

Optimal and Heuristic Echelon (r, nQ, T) Policies in Serial Inventory Systems with Fixed Costs

Kevin H. Shang

Fuqua School of Business, Duke University, Durham, North Carolina 27708,
khshang@duke.edu

Sean X. Zhou

Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong,
Shatin, N.T., Hong Kong, zhoux@se.cuhk.edu.hk

This paper studies a periodic-review, serial inventory system in which echelon (r, nQ, T) policies are implemented. Under such a policy, each stage reviews its inventory in every T period and orders according to an echelon (r, nQ) policy. Two types of fixed costs are considered: one is associated with each order batch Q , and the other is incurred for each inventory review. The objective is to find the policy parameters such that the average total cost per period is minimized. This paper provides a method for obtaining heuristic and optimal policy parameters. The heuristic is based on minimizing lower and upper bounds on the total cost function. These total cost bounds, which are separable functions of the policy parameters, are obtained in two steps: First, we decompose the total cost into costs associated with each stage, which include a penalty cost for holding inadequate stock. Second, we construct lower and upper bounds for the penalty cost by regulating downstream policy parameters. To find the optimal solution, we further construct cost bounds for each echelon (a subsystem that includes a stage and all of its downstream stages) by regulating holding and backorder cost parameters. The echelon lower-bound cost functions, as well as the stage cost bounds, generate bounds for the optimal solution. In a numerical study, we find that the heuristic is near optimal when the ratio of the fixed cost to the holding cost at the most downstream stage is large. We also find that changing the optimal batch sizes may not affect the optimal reorder intervals or, equivalently, the delivery schedules under some conditions.

Subject classifications: multiechelon inventory systems; optimal policy; heuristic; lower and upper bounds; batch ordering; replenishment interval.

Area of review: Manufacturing, Service, and Supply Chain Operations.

History: Received January 2006; revisions received November 2006, December 2007, August 2008, November 2008, March 2009; accepted April 2009. Published online in *Articles in Advance* November 19, 2009.

1. Introduction

In production and distribution systems, materials are often ordered in batches, such as cases and pallets, at fixed intervals. For instance, most manufacturers order batches of materials from suppliers when they run their material requirements planning (MRP) systems. MRP systems often run on weekly, biweekly or monthly schedules, resulting in periodic batch-ordering from suppliers. Large retail chains replenish inventory in the same fashion; i.e., the products are often ordered in batches at fixed intervals. Periodic batch-ordering is also commonly seen in firms that contract with third-party logistics providers. EMC², a leading manufacturer for database servers in the United States, collaborates with United Parcel Service (UPS) for inventory replenishments. The UPS trucks deliver batches of materials to EMC² from its suppliers and from EMC² to its customers on fixed days of the week.¹

This paper considers an inventory system that models periodic batch-ordering. Specifically, we consider an N -stage, serial inventory system in which random customer demand occurs at stage 1; stage 1 orders from stage 2,

stage 2 orders from stage 3, etc., and stage N orders from an outside supplier that has ample supply. Each stage implements an echelon (r, nQ, T) policy. Under such a policy, stage j reviews its echelon inventory order position (defined in §2) every T_j periods and orders according to an echelon (r, nQ) policy. That is, if the echelon inventory order position is less than or equal to the echelon reorder point r_j , the stage orders a quantity of the smallest integer multiple of batch size Q_j so as to bring the echelon inventory order position above r_j . We call T_j the *reorder interval*. For order coordination, we assume that the order batches and the reorder intervals satisfy integer-ratio relations.

Two types of fixed costs are often incurred by companies that order batches at periodic intervals. The first type, referred to as *review cost*, is associated with an inventory review. For instance, a manager may have to physically review the inventory status at each inventory order period. In some cases, the review cost may include a shipping cost. At EMC², the shipping cost per truck is specified in a contract with UPS. Consequently, the shipping cost is always incurred in a customer's inventory review period even if

there is no material to ship from EMC² (e.g., when there are no orders or insufficient supply). The second type of fixed cost, referred to as *setup cost*, is associated with processing a batch. This may include order costs, setup costs, material handling costs, and quality control costs. At EMC², some machines need to be set up for assembling final products. These products are then inspected and packaged in batches, and loaded onto the UPS truck. In this paper, we consider these two sources of fixed costs. In particular, we assume that there is a review cost K_j incurred for each T_j and a setup cost k_j incurred for each Q_j . In addition to these fixed costs, there are linear holding and backorder costs. The objective is to find the (r, nQ, T) policy such that the total supply chain cost per period is minimized.

The single-stage (r, nQ, T) model was first studied by Hadley and Whitin (1963). They provided an approach to evaluate the total cost per period. Kiesmüller and Kok (2006) analyzed the waiting time in this model. Larson and Kiesmüller (2007) developed a closed-form cost expression when the demand process is compound generalized Erlang. Our model can be viewed as a generalization of this single-stage model. It also can be viewed as a generalization of the model studied by Maxwell and Muckstadt (1985) and Roundy (1985), who assumed that demand is deterministic with a stationary rate. They showed that the well-known power-of-two policies, under which reorder intervals at all stages are power-of-two multiples of a base time unit, are near optimal. Assuming deterministic demand, the problem can be reformulated as that of finding optimal batch sizes. These different approaches to modeling control variables lead to two different policies for the stochastic demand model in the literature, namely, the (r, nQ) policy and the (s, T) policy.

The (r, nQ) policy is a special case of the (r, nQ, T) policy, a case in which the reorder intervals are equal to one period for all stages.² The (r, nQ) policy has been studied extensively in the literature. Available results include policy evaluation, optimization, and approximations. See, for example, Axsäter and Rosling (1993), Chen (2000), Chen and Zheng (1994, 1998), De Bodt and Graves (1985), Shang (2008), and Shang and Song (2007). We refer the reader to Axsäter (2003) and Chen (1998a) for a complete review.

For the (s, T) policy, stage j reviews its echelon inventory order position at the beginning of each T period. If the inventory order position is less than the echelon base-stock level s , the stage orders up to s . In other words, this policy is a special case of the (r, nQ, T) policy in which the batch sizes are equal to one for all stages. The (s, T) policy was first discussed by Hadley and Whitin (1963). Since then, the policy has not attracted much attention from researchers. Recently, Rao (2003) studied a single-stage system, in which fixed costs are incurred for each inventory reorder. He showed that the total cost is jointly convex in the policy parameters, and he developed a worst-case bound on the optimal cost. For multistage systems, Cachon (1999)

studied the reorder-interval policy in a one-supplier, multi-retailer system. He showed that the supplier's demand variance will decline as the retailers' reorder interval becomes longer. Graves (1996) provided a new approach to evaluating the cost for distribution systems under the so-called virtual allocation rule. Van Houtum et al. (2007) studied a serial model and showed that the echelon (s, T) policies are optimal when the reorder intervals are fixed. They also provided an algorithm to obtain the optimal basestock levels. Chao and Zhou (2009) extended these results to the echelon (r, nQ, T) policies that we consider here. They provided an algorithm to obtain the optimal reorder points with fixed batch sizes and reorder intervals. However, it is not clear how to jointly optimize batch sizes and reorder intervals. To our knowledge, the only paper that attempts to find optimal reorder intervals is Feng and Rao (2007). They studied a two-stage system with echelon (s, T) policies and derived the average total cost function. They used the golden section search to obtain heuristic reorder intervals.

The present paper provides an approach for obtaining effective heuristic and optimal batch sizes and reorder intervals. For brevity, hereafter we call a set of feasible batch sizes and reorder intervals *a solution*. The heuristic solution is obtained by minimizing the lower and upper bounds of the total cost function. These total cost bounds, which are a sum of N separable functions of policy parameters, are obtained in two steps. First, we decompose the total cost into costs associated with each stage, which include a penalty cost for holding inadequate stock, referred to as *induced-penalty cost*. Second, we construct lower and upper bounds for the induced-penalty cost by regulating downstream policy parameters. More specifically, the penalty cost charged to stage j is smallest when all of its downstream stages set their reorder intervals and batch sizes equal to those used at stage j . Conversely, the penalty cost is highest when all of the downstream stages set their reorder intervals and batch sizes equal to one. By substituting these penalty cost bounds for the exact induced-penalty cost function, in effect we construct lower and upper bounds on the stage cost. We then minimize the sum of these stage cost bounds to obtain a heuristic solution.

Finding the optimal solution is more difficult. We propose a complete enumeration, which is facilitated with upper and lower bounds on the optimal solution. To obtain these solution bounds, we derive bounds on the cost of each echelon. The echelon lower-bound cost function and the minimum stage costs obtained from the stage lower-bound cost functions generate the solution bounds for each stage.

Our results enable us to observe the optimal solution and draw insights on how to manage the (r, nQ, T) system. In a numerical study, we find that a stage's optimal batch size and reorder interval tend to (1) increase in its setup cost and review cost, respectively; (2) decrease in its holding cost and the backorder cost, and (3) be insensitive to system lead times. One interesting finding is that the optimal batch sizes seem sensitive to K_j , but the optimal reorder intervals

may not be sensitive to k_j , especially when a change of k_j occurs at an upstream stage. This suggests that changing the batch sizes (due to a change of k_j) may not affect the optimal reorder intervals or, equivalently, the order and delivery schedules. Finally, we find that our heuristic is particularly effective when K_j/h_j and k_j/h_j decrease in j , where h_j is the echelon holding cost per unit per period for stage j . Under such circumstances, the heuristic batch sizes and reorder intervals tend to be equal between stages. This behavior is consistent with that of the optimal solution.

Our model assumes that K_j is incurred for each T_j and k_j for each Q_j . Such assumptions are appropriate for some companies, such as EMC², and are commonly seen in the literature. See, for example, Hadley and Whitin (1963) and Chen and Zheng (1998). In §5, we consider two alternative fixed cost assumptions. One is the assumption that the review cost K_j is incurred only when a stage places an order. For example, some companies require a physical inventory count after an order is placed. The other assumption is that k_j is incurred for each order (instead of for each batch). This fits a situation where, for example, a machine requires only one setup for making multiple batches of items. As we shall see, our analytical results can be applied to find the optimal solution under these assumptions.

The rest of the paper is organized as follows: Section 2 describes the model and presents a bottom-up recursion to evaluate the echelon (r, nQ, T) policy. This evaluation scheme sets the stage for the subsequent analysis. Section 3 constructs lower and upper bounds on the cost of each stage for a given set of batch sizes and reorder intervals. We provide a heuristic based on solving the sum of these cost-bound functions. Section 4 constructs bounds on the cost of each echelon, and presents an approach for obtaining the optimal solution. Section 5 discusses alternative fixed cost assumptions. Section 6 performs a numerical study to examine the optimal solution and test the heuristic. Section 7 concludes. All proofs are provided in the online appendix at <http://or.journal.informs.org>.

2. The Model and Preliminaries

We consider a single-item, periodic-review inventory system with N stages. Customer demand occurs at stage 1. Stage 1 obtains supplies from stage 2, stage 2 from stage 3, etc., and stage N is replenished by an outside source that has ample supply. Demands in different periods are independent, identically distributed, nonnegative, and integer valued. Let μ denote the mean one-period demand. Denote by $D[t, t + \tau)$ and $D[t, t + \tau]$ the total demand over periods $t, t + 1, \dots, t + \tau - 1$ and $t, t + 1, \dots, t + \tau$, respectively. Let $D[\tau)$ and $D[\tau]$ be the total demand in τ and $\tau + 1$ periods if the period index t is omitted. Unsatisfied demand is backlogged. Echelon holding cost h_j is incurred for each unit of on-hand inventories held in echelon j per period, and backorder cost b is incurred for each unit of backlogs occurring at stage 1 per period. Define $h_{[i, j]} = \sum_{k=i}^j h_k$. The transportation lead time L_j between stage $j + 1$ and stage

j is constant, and $L_j \in \mathbb{N}$ is the set of positive integers. Define $L_{[i, j]} = \sum_{k=i}^j L_k$.

Each stage implements an echelon (r, nQ, T) policy. The policy operates as follows: Stage j orders at the beginning of every T_j th period. If the echelon inventory order position (= inventory on order + inventory on hand + inventory in transit to and at its downstream stages—stage 1's backorders) is less than or equal to the echelon reorder point r_j , the stage orders an integer multiple of batch size Q_j to raise the echelon inventory order position back to the interval of $\{r_j + 1, r_j + 2, \dots, r_j + Q_j\}$. We refer to these T_j th periods as *order periods*, and T_j as the reorder interval. The batch sizes and reorder intervals satisfy integer-ratio relations, i.e., $Q_{j+1} = q_j Q_j$, $T_{j+1} = n_j T_j$, where $Q_j, T_j, q_j, n_j \in \mathbb{N}$, $j = 1, \dots, N - 1$. A fixed review cost K_j is incurred in each order period. Define $K_{[i, j]} = \sum_{v=i}^j K_v$. Also, a fixed setup cost k_j is incurred for ordering a batch. Let $k_{[i, j]} = \sum_{v=i}^j k_v$.

We assume that the system starts with a plausible inventory state in which stage j 's local on-hand inventory is an integer multiple of Q_{j-1} , $j = 2, \dots, N$ (Chen and Zheng 1994, Chao and Zhou 2009). In addition, all the replenishment activities in a period occur at the beginning of the period. At stage $j > 1$, they occur in the following sequence: (1) an order, if any, from stage $j - 1$ is received; (2) an order is placed with stage $j + 1$ if the period is stage j 's order period; (3) a shipment sent from stage $j + 1$ L_j periods earlier is received; and (4) a shipment is sent to stage $j - 1$. For stage 1, order placement occurs at the beginning of stage 1's order periods, while customer demand arrives during a period. Costs are evaluated at the end of a period. We assume that all shipments are synchronized. That is, a downstream stage, whenever possible, places an order when its upstream stage receives a shipment. (A synchronized shipping policy dominates a non-synchronized one; see Chao and Zhou 2009.) The objective is to minimize the average total system cost per period.

Define the following inventory variables:

$IOP_j(t)$ = echelon inventory order position after ordering and before demand at stage j at the beginning of an order period t ,

$IP_j(t)$ = echelon inventory in-transit position after ordering and before demand at stage j at the beginning of an order period t ,

$IL_j^-(t)$ = echelon inventory level at stage j at the beginning of a period t ,

$IL_j(t)$ = echelon inventory level at stage j at the end of a period t .

Here, IOP_j and IP_j are defined for each order period; IL_j^- and IL_j are defined for all periods. Under the (r, nQ, T) policy, $IOP_j(t)$ is uniformly distributed over $\{r_j + 1, r_j + 2, \dots, r_j + Q_j\}$ (Zipkin 1986, Chao and Zhou 2009), which may be viewed as the “target” echelon inventory quantity that stage j aims to achieve; $IP_j(t)$ may be viewed as the “physical” echelon inventory quantity for

stage j . The difference between $IOP_j(t)$ and $IP_j(t)$ is the number of outstanding orders for stage j (i.e., orders not yet filled by stage $j + 1$). The inventory cost is determined by the $IL_j(t)$ values.

We now discuss how to evaluate the average total cost per period under the (r, nQ, T) policy. This total cost includes two parts: the average inventory holding and backorder costs per period and the average review and setup costs per period. We first show how to evaluate the former. For brevity, hereafter we refer to “inventory holding and backorder costs” as “inventory-related costs.”

Consider the dynamics of the echelon inventory variables under the (r, nQ, T) policies. Suppose that stage N places an order at the beginning of an order period t . Define a *cycle* for stage j , $j = 1, \dots, N$, with respect to t as a time interval that includes periods $t + L_{[j, N]} + \tau$, $\tau = 0, \dots, T_N - 1$. As we shall see below, this order will directly or indirectly determine IL_j^- and IL_j within the stage j 's cycle. Since the system repeats itself when stage N places an order in every T_N period, it is a regenerative process with a cycle length of T_N periods. Thus, the long-run average inventory-related costs per period are equal to the expected inventory-related costs incurred in the cycle divided by T_N . Since these expected costs are determined by IL_j , we show below how to derive IL_j within the cycle.

We start from stage N . Suppose that stage N orders at the beginning of an order period t , and the echelon inventory order position after ordering is $IOP_N(t)$. Because stage N has ample supply, $IP_N(t) = IOP_N(t)$. This order will arrive at stage N at period $t + L_N$. Since there will be no other order periods until period $t + T_N$, $IP_N(t)$ will determine both $IL_N^-(t + L_N + \tau)$ and $IL_N(t + L_N + \tau)$ for $\tau = 0, \dots, T_N - 1$. That is,

$$IL_N^-(t + L_N + \tau) = IP_N(t) - D[t, t + L_N + \tau],$$

and

$$IL_N(t + L_N + \tau) = IP_N(t) - D[t, t + L_N + \tau].$$

Now consider stage $j = N - 1, N - 2, \dots, 1$ sequentially. Define $\lfloor a \rfloor$ as a roundoff operator, which returns the greatest integer less than or equal to a , a real number. Let $\mathbb{M}_x(y)$ be an operator that returns the remainder of y divided by x , $x \in \mathbb{N}$, and $y \in \{0, \mathbb{N}\}$. According to the synchronized replenishment rule, stage j will order in periods $t + L_{[j+1, N]} + \lfloor \tau/T_j \rfloor T_j$, for $\tau = 0, \dots, T_N - 1$. Since stage j may not have ample supply, IP_j is determined jointly by its echelon inventory order position IOP_j and stage $j + 1$'s net echelon inventory level IL_{j+1}^- . That is, for $\tau = 0, \dots, T_N - 1$,

$$\begin{aligned} IP_j \left(t + L_{[j+1, N]} + \left\lfloor \frac{\tau}{T_j} \right\rfloor T_j \right) \\ = O_j \left[IL_{j+1}^- \left(t + L_{[j+1, N]} + \left\lfloor \frac{\tau}{T_j} \right\rfloor T_j \right) \right], \end{aligned} \quad (1)$$

where

$$O_j[x] = \begin{cases} x & x \leq r_j, \\ x - mQ_j & \text{otherwise,} \end{cases} \quad (2)$$

and $m \in \{0, \mathbb{N}\}$ such that $r_j + 1 \leq x - mQ_j \leq r_j + Q_j$. Note that $x - mQ_j$ is IOP_j . Equation (1) thus means that if stage $j + 1$ has sufficient stock such that $IL_{j+1}^- > r_j$, IP_j is equal to IOP_j . Otherwise, stage $j + 1$ will ship as much as possible, in which case $IP_j = IL_{j+1}^-$.

The IP_j in the order periods will further determine IL_j^- and IL_j within periods $t + L_{[j, N]} + \tau$, $\tau = 0, \dots, T_N - 1$:

$$\begin{aligned} IL_j^-(t + L_{[j, N]} + \tau) \\ = IP_j \left(t + L_{[j+1, N]} + \left\lfloor \frac{\tau}{T_j} \right\rfloor T_j \right) - D \left[t + L_{[j+1, N]} + \left\lfloor \frac{\tau}{T_j} \right\rfloor T_j, \right. \\ \left. t + L_{[j, N]} + \left\lfloor \frac{\tau}{T_j} \right\rfloor T_j + \mathbb{M}_{T_j}(\tau) \right], \end{aligned} \quad (3)$$

$$\begin{aligned} IL_j(t + L_{[j, N]} + \tau) \\ = IP_j \left(t + L_{[j+1, N]} + \left\lfloor \frac{\tau}{T_j} \right\rfloor T_j \right) - D \left[t + L_{[j+1, N]} + \left\lfloor \frac{\tau}{T_j} \right\rfloor T_j, \right. \\ \left. t + L_{[j, N]} + \left\lfloor \frac{\tau}{T_j} \right\rfloor T_j + \mathbb{M}_{T_j}(\tau) \right]. \end{aligned} \quad (4)$$

We write $IL_j(\tau)$ to represent $IL_j(t + L_{[j, N]} + \tau)$ at steady state. The long-run average inventory-related costs per period are equal to

$$\begin{aligned} G(\mathbf{r}, \mathbf{Q}, \mathbf{T}) \\ = \frac{1}{T_N} \mathbb{E} \left[\sum_{\tau=0}^{T_N-1} \left(\sum_{j=1}^N h_j IL_j(\tau) + (b + h_{[1, N]}) [IL_1(\tau)]^- \right) \right], \end{aligned} \quad (5)$$

where $\mathbf{r} = (r_1, \dots, r_N)$, $\mathbf{Q} = (Q_1, \dots, Q_N)$, $\mathbf{T} = (T_1, \dots, T_N)$, and $[x]^- = \max\{0, -x\}$.

Below we provide a convenient recursion to evaluate $G(\mathbf{r}, \mathbf{Q}, \mathbf{T})$. Define *echelon* j as a subsystem that includes stage i , $i \leq j$. The idea behind this scheme is that, at each iteration, we evaluate the average inventory-related costs for echelon j , referred to as $G_j(y)$, provided that stage j 's echelon inventory order position $IOP_j(t)$ is equal to y and its downstream stage i ($< j$) follows an (r, nQ, T) policy with parameters (r_i, Q_i, T_i) .

PROPOSITION 1. *Define*

$$\begin{aligned} G_1(y) = \frac{1}{T_1} \left(\sum_{\tau=0}^{T_1-1} \mathbb{E}[h_1(y - D[L_1 + \tau])] \right. \\ \left. + (b + h_{[1, M]})(y - D[L_1 + \tau])^- \right). \end{aligned} \quad (6)$$

For $j = 2, \dots, N$, define recursively

$$\begin{aligned} G_j(y) = \frac{1}{T_j} \sum_{\tau=0}^{T_j-1} \mathbb{E} \left[h_j(y - D[L_j + \tau]) \right. \\ \left. + G_{j-1} \left(O_{j-1} \left[y - D \left[L_j + \left\lfloor \frac{\tau}{T_{j-1}} \right\rfloor T_{j-1} \right] \right) \right) \right]. \end{aligned} \quad (7)$$

Then, $G(\mathbf{r}, \mathbf{Q}, \mathbf{T}) = \widehat{G}_N(r_N)$, where $\widehat{G}_N(r_N) = (1/Q_N) \cdot \sum_{x=1}^{Q_N} G_N(r_N + x)$.

We next determine the average fixed costs per period. For stage j , the review cost K_j is incurred for each T_j . So the average cost per period is K_j/T_j . The setup cost k_j is incurred for each batch Q_j . That is, the setup cost of no order, ordering one batch, ordering two batches, ... are 0, k_j , $2k_j$, ... In the long run, all demand will be fulfilled, so the average setup cost per period is $(k_j\mu)/Q_j$. With these fixed cost terms, the total cost per period is

$$C(\mathbf{r}, \mathbf{Q}, \mathbf{T}) = \sum_{j=1}^N \left(\frac{K_j}{T_j} + \frac{k_j\mu}{Q_j} \right) + \widehat{G}_N(r_N). \quad (8)$$

We now turn to optimization. For fixed \mathbf{Q} and \mathbf{T} , Chao and Zhou (2009) provided a recursion to find the optimal reorder points. Their algorithm sequentially minimizes the inventory-related costs per period for echelon j , $j = 1, 2, 3, \dots$, assuming that stage i , $i < j$ implements the optimal reorder point. The minimizer is the optimal reorder point for stage j . Specifically, define $\mathbf{T}_j = (T_1, T_2, \dots, T_j)$ and $\mathbf{Q}_j = (Q_1, Q_2, \dots, Q_j)$ (so $\mathbf{T} = \mathbf{T}_N$ and $\mathbf{Q} = \mathbf{Q}_N$). Starting from stage 1, let

$$\widehat{G}_1(r_1, \mathbf{Q}_1, \mathbf{T}_1) = \frac{1}{Q_1} \sum_{x=1}^{Q_1} G_1(r_1 + x),$$

which is convex in r_1 . Let

$$r_1(\mathbf{Q}_1, \mathbf{T}_1) = \arg \min_{r_1} \widehat{G}_1(r_1, \mathbf{Q}_1, \mathbf{T}_1).$$

For $j = 2, \dots, N$, suppose that $r_{j-1}(\mathbf{Q}_{j-1}, \mathbf{T}_{j-1})$ is known. Substitute $r_{j-1}(\mathbf{Q}_{j-1}, \mathbf{T}_{j-1})$ for r_{j-1} in the O_{j-1} function defined in (2), and denote the new function as

$$O_{j-1}^*[x] = \begin{cases} x & x \leq r_{j-1}(\mathbf{Q}_{j-1}, \mathbf{T}_{j-1}), \\ x - mQ_{j-1} & \text{otherwise,} \end{cases} \quad (9)$$

where $m \in \{0, \mathbb{N}\}$ such that $r_{j-1}(\mathbf{Q}_{j-1}, \mathbf{T}_{j-1}) + 1 \leq x - mQ_{j-1} \leq r_{j-1}(\mathbf{Q}_{j-1}, \mathbf{T}_{j-1}) + Q_{j-1}$. This O_{j-1}^* function has the same meaning as the O_{j-1} function defined in (2), except that the optimal reorder point is in place.

Replace the O_{j-1} function in (7) with O_{j-1}^* , and let the resulting function be

$$\begin{aligned} & G_j(y, \mathbf{Q}_{j-1}, \mathbf{T}_j) \\ &= \frac{1}{T_j} \sum_{\tau=0}^{T_j-1} \mathbb{E} \left[h_j(y - D[L_j + \tau]) \right. \\ & \quad \left. + G_{j-1} \left(O_{j-1}^* \left[y - D \left[L_j + \left\lfloor \frac{\tau}{T_{j-1}} \right\rfloor T_{j-1} \right] \right), \mathbf{Q}_{j-2}, \mathbf{T}_{j-1} \right) \right], \end{aligned}$$

where $G_1(y, \mathbf{Q}_0, \mathbf{T}_1) = G_1(y)$. Define the average inventory-related cost per period for echelon j as

$$\widehat{G}_j(r_j, \mathbf{Q}_j, \mathbf{T}_j) = \frac{1}{Q_j} \sum_{x=1}^{Q_j} G_j(r_j + x, \mathbf{Q}_{j-1}, \mathbf{T}_j). \quad (10)$$

It can be shown that $\widehat{G}_j(r_j, \mathbf{Q}_j, \mathbf{T}_j)$ is convex in r_j . Let $r_j(\mathbf{Q}_j, \mathbf{T}_j) = \arg \min_{r_j} \widehat{G}_j(r_j, \mathbf{Q}_j, \mathbf{T}_j)$. Continue this procedure until stage N . Then $(r_1(\mathbf{Q}_1, \mathbf{T}_1), \dots, r_N(\mathbf{Q}_N, \mathbf{T}_N))$ are optimal reorder points for fixed \mathbf{Q} and \mathbf{T} .

With this result, the problem of minimizing $C(\mathbf{r}, \mathbf{Q}, \mathbf{T})$ can be reduced to finding the optimal batch sizes and reorder intervals. That is, let

$$\widehat{G}_j(\mathbf{Q}_j, \mathbf{T}_j) \stackrel{\text{def}}{=} \widehat{G}_j(r_j(\mathbf{Q}_j, \mathbf{T}_j), \mathbf{Q}_j, \mathbf{T}_j), \quad j = 1, \dots, N.$$

We aim to solve

$$\begin{aligned} (P) \quad & \min_{\mathbf{Q}, \mathbf{T}} C(\mathbf{Q}, \mathbf{T}) = \sum_{j=1}^N \left(\frac{K_j}{T_j} + \frac{k_j\mu}{Q_j} \right) + \widehat{G}_N(\mathbf{Q}_N, \mathbf{T}_N) \\ & \text{s.t. } Q_{j+1} = q_j Q_j, \\ & T_{j+1} = n_j T_j, \\ & Q_j, T_j, q_j, n_j \in \mathbb{N}, \quad j = 1, \dots, N-1. \end{aligned}$$

In §3, we develop bounds for $C(\mathbf{Q}, \mathbf{T})$. These bounds will lead to an effective heuristic solution for (P). In §4, we provide an approach to solve (P). Throughout the paper, the convexity results are based on continuous approximation on demand and control variables.

3. Cost Bounds and Heuristic

This section develops lower and upper bounds on $C(\mathbf{Q}, \mathbf{T})$. In §3.1, we first decompose $C(\mathbf{Q}, \mathbf{T})$ into costs associated with each stage by using an induced-penalty cost function. In §3.2, we construct bounds for the induced-penalty function. By substituting these bounds for the exact induced-penalty function, we effectively establish bounds for $C(\mathbf{Q}, \mathbf{T})$. These total cost bounds will be used to derive a heuristic presented in §3.3.

3.1. Decomposition of the Total Cost Function

The decomposition of $C(\mathbf{Q}, \mathbf{T})$ is based on the construction of an induced-penalty cost function, which is the penalty cost charged to an upstream stage if the upstream stage cannot fulfill an order from its downstream stage. Thus, the induced-penalty cost is incurred in each downstream stage's order period.

Let us start by computing the induced-penalty cost charged to stage 2. Consider an order period t for stage 1. Conditioning on $IP_1(t)$, stage 1's inventory holding and backorder cost per period is

$$\begin{aligned} g_1(IP_1(t), T_1) &= \frac{1}{T_1} \left(\sum_{\tau=0}^{T_1-1} \mathbb{E} [h_1(IP_1(t) - D[L_1 + \tau]) \right. \\ & \quad \left. + (b + h_{[1, N]})(IP_1(t) - D[L_1 + \tau])^- \right]. \end{aligned}$$

Since stage 2 will be charged for unfilled orders placed by stage 1, the optimal reorder point for stage 1 is a solution that

minimizes stage 1's cost, assuming that it has ample supply from stage 2. In such case, $IP_1(t)$ is equal to $IOP_1(t)$, which is uniformly distributed between $\{r_1 + 1, r_1 + 2, \dots, r_1 + Q_1\}$. The average inventory holding and backorder cost per period for stage 1 is

$$E[g_1(IOP_1(t), T_1)] = \frac{1}{Q_1} \sum_{x=1}^{Q_1} g_1(r_1 + x, T_1) \stackrel{\text{def}}{=} \hat{g}_1(r_1, \mathbf{Q}_1, \mathbf{T}_1),$$

which is equal to $\hat{G}_1(r_1, \mathbf{Q}_1, \mathbf{T}_1)$. Thus, stage 1 will choose the optimal reorder point $r_1(\mathbf{Q}_1, \mathbf{T}_1)$ as its reorder point.

However, stage 1 may not be able to obtain ample supply from stage 2, namely, $IP_1(t)$ is constrained by $IL_2^-(t)$. Specifically, $IP_1(t) = O_1^*[IL_2^-(t)]$, where $O_1^*[\cdot]$ is defined in (9). The induced-penalty cost charged to stage 2 is

$$g_{1,2}(IL_2^-(t), \mathbf{Q}_1, \mathbf{T}_1) = g_1(O_1^*[IL_2^-(t)], \mathbf{T}_1) - g_1(IOP_1(t), \mathbf{T}_1).$$

To see why this is the penalty cost charged to stage 2 for holding inadequate stock, let us recall the definition of the $O_1^*[\cdot]$ function. If stage 2 has sufficient stock such that $IL_2^-(t) > r_1(\mathbf{Q}_1, \mathbf{T}_1)$, then $O_1^*[IL_2^-(t)] = IL_2^-(t) - mQ_1$, which is equal to $IOP_1(t)$. In this case, there is no induced-penalty cost charged to stage 2. On the other hand, if $IL_2^-(t) \leq r_1(\mathbf{Q}_1, \mathbf{T}_1)$, then $O_1^*[IL_2^-(t)] = IL_2^-(t)$, and $g_1(IL_2^-(t), \mathbf{T}_1) > g_1(IOP_1(t), \mathbf{T}_1)$. The difference will be the induced-penalty cost charged to stage 2.

Now we can compute stage 2's average inventory and penalty cost per period. Stage 2 orders every T_2 periods. Consider an order period t for stage 2. Conditioning on $IP_2(t)$, the inventory and penalty cost per period for stage 2 is

$$g_2(IP_2(t), \mathbf{Q}_1, \mathbf{T}_2) = \frac{1}{T_2} \left(\sum_{\tau=0}^{T_2-1} E \left[h_2(IP_2(t) - D[L_2 + \tau]) + g_{1,2} \left(IP_2(t) - D \left[L_2 + \left\lfloor \frac{\tau}{T_1} \right\rfloor T_1 \right), \mathbf{Q}_1, \mathbf{T}_1 \right) \right] \right).$$

Here, $IP_2(t) - D[L_2 + \lfloor \tau/T_1 \rfloor T_1]$ is IL_2^- at the beginning of stage 1's order period $t + L_2 + \lfloor \tau/T_1 \rfloor$, $\tau = 0, \dots, T_2 - 1$. If stage 2 has ample supply from stage 3, $IP_2(t)$ will be uniformly distributed between $\{r_2 + 1, r_2 + 2, \dots, r_2 + Q_2\}$. Stage 2's average inventory and penalty cost per period is

$$E[g_2(IP_2(t), \mathbf{Q}_1, \mathbf{T}_2)] = \frac{1}{Q_2} \sum_{x=1}^{Q_2} g_2(r_2 + x, \mathbf{Q}_1, \mathbf{T}_2) \stackrel{\text{def}}{=} \hat{g}_2(r_2, \mathbf{Q}_2, \mathbf{T}_2).$$

Following the same logic, since $IP_2(t)$ is in fact constrained by $IL_3^-(t)$, we can derive the induced-penalty cost charged to stage 3 in each stage 2's order period t , and obtain the average inventory holding and penalty cost per

period for stage 3. The same procedure can be carried out for the rest of the chain.

We generalize the above procedure below. For stage j , $j = 2, 3, \dots, N$, and let $IL_j^-(t) = y$, where t is stage $j-1$'s order period. Suppose that $r_{j-1}(\mathbf{Q}_{j-1}, \mathbf{T}_{j-1})$ is known. The induced-penalty cost charged to stage j is

$$g_{j-1,j}(y, \mathbf{Q}_{j-1}, \mathbf{T}_{j-1}) = g_{j-1}(O_{j-1}^*[y], \mathbf{Q}_{j-2}, \mathbf{T}_{j-1}) - g_{j-1}(y - zQ_{j-1}, \mathbf{Q}_{j-2}, \mathbf{T}_{j-1}), \quad (11)$$

where $z \in \mathbb{Z}$, the set of integers, such that $r_{j-1}(\mathbf{Q}_{j-1}, \mathbf{T}_{j-1}) + 1 \leq y - zQ_{j-1} \leq r_{j-1}(\mathbf{Q}_{j-1}, \mathbf{T}_{j-1}) + Q_{j-1}$. Note that z is an integer to bring $y - zQ_{j-1}$ back to the interval of $\{r_{j-1}(\mathbf{Q}_{j-1}, \mathbf{T}_{j-1}) + 1, \dots, r_{j-1}(\mathbf{Q}_{j-1}, \mathbf{T}_{j-1}) + Q_{j-1}\}$, so $y - zQ_{j-1}$ is $IOP_{j-1}(t)$.

Conditioning on $IP_j(t) = y$, the inventory and penalty cost per period for stage j is

$$g_j(y, \mathbf{Q}_{j-1}, \mathbf{T}_j) = \frac{1}{T_j} \left(\sum_{\tau=0}^{T_j-1} E \left[h_j(y - D[L_j + \tau]) + g_{j-1,j} \left(y - D \left[L_j + \left\lfloor \frac{\tau}{T_{j-1}} \right\rfloor T_{j-1} \right), \mathbf{Q}_{j-1}, \mathbf{T}_{j-1} \right) \right] \right). \quad (12)$$

If stage j has ample supply from stage $j+1$, its average inventory and penalty cost per period is

$$\begin{aligned} \hat{g}_j(r_j, \mathbf{Q}_j, \mathbf{T}_j) &= \frac{1}{Q_j} \sum_{x=1}^{Q_j} g_j(r_j + x, \mathbf{Q}_{j-1}, \mathbf{T}_j), \\ &= \frac{1}{Q_j T_j} \sum_{x=1}^{Q_j} \sum_{\tau=0}^{T_j-1} E[h_j(r_j + x - D[L_j + \tau])] \\ &\quad + P_j(r_j, \mathbf{Q}_j, \mathbf{T}_j), \end{aligned} \quad (13)$$

where the average penalty cost per period is

$$P_j(y, \mathbf{Q}_j, \mathbf{T}_j) = \frac{1}{Q_j T_j} \sum_{x=1}^{Q_j} \sum_{\tau=0}^{T_j-1} E \left[g_{j-1,j} \left(y + x - D \left[L_j + \left\lfloor \frac{\tau}{T_{j-1}} \right\rfloor T_{j-1} \right), \mathbf{Q}_{j-1}, \mathbf{T}_{j-1} \right) \right]. \quad (14)$$

As we shall show in Proposition 2 below, the optimal reorder point $r_j(\mathbf{Q}_j, \mathbf{T}_j)$ discussed in §2 will minimize $\hat{g}_j(r_j, \mathbf{Q}_j, \mathbf{T}_j)$, i.e.,

$$r_j(\mathbf{Q}_j, \mathbf{T}_j) = \arg \min_{r_j} \hat{g}_j(r_j, \mathbf{Q}_j, \mathbf{T}_j). \quad (15)$$

PROPOSITION 2. For fixed \mathbf{Q} and \mathbf{T} , the optimal reorder point $r_j(\mathbf{Q}_j, \mathbf{T}_j)$ minimizes $\hat{g}_j(r_j, \mathbf{Q}_j, \mathbf{T}_j)$, $j = 1, \dots, N$.

In other words, the optimal reorder point may be interpreted as an echelon inventory level that a stage tries to maintain in order to minimize its stage cost, while all of its downstream stages use the best reorder points.

Define the cost for stage j in which the optimal reorder point $r_j(\mathbf{Q}_j, \mathbf{T}_j)$ is implemented as

$$\hat{g}_j(\mathbf{Q}_j, \mathbf{T}_j) = \hat{g}_j(r_j(\mathbf{Q}_j, \mathbf{T}_j), \mathbf{Q}_j, \mathbf{T}_j).$$

Proposition 3 states that the inventory-related costs for echelon j can be decomposed into costs associated with each stage within the echelon.

PROPOSITION 3. $\hat{G}_j(\mathbf{Q}_j, \mathbf{T}_j) = \sum_{i=1}^j \hat{g}_i(\mathbf{Q}_i, \mathbf{T}_i)$, for $j = 1, \dots, N$,

Let $c_j(\mathbf{Q}_j, \mathbf{T}_j) \stackrel{\text{def}}{=} K_j/T_j + (k_j\mu)/Q_j + \hat{g}_j(\mathbf{Q}_j, \mathbf{T}_j)$, the average total cost per period for stage j , $j = 1, \dots, N$. We have

$$\begin{aligned} C(\mathbf{Q}, \mathbf{T}) &= \sum_{j=1}^N \left(\frac{K_j}{T_j} + \frac{k_j\mu}{Q_j} \right) + \hat{G}_N(\mathbf{Q}_N, \mathbf{T}_N) \\ &= \sum_{j=1}^N \left(\frac{K_j}{T_j} + \frac{k_j\mu}{Q_j} \right) + \sum_{j=1}^N \hat{g}_j(\mathbf{Q}_j, \mathbf{T}_j) \\ &= \sum_{j=1}^N c_j(\mathbf{Q}_j, \mathbf{T}_j). \end{aligned}$$

This completes the decomposition of $C(\mathbf{Q}, \mathbf{T})$.

3.2. Bounds for the Stage Cost Functions

This section derives lower and upper bounds for the stage cost $c_j(\mathbf{Q}_j, \mathbf{T}_j)$. At the end, we shall see that these cost bounds are a function of stage j 's control variables and are independent of stage i 's, $i \neq j$.

Consider any given feasible solution (\mathbf{Q}, \mathbf{T}) with \mathbf{Q}_j and \mathbf{T}_j defined as in §2. For $j = 1, \dots, N$, define

- \mathbf{T}_j^j = a vector with j components whose values are T_j ,
- \mathbf{Q}_j^j = a vector with j components whose values are Q_j ,
- \mathbf{T}_j^1 = a vector with j components whose j th component is T_j and the other components are 1,
- \mathbf{Q}_j^1 = a vector with j components whose j th component is Q_j and the other components are 1.

For example, $\mathbf{T}_3^3 = (T_3, T_3, T_3)$ and $\mathbf{T}_4^1 = (1, 1, 1, T_4)$. Recall the average induced-penalty cost function $P_j(y, \mathbf{Q}_j, \mathbf{T}_j)$ in (14). We first state the main result of this section.

PROPOSITION 4.

- (1) With fixed batch sizes \mathbf{Q} , $P_j(y, \mathbf{Q}_j, \mathbf{T}_j^j) \leq P_j(y, \mathbf{Q}_j, \mathbf{T}_j) \leq P_j(y, \mathbf{Q}_j, \mathbf{T}_j^1)$, for all y and $j = 2, \dots, N$.
- (2) With fixed reorder intervals \mathbf{T} , $P_j(y, \mathbf{Q}_j^j, \mathbf{T}_j) \leq P_j(y, \mathbf{Q}_j, \mathbf{T}_j) \leq P_j(y, \mathbf{Q}_j^1, \mathbf{T}_j)$, for all y and $j = 2, \dots, N$.

Proposition 4(1) states that for stage j , when its downstream stage i uses the same reorder interval length as stage j (i.e., $T_i = T_j$, $i < j$), the resulting average induced-penalty cost function $P_j(y, \mathbf{Q}_j, \mathbf{T}_j^j)$ is a lower bound to $P_j(y, \mathbf{Q}_j, \mathbf{T}_j)$ for all y . On the other hand, when a downstream stage i uses the smallest reorder interval length (i.e., $T_i = 1$, $i < j$), the resulting induced-penalty cost function $P_j(y, \mathbf{Q}_j, \mathbf{T}_j^1)$ is an upper bound. Proposition 4(2) states the same result for regulating downstream batch sizes.

With Proposition 4, we can construct cost bounds to $\hat{g}_j(y, \mathbf{Q}_j, \mathbf{T}_j)$ for any y by first regulating downstream reorder intervals and then regulating downstream batch sizes. More specifically, since

$$\begin{aligned} \hat{g}_j(y, \mathbf{Q}_j, \mathbf{T}_j) &\geq \hat{g}_j(y, \mathbf{Q}_j, \mathbf{T}_j^j) \geq \hat{g}_j(y, \mathbf{Q}_j^j, \mathbf{T}_j^j) \quad \text{and} \\ \hat{g}_j(y, \mathbf{Q}_j, \mathbf{T}_j) &\leq \hat{g}_j(y, \mathbf{Q}_j, \mathbf{T}_j^1) \leq \hat{g}_j(y, \mathbf{Q}_j^1, \mathbf{T}_j^1), \end{aligned}$$

we have $\hat{g}_j(y, \mathbf{Q}_j^j, \mathbf{T}_j^j) \leq \hat{g}_j(y, \mathbf{Q}_j, \mathbf{T}_j) \leq \hat{g}_j(y, \mathbf{Q}_j^1, \mathbf{T}_j^1)$ for all y . Consequently,

$$\begin{aligned} \hat{g}_j(r_j(\mathbf{Q}_j^1, \mathbf{T}_j^1), \mathbf{Q}_j^1, \mathbf{T}_j^1) &\geq \hat{g}_j(r_j(\mathbf{Q}_j^1, \mathbf{T}_j^1), \mathbf{Q}_j, \mathbf{T}_j) \\ &\geq \hat{g}_j(r_j(\mathbf{Q}_j, \mathbf{T}_j), \mathbf{Q}_j, \mathbf{T}_j). \end{aligned}$$

Similarly, we can show that $\hat{g}_j(r_j(\mathbf{Q}_j, \mathbf{T}_j), \mathbf{Q}_j, \mathbf{T}_j) \geq \hat{g}_j(r_j(\mathbf{Q}_j^j, \mathbf{T}_j^j), \mathbf{Q}_j^j, \mathbf{T}_j^j)$. Together, we have

$$\begin{aligned} \hat{g}_j(r_j(\mathbf{Q}_j^j, \mathbf{T}_j^j), \mathbf{Q}_j^j, \mathbf{T}_j^j) &\leq \hat{g}_j(r_j(\mathbf{Q}_j, \mathbf{T}_j), \mathbf{Q}_j, \mathbf{T}_j) \\ &\leq \hat{g}_j(r_j(\mathbf{Q}_j^1, \mathbf{T}_j^1), \mathbf{Q}_j^1, \mathbf{T}_j^1) \end{aligned}$$

or, equivalently, $\hat{g}_j(\mathbf{Q}_j^j, \mathbf{T}_j^j) \leq \hat{g}_j(\mathbf{Q}_j, \mathbf{T}_j) \leq \hat{g}_j(\mathbf{Q}_j^1, \mathbf{T}_j^1)$, which implies

$$c_j(\mathbf{Q}_j^j, \mathbf{T}_j^j) \leq c_j(\mathbf{Q}_j, \mathbf{T}_j) \leq c_j(\mathbf{Q}_j^1, \mathbf{T}_j^1). \quad (16)$$

This completes the construction of the lower- and upper-bound functions. Note that $c_j(\mathbf{Q}_j^j, \mathbf{T}_j^j)$ and $c_j(\mathbf{Q}_j^1, \mathbf{T}_j^1)$ are a function of stage j 's control variables (Q_j, T_j). They are independent of downstream stage i 's decision variables, $i < j$.

To explain why regulating downstream policy parameters leads to cost bounds for the exact induced-penalty function, we take regulating reorder intervals as an example. Let us consider a simple two-stage example by fixing $Q_2 = Q_1 = 1$ and letting $T_2 = 2$. Thus, \mathbf{T}_2 can be either $(1, 2)$ or $(2, 2)$. Since \mathbf{Q} is fixed, we omit \mathbf{Q}_j in all cost functions in (11)–(14). Note that when $Q_j = 1$, the (r, nQ, T) policy reduces to the (s, T) policy, where $s_j = r_j + 1$. Also, the O_j^* function defined in (9) becomes

$$O_j^*[x] = \min\{r_j(\mathbf{T}_j) + 1, x\}.$$

For convenience, set $\mathbf{T}_2 = (1, 2)$ and $\mathbf{T}_2^2 = (2, 2)$. For $\mathbf{T}_2 = (1, 2)$, stage 1 orders twice in $T_2 = 2$ periods, each order

resulting in an induced-penalty cost. Thus, for any $IP_2 = y$, the average induced-penalty cost per period is

$$\begin{aligned} P_2(y, \mathbf{T}_2) &= \frac{1}{2} \mathbb{E}[g_{1,2}(y - D[L_2], 1) + g_{1,2}(y - D[L_2 + 1], 1)] \\ &= \frac{1}{2} (\mathbb{E}[g_1(\min\{r_1(1) + 1, y - D[L_2]\}, 1) - g_1(r_1(1) + 1, 1)] \\ &\quad + \mathbb{E}[g_1(\min\{r_1(1) + 1, y - D[L_2 + 1]\}, 1) \\ &\quad - g_1(r_1(1) + 1, 1)]). \end{aligned}$$

On the other hand, for $\mathbf{T}_2^2 = (2, 2)$, stage 1 orders once in $T_2 = 2$ periods, and the average penalty cost per period is

$$\begin{aligned} P_2(y, \mathbf{T}_2^2) &= \frac{1}{2} \mathbb{E}[g_{1,2}(y - D[L_2], 2) + g_{1,2}(y - D[L_2], 2)] \\ &= \mathbb{E}[g_{1,2}(y - D[L_2], 2)] \\ &= \mathbb{E}[g_1(\min\{r_1(2) + 1, y - D[L_2]\}, 2) - g_1(r_1(2) + 1, 2)]. \end{aligned}$$

Figure 1(a) shows these $\mathbb{E}[g_{1,2}(\cdot, \cdot)]$ functions. It is interesting to observe that

$$\begin{aligned} \mathbb{E}[g_{1,2}(y - D[L_2], 1)] &\leq \mathbb{E}[g_{1,2}(y - D[L_2], 2)] \\ &\leq \mathbb{E}[g_{1,2}(y - D[L_2 + 1], 1)], \end{aligned}$$

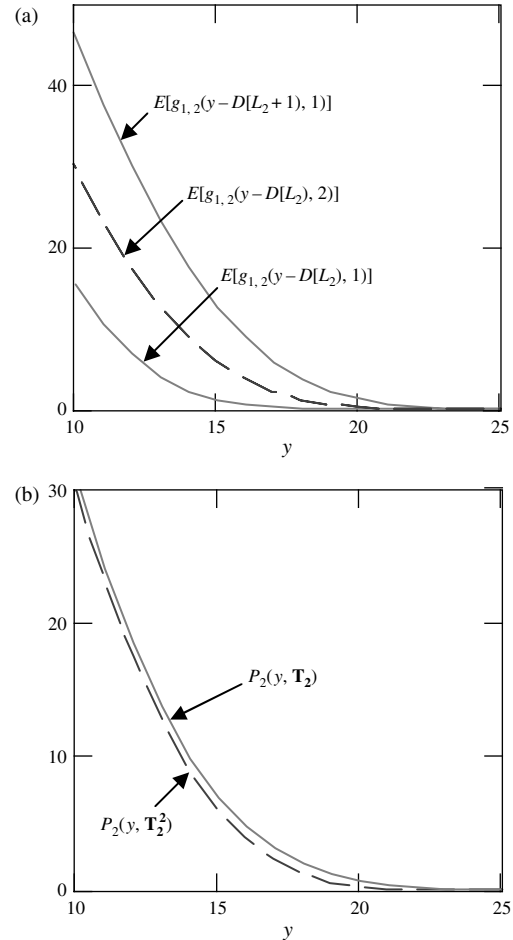
but $P_2(y, \mathbf{T}_2^2) \leq P_2(y, \mathbf{T}_2)$, as shown in Figure 1(b). This observation suggests that given the same $IP_2 = y$, stage 2 is more likely to fulfill a stage 1 order generated by two periods of demand than to fulfill two stage 1 orders, each generated by one period of demand. We refer to this fact as *demand aggregation effect*. Thus, the resulting penalty cost charged to stage 2 is lower in the former case.

The cost bound constructed by regulating downstream batch sizes is easier to explain. We use the lower bound as an example. Consider a two-stage system with batch sizes Q_1 and Q_2 . If stage 1 uses Q_2 instead of Q_1 as its batch size, the resulting best reorder point will be smaller (see Lemma 7(2) in the electronic companion). Note that the penalty cost is incurred when $IL_2^- (= IP_2 - D[L_2])$ is less than or equal to the reorder point. Thus, given the same IP_2 , a smaller reorder point will generate a smaller induced-penalty cost.

3.3. Heuristic

We now propose a heuristic for (P). We first introduce two simplified versions of (P) by assuming that either \mathbf{Q} or \mathbf{T} is fixed. We call these problems the *T*-problem and the *Q*-problem, respectively.

Figure 1. A two-stage example with $h_1 = h_2 = 0.5$, $L_1 = L_2 = 1$, $b = 9$.



Notes. Demand is Poisson with mean $\mu = 4$ units/period. (a) The expected induced-penalty cost functions incurred at each order period for $\mathbf{T}_2 = (1, 2)$ (solid lines) and for $\mathbf{T}_2^2 = (2, 2)$ (dashed line). (b) The average induced-penalty cost per period for $\mathbf{T}_2 = (1, 2)$ (solid line) and for $\mathbf{T}_2^2 = (2, 2)$ (dashed line).

3.3.1. The *T*-problem. This subproblem assumes that Q_j are fixed in (P) and the decision variables are T_j . For notational simplicity, we omit \mathbf{Q}_j in the related cost functions, such as $c_j(\mathbf{Q}_j, \mathbf{T}_j)$ and $\hat{g}_j(\mathbf{Q}_j, \mathbf{T}_j)$, and omit the fixed cost term $\sum_{i=1}^N (k_i \mu) / Q_i$ because it is a constant. Thus, the problem (P) becomes

$$\begin{aligned} (TP) \quad & \min_{\mathbf{T}} \sum_{j=1}^N c_j(\mathbf{T}_j), \\ & \text{s.t. } T_{j+1} = n_j T_j, \quad n_j, T_j \in \mathbb{N}, \quad j = 1, \dots, N-1, \end{aligned}$$

where $c_j(\mathbf{T}_j) = (K_j / T_j) + \hat{g}_j(\mathbf{T}_j)$.

Now consider the lower- and upper-bound functions $\hat{g}_j(\mathbf{T}_j^l)$ and $\hat{g}_j(\mathbf{T}_j^u)$. Define $\underline{g}_j(T_j) = \hat{g}_j(\mathbf{T}_j^l)$, $\bar{g}_j(T_j) = \hat{g}_j(\mathbf{T}_j^u)$, and $\underline{c}_j(T_j) = (K_j / T_j) + \underline{g}_j(T_j)$, $\bar{c}_j(T_j) = (K_j / T_j) + \bar{g}_j(T_j)$.

PROPOSITION 5. $\bar{c}_j(T_j)$ is convex in T_j .

We propose a simple heuristic for (TP) by solving the sum of the stage cost bound functions, subject to relaxed

constraints. We use the upper-bound cost functions to illustrate the idea. We aim to solve the following problem:

$$\begin{aligned} \min_T \quad & \sum_{j=1}^N \bar{c}_j(T_j) \\ \text{s.t.} \quad & T_{j+1} \geq T_j, \quad j = 1, \dots, N-1. \end{aligned}$$

Since the objective function is the sum of N separable, convex functions, a clustering algorithm (e.g., Maxwell and Muckstadt 1985, pp. 1325–1334) can solve the relaxed problem efficiently. The output of the algorithm is an optimal partition that includes disjoint clusters, such as $\{c(1), c(2), \dots, c(M)\}$, where M is the number of clusters in the optimal partition. (A cluster is a set of consecutive stages that use the same reorder interval.)

After the optimal partition is identified, we find a solution that satisfies the integer-ratio constraints for each cluster. More specifically, let $T_{c(1)} = \arg \min_T \sum_{i \in c(1)} \bar{c}_i(T)$. For $m = 2, \dots, M$, we solve the following problem sequentially:

$$T_{c(m)} = \arg \min_T \sum_{i \in c(m)} \bar{c}_i(T), \quad \text{s.t. } T = nT_{c(m-1)}, \quad n \in \mathbb{N}. \quad (17)$$

In other words, we restrict $T_{c(m)}$ to be an integer multiple of $T_{c(m-1)}$. Let $T'_j = T_{c(m)}$ for $j \in c(m)$, $m = 1, \dots, M$. We then obtain one feasible reorder interval solution (T'_1, \dots, T'_N) for (TP) . With these reorder intervals, we can find the best reorder points through the procedure in §2. Let the resulting total cost be C' .

Similarly, we can apply the same procedure to minimize $\sum_{j=1}^N \underline{c}_j(T_j)$. However, since $\underline{c}_j(T_j)$ may not be a convex function, we cannot apply the clustering algorithm directly. Thus, we use the same partition found in the upper-bound problem and then find the reorder intervals in the same fashion as in (17) except replacing $\bar{c}_j(T_j)$ with $\underline{c}_j(T_j)$. In this case, we use the first local minimizer as the reorder interval solution for a cluster. Define the resulting feasible reorder intervals as (T''_1, \dots, T''_N) and the resulting optimal cost as C'' .

The heuristic solution for (TP) is either T'_j or T''_j , $j = 1, \dots, N$, whichever yields a smaller total cost.

3.3.2. The Q -Problem. The Q -problem assumes that T_j are fixed and the decision variables are Q_j . Again, we omit \mathbf{T}_j in the related cost functions and omit the fixed cost term K_j/T_j . The problem (P) becomes

$$\begin{aligned} (QP) \quad \min_Q \quad & \sum_{j=1}^N c_j(\mathbf{Q}_j), \\ \text{s.t.} \quad & Q_{j+1} = q_j Q_j, \quad q_j, Q_j \in \mathbb{N}, \quad j = 1, \dots, N-1, \end{aligned}$$

where $c_j(\mathbf{Q}_j) = (k_j \mu / Q_j) + \hat{g}_j(\mathbf{Q}_j)$. Also, define $\underline{g}_j(Q_j) = \hat{g}_j(\mathbf{Q}_j^1)$, $\bar{g}_j(Q_j) = \hat{g}_j(\mathbf{Q}_j^M)$, $\underline{c}_j(Q_j) = (k_j \mu / Q_j) + \underline{g}_j(Q_j)$, and $\bar{c}_j(Q_j) = (k_j \mu / Q_j) + \bar{g}_j(Q_j)$. It can be shown that $\bar{c}_j(Q_j)$ is convex in Q_j . (The proof is available from the authors upon request.) Thus, the same procedure for generating candidate

solutions T'_j and T''_j can be applied to (QP) . Define the resulting candidate solutions as Q'_j and Q''_j . The heuristic solution is either Q'_j or Q''_j , whichever generates a smaller total cost.

The heuristic for (P) utilizes the proposed heuristics for (TP) and (QP) . Specifically, it includes three steps. First, we generate initial reorder intervals from the corresponding *deterministic* model, i.e., by assuming deterministic demand with a rate equal to μ and setting K_j as the fixed costs (see, e.g., Chen 1998b for the cost formulation). An optimal partition for the relaxed problem can be determined by the cost parameters. Then, the initial reorder intervals can be determined by applying the same procedure as in (17). Second, we use these reorder intervals as a “seed” and apply the heuristic for (QP) to generate candidate batch sizes Q'_j and Q''_j , $j = 1, \dots, N$. Third, we fix batch sizes to either Q'_j or Q''_j and apply the heuristic for (TP) to generate the candidate reorder intervals. In other words, at the end of the procedure, we obtain four candidate solutions. The final heuristic solution for (P) is the one that yields the smallest total cost. As we show in §6, this heuristic performs well in general.

4. The Optimal Solution

To find the optimal solution (Q_j^*, T_j^*) , $j = 1, \dots, N$, we propose a complete enumeration, which is facilitated with bounds on the optimal solution. Below we first construct bounds on $\hat{G}_j(y, \mathbf{Q}_j, \mathbf{T}_j)$, the average inventory-related costs for echelon j , $j = 1, \dots, N$. We then demonstrate how to use these echelon cost bounds to obtain the solution bounds.

The cost bounds for echelon j are obtained from a single-stage (r, nQ, T) system that has the original system parameters. The rationale for constructing these echelon cost bounds is the same as that for the system with (r, nQ) policies (i.e., $T_j = 1$) described in Shang and Song (2007). Briefly, consider echelon j that includes stages $1, 2, \dots, j$. If we restrict the local holding cost rate for each stage in this echelon to the same value, there will be no incentive to stock inventories in stages $i = 2, \dots, j$. Thus, the echelon system collapses into a single-stage system with lead time $L_{[1,j]}$. The lower- (upper-) bound is obtained by undercharging (overcharging) the holding cost rate h_j (resp., $h_{[1,j]}$) to each stage. We refer the reader to Shang and Song (2007) for a detailed discussion.

The resulting echelon cost bound functions are defined below. For $j = 1, \dots, N$, let

$$G_j^l(y, T_j) = \frac{1}{T_j} \sum_{\tau=0}^{T_j-1} \mathbb{E}[h_j(y - D[L_{[1,j]} + \tau]) + (b + h_{[j,N]})(y - D[L_{[1,j]} + \tau])^-], \quad (18)$$

$$G_j^u(y, T_j) = \frac{1}{T_j} \sum_{\tau=0}^{T_j-1} \mathbb{E}[h_{[1,j]}(y - D[L_{[1,j]} + \tau]) + (b + h_{[1,M]})(y - D[L_{[1,j]} + \tau])^-]. \quad (19)$$

The lower-bound function is

$$\widehat{G}_j^l(y, Q_j, T_j) = \frac{1}{Q_j} \sum_{x=1}^{Q_j} G_j^l(y+x, T_j), \quad (20)$$

and the upper-bound function is

$$\widehat{G}_j^u(y, Q_j, T_j) = \frac{1}{Q_j} \sum_{x=1}^{Q_j} G_j^u(y+x, T_j). \quad (21)$$

Note that \widehat{G}_j^l (resp., \widehat{G}_j^u) is the average inventory-related costs per period for a single-stage (r, nQ, T) system with holding cost rate h_j (resp., $h_{[1,j]}$), backorder cost rate $(b + h_{[j+1,N]})$, and lead time $L_{[1,j]}$.

Let

$$r_j^u(Q_j, T_j) = \arg \min_y \widehat{G}_j^u(y, Q_j, T_j),$$

$$r_j^l(Q_j, T_j) = \arg \min_y \widehat{G}_j^l(y, Q_j, T_j),$$

and

$$\widehat{G}_j^l(Q_j, T_j) = \widehat{G}_j^l(r_j^u(Q_j, T_j), Q_j, T_j),$$

$$\widehat{G}_j^u(Q_j, T_j) = \widehat{G}_j^u(r_j^l(Q_j, T_j), Q_j, T_j).$$

Also, define the average pipeline inventory cost (or the average inventory in-transit cost) as $\pi_j = \sum_{i=2}^j (h_i E[D[L_{[1,i-1]}]])$. We have

PROPOSITION 6. For $j = 1, \dots, N$,

- (1) $\widehat{G}_j^l(y, Q_j, T_j) + \pi_j \leq \widehat{G}_j(y, \mathbf{Q}_j, \mathbf{T}_j) \leq \widehat{G}_j^u(y, Q_j, T_j) + \pi_j$.
- (2) $r_j^l(Q_j, T_j) \leq r_j(\mathbf{Q}_j, \mathbf{T}_j) \leq r_j^u(Q_j, T_j)$.
- (3) $\widehat{G}_j^l(Q_j, T_j) + \pi_j \leq \widehat{G}_j(\mathbf{Q}_j, \mathbf{T}_j) \leq \widehat{G}_j^u(Q_j, T_j) + \pi_j$.
- (4) $\widehat{G}_j^l(Q_j, T_j)$ and $\widehat{G}_j^u(Q_j, T_j)$ are jointly convex and increasing in Q_j and T_j .

Let C^h be any heuristic cost and $c_j(Q_j, T_j) \stackrel{\text{def}}{=} c_j(\mathbf{Q}_j^j, \mathbf{T}_j^j)$. For any j , we have

$$\begin{aligned} C^h &\geq C(\mathbf{Q}^*, \mathbf{T}^*) = \sum_{i=1}^N \frac{K_i}{T_i^*} + \sum_{i=1}^N \frac{k_i \mu}{Q_i^*} + \widehat{G}_N(\mathbf{Q}_N^*, \mathbf{T}_N^*) \\ &= \sum_{i=1}^j \frac{K_i}{T_i^*} + \sum_{i=1}^j \frac{k_i \mu}{Q_i^*} + \sum_{i=j+1}^N c_i(\mathbf{Q}_i^*, \mathbf{T}_i^*) + \widehat{G}_j(\mathbf{Q}_j^*, \mathbf{T}_j^*) \\ &\geq \frac{K_{[1,j]}}{T_j^*} + \frac{k_{[1,j]} \mu}{Q_j^*} + \sum_{i=j+1}^N c_i(Q_i^*, T_i^*) \\ &\quad + \widehat{G}_j^l(Q_j^*, T_j^*) + \pi_j. \end{aligned} \quad (22)$$

Inequality (22) is due to the fact that $T_j^* \geq T_{j-1}^* \geq \dots \geq T_1^*$ and $Q_j^* \geq Q_{j-1}^* \geq \dots \geq Q_1^*$, Equation (16), and Proposition 6(3).

We can construct solution bounds for the optimal solution (Q_j^*, T_j^*) recursively for stage $j = N, N-1, \dots, 1$ from

the inequality (22). The construction starts from stage N . When $j = N$, (22) becomes

$$\frac{K_{[1,N]}}{T_N^*} + \frac{k_{[1,N]} \mu}{Q_N^*} + \widehat{G}_N^l(Q_N^*, T_N^*) \leq C^h - \pi_N. \quad (23)$$

Since the left-hand side of the inequality is jointly convex in (Q_N, T_N) , we can identify the solution bounds for (Q_N^*, T_N^*) such that the inequality holds. Let $(\underline{T}_N, \bar{T}_N)$ and $(\underline{Q}_N, \bar{Q}_N)$ denote the bounds for T_N^* and Q_N^* , respectively.³

Next, when $j = N-1$, (22) becomes

$$\begin{aligned} \frac{K_{[1,N-1]}}{T_{N-1}^*} + \frac{k_{[1,N-1]} \mu}{Q_{N-1}^*} + \widehat{G}_{N-1}^l(Q_{N-1}^*, T_{N-1}^*) \\ \leq C^h - \pi_{N-1} - c_N(Q_N^*, T_N^*). \end{aligned} \quad (24)$$

For $T_N \in \{\underline{T}_N, \underline{T}_N + 1, \dots, \bar{T}_N\}$ and $Q_N \in \{\underline{Q}_N, \underline{Q}_N + 1, \dots, \bar{Q}_N\}$, we can search for the minimum value of $c_N(Q_N, T_N)$, which we refer to as c_N . Thus, the right-hand side of (24) is less than or equal to $C^h - \pi_{N-1} - c_N$. From here, we can search for the solution bounds for (Q_{N-1}^*, T_{N-1}^*) because the left-hand side of (24) is jointly convex. The same procedure repeats until $j = 1$. At the end of the procedure, we obtain the solution bounds for (Q_j^*, T_j^*) , $j = 1, \dots, N$. The optimal solution can be determined by enumerating all feasible solutions that satisfy the integer-ratio constraints within the computed solution bounds.

5. Alternative Fixed Cost Assumptions

In our model, we assume that K_j is incurred in each order period, regardless of whether or not an order is placed, and k_j is incurred for each order batch. As discussed in the Introduction, these assumptions do indeed characterize some companies, such as EMC². Below we consider alternative fixed cost assumptions that model other situations. As we shall see, the approach for finding the optimal solution in §4 can be applied with minor modifications.

K_j Incurred Only When Placing an Order

This assumption applies to a situation in which K_j is incurred only when an order is placed. For example, some companies require a physical inventory count when an order is placed. Or, a company may have fleet trucks that ship materials only when its buyer places an order. (K_j is the shipping cost in the latter example.)

This assumption only affects the first fixed cost term in (8). More specifically, for stage j with the reorder interval T_j , an order will be placed when $IOP_j - D[T_j] \leq r_j$. Since IOP_j is uniformly distributed over $\{r_j + 1, \dots, r_j + Q_j\}$, the probability of placing an order in an order period is

$$p(Q_j, T_j) = \frac{1}{Q_j} \sum_{x=1}^{Q_j} P(D[T_j] \geq x). \quad (25)$$

Table 1. Fixed cost terms under alternative assumptions.

	K_j incurred for each T_j	K_j incurred when placing an order
k_j incurred per batch	Type I: $\sum_{j=1}^N \left(\frac{K_j}{T_j} + \frac{k_j \mu}{Q_j} \right)$	Type II: $\sum_{j=1}^N \left(\frac{K_j p(Q_j, T_j)}{T_j} + \frac{k_j \mu}{Q_j} \right)$
k_j incurred per order	Type III: $\sum_{j=1}^N \left(\frac{K_j}{T_j} + \frac{k_j p(Q_j, T_j)}{T_j} \right)$	Type IV: $\sum_{j=1}^N \left(\frac{K_j + k_j}{T_j} \right) p(Q_j, T_j)$

Thus, the average total order cost per period for all stages is

$$\sum_{j=1}^N \left(\frac{K_j p(Q_j, T_j)}{T_j} \right). \tag{26}$$

We only need to replace the first term in (8) with (26).

k_j Incurred for Each Order, Not for Each Batch

This assumption assumes that the setup cost k_j is incurred for each order that may include several batches. This is equivalent to saying that k_j is incurred only when placing an order. Thus, we only need to replace the second fixed cost term in (8) with $\sum_{j=1}^N (k_j p(Q_j, T_j)/T_j)$.

Table 1 presents a summary of four possible combinations of fixed costs. Type I is our current model. By defining an aggregate fixed cost term, say, $K'_j = k_j + K_j$, and setting $T_j = 1$, Type IV reduces to the classic (r, nQ) model by assuming that there is a single fixed cost K'_j incurred per order (e.g., Zheng and Chen 1992). For Type II, III, and IV, we can use the same approach as in §4 to find the optimal solution. However, since $p(Q_j, T_j)/T_j$ is not necessarily convex, we need to exclude the fixed cost terms in (22) to find the upper bound. That is,

$$(22) \geq \sum_{i=j+1}^N \underline{c}_i(Q_i^*, T_i^*) + \widehat{G}_j^l(Q_j^*, T_j^*) + \pi_j.$$

Since $\widehat{G}_j^l(Q_j, T_j)$ is convex and increasing in (Q_j, T_j) , we can follow the same procedure to find the solution bounds from stage $N, N - 1$, until stage 1. In this case, clearly, the lower bounds for both Q_j^* and T_j^* are one.

6. Numerical Study

We perform a three-part numerical study. In §6.1, we conduct a sensitivity analysis on the optimal solution. The objective is to examine how the optimal policy changes with respect to a change in system parameters. In §6.2, we examine the effectiveness of the heuristic. From this examination, we identify the situations under which the heuristic performs effectively and ineffectively. Finally, in §6.3, we provide a numerical study for the system with a fixed setup cost incurred per order, i.e., the Type III model in Table 1.

6.1. Observations on the Optimal Solution

We observe the optimal solution of instances that have the following parameters:

$$N = 3, \quad K_1, K_3 \in \{5, 50\}, K_2 = 20,$$

$$k_1, k_3 \in \{1, 20\}, k_2 = 10, \quad h_1, h_3 \in \{0.1, 1\}, h_2 = 1, \\ L_1, L_3 \in \{1, 3\}, L_2 = 2, \quad b = \{30, h_{[1,3]}\}.$$

We assume a Poisson demand with rate $\mu = 4$. We fix stage 2's parameters, and change the parameters of stage 1 and stage 3 as shown. We also consider small $b = h_{[1,3]}$ and large $b = 30$. The total number of instances is 512. For each instance, we compute the optimal solution $(\mathbf{Q}^*, \mathbf{T}^*)$ and the optimal cost C^* .

We summarize several numerical observations below.

(1) T_j^* and Q_j^* increase in K_j and k_j , respectively. Both decrease in h_j . This observation is intuitive and is consistent with the observation found in the single-stage (s, T) and (r, nQ) systems (Rao 2003, Zheng 1992). When K_j is large, the stage should use a larger T_j to reduce the order frequency. Also, when h_j is small, the stage should stock more inventory by increasing T_j . Similar arguments apply to the effect of k_j on Q_j .

Both T_j^* and Q_j^* decrease in b . (In fact, most T_j^* and Q_j^* decrease in b .) This finding may be explained as follows: When the backorder penalty is large, stage 1 needs to maintain a small batch size and a shorter reorder interval so as to be more responsive to the demand. On the other hand, the optimal reorder points $r_j(\mathbf{Q}^*, \mathbf{T}^*)$ increase in b , because each stage has to increase its safety stock level in order to avoid a higher backorder penalty.

Lead times seem to have little effect on T_j^* and Q_j^* , and even less effect on T_j^* .

(2) As stated, when K_j (k_j) changes, the change will affect T_j^* (Q_j^*). However, it is not clear how K_j (k_j) affects Q_j^* (T_j^*). We provide observations for this question.

First, consider 256 pairs of instances that differ only in the value of K_3 . Q_j^* changes in 121 (i.e., 47.3%) of these pairs. When K_1 rather than K_3 is allowed to vary, Q_j^* changes in 186 (i.e., 72.7%) of the pairs. Thus, K_1 has a greater impact on the optimal batch sizes than does K_3 .

On the other hand, when k_3 increases from 1 to 20, the value of T_j^* changes in 10 pairs (i.e., 3.9%). However, when k_1 rather than k_3 varies, the value of T_j^* changes in 38 pairs (i.e., 14.8%). Thus, k_1 has a greater impact on T_j than does k_3 .

These observations suggest two conclusions: (i) The optimal batch sizes and reorder intervals are more sensitive to downstream fixed costs than to upstream fixed costs. (ii) It is necessary to re-examine both optimal reorder intervals and batch sizes when K_j changes; however, it may not be

Table 2. Sensitivity of Q_j^* and T_j^* to value of fixed cost parameters.

Fixed cost parameters	Ratio of Q_j^* change (due to K_3 change) (%)	Ratio of T_j^* change (due to k_3 change) (%)
$K_3 = \{5, 50\}, k_3 = \{1, 20\}$	$30/64 = 46.9$	$3/64 = 4.7$
$K_3 = \{1, 20\}, k_3 = \{5, 50\}$	$27/64 = 42.2$	$13/64 = 20.3$

necessary to change reorder intervals when k_j changes at an upstream stage.

To assess whether the above observations hold true with bigger k_j and smaller K_j , we swap the parameter sets of k_j and K_j , i.e., $k_j = \{5, 50\}$, and $K_j = \{1, 20\}$ for the 128 instances (i.e., 64 pairs), where $(L_1, L_2, L_3) = (1, 2, 1)$ and hold other parameters the same. We compare this new set with the original 128 instances that have the same lead times. Table 2 presents the results for both the new set of 64 pairs and the original set. While the impact of k_3 on T_j^* is greater in the new set, the impact of K_3 on Q_j^* is only slightly reduced. We also observe in the new set that the impacts of K_1 and k_1 are larger than the impacts of K_3 and k_3 . These findings provide further support for the above two conclusions.

(3) The cost ratios K_j/h_j and k_j/h_j are strongly related to T_j^* and Q_j^* , respectively. More specifically, when K_j/h_j (k_j/h_j) decreases in j , the optimal reorder intervals T_j^* (Q_j^*) tend to be the same. This is because when the ratio of K_1/h_1 (k_1/h_1) is high, stage 1 will choose a larger T_1^* (Q_1^*). Due to the integer-ratio constraints, an upstream stage would tend to use the value of T_1^* (Q_1^*) as its reorder interval (batch size).

REMARK 1. The range of the solution bounds varies. In the 512 instances we tested, the means (standard deviations) of the range are 48.4 (24.2) and 17.9 (13.3) for Q_1^* and T_1^* , 47.1 (19.5) and 11.6 (5.0) for Q_2^* and T_2^* , and 73.0 (6.5) and 28.8 (9.1) for Q_3^* and T_3^* . We observe that the optimal solution can locate anywhere within these ranges. □

6.2. Effectiveness of the Heuristic

We next examine the effectiveness of the proposed heuristic for the (r, nQ, T) policy. We compute the heuristic solution and the corresponding heuristic cost C^h for each of the original 512 instances tested in §6.1. Define the percentage error as

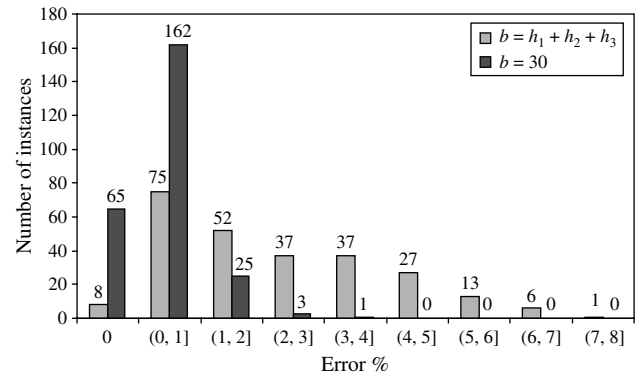
$$\epsilon\% = \frac{(C^h - C^*)}{C^*} \times 100\%.$$

The average percentage error is 1.31% with a maximum of 7.67%. Figure 2 shows the distribution of the errors.

Below we provide several observations regarding the heuristic.

(1) The heuristic is more effective when the backorder cost rate b is large. For example, when $b = 30$, the average error is 0.32% with a maximum of 3.22%. The heuristic

Figure 2. The distributions of the percentage errors for the 256 instances that have $b = h_{[1,3]}$ (gray bars) and $b = 30$ (black bars).



generates the optimal solution in 65 out of 256 instances. On the other hand, when b is small, i.e., $b = h_{[1,3]}$, the average error is 2.30% with a maximum of 7.67%.

(2) For a given b , the heuristic is more effective when K_1/h_1 and k_1/h_1 are larger than K_j/h_j and k_j/h_j , respectively, $j \neq 1$. For the 32 instances with $b = h_{[1,3]}$ that satisfy the above condition, the average percentage error is 0.41%, which is significantly lower than the average error of 2.30% of all 256 instances with $b = h_{[1,3]}$. We observe that the optimal batch sizes and reorder intervals tend to be equal among all stages in these 32 instances. For the special case where K_j/h_j and k_j/h_j decrease in j (12 instances), the heuristic is surprisingly effective. The heuristic generates eight optimal solutions, which comprise the total optimal solutions found by the heuristic when $b = h_{[1,3]}$.

To explain why the heuristic is effective under such cost structure, we first review an existing result. For the deterministic demand model with reorder intervals (batch sizes) as control variables, it is well known that the optimal partition obtained from solving the relaxed problem can be determined by the cost ratios K_j/h_j (k_j/h_j). See Zipkin (2000, pp. 125–130) for an explanation. Here, we observe that the partition obtained from the cost ratios is mostly consistent with that obtained from solving (QP) in step 2 and (TP) in step 3 in our heuristic. For example, when K_1/h_1 and k_1/h_1 are significantly larger than the others, we observe that the heuristic tends to group all stages into one cluster, which results in all stages having the same heuristic reorder intervals and batch sizes. This behavior is consistent with that of the optimal solution as explained in the third observation of §6.1.

(3) On the other hand, for a given b , the heuristic performs less effectively when K_j/h_j increases in j . For example, consider the 48 instances that have $b = h_{[1,3]}$ and $(K_1/h_1, K_2/h_2, K_3/h_3)$ equal to either $(5, 20, 50)$ or $(5, 20, 500)$. The average percentage error is 4.19%, which is larger than the average percentage error of 2.30%. If we only consider the 32 instances that have $(K_1/h_1, K_2/h_2, K_3/h_3) = (5, 20, 50)$, the average percentage error is 4.65%. Interestingly, the performance does not

seem worse if we only consider the instances where both K_j/h_j and k_j/h_j increase in j .

This again may be explained by the clusters generated from the heuristic. Under such cost structure, the heuristic tends to generate three clusters after solving (TP) in step 3. Thus, stage 1 in cluster $c(1)$ will generate a small T_1 (because K_1/h_1 is small, which implies either K_1 is small and/or h_1 is large). This small T_1 restricts the possibility that T_1^* could be a large value. For example, for the instance with the maximum percentage error of 7.67%, the parameters are $(K_1, K_2, K_3) = (5, 20, 50)$, $(k_1, k_2, k_3) = (20, 10, 20)$, $(h_1, h_2, h_3) = (1, 1, 1)$, $(L_1, L_2, L_3) = (1, 2, 1)$, and $b = h_{[1,3]}$. The optimal solution is $(T_1^*, T_2^*, T_3^*) = (6, 6, 6)$ and $(Q_1^*, Q_2^*, Q_3^*) = (22, 22, 22)$. The heuristic groups stages into three clusters, which results in a heuristic solution $(T_1, T_2, T_3) = (2, 4, 8)$ and $(Q_1, Q_2, Q_3) = (16, 16, 16)$.

(4) The performance of the heuristic seems insensitive to lead times. That is, for the same lead time parameters, we can find both high and low percentage error cases. This observation is intuitive: Lead times have a direct impact on the optimal reorder points, but have less impact on the optimal batch sizes or reorder intervals. Thus, they have little impact on the performance of the heuristic.

REMARK 2. Our heuristic solution is determined by comparing the four candidate solutions, i.e., (Q', T'_j) , (Q'', T'') , (Q''', T''') , and (Q''', T''') , $j = 1, \dots, N$. The number of instances that the heuristic solution is yielded from (Q', T'_j) , (Q'', T'') , (Q''', T''') , and (Q''', T''') is 263, 188, 131, and 155, respectively. (Note that the sum is not equal to 512 because some of the candidate solutions are the same.) Thus, we do not observe that one solution dominates the others. □

To examine whether the heuristic still performs well when N increases, we consider $N = 2, 4, 6$. The system parameters are $K_j = 20$, $k_j = 5$, $b \in \{9, 39\}$ and demand is Poisson with $\mu = 4$. We fix $h_{[1,N]} = 1$ and test four different forms of holding costs following Zipkin (2000, p. 313). Specifically, for the linear form, we set $h_j = 1/N$; for affine, we set $h_N = \alpha + (1 - \alpha)/N$, $h_j = (1 - \alpha)/N$; for kink, we set $(1 - \alpha)/N$ for $j \leq N/2$ and $(1 + \alpha)/N$ for $j > N/2$; finally, for the jump form, we set $h_j = \alpha + (1 - \alpha)/N$ for $j = N/2 + 1$ and $h_j = (1 - \alpha)/N$ otherwise. Here, we set $\alpha = 0.75$ for all cases. The total number of instances is 24. Intuitively, the heuristic should be less effective when N increases because the chance that the heuristic chooses ineffective clusters is higher. This intuition holds true when $b = 39$: The average percentage error for $N = 2, 4$, and 6 is 0.00%, 0.04%, and 0.11%, respectively. Among all instances, the average percentage error for $N = 2, 4$, and 6 is 0.00%, 0.14%, and 0.12%, respectively. This suggests that our heuristic does not seem to deteriorate much when N increases. Similar to the $N = 3$ cases, the heuristic is particularly effective when k_1/h_1 and K_1/h_1 are large. For example, for the instances with the affine, kink, and jump forms (18 instances), the average percentage error is 0.01%, with a maximum of 0.04%.

As stated, the (r, nQ) and (s, T) policies are special cases of the (r, nQ, T) policy. Specifically, the (r, nQ) and (s, T) models are the Q -problem and the T -problem formulated in §3.3 when $T_j = 1$ and $Q_j = 1$ for all j , respectively. We can use the proposed heuristic for each of the problems to generate a heuristic solution. We provide two tests that examine the effectiveness of these heuristics. For the (r, nQ) model, we compare our heuristic with the best heuristic proposed by Chen and Zheng (1998). We consider the same parameters as those in Tables 1 and 3 of Chen and Zheng. The average (maximum) percentage error of our heuristic is 0.4% (3.4%), which is less than the average error of 1.7% (5.1%) obtained by using Chen and Zheng’s heuristic. For the (s, T) policy, we test the following parameters: $N = 3$, $K_j \in \{1, 10\}$, $h_j \in \{0.1, 1\}$, $b \in \{30, h_{[1,N]}\}$, $L_j \in \{1, 3\}$, and $\lambda = 4$. The total number of instances is 1024. The average percentage error is 0.67%, with a maximum of 7.30%.

6.3. Comparison of Type I and Type III Models

This section presents a numerical study comparing the models where setup cost k_j is incurred per batch (i.e., Type I in Table 1) and incurred per order (i.e., Type III). The purpose of this comparison is to examine the sensitivity of how the optimal solution in these two models changes to the changes in the quotient K_j/k_j . For both models, we use the following parameters: $N = 3$, $h_j = 0.1$, $L_j = 1$, $b = 3$, and $k_j = 40$, for $j = 1, 2, 3$. We set $K_1 = K_2 = K_3 \in \{1, 5, 20, 50\}$. The demand is Poisson with $\mu = 5$. Table 3 shows the optimal solution for these eight instances.

For the Type I model, T_j^* and Q_j^* increase in K_j . For the Type III model, while T_j^* increases in K_j , interestingly, Q_j^* remains equal to one except that $Q_3^* = 2$ when $K_j = 1$. The fact that Q_j^* is smaller in the Type III model may be explained as follows: When more batches share the same setup cost k_j , Q_j^* tends to decline. However, it is surprising to observe that even when k_j is much bigger than K_j , e.g., $k_j = 40$ and $K_j = 5$, the optimal batch sizes are still equal to one. This result suggests that (s, T) policies can be very effective for companies that order periodically, as advanced technology, such as flexible manufacturing systems, has significantly reduced the number of setups in production.

Table 3. Comparison of the Type I and Type III models.

Type	$K_j = 1$		$K_j = 5$		$K_j = 20$		$K_j = 50$	
	I	III	I	III	I	III	I	III
(Q_1^*, T_1^*)	(69, 3)	(1, 7)	(71, 6)	(1, 10)	(74, 11)	(1, 12)	(78, 16)	(1, 13)
(Q_2^*, T_2^*)	(69, 3)	(1, 7)	(71, 6)	(1, 10)	(74, 11)	(1, 12)	(78, 16)	(1, 13)
(Q_3^*, T_3^*)	(69, 3)	(2, 7)	(71, 6)	(1, 10)	(74, 11)	(1, 12)	(78, 16)	(1, 13)

7. Conclusion

This paper analyzes a serial inventory system with echelon (r, nQ, T) policies. Previous studies showed how to evaluate the policy and optimize the echelon reorder points. This paper provides a simple heuristic for obtaining effective batch sizes and reorder intervals. This heuristic is based on solving lower and upper bounds of the total cost function. We also provide an approach for finding the optimal batch sizes and reorder intervals. This is achieved by constructing cost bounds for each echelon. The stage and echelon cost bounds together generate the bounds for the optimal solution. We then conduct a complete enumeration to obtain the optimal solution. In a numerical study, we find that the heuristic performs well in general and is especially effective when (1) backorder cost is high, or (2) the ratio of the fixed cost to the echelon holding cost at the most downstream stage is large. We also find that when fixed order costs change, it is often necessary to re-examine both optimal batch sizes and reorder intervals; however, when fixed setup costs change, it may suffice to adjust batch sizes, not reorder intervals. This finding suggests that a company may not need to re-evaluate the ordering/shipping schedule when the setup costs change due to, say, new technology. Finally, by considering a special case where the setup cost is incurred per order, our numerical result suggests that the (s, T) policy can be very effective. This may explain why such a policy is prevalent in practice.

8. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at <http://orjournal.informs.org/>.

Endnotes

1. This description is based on private conversations with EMC² managers.
2. Due to the PASTA (Wolff 1982) property, the analysis for the periodic-review (r, nQ) model is essentially the same as that for the continuous-review model with compound Poisson demand. See Chen and Zheng (1994).
3. More specifically, for fixed T_N , let $Q_N(T_N) = \arg \min_{Q_N} \{(k_{[1, N]}\mu)/Q_N + \widehat{G}_N^l(Q_N, T_N)\}$. Thus, the left-hand-side of (23) is further greater than or equal to $f(T_N^*) \stackrel{\text{def}}{=} (K_{[1, N]}/T_N^*) + (k_{[1, N]}\mu)/Q_N(T_N^*) + \widehat{G}_N^l(Q_N(T_N^*), T_N^*)$. We then use $f(T_N)$ to construct the solution bounds for T_N^* , i.e., $\overline{T}_N = \max\{T_N \mid f(T_N) \leq C^h - \pi_N\}$ and $\underline{T}_N = \min\{T_N \mid f(T_N) \leq C^h - \pi_N\}$. We can use the same idea to construct the solution bounds for Q_N^* and for (Q_j^*, T_j^*) , $j < N$.

Acknowledgments

The authors are grateful to Xiuli Chao, Izak Duenyas, Jing-Sheng Song, Paul Zipkin, and the reviewers for helpful comments. The second author is supported in part by Hong Kong GRF grant CUHK 419408.

References

- Axsäter, S. 2003. Supply chain operations: Serial and distribution inventory systems. S. C. Graves, T. de Kok, eds. Chapter 10, *Supply Chain Management: Design, Coordination and Operation*, Vol. 11: *Handbooks in Operations Research and Management Science*. Elsevier, Amsterdam, 525–559.
- Axsäter, S., K. Rosling. 1993. Installation vs. echelon stock policies for multi-level inventory control. *Oper. Res.* **39** 1274–1280.
- Cachon, G. 1999. Managing supply chain demand variability with scheduled ordering policies. *Management Sci.* **45** 843–856.
- Chao, X., S. X. Zhou. 2009. Optimal policy for a multiechelon inventory system with batch ordering and fixed replenishment intervals. *Oper. Res.* **57**(2) 377–390.
- Chen, F. 1998a. On (R, nQ) policies in serial inventory systems. S. Tayur, R. Ganeshan, M. Magazine, eds. *Quantitative Models for Supply Chain Management*. Kluwer Academic Publishers, Norwell, MA, 71–110.
- Chen, F. 1998b. Stationary policies in multiechelon inventory systems with deterministic demand and backlogging. *Oper. Res.* **46** S26–S34.
- Chen, F. 2000. Optimal policies for multi-echelon inventory problems with batch ordering. *Oper. Res.* **48** 376–389.
- Chen, F., Y.-S. Zheng. 1994. Evaluating echelon stock (R, nQ) policies in serial production/inventory systems with stochastic demand. *Management Sci.* **40** 1262–1275.
- Chen, F., Y.-S. Zheng. 1998. Near-optimal echelon-stock (R, nq) policies in multistage serial systems. *Oper. Res.* **46** 592–602.
- De Bodt, M., S. Graves. 1985. Continuous review policies for a multi-echelon inventory problem with stochastic demand. *Management Sci.* **31** 1286–1295.
- Feng, K., U. Rao. 2007. Echelon-stock (R, nT) control in two-stage serial stochastic inventory systems. *Oper. Res. Lett.* **35** 95–104.
- Graves, S. 1996. A multiechelon inventory model with fixed replenishment intervals. *Management Sci.* **42** 1–18.
- Hadley, G., T. Whitin. 1963. *Analysis of Inventory Systems*. Prentice-Hall, Englewood Cliffs, NJ.
- Kiesmüller, G. P., A. G. de Kok. 2006. The customer waiting time in an (R, s, Q) inventory system. *Internat. J. Production Econom.* **104** 354–364.
- Larson, C., G. P. Kiesmüller. 2007. Developing a closed-form cost expression for an (R, s, nQ) policy where the demand process is compound generalized Erlang. *Oper. Res. Lett.* **35** 567–572.
- Maxwell, W., J. Muckstadt. 1985. Establishing consistent and realistic reorder intervals in production-distribution systems. *Oper. Res.* **33** 1316–1341.
- Rao, U. 2003. Properties of the period review (R, T) inventory control policy for stationary, stochastic demand. *Manufacturing Service Oper. Management* **5** 37–53.
- Roundy, R. 1985. 98%-effective integer-ratio lot-sizing for one-warehouse multi-retailer systems. *Management Sci.* **31** 1416–1430.
- Shang, K. 2008. Note: A simple heuristic for serial inventory systems with fixed order costs. *Oper. Res.* **56** 1039–1043.
- Shang, K., J.-S. Song. 2007. Serial supply chains with economies of scale: Bounds and approximations. *Oper. Res.* **55** 843–853.
- van Houtum, G.-J., A. Scheller-Wolf, J. Yi. 2007. Optimal control of serial inventory systems with fixed replenishment intervals. *Oper. Res.* **55** 674–687.
- Wolff, R. W. 1982. Poisson arrivals see time averages. *Oper. Res.* **30** 223–231.
- Zheng, Y.-S. 1992. On properties of stochastic inventory systems. *Management Sci.* **38** 87–103.
- Zheng, Y.-S., F. Chen. 1992. Inventory policies with quantized ordering. *Naval Res. Logist.* **39** 285–305.
- Zipkin, P. 1986. Inventory service-level measures: Convexity and approximation. *Management Sci.* **32** 975–981.
- Zipkin, P. 2000. *Foundations of Inventory Management*. McGraw-Hill, New York.