015	0.000		33.188	33.204	0.047
2		33.916	33.916	0.000		19.370	19.389	0.100

Table 3 Optimal and Heuristic Policies: Affine Holding Costs

N	Form	$C(\mathbf{s}^*)$	$C(\mathbf{s}^a)$	Error (%)	Form	$C(\mathbf{s}^*)$	$C(\mathbf{s}^a)$	Error (%)
64	Affine	56.707	56.877	0.299	Affine	74.085	74.105	0.027
32	$\alpha = 0.25$	56.12	56.14	0.036	$\alpha = 0.75$	73.222	73.240	0.025
16		54.954	54.966	0.022		71.495	71.514	0.027
8		52.609	52.615	0.011		68.042	68.065	0.034
4		47.921	47.931	0.021		61.128	61.16	0.051
2		38.457	38.475	0.000		47.267	47.314	0.101

Table 4 Optimal and Heuristic Policies: Kink Holding Costs

N	Form	$C(\mathbf{s}^*)$	$C(\mathbf{s}^a)$	Error (%)	Form	$C(\mathbf{s}^*)$	$C(\mathbf{s}^a)$	Error (%)
64	Kink	52.352	52.742	0.745	Kink	61.622	62.378	1.227
32	$k = 0.25$	51.903	52.086	0.353	$k = 0.75$	61.157	61.610	0.741
16		51.011	51.099	0.174		60.226	60.510	0.472
8		49.223	49.259	0.073		58.381	58.484	0.177
4		45.660	45.660	0.000		54.675	54.773	0.179
2		38.457	38.457	0.000		47.267	47.314	0.101

cases. More importantly, all s_j^l, s_j^u, and s_j^a move in the same pattern as s_j^*. Similar results are found in the affine and kink holding cost forms.

We finally examine the performance of the heuristic when leadtimes are not equal. Consider a four-stage benchmark system with $L_1 = L_2 = L_3 = L_4 = 1.5$, $h_1 = h_2 = h_3 = h_4 = 0.25$, $b = 39$, and $\lambda = 4$. We reduce the leadtime from 1.5 to 0.5 units for stage 1, 2, 3, or 4, respectively.

Table 5 is the comparison of the optimal and heuristic solutions and their corresponding costs. The max-

imum percentage error to the optimal cost is only 0.004%. Thus, the heuristic policy is highly adaptive to the optimal one under different system parameters.

5. Parametric Analysis and Managerial Insights

The simplicity of the bounding cost functions and the closed-form expressions of the solution bounds and the heuristic allow us to analyze the effect of system

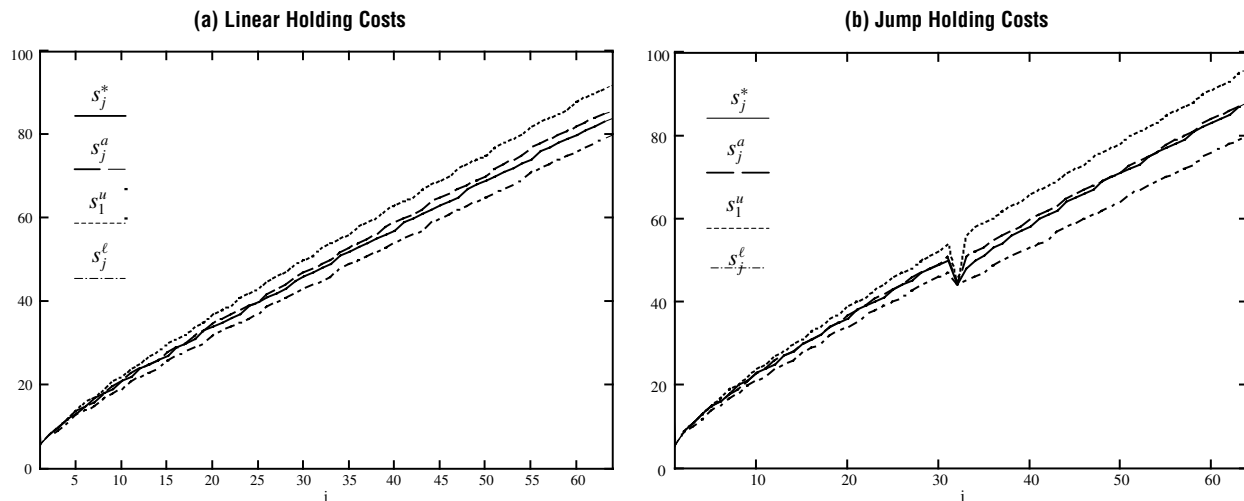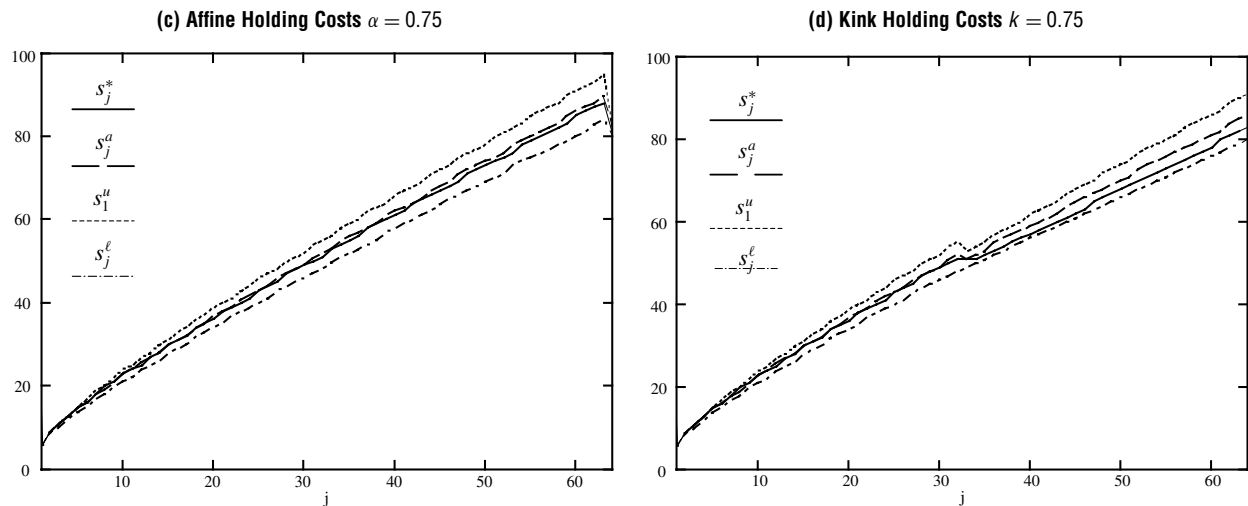**Figure 3 The Comparison of Heuristic and Optimal Policies**

(a) Linear Holding Costs

(b) Jump Holding Costs

Figure 3   Continued

(c) Affine Holding Costs $\alpha = 0.75$

(d) Kink Holding Costs $k = 0.75$

parameters easily. This is in sharp contrast to the not-so-transparent exact recursive procedure. Given the good performance of the heuristic, and given the fact that both $\mathbf{s}^*$ and C^* are "wrapped" by the bounds, it is reasonable to believe the parametric effects on the bounds hold for the true optimal policy and optimal cost as well. This section is devoted to the parametric analysis of the bounds and related managerial implications. When investigating the effect of changing one parameter, we assume the other parameters remain unchanged.

For convenience, denote the lower and upper bound cost ratios as

$$\theta_j^l = \frac{b + \sum_{i=j+1}^{N} h_i}{b + \sum_{i=1}^{N} h_i}, \quad \theta_j^u = \frac{b + \sum_{i=j+1}^{N} h_i}{b + \sum_{i=j}^{N} h_i}. \quad (26)$$

Thus, $s_j^l = F_j^{-1}(\theta_j^l)$ and $s_j^u = F_j^{-1}(\theta_j^u)$, $j = 1, \ldots, N$.

To see the magnitude of the changes, it is more convenient to work with normal distributions. For this purpose, we approximate the leadtime demand $\widetilde{D}_j$ by a normal distribution with mean $\mathsf{E}[\widetilde{D}_j]$ and $\mathsf{Var}[\widetilde{D}_j]$. Let λ be the demand arrival rate, and let Z denote the demand batch size with mean μ and variance σ^2. Then, $\widetilde{D}_j = \sum_{k=1}^{N(\widetilde{L}_j)} Z_k$, where $N(\widetilde{L}_j)$ is the total number of demand arrivals during $\widetilde{L}_j (= \sum_{i=1}^{j} L_i)$, which has a Poisson distribution with mean $\lambda \widetilde{L}_j$. Z_k is the batch size of the kth demand, an independent copy of Z. Thus,

$$\mathsf{E}[\widetilde{D}_j] = \lambda \mu \widetilde{L}_j,$$
$$\mathsf{Var}[\widetilde{D}_j] = \lambda(\mu^2 + \sigma^2)\widetilde{L}_j.$$

Let $\phi(\cdot)$ and $\Phi(\cdot)$ denote standard normal pdf and cdf, respectively, and define $z_j^l = \Phi^{-1}(\theta_j^l)$ and $z_j^u = \Phi^{-1}(\theta_j^u)$. Then, following the standard procedure (see

Table 5   Optimal and Heuristic Policies: Effect of Leadtimes

Leadtime	s_1^*	s_2^l	s_2^*	s_2^a	s_2^u	s_3^l	s_3^*	s_3^a	s_3^u	s_4^l	s_4^*	s_4^a	s_4^u	$C(\mathbf{s}^*)$	Error (%)
Benchmark	13	20	21	21	21	27	28	28	29	34	36	36	37	19.755	0.000
$L_1 = 0.5$	6	15	16	16	16	22	24	23	24	29	31	31	32	15.198	0.000
$L_2 = 0.5$	13	15	15	16	16	22	23	23	24	29	31	31	32	16.705	0.002
$L_3 = 0.5$	13	20	21	21	21	22	23	23	24	29	31	31	32	18.011	0.000
$L_4 = 0.5$	13	20	21	21	21	27	28	28	29	29	30	31	32	19.287	0.004

Zipkin 2000, pp. 215–216), we have

$$s_j^l = \lambda\mu\widetilde{L}_j + z_j^l\sqrt{\lambda(\mu^2+\sigma^2)\widetilde{L}_j}, \qquad (27)$$

$$s_j^u = \lambda\mu\widetilde{L}_j + z_j^u\sqrt{\lambda(\mu^2+\sigma^2)\widetilde{L}_j}, \qquad (28)$$

$$s_j^a = \lambda\mu\widetilde{L}_j + \frac{1}{2}(z_j^l+z_j^u)\sqrt{\lambda(\mu^2+\sigma^2)\widetilde{L}_j}, \qquad (29)$$

$$C_j^u(s_j^l) = \left(b+\sum_{i=1}^{N} h_i\right)\phi(z_j^l)\sqrt{\lambda(\mu^2+\sigma^2)\widetilde{L}_j}$$
$$+ \sum_{i=1}^{j-1}(h_{i+1}\lambda\mu\widetilde{L}_i), \qquad (30)$$

$$C_j^l(s_j^u) = \left(b+\sum_{i=j}^{N} h_i\right)\phi(z_j^u)\sqrt{\lambda(\mu^2+\sigma^2)\widetilde{L}_j}$$
$$+ \sum_{i=1}^{j-1}(h_{i+1}\lambda\mu\widetilde{L}_i). \qquad (31)$$

The closed-form expressions of the approximate echelon costs in (30) and (31) allow us to view the echelon-j cost in two parts: (1) the safety-stock cost due to the randomness of echelon-j leadtime demand $\widetilde{D}_j$, and (2) the average holding cost of pipeline inventory. Recall that we have excluded the average shipping/processing cost from consideration.

5.1. Effect of Cost Parameters
From (26) it is easy to see that as h_j increases, θ_i^l decreases for $i \geq j$ and increases for $i < j$. Also, as h_j increases, θ_i^u decreases for $i = j$, increases for $i < j$, and remains unchanged for $i > j$. This implies the same directional change for s_i^l, s_i^u, and s_i^a. On the other hand, as b increases, all cost ratios increase, implying higher values of all base-stock levels.

Examining the constructions of the bounding functions, it is also easy to see that for any fixed y, as either h_j or b increases, $C_i^l(y)$ and $C_i^u(y)$ for all i increase due to the increased coefficients. This in turn leads to higher minimum values $C_i^l(s_i^u)$ and $C_i^u(s_i^l)$ for all i. The same argument also applies to the decomposition of the optimal cost $C_i(y)$ for all i.

To summarize, we have:

PROPOSITION 4. *For $j = 1, 2, \ldots, N$,*
(a) *As h_j increases, s_i^l increases for $i = 1, \ldots, j-1$, but decreases for $i = j, \ldots, N$; s_i^u increases for $i = 1, \ldots, j-1$, decreases for $i = j$, and remains unchanged for $i = j+1, \ldots, N$. Consequently, s_i^a increases for $i = 1, \ldots, j-1$, decreases for $i = j, \ldots, N$.*
(b) *As b increases, s_i^l, s_i^u, and s_i^a increase for all i.*
(c) *As either h_j or b increases, $C_i^u(s_i^u)$, $C_i^u(s_i^l)$, and $C_i(s_i^*)$ increase for all i. In particular, $C^* = C(\mathbf{s}^*)$ increases.*

Thus, roughly speaking, increasing the local holding cost rate at one stage leads to increased downstream echelon base stocks but decreased upstream (including that stage itself) echelon base stocks. However, the optimal system cost always increases. Also, as the backorder cost rate increases, the optimal echelon base-stock levels all increase, leading to both increased system stock and increased system cost.
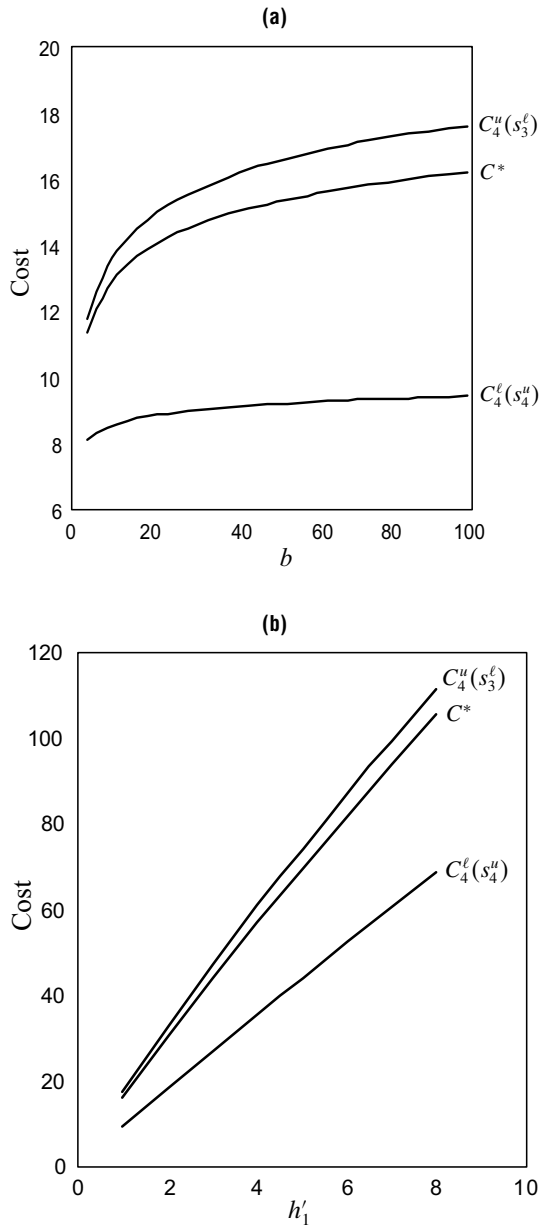
From the normal approximation, we can see that the cost parameters affect the optimal policies only through the ratios θ_j^l and θ_j^u, which determine z_j^l and z_j^u, while affecting the optimal cost with an additional factor $b+\sum_{i=j+1}^{N} h_i$.

We now illustrate these properties by some numerical examples. Figure 4(a) is a four-stage system with the Poisson demand. The system parameters are $h_1 = h_2 = h_3 = h_4 = 0.25$, $L_1 = L_2 = L_3 = L_4 = 0.25$, $\lambda = 16$. We range b from 9 to 99. It is clear that as b increases, $C_4^u(s_4^l)$, C^*, and $C_4^l(s_4^u)$ are increasing. Notice that C^* is closer to $C_4^u(s_4^l)$ when b is smaller. This is not intuitive because when b is sufficiently large, the serial system will allocate more stocks to stage 1 (to avoid high penalty cost) so that the system should perform more like a NV(h_1', b, D_4) system.

Figure 4(b) is the same system with $b = 99$, but we range h_1' from 1 to 8. We assume each $h_j = 0.25h_1'$. Again, as h_1' increases, $C_4^u(s_4^l)$, C^*, and $C_4^l(s_4^u)$ all increase. In this case, C^* is also closer to $C_4^u(s_4^l)$ when h_1' is smaller. Note that if the average shipment/precessing cost were included in the total cost expression, the performance of the approximation would be more attractive, as shown in Dong and Lee (2001).

The following proposition can be easily verified from Equations (27) and (28). It identifies the condi-

Figure 4   Optimal and Heuristic Costs

when $h_N \ll h_{N-1} \ll \cdots \ll h_1$, then the bounds for s_j^* are loose.

From Proposition 5 we can speculate that $C_N^u(s_N^l)$ approximates C^* well if h_N is large relative to the other echelon holding costs. In this case, according to the proposition, the gap between s_N^l and s_N^* is small. Because there is less incentive to hold inventory at stage N due to its higher holding cost, a serial system will perform just like a newsvendor system. For example, in the above four-stage system (Figure 4(b) system), if $h_4 = 10$, $h_3 = 1$, $h_2 = 1$, and $h_1 = 1$, then $C^* = 222.367$ and $C_4^u(s_4^l) = 223.382$ (error ≈ 0). However, if $h_4 = 5$, $h_3 = 4$, $h_2 = 3$, and $h_1 = 1$, then $C^* = 192.417$ and $C_4^u(s_4^l) = 195.382$ (error $= 1.54\%$).

Note that holding costs consist of several components, such as cost of capital, facility, maintenance, and leakage/spoilage. Holding costs can be reduced by introducing new technology, through better management, or by outsourcing. To better allocate resources, it is interesting to know which stage can lead to the greatest benefit of holding cost reduction. This is equivalent to identifying the bottleneck stage such that by reducing its echelon holding cost the total optimal cost is minimized. Let $C_N^u(y \mid h_j)$ $(C_N^l(y \mid h_j))$ be the system upper (lower) bound cost function when echelon holding cost rate at stage j is h_j. Proposition 6 provides insights on this issue.

PROPOSITION 6. *Consider an N-stage system. Define*
$\Delta_{h_j} C_N^u = \min C_N^u(y \mid h_j) - \min C_N^u(y \mid h_j - \Delta h)$ *and*
$\Delta_{h_j} C_N^l = \min C_N^l(y \mid h_j) - \min C_N^l(y \mid h_j - \Delta h)$. *Then we have*
   (a) $\Delta_{h_1} C_N^u \le \Delta_{h_2} C_N^u \le \cdots \le \Delta_{h_N} C_N^u$,
   (b) $\Delta_{h_1} C_N^l \le \Delta_{h_2} C_N^l \le \cdots \le \Delta_{h_N} C_N^l$, *provided* $b \ge h_N$.

Proposition 6 implies that reducing echelon holding cost at stage N is most effective. Because the optimal cost is bounded by C_N^l and C_N^u, we conjecture that it is also the most effective to reduce holding cost at stage N for C_N. The following numerical example demonstrates this point.

Consider a four-stage benchmark system with $L_1 = L_2 = L_3 = L_4 = 0.25$, $h_1 = h_2 = h_3 = h_4 = 2.5$, $b = 99$, and $\lambda = 16$. In this case, we reduce the holding cost from 2.5 to 0.25 for stage 1, 2, 3, or 4, respectively. Tables 6

tions under which the difference between the bounds is smaller so the heuristic is more accurate.

PROPOSITION 5. *Using normal approximation, as h_j increases, the distance between s_i^l and s_i^u will be larger for $i = j+1, j+2, \ldots, N$, and will be smaller for $i = 2, 3, \ldots, j$. In particular, when $h_N \gg h_{N-1} \gg \cdots \gg h_1$, the difference between s_j^l and s_j^u is very small. On the contrary,*

Table 6   Optimal and Heuristic Solutions: Holding Cost Reduction

Leadtime	s_1^*	s_2^l	s_2^*	s_2^a	s_2^u	s_3^l	s_3^*	s_3^a	s_3^u	s_4^l	s_4^*	s_4^a	s_4^u	$C(\mathbf{s}^*)$	$C(\mathbf{s}^a)$
Benchmark	8	13	14	14	14	17	18	18	19	21	23	23	24	128.591	128.591
$h_1 = 0.25$	11	14	14	14	14	18	18	19	19	22	23	23	24	119.151	119.227
$h_2 = 0.25$	8	14	16	16	17	18	19	19	19	22	23	23	24	107.532	107.573
$h_3 = 0.25$	8	13	14	14	14	18	21	21	23	22	23	23	24	96.373	96.373
$h_4 = 0.25$	8	13	14	14	14	17	18	18	19	22	25	25	28	84.263	84.263

and 7 are the optimal and heuristic (rounding-up) solutions and their corresponding costs.

From the optimal solution in Table 6, we notice that s_j^* *increases when the holding cost reduction occurs at stage j, but the other optimal echelon base-stock levels remain stable* (except when $h_2 = 0.25$, both s_2^* and s_3^* increase). Thus, when h_j decreases, there is an *inventory shift* between stage j and all its upstream stages— while the local inventory at stage j increases, the total amount of inventory holding at stage $j+1, \dots, N$ tends to decrease. For example, in the benchmark case the optimal installation base-stock level $(s_1^{\prime*}, s_2^{\prime*}, s_3^{\prime*}, s_4^{\prime*})$ is $(8, 6, 4, 5)$. When h_2 decreases from 2.5 to 0.25, the optimal installation policy is $(8, 8, 3, 4)$.

In Table 7, the benchmark example has the largest average cost of 128.591, which includes $\tau_4 = 60$ and $G_4(s_4^*) = 68.597$. As we reduce the holding cost on h_1, h_2, h_3, and h_4, the optimal cost is decreased to 119.151, 107.532, 96.373, and 84.263, respectively. Note that this decreased optimal cost occurs on *both* τ_4 and $G_4(s_4^*)$. On the other hand, the heuristic cost $C_4^l(s_4^u)$ and $C_4^u(s_4^l)$ are both decreasing in the same pattern as well, but the decreased cost only occurs at τ_4. Nevertheless, we can still identify the bottleneck stage through the comparison of all $C_4^l(s_4^u)$ or $C_4^u(s_4^l)$ values. Thus, *reduction of holding cost at upper stages is more effective than at lower stages*.

In fact, we can intuitively determine the effect of reduction on the echelon holding cost. If we can only reduce a fixed amount of echelon holding cost for any stage, it is intuitive that reducing the echelon holding cost at stage N will be the most effective, because every unit of inventory that enters into this serial system will be beneficial from cost reduction. Take the above four-stage system for example; if we reduce h_4 from 2.5 to 0.25, then every unit from stage 4 to stage 1 will be beneficial from cost reduction. If we reduce h_3, then some inventories at stage 4 cannot take the advantage of h_3 reduction.

5.2. Effect of Leadtimes

For any j, if L_j increases, $\widetilde{D}_j$ becomes stochastically larger, leading to stochastically larger $\widetilde{D}_i$ for all $i \geq j$ and leaving $\widetilde{D}_i$ unchanged for $i < j$. This implies increased s_i^l, s_i^u, and s_i^a for $i \geq j$ while leaving those quantities unchanged for $i < j$. See Song (1994). The effect on the cost bounds is not easy to see by using the stochastic comparison technique. It is, however, quite transparent from the normal approximation (30) and (31). To summarize, we have:

PROPOSITION 7. *As L_j, $j = 1, 2, \dots, N$, increases,*
(a) s_i^l, s_i^u, and s_i^a *increase for $i = j, \dots, N$, but remain the same for $i = 1, \dots, j-1$.*
(b) *Using normal approximation, both the lower and upper bounds for the optimal system costs, $C_N^l(s_N^u)$ and $C_N^u(s_N^l)$, increase.*

Next, we address the following question: If we can reduce one unit of leadtime, at which stage should we perform the reduction to achieve the maximum cost savings?

Let $C_N^u(y \mid L_j)$ $(C_N^l(y \mid L_j))$ be the system upper (lower) bound cost function when leadtime at stage j is L_j.

PROPOSITION 8. *Consider an N-stage system. Define* $\Delta_{L_j} C_N^u = \min C_N^u(y \mid L_j) - \min C_N^u(y \mid L_j - \Delta L)$ *and*

Table 7   Optimal and Heuristic Costs: Holding Cost Reduction

Leadtime	τ_4	$G_4^l(s_4^u)$	$G_4^u(s_4^l)$	$G_4(s_4^*)$	$C_4^l(s_4^u)$	$C_4^u(s_4^l)$	$C^* = C(\mathbf{s}^*)$
Benchmark	60	25.217	75.675	68.591	85.217	135.675	128.591
$h_1 = 0.25$	60	25.217	62.120	59.151	85.217	122.12	119.151
$h_2 = 0.25$	51	25.217	62.120	56.532	76.217	113.120	107.532
$h_3 = 0.25$	42	25.217	62.120	54.373	67.217	104.120	96.373
$h_4 = 0.25$	33	3.437	62.120	51.263	36.437	95.120	84.263

$\Delta_{L_j} C_N^l = \min C_N^l(y \mid L_j) - \min C_N^l(y \mid L_j - \Delta L)$. *Then we have*

(a) $\Delta_{L_1} C_N^u \geq \Delta_{L_2} C_N^u \geq \cdots \geq \Delta_{L_N} C_N^u$, *and*
(b) $\Delta_{L_1} C_N^l \geq \Delta_{L_2} C_N^l \geq \cdots \geq \Delta_{L_N} C_N^l$.

This implies that according to the heuristic solution, stage 1 is the bottleneck stage for leadtime reduction. Because the optimal system cost C^* is bounded by $C_N^l(s_N^u)$ and $C_N^u(s_N^l)$, we expect this conclusion will also apply to the optimal cost C^*. Indeed, in Table 5, as we shorten the leadtimes L_1, L_2, L_3, and L_4 by the same amount, the optimal cost is decreased from the benchmark 19.755 to 15.198, 16.705, 18.011, and 19.287, respectively. Thus, *reduction of leadtime at lower stages is more effective than at upper stages.*

5.3. Effect of Demand Variability

Suppose we face a more variable demand size, Z', than Z in the sense of increasing convex ordering, i.e., $Z' \geq_{ic} Z$ (see, e.g., Ross 1983), but $\mathsf{E}[Z'] = \mathsf{E}[Z]$. That is, $\mathsf{E}[g(Z')] \geq \mathsf{E}[g(Z)]$ for all convex $g(\cdot)$. In particular, $\mathsf{Var}[Z'] \geq \mathsf{Var}[Z]$. Then it can be shown that $\widetilde{D}_j' \geq_{ic} \widetilde{D}_j$ for all j, where $\widetilde{D}_j'$ is the counterpart of $\widetilde{D}_j$ with batch size Z'. Applying the results in Song (1994) and the normal approximation, we have

PROPOSITION 9. *If the demand batch size Z is more variable but has the same mean, then*

(a) *both the lower and upper bounds for the optimal system costs, $C_N^l(s_N^u)$ and $C_N^u(s_N^l)$, increase, and*
(b) *using normal approximation, s_i^l, s_i^u, and s_i^a increase for all i due to increased σ.*

Because system optimal cost and optimal solution s_j^* both increase in demand variability, more inventory should be held when demand variance increases. A natural question is, how should we allocate additional units of inventory to mitigate the increased cost? This question is equivalent to examining the optimal *installation* base-stock levels. The reason is that we should always allocate additional inventory to the stage with the maximum increase in optimal installation base-stock levels when the demand variability increases. Because it is not possible to obtain a closed-form solution for optimal installation base-stock levels, we can examine their behaviors through our heuristic.

We first provide a proposition to summarize our findings. Let superscript " $'$ " denote the installation terms for all variables. Define $\Delta s_j^{'l} = s_j^{'l}(\sigma + \Delta\sigma) -$

$s_j^{'l}(\sigma)$, $\Delta s_j^{'u} = s_j^{'u}(\sigma + \Delta\sigma) - s_j^{'u}(\sigma)$, and $\Delta s_j^{'a} = s_j^{'a}(\sigma + \Delta\sigma) - s_j^{'a}(\sigma)$. We can conclude the following results.

PROPOSITION 10. *For an N-stage system, assume all leadtimes are equal. Then*

(a) $\Delta s_1^* \geq \Delta s_j^{'l}$, *for $j = 2, \ldots, N$, and*
(b) $\Delta s_1^* \geq \Delta s_j^{'u}$, *for $j = 2, \ldots, N$, if $b \gg h_1'$ or $h_1 \leq h_j$, for $j = 2, \ldots, N$.*

Perhaps it is not possible that $h_N \geq h_{N-1} \geq \cdots \geq h_1$ in the serial systems. However, the backorder cost rate is generally much larger than the holding cost parameters. Therefore, from the heuristic, stage 1 is the bottleneck stage when demand variance increases, because s_1^* increases faster than both lower and upper bound solutions for upper stages. We verify this conclusion through a numerical example below.
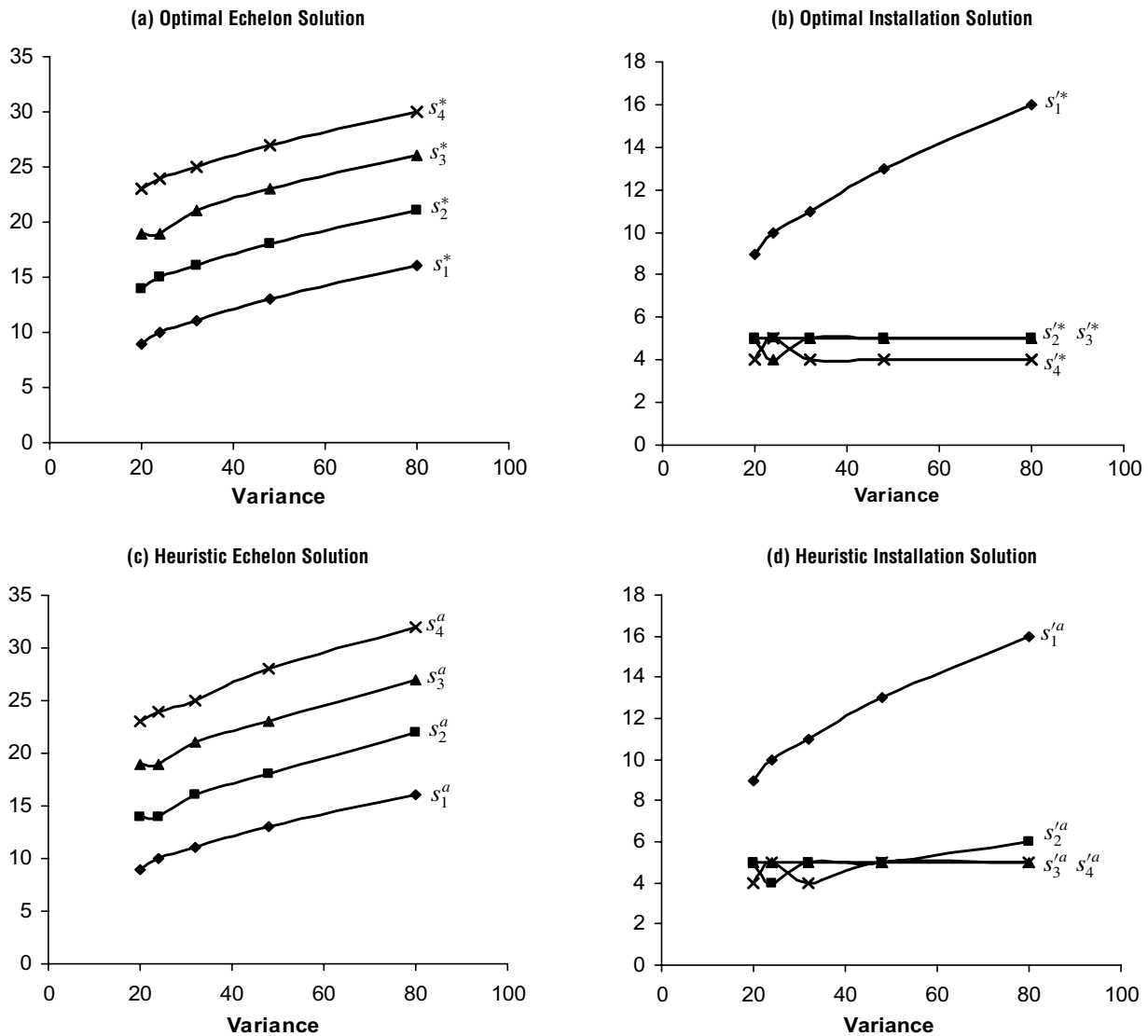
Consider the following four-stage system with negative binomial demand. The system parameters are $h_1 = h_2 = h_3 = h_4 = 0.25$, $L_1 = L_2 = L_3 = L_4 = 0.25$, $b = 9$. We fix the mean of demand to 16 and change the variance from 20, 24, 36, 48, and to 80. Figures 5(a) through 5(d) show the optimal and heuristic echelon and installation base-stock levels.

From Figure 5(a), when the demand variance increases, the optimal echelon base-stock levels increase for all stages because of the increasing base-stock level at stage 1. However, the increase in installation inventory at stage 1 is higher than that at the other stages. It can be illustrated in Figure 5(b). In this graph, the installation base-stock level at stage 2, 3, or 4 is relatively stable, but increases dramatically at stage 1 as demand variance increases. Figures 5(c) and 5(d) are the heuristic solutions in this example. Clearly, we can also draw the same conclusion as above from the heuristic.

6. Incentives and Cooperation in Decentralized Supply Chains

So far, we have focused only on the centralized control mechanism—a central planner or an owner knows the information for the entire system and calculates the optimal base-stock level for each stage (e.g., division), assuming the division manager would follow this policy to control local inventory. In reality, however, without appropriate incentives division

Figure 5   The Effect of Demand Variance to Optimal Base-Stock Levels

(a) Optimal Echelon Solution

(b) Optimal Installation Solution

(c) Heuristic Echelon Solution

(d) Heuristic Installation Solution

managers may pursue their own benefits and operate according to local optimal policies, which may not lead to optimal system performance. It is therefore important to identify incentive-compatible schemes to facilitate coordination so that, while each division manager minimizes her own cost, the team as a whole would achieve the system optimality. Several authors have proposed possible ways to accomplish this, such as Chen (1999), Lee and Whang (1999), and Porteus (2000). In this section, we relate our results to some of these works.

First, the results developed in the paper can be used to simplify the calculation of Chen's incentive-compatible scheme, which is based on the *accounting inventory levels* (the installation inventory levels that would be experienced if the stage had an ample supply). We use the stationary beer game in Chen (1999) as an example to illustrate this.

In the stationary beer game, there are four stages. Stages 1 through 4 are referred to retailer, wholesaler, distributor, and factory, respectively. Stage j has a delivery leadtime L_j and information leadtime l_j,

where $L_j = 2$ for all j, $l_4 = 1$, and $l_j = 2$ for $j < 4$. Each stage implements an installation base-stock policy. Demand at stage 1 is normally distributed with $\mu = 50$, $\sigma = 10$. Echelon holding cost rates are $h_j = 0.25$ for all j; backorder cost rate is $b = 10$. Chen shows that the information leadtimes can be treated as the transportation leadtimes when searching for the optimal policies. As mentioned in §3, because inventory costs are assessed at the end of each period in the beer game, the accumulated leadtimes used in our heuristic should be $\widetilde{L}_1 = 5$, $\widetilde{L}_2 = 9$, $\widetilde{L}_3 = 13$, and $\widetilde{L}_4 = 16$. Table 8 shows the optimal policy and cost as well as the newsvendor approximation. The relative error of the heuristic is less than 0.05%. (Here, we take the information leadtime as the physical transportation leadtime so that the optimal cost is 399.501 rather than 65 as mentioned in Chen. Also, there is a discrepancy on s_4^* solution: Our answer is 862 rather than 863 as reported in the same paper.)

The incentive-compatible scheme introduced in Chen works as follows. Suppose the optimal installation policy is $(s_1'^*, \ldots, s_N'^*)$. Assume each division j has an ample supply. Because the holding cost rate is h_j and demand information is known to every division, the division manager would choose $s_j'^*$ to minimize his own cost if his backorder cost rate is b_j, where

$$b_j = \frac{h_j \Psi_j(s_j'^*)}{1 - \Psi_j(s_j'^*)}, \tag{32}$$

and $\Psi_j(\cdot)$ is the cdf of D_j. Therefore, the owner only needs to tell each division manager j a set of holding and backorder cost rates (h_j, b_j). As a result, the entire system operates optimally.

Note that in implementing this scheme, the key is devising b_j, for which the owner first needs to compute $(s_1'^*, \ldots, s_N'^*)$ recursively from the optimality recursion. By replacing $s_j'^*$ with $s_j'^a$ in (32), the entire calculation can be simplified substantially, without involving any recursion. Continuing the stationary beer game example, the installation base-stock levels $(s_1'^*, s_2'^*, s_3'^*, s_4'^*) = (295, 210, 206, 152)$. From (32), the incentive-compatible penalty rates are $(b_1, b_2, b_3, b_4) = (10.75, 0.5603, 0.4043, 0.3006)$. By using $s_j'^a$ in (32), we have $(b_1^a, b_2^a, b_3^a, b_4^a) = (10.75, 0.5603, 0.4755, 0.3297)$. Now, with the set of holding and backorder cost rates $(h_j, b_j^a)_{j=1}^4$, the division managers will choose the heuristic solution $(s_1'^a, s_2'^a, s_3'^a, s_4'^a) = (295, 210, 208, 153)$. Thus, our heuristic yields an easy-to-compute and near-optimal incentive scheme, under which the overall system cost is only slightly higher than the optimal cost.

Porteus (2000) proposes responsibility tokens as a way to implement the decentralized supply-chain coordination scheme of Lee and Whang (1999), which in turn can be viewed as proposing a way to operationalize the decentralized management scheme implicit in Clark and Scarf (1960). If we revise the original scheme of Porteus by sending all tokens all the way to stage 1, then the backorder cost incurred at stage 1 will be completely transferred/charged to stage N. Thus, there would be no incentive for stage $N-1$ to stage 1 to hold inventories. As a result, the original N-stage system would be collapsed into a single-stage system where the holding cost rate is h_N, backorder cost rate is b, leadtime demand is $\widetilde{D}_N$, and the echelon-N manager has full responsibility for the inventory in the entire system. The resulting cost function is exactly the same as C_N^l, the lower bound cost function at stage N. Similarly, by replacing h_N with $\sum_{i=1}^N h_i$ in the above scheme, we can obtain C_N^u.

7. Concluding Remarks

In this paper, we have developed an easily implementable heuristic to the optimal echelon base-stock levels for an N-stage serial system by solving $2N$ single-stage newsvendor-type problems. The analysis and the closed-form expressions revealed insights into the key drivers of the optimal policy. It sheds light on how system parameters are interrelated, and the corresponding physical meanings. We observe that the cost parameters determine the shape of the echelon cost functions, while the leadtimes and the demand rate mainly influence the position of the echelon cost

Table 8   The Stationary Beer Game Solutions

Stage	s_j^l	s_j^*	s_j^a	s_j^u	$s_j'^*$	$s_j'^a$	$C(\mathbf{s}^*)$	$C(\mathbf{s}^a)$
$j = 1$	295	295	295	295	295	295	—	—
$j = 2$	501	505	505	510	210	210	—	—
$j = 3$	704	711	713	722	206	208	—	—
$j = 4$	854	862	866	879	151	153	399.501	399.667

functions. The results presented in this paper will ease the classroom teaching and real-world implementation. They also allow us to derive various managerial insights which would be difficult, if not impossible, to obtain analytically from the exact algorithm.

The results developed here can be easily adapted for assembly systems, following the approach of Rosling (1989) to convert an assembly system into an equivalent serial system. Future research directions will include extending the analysis to systems with other structures, such as the general (Q, r) systems, the distribution systems, as well as systems with nonstationary demands.

Acknowledgments

Appendix

PROOF OF PROPOSITION 2. For a given $IP_j = y$ at stage j and the known $s_1^*, \ldots, s_{j-1}^*$, then

$$
\begin{aligned}
C_j(y) &= \mathsf{E}[h_j IN_j + C_{j-1}(IP_{j-1}) \mid IP_j = y] \\
&= \mathsf{E}[h_j IN_j + h_{j-1} IN_{j-1} + C_{j-2}(IP_{j-2}) \mid IP_j = y] \\
&= \cdots \\
&= \mathsf{E}\left[\sum_{i=2}^{j} h_i IN_i + C_1(IP_1) \mid IP_j = y\right] \\
&= \mathsf{E}\left[\sum_{i=1}^{j} h_i IN_i + (b+h_1')[IN_1]^- \mid IP_j = y\right] \\
&= \mathsf{E}\left[\sum_{i=1}^{j} h_i I_i + \left(b+h_1' - \sum_{i=1}^{j} h_i\right)B \mid IP_j = y\right] \\
&= \mathsf{E}\left[\sum_{i=1}^{j} h_i\left(\sum_{k=1}^{i} I_k' + \sum_{k=1}^{i-1} D_k\right) + \left(b+\sum_{i=j+1}^{N} h_i\right)B \mid IP_j = y\right] \\
&= \mathsf{E}\left[h_j I_j' + (h_j + h_{j-1})I_{j-1}' + \cdots + \left(\sum_{i=1}^{j} h_i\right)I_1'\right. \\
&\quad \left. + \left(b+\sum_{i=j+1}^{N} h_i\right)B \mid IP_j = y\right] + \tau_j. \quad \square
\end{aligned}
$$

PROOF OF THEOREM 3. Because part (a) follows immediately from part (c), and part (d) follows immediately from (13), we only need to prove parts (b) and (c).

We first prove (b) by induction. To show $s_j^* \le s_j^u$, it is sufficient to show $\Delta G_j(y) \ge \Delta G_j^l(y)$ or, equivalently, $\Delta C_j(y) \ge \Delta C_j^l(y)$, for all y and $j \ge 1$. Note that

$$
\Delta C_j^l(y) = h_j - \left(b+\sum_{i=j}^{N} h_i\right) \mathsf{P}(\widetilde{D}_j > y).
$$

When $k = 1$, from (d), $\Delta C_1(y) = \Delta C_1^l(y)$. Assuming $k = j-1$ is true, i.e., $\Delta C_{j-1}(y) \ge \Delta C_{j-1}^l(y)$. When $k = j$,

$$
C_j(y) = \mathsf{E}[h_j(y-D_j) + C_{j-1}(s_{j-1}^* \wedge (y-D_j))].
$$

Conditioning on $D_j = d_j$, if $s_{j-1}^* \le y - d_j$,

$$
\Delta C_j(y \mid d_j) = h_j \ge h_j - (b+h_j)\mathsf{P}(\widetilde{D}_{j-1} > y - d_j) = \Delta C_j^l(y \mid d_j). \quad \text{(A1)}
$$

If $s_{j-1}^* > y - d_j$ (so $s_{j-1}^* \ge y + 1 - d_j$),

$$
\begin{aligned}
\Delta C_j(y \mid d_j) &= h_j + C_{j-1}(y+1-d_j) - C_{j-1}(y-d_j) \\
&= h_j + \Delta C_{j-1}(y-d_j) \\
&\ge h_j + \Delta C_{j-1}^l(y-d_j) \quad \text{(by induction assumption)} \\
&= h_j + h_{j-1} - \left(b + \sum_{i=j-1}^{N} h_i\right)\mathsf{P}(\widetilde{D}_{j-1} > y - d_j) \\
&= h_j - \left(b + \sum_{i=j}^{N} h_i\right)\mathsf{P}(\widetilde{D}_{j-1} > y - d_j) \\
&\quad + h_{j-1}(1 - \mathsf{P}(\widetilde{D}_{j-1} > y - d_j)) \\
&= \Delta C_j^l(y \mid d_j) + h_{j-1}(1 - \mathsf{P}(\widetilde{D}_{j-1} + d_j > y)) \\
&\ge \Delta C_j^l(y \mid d_j). \quad \text{(A2)}
\end{aligned}
$$

Therefore, from (A1) and (A2), we have

$$
\Delta C_j(y) \ge \Delta C_j^l(y).
$$

Thus, by induction, we know $\Delta C_j(y) \ge \Delta C_j^l(y)$ for all j and $s_j^* \le s_j^u$.

We next show $s_j^* \ge s_j^l$. Similarly, we need to show $\Delta C_j^u(y) \ge \Delta C_j(y)$. Below we show $\Delta C_k^u(y) \ge \Delta C_k(y) \ge \Delta \underline{C}_k(y)$ for $k \ge 1$ by induction.

For $k = 1$, because $C_1^u(y) = C_1(y)$, the first inequality is immediate. To see the second inequality, note that

$$
\text{if } y \le s_1^*, \text{ then } C_1(y) = \underline{C}_1(y), \text{ so } \Delta C_1(y) = \Delta \underline{C}_1(y), \quad \text{(A3)}
$$

and

$$
\text{if } y > s_1^*, \text{ then } C_1(y) \ge \underline{C}_1(y), \text{ so } \Delta C_1(y) \ge \Delta \underline{C}_1(y). \quad \text{(A4)}
$$

Hence, we have $\Delta C_1^u(y) \ge \Delta C_1(y) \ge \Delta \underline{C}_1(y)$.

Assume $k = j-1$ is true; i.e., $\Delta C_{j-1}^u(y) \ge \Delta C_{j-1}(y) \ge \Delta \underline{C}_{j-1}(y)$. For $k = j$,

$$
C_j(y) = \mathsf{E}[h_j(y-D_j) + \underline{C}_{j-1}(y-D_j)]
$$

and

$$
C_j^u(y) = \mathsf{E}[h_j(y-D_j) + C_{j-1}(y-D_j)],
$$

so the derivatives of $C_j(y)$ and $C_j^u(y)$ can be expressed as

$$\Delta C_j(y) = h_j + \mathsf{E}[\Delta \underline{C}_{j-1}(y - D_j)]$$

and

$$\Delta C_j^u(y) = h_j + \mathsf{E}[\Delta C_{j-1}(y - D_j)].$$

Applying the second inequality in the induction assumption, we have

$$\mathsf{E}[\Delta C_{j-1}(y - D_j)] \ge \mathsf{E}[\Delta \underline{C}_{j-1}(y - D_j)].$$

Therefore,

$$\Delta C_j^u(y) = h_j + \mathsf{E}[\Delta C_{j-1}(y - D_j)]$$
$$\ge h_j + \mathsf{E}[\Delta \underline{C}_{j-1}(y - D_j)] = \Delta C_j(y). \qquad (A5)$$

Finally, by replacing subscript 1 with subscript j in (A3) and (A4), we have $\Delta C_j(y) \ge \Delta \underline{C}_j(y)$ for every y. As a result, together with (A5), we have $\Delta C_j^u(y) \ge \Delta C_j(y) \ge \Delta \underline{C}_j(y)$. That is, $s_j^l \le s_j^*$ for all $j \ge 1$.

We proceed to show (c). We first show that $C_j^l(y) \le C_j(y)$ for all j. Note that from Proposition 2, when $y = 0$, $\mathsf{E}[I_i'] = 0$ for all $i \le j$ and $\mathsf{E}[B] = \mathsf{E}[\widetilde{D}_j]$. So,

$$C_j(0) = \tau_j + G_j(0) = \tau_j + \left(b + \sum_{i=j+1}^{N} h_i\right)\mathsf{E}[\widetilde{D}_j] = \tau_j + G_j^l(0) = C_j^l(0).$$

Therefore,

$$C_j(y) = C_j(0) + \sum_{x=1}^{y-1} \Delta C_j(x)$$
$$\ge C_j^l(0) + \sum_{x=1}^{y-1} \Delta C_j^l(x)$$
$$= C_j^l(y).$$

Finally, we show $C_j(y) \le C_j^u(y)$ for all j. Recall $s_{j-1}^* = \arg\min C_{j-1}(y)$ for all j. So

$$C_j(y) = \mathsf{E}[h_j(y - D_j) + C_{j-1}(s_{j-1}^* \wedge (y - D_j))] \qquad (A6)$$
$$\le \mathsf{E}[h_j(y - D_j) + C_{j-1}(y - D_j)] \qquad (A7)$$
$$= \mathsf{E}[h_j(y - D_j) + h_{j-1}(y - D_j - D_{j-1})$$
$$+ C_{j-2}(s_{j-2}^* \wedge (y - D_j - D_{j-1}))]$$
$$\le \mathsf{E}[h_j(y - D_j) + h_{j-1}(y - D_j - D_{j-1})$$
$$+ C_{j-2}(y - D_j - D_{j-1})]$$
$$= \cdots$$
$$= \mathsf{E}\left[\sum_{i=2}^{j} h_i\left(y - \sum_{k=i}^{j} D_k\right) + C_1\left(s_1^* \wedge \left(y - \sum_{i=2}^{j} D_i\right)\right)\right]$$
$$\le \mathsf{E}\left[\sum_{i=2}^{j} h_i\left(y - \sum_{k=i}^{j} D_k\right) + C_1\left(y - \sum_{i=2}^{j} D_i\right)\right]$$

$$= \sum_{i=2}^{j} h_i\mathsf{E}[\widetilde{D}_{i-1}] + \mathsf{E}\left[\left(\sum_{i=1}^{j} h_i\right)(y - \widetilde{D}_j) + (b + h_1')(y - \widetilde{D}_j)^-\right]$$
$$= \tau_j + G_j^u(y)$$
$$= C_j^u(y). \qquad \square$$

PROOF OF PROPOSITION 6. From Equations (30) and (31), we have

$$\Delta_{h_j} C_N^u = \left[\left(b + \sum_{i=1}^{N} h_i\right)\phi\left(\Phi^{-1}\left(\frac{b}{b + \sum_{i=1}^{N} h_i}\right)\right)\right.$$
$$\left. - \left(b + \sum_{i=1}^{N} h_i - \Delta h\right)\phi\left(\Phi^{-1}\left(\frac{b}{b + \sum_{i=1}^{N} h_i - \Delta h}\right)\right)\right]$$
$$\times \sqrt{\lambda(\mu^2 + \sigma^2)\widetilde{L}_j} + \Delta h \lambda \mu \widetilde{L}_{j-1}.$$

Hence,

$$\Delta_{h_j} C_N^u - \Delta_{h_{j+1}} C_N^u = \Delta h \lambda \mu(\widetilde{L}_{j-1} - \widetilde{L}_j) < 0$$

so $\Delta_{h_{j+1}} C_N^u > \Delta_{h_j} C_N^u$, for $j = 1, \dots, N-1$.

On the other hand,

$$\Delta_{h_j} C_N^l = \left[(b + h_N)\phi\left(\Phi^{-1}\left(\frac{b}{b + h_N}\right)\right)\right.$$
$$\left. - (b + h_N)\phi\left(\Phi^{-1}\left(\frac{b}{b + h_N}\right)\right)\right]$$
$$\times \sqrt{\lambda(\mu^2 + \sigma^2)\widetilde{L}_j} + \Delta h \lambda \mu \widetilde{L}_{j-1}$$
$$= \Delta h \lambda \mu \widetilde{L}_{j-1} \quad \text{for } j < N. \qquad (A8)$$

Thus,

$$\Delta_{h_j} C_N^l > \Delta_{h_{j-1}} C_N^l \quad \text{for } j = 1, \dots, N-1. \qquad (A9)$$

For $j = N$,

$$\Delta_{h_N} C_N^l = \left[(b + h_N)\phi\left(\Phi^{-1}\left(\frac{b}{b + h_N}\right)\right)\right.$$
$$\left. - (b + h_N - \Delta h)\phi\left(\Phi^{-1}\left(\frac{b}{b + h_N - \Delta h}\right)\right)\right]$$
$$\times \sqrt{\lambda(\mu^2 + \sigma^2)\widetilde{L}_j} + \Delta h \lambda \mu \widetilde{L}_{j-1}. \qquad (A10)$$

Because $b \ge h$,

$$\Phi^{-1}\left(\frac{b}{b + h_N - \Delta h}\right) \ge \Phi^{-1}\left(\frac{b}{b + h_N}\right) \ge 0.5.$$

Thus,

$$\phi\left(\Phi^{-1}\left(\frac{b}{b + h_N}\right)\right) \ge \phi\left(\Phi^{-1}\left(\frac{b}{b + h_N - \Delta h}\right)\right) \ge 0,$$

and hence (A10) $\ge$ (A8), or equivalently,

$$\Delta_{h_N} C_N^l \ge \Delta_{h_j} C_j^l \quad \text{for } j < N. \qquad (A11)$$

From (A9) and (A11), the results in Proposition 6 immediately follow. $\square$

PROOF OF PROPOSITION 8. From Equations (30) and (31), we obtain

$$\Delta_{L_j} C_N^u = \Delta L \left(\sum_{i=j+1}^{N} h_i \right) \lambda \mu$$
$$+ \left(b + \sum_{i=1}^{N} h_i \right) \phi(z_N^l) \sqrt{\lambda(\mu^2 + \sigma^2)} \left[\sqrt{\widetilde{L}_N} - \sqrt{\widetilde{L}_N - \Delta L} \right].$$

Therefore, $\Delta_{L_{j+1}} C_N^u - \Delta_{L_j} C_N^u < 0$ for $j = 1, \ldots, N-1$. Similarly, we can show $\Delta_{L_{j+1}} C_N^l - \Delta_{L_j} C_N^l < 0$ for $j = 1, \ldots, N-1$ in the same manner. $\square$

PROOF OF PROPOSITION 10. For notational simplicity, denote $\Delta \hat{\sigma} = \sqrt{\mu^2 + (\sigma + \Delta\sigma)^2} - \sqrt{\mu^2 + \sigma^2}$. Because all leadtimes are equally divided, we can change $\widetilde{L}_j$ to jL, where L is the leadtime for each stage. We first prove part (a). From Equation (27), we obtain

$$s_1^*(\sigma) = \lambda \mu \widetilde{L}_1 + z_1^l \sqrt{\lambda(\mu^2 + \sigma^2)\widetilde{L}_1},$$

and

$$s_j'^l(\sigma) = s_j^l - s_{j-1}^l$$
$$= \lambda\mu L_j + \sqrt{\lambda(\mu^2 + \sigma^2)L} \left(z_j^l \sqrt{j} - z_{j-1}^l \sqrt{j-1} \right).$$

Thus, we have

$$\Delta s_1^* = z_1^l \sqrt{\lambda L} \Delta \hat{\sigma}$$

and

$$\Delta s_j'^l = \sqrt{\lambda L} \left(z_j^l \sqrt{j} - z_{j-1}^l \sqrt{j-1} \right) \Delta \hat{\sigma}.$$

Note that $z_1^l \geq z_j^l$, for $j = 2, \ldots, N$, thus

$$\Delta s_1^* = \sqrt{\lambda L} z_1^l (1) \Delta \hat{\sigma}$$
$$\geq \sqrt{\lambda L} z_1^l (\sqrt{j} - \sqrt{j-1}) \Delta \hat{\sigma}$$
$$\geq \sqrt{\lambda L} z_j^l (\sqrt{j} - \sqrt{j-1}) \Delta \hat{\sigma}$$
$$= \sqrt{\lambda L} z_j^l \sqrt{j} \Delta \hat{\sigma} - \sqrt{\lambda L} z_j^l \sqrt{j-1} \Delta \hat{\sigma}$$
$$\geq \sqrt{\lambda L} z_j^l \sqrt{j} \Delta \hat{\sigma} - \sqrt{\lambda L} z_{j-1}^l \sqrt{j-1} \Delta \hat{\sigma}$$
$$= \Delta s_j'^l.$$

Therefore, part (a) is proved.

A similar proof works for part (b). However, to guarantee $z_1^u \geq z_j^u$ for $j = 2, \ldots, N$, the condition of $h_1 \leq h_j$ for $j = 2, \ldots, N$ must hold. Also, when b is sufficiently large, $z_1^u \approx z_2^u \approx \cdots \approx z_N^u$. Hence the result of part (b) still holds. $\square$

References

Axsäter, S., K. Rosling. 1993. Installation vs. echelon stock policies for multi-level inventory control. *Management Sci.* **39** 1274–1280.

Cachon, G., P. H. Zipkin. 1999. Competitive and cooperative inventory policies in a two-stage supply chain. *Management Sci.* **45** 936–953.

Chen, F. 1999. Decentralized supply chains subject to information delays. *Management Sci.* **45** 1076–1090.

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

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

Cohen, M., Y. S. Zheng, V. Agrawal. 1994. Service parts logistics benchmark study. Final project report, Center for Manufacturing and Logistics Research, Wharton School, University of Pennsylvania, Philadelphia, PA.

Dong, L., H. Lee. 2001. Optimal policies and approximations for a serial multi-echelon inventory system with time-correlated demand. *Oper. Res.* Forthcoming.

Federgruen, A., P. Zipkin. 1984. Computational issues in an infinite horizon multi-echelon inventory problem with stochastic demand. *Oper. Res.* **32** 818–836.

Gallego, G. 1998. New bounds and heuristics for (Q, r) policies. *Management Sci.* **44** 219–233.

——, P. Zipkin. 1999. Stock positioning and performance estimation in serial production-transportation systems. *Manufacturing Service Oper. Management* **1** 77–88.

Geoffrion, A. M. 1976. The purpose of mathematical programming is insight, not numbers. *Interfaces* **7** 81–92.

Glasserman, P. 1997. Bounds and asymptotics for planning critical safety stocks. *Oper. Res.* **45** 244–257.

Hopp, W. J., M. L. Spearman, R. Q. Zhang. 1997. Easily implementable inventory control policies. *Oper. Res.* **45** 327–340.

Lee, H., S. Whang. 1999. Decentralized multi-echelon supply chains: Incentives and information. *Management Sci.* **45** 633–640.

Porteus, E. L. 2000. Responsibility tokens in supply chain management. *Manufacturing Service Oper. Management* **2** 203–219.

Rosling, K. 1989. Optimal inventory policies for assembly systems under random demands. *Oper. Res.* **37** 565–579.

Ross, S. 1983. *Stochastic Processes*. Wiley, New York.

Song, J. S. 1994. The effect of leadtime uncertainty in a simple stochastic inventory model. *Management Sci.* **40** 603–613.

van Houtum, G. J., K. Inderfurth, W. H. M. Zijm. 1996. Material coordination in stochastic multi-echelon systems. *Eur. J. Oper. Res.* **95** 1–23.

Zipkin, P. H. 2000. *Foundations of Inventory Management*. McGraw-Hill, New York.