NEAR-OPTIMAL PRICING AND REPLENISHMENT STRATEGIES FOR A RETAIL/DISTRIBUTION SYSTEM

FANGRUO CHEN
Graduate School of Business, Columbia University, New York, New York 10027, fc26@columbia.edu

AWI FEDERGRUEN
Graduate School of Business, Columbia University, New York, New York 10027, af7@columbia.edu

YU-SHENG ZHENG
The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, zheng@wharton.upenn.edu
(Received July 1998; revision received September 1999; accepted May 2000)

This paper integrates pricing and replenishment decisions for the following prototypical two-echelon distribution system with deterministic demands. A supplier distributes a single product to multiple retailers, who in turn sell it to consumers. The retailers serve geographically dispersed, heterogeneous markets. The demand in each retail market arrives continuously at a constant rate, which is a general decreasing function of the retail price in the market. The supplier replenishes its inventory through orders (purchases, production runs) from a source with ample capacity. The retailers replenish their inventories from the supplier. We develop efficient algorithms to determine optimal pricing and replenishment strategies for the following three channel structures. The first is the vertically integrated channel, where the system-wide pricing and replenishment strategies are determined by a central planner whose objective is to maximize the system-wide profits. The second structure is that of a vertically integrated channel in which pricing and operational decisions are made sequentially by separate functional departments. The third channel structure is decentralized, i.e., the supplier and the retailers are independent, profit-maximizing firms with the supplier acting as a Stackelberg game leader. We apply our algorithms to a set of numerical examples to quantify the supply chain inefficiencies due to functional segregation or uncoordinated decision making in a decentralized channel. We also gain insight into systematic differences in the associated pricing and operational patterns.

1. INTRODUCTION

Traditional inventory planning models for supply chains assume that all demand processes and hence all revenue streams are exogenously determined. As a consequence, such models focus on the minimization of the logistics costs in the supply chain. Marketing models, on the other hand, focus on pricing strategies and their impact on sales volumes and revenues, typically with a rudimentary and simplistic treatment of the operational costs in the system.

This paper integrates pricing decisions with the determination of replenishment strategies for the following prototypical two-echelon distribution system with deterministic demands. A supplier distributes a single product to multiple retailers who in turn sell it to consumers. The retailers serve geographically dispersed, heterogeneous markets (i.e., the retailers are local monopolists). (A much more complex model arises when the retailers act as oligopolists, where the demand at each retailer is a function of not only its own price but also the prices charged by the other retailers. In a decentralized setting, the problem facing the retailers can be modeled as a noncooperative game, with a solution being based on a particular equilibrium concept. We do not address this type of model, which requires entirely different algorithmic approaches.) The demand in each retail market arrives continuously at a constant rate which is a general decreasing function of the retail price in the market. The supplier replenishes its inventory through orders (purchases, production runs) from a source with ample capacity. The retailers replenish their inventories from the supplier. The logistics costs consist of (1) holding costs for the inventories carried at the supplier and the retailers, (2) fixed and variable costs for orders placed by the supplier and those placed by each of the retailers, and (3) account management costs, i.e., the costs incurred by the supplier for managing each retailer’s needs and transactions.

Channel-wide profits are most straightforwardly optimized, if there is a central planner who makes all the pricing and replenishment decisions in the system. A recent paper by Chen et al. (2001) has identified that this centralized solution can in fact be realized in a decentralized system if the channel implements a wholesale pricing scheme that offers discounts based on a number of the retailers’ characteristics. However, whether the centralized solution is implemented by a central decision maker specifying all pricing and operational decisions throughout the system, or in a decentralized manner via a properly designed discount scheme, it is necessary to determine the Centralized Solution efficiently. This represents one of the main objectives of our paper.

Inefficiencies in the above supply chain may arise in a variety of ways. In a vertically integrated channel (i.e., the
supplier and the retailers belong to the same firm), one often sees that the pricing decisions are made by the marketing department to maximize revenues minus variable production/transportation costs, ignoring less apparent operational (for instance, setup and inventory) costs. Taking these pricing decisions (and the resulting demand rates) as given, the production/distribution department subsequently determines a replenishment policy minimizing total operational costs. In short, the pricing and replenishment decisions are made sequentially. This artificial separation of the marketing and operational decisions, typical of many organizations in which different functional departments act as isolated silos, causes supply chain inefficiencies. We refer to the resulting solution as the silo solution.

In a decentralized channel where the supplier and the retailers are independent, profit-maximizing firms, supply chain inefficiencies may be even more severe. Consider the case where the supplier is able to specify the terms (i.e., the wholesale price) unilaterally to maximize its own performance. This situation can be modeled as a Stackelberg game, with the supplier as the leader and the retailers as followers. More specifically, the supplier sets a wholesale price and a replenishment strategy for itself with the objective of maximizing its own profit. Each of the retailers takes the wholesale price as given and determines a retail price and a replenishment strategy so as to maximize its own profits. The resulting solution for the supply chain is referred to as the Stackelberg solution.

To understand the magnitudes of the supply chain inefficiencies in the above scenarios as well as any systematic differences in the associated pricing and operational patterns, one must determine the resulting solutions and compare them with the centralized solution. Note that in the silo solution, the channel-wide profits can easily be determined by using existing results (see the literature review below); in fact, this problem mirrors the current academic literature in marketing and operations. However, the problems of obtaining the centralized solution and the Stackelberg solution are considerably harder and to our knowledge, untackled. The main contribution of this paper is to provide efficient algorithms to solve these problems. While not strictly polynomial in \( N \), the number of retailers, we demonstrate that all proposed algorithms are in practice of linear or \( O(N \log N) \) time, with instances with 10 retailers solvable in approximately 0.02 CPU seconds on a PC with a Pentium 450-MHz processor.

We apply our algorithms to a set of numerical examples to gain insight into the above mentioned questions and comparisons. Based on the results, we make the following observations. First, the percentage decrease in channel-wide profits when going from the centralized solution to the Stackelberg solution is significant, ranging from around 30% to 100%. When the percentage reaches 100%, it means that the supplier cannot ensure itself of a positive profit, even though a vertically integrated system results in healthy profits. In such cases, the supplier, despite being the price leader, is forced to withdraw from the market. Second, sequential (as opposed to joint) decision making can be quite costly. In some cases, sequential decision making results in negative profits even though the maximum channel profits are positive. Third, compared with the centralized solution, the Stackelberg solution results in higher retail prices and lower sales volumes. Therefore, the double marginalization phenomenon, identified in many industrial organization models since Spengler’s (1950) seminal paper, also arises in our Stackelberg game. Finally, we have observed that the centralized solution generally prescribes longer replenishment intervals than the Stackelberg solution. Therefore, shorter replenishment intervals are not automatically the result of the mere integration or coordination of decision making processes, contrary to what some have suggested (Eckstut and Tang 1996).

The remainder of this paper is organized as follows. Section 2 briefly reviews the relevant literature. Section 3 specifies the model and basic notation. The solution methods for the centralized solution, the silo solution, and the Stackelberg solution are developed in §§4, 5, and 6, respectively. Section 7 reports on the numerical examples.

2. LITERATURE REVIEW

It is only slightly more than a decade ago that Roundy (1985) identified effective strategies for the one-warehouse multi-retailer model considered here, under the assumption of centralized control and exogenously given demand rates for all retailers. Even in this simplified setting, the structure of an optimal strategy is exceedingly complex; such fully optimal strategies are therefore of no practical use. Roundy showed, however, that a simple so-called power-of-two policy comes within 2% of being optimal. Under a power-of-two policy, each facility replenishes its inventory with intervals of constant length, which is a power-of-two multiple of a common base-period. Subsequent work (e.g., Roundy 1986, Federgruen et al. 1992, and Federgruen and Zheng 1995) showed that power-of-two policies continue to come within 2% of optimality for a much larger class of production/distribution networks with general cost structures.

The marketing literature on channel coordination focuses on pricing decisions. Jeuland and Shugan (1983) consider a simple channel with one supplier and one retailer. Their model does not consider any inventory replenishment decisions and resulting setup and inventory carrying costs. This single-retailer model is generalized by Ingene and Parry (1995) to multi-retailer settings. The operations literature dealing with channel coordination, on the other hand, has until recently been confined to replenishment decisions, assuming that all demand processes are exogenously given. With the exception of Lal and Staelin (1984), this literature restricts itself to channels with a single retailer (or multiple, identical retailers). These papers compute an optimal centralized solution and discuss various alternatives to implement this solution. Monahan (1984) determines the centralized solution with the restriction that the
supplier and retailer use identical order intervals, while Lee and Rosenblatt (1986) relax this assumption. Lal and Staelin (1984)—dealing with multiple, nonidentical retailers—compute a centralized solution implicitly assuming the supplier replenishes its stock infrequently, hence in large quantities. This paper also discusses a heuristic coordination mechanism via an order quantity discount scheme. We refer to Chen et al. (2001) for a more detailed review of this part of the literature.

The need to integrate inventory control and pricing strategies was first advocated by Whitin (1955). Both Whitin (1955) and later Mills (1959, 1962) address a single-period, single-location model in which a single price and supply quantity need to be determined (under uncertain demands). Subsequent work by Karlin and Carr (1962) considers an infinite horizon model, again for a single item and under the assumption that a single constant price is to be specified at the beginning of the planning horizon. The first treatments of dynamic integrated pricing and inventory strategies (i.e., in a multi-period setting) are due to Thomas (1970) and Kuenreuther and Schrage (1973), developing variants of the Wagner and Whitin (1958) dynamic lot-sizing model for settings where the demands can be controlled by selecting appropriate price levels. Rajan et al. (1992) analyze a continuous-time version of the same model. Gilbert (1997) extends the results in Kuenreuther and Schrage (1973). See Eliaishberg and Steinberg (1993) for a comprehensive survey of this area of research. Federgruen and Heching (1999) characterize optimal strategies for a periodic-review model with stochastic demand; see the references therein for earlier treatments of special cases. Finally, there is a stream of recent papers characterizing dynamic pricing strategies in settings with a single initial procurement or perhaps two procurements during the planning horizon. These include Bitran and Mondshein (1993, 1997), Bitran et al. (1998), Gallego and van Ryzin (1994), Fisher and Raman (1996), Eppen and Iyer (1997a, 1997b) and Zhao and Zheng (2000). All these models deal with a single item held at a single location only. Most recently, Gilbert (1998) extended the Kuenreuther-Schrage model, without setup costs, to allow for multiple items tied together by a joint capacity constraint.

3. MODEL

Consider a distribution system consisting of a supplier who distributes a single product to N geographically dispersed retailers. The consumer demand in each retailer's market occurs continuously at a constant rate, which is determined by the retail price charged in this market, in accordance with a general, time-invariant (i.e., stationary) demand function. All demands must be satisfied without backlogging.

Each firm (i.e., the retailers and the supplier) incurs inventory carrying costs which at any point in time are proportional to its inventory level, at a facility-specific cost rate. The supplier replenishes its inventory through orders from an outside source with unlimited stock, and the retailers place orders with the supplier. All such orders incur fixed and variable order costs; the fixed cost components as well as the variable order cost rates may all be facility-dependent. As with the outside consumer demand, retail orders cannot be backlogged.

In addition, our model allows for a cost component that is often ignored in many inventory models: the supplier may incur a specific annual cost for managing each retailer's needs and transactions. In the consumer electronics industry, for example, one finds that some suppliers establish a "management team" for each of its major retailer accounts, to monitor the retailers' needs, transactions and forecasts, to negotiate and implement sales and logistical terms, etc. Such account teams may comprise logistics managers, and sales and production representatives. The same management team may be devoted exclusively to a single retailer account, for the largest among them, or to multiple accounts, for the smaller ones. For settings where these costs arise, we model the "management costs" associated with a retailer account by a concave function of the retailer's annual sales volume, reflecting economies of scale.

All cost parameters are stationary. Similarly, each retailer maintains a constant retail price. The planning horizon is infinite. For the centralized system, the objective is to optimize long-run average system-wide profits, whereas for the decentralized system, each firm maximizes its own long-run average profits.

Define for \( i = 1, \ldots, N \),

\[
\begin{align*}
    p_i &= \text{retail price charged by retailer } i \\
    d_i(p_i) &= \text{annual consumer demand in the market served by retailer } i, \text{ a strictly decreasing function of } p_i \\
    K_0 &= \text{fixed (setup) cost per order placed by the supplier} \\
    K_i &= \text{fixed (setup) cost per order placed by retailer } i \\
    h_0 &= \text{basic annual holding cost per unit carried in inventory anywhere in the system} \\
    h_i &= \text{incremental (or echelon) annual holding cost per unit carried in inventory at retailer } i \\
    \tilde{h}_i &= h_0 + h_i \\
    c_0 &= \text{variable cost per unit ordered by the supplier} \\
    c_i &= \text{variable cost per unit ordered by retailer } i \\
    \psi(d_i) &= \text{annual cost for managing retailer } i's \text{ account, with } \Psi(\cdot) \text{ increasing, concave and } \Psi(0) = 0
\end{align*}
\]

(In this paper, we use "increasing" and "decreasing" in the weak sense.) Let \( p(d_i) \) be the inverse demand function, since the demand function \( d(\cdot) \) is strictly decreasing. We assume that \( p(d_i) d_i \) is strictly concave and differentiable in \( d_i \) and that the price \( p_i \) is allowed to vary over a finite interval \([p_i^\text{min}, p_i^\text{max}]\) only.

Note that the quantities \( \{K_i, h_i, c_i; i = 0, \ldots, N\} \) are given parameters while the prices \( \{p_i; i = 1, \ldots, N\} \) or alternatively the demand rates \( \{d_i; i = 1, \ldots, N\} \) are decision variables. We assume, as in standard inventory models,
that for all $i = 1, \ldots, N$, $h_i > 0$, i.e., the cost of carrying a unit at retailer $i$ is at least as large as the cost of carrying it in the supplier's warehouse. In addition, the above specified costs for retailers' orders, as represented by the parameters $\{K_i, c_i; i = 1, \ldots, N\}$, reflect the total expenditure associated with retailer orders for the system as a whole, whether expended by the supplier, the retailer or both. Note that these costs include all logistics costs such as handling and transportation costs, but exclude any transfer payments between the retailers and the supplier. (For each order by retailer $i$, it is often the case that the supplier incurs a fixed order-processing cost $K_i$ and the retailer incurs a fixed setup cost $K'_i$ with $K_i + K'_i = K_i$. The separation of these cost components of $K_i$ is clearly unnecessary in analyzing the centralized system, and it does not fundamentally affect the analysis of the decentralized system, see below.)

We assume, for convenience, that all orders are received instantaneously upon placement. Positive, deterministic leadtimes can be handled by a simple shift in time of all desired replenishment epochs, as determined under zero leadtimes, by the actual leadtimes.

Assume, for a moment, that all retailer prices and hence all demand rates and revenues are given, and consider the problem of finding an integrated replenishment strategy minimizing system-wide long-run average setup and holding costs. As mentioned in the introduction, Roundy (1985) showed that even for this limited problem an exact optimal strategy is exceedingly complex and entirely intractable. On the other hand, he showed that a simple power-of-two policy is guaranteed to come within 2% of optimality and that the parameters of this policy can be computed by a simple algorithm in $O(N \log N)$ time. (A later refinement by Queyranne 1987 reduces the complexity to $O(N)$.) Under such a power-of-two policy a base period $T_L$ is chosen, and each of the facilities chooses a constant replenishment interval from the set of the power-of-two multiples of $T_L$, receiving replenishments when its inventory is down to zero. Let

$$T_i = \text{replenishment interval for retailer } i, \ i = 1, \ldots, N$$

$$T_0 = \text{replenishment interval for the supplier.}$$

Throughout this paper, we assume that $T_L$ is fixed. A power-of-two policy is one where

$$T_i = 2^{m_i} T_L, \quad i = 0, \ldots, N, \quad (1)$$

with $m_i$ a (positive or negative) integer. (The above 2% optimality gap is obtained when $T_L$ is allowed to vary. If $T_L$ is fixed, the guaranteed optimality gap increases to 6%.)

Assume that the system starts out empty (i.e., without inventories). Thus all facilities place an order at time 0. Note that thereafter all retailers with a given replenishment interval $T_i = t$ place all of their orders at the same replenishment epochs along with all retailers with a smaller replenishment interval. (A replenishment interval $t'$ smaller than $t$ satisfies $t = 2^mt'$, due to Equation (1), with $m$ a positive integer.) The power-of-two structure thus induces maximum coordination between the replenishment orders of different facilities, providing an intuitive explanation for its excellent performance.

4. THE CENTRALIZED SYSTEM

This section considers the problem facing a central planner for the supply chain, who determines the retail prices and the system-wide replenishment strategy, and whose objective is to maximize the system-wide profits. We develop an efficient algorithm for solving this problem.

The problem considered here is considerably harder than that treated in Roundy (1985), who showed that for exogenously given demand rates, there exists a power-of-two policy that comes within 2% of optimality (or 6% if the base period is fixed). This negligible optimality gap, when measured in terms of cost differentials along with other attractive features discussed above and elsewhere (e.g., Roundy 1985 and Federgruen et al. 1992), motivates us to restrict ourselves to replenishment strategies which are of the power-of-two type. Below, we examine the optimality gap (in terms of system-wide profits) resulting from this restriction. Let $\text{profit}^* = \text{revenue}^* - \text{cost}^*$ be the maximum system-wide profits under an optimal pricing/replenishment strategy without the power-of-two restriction, and $\text{profit}^0$ the maximum system-wide profits under the restriction. (Here the revenue term is defined to be sales minus variable costs minus account management costs, and the cost term refers to setup and inventory holding costs only. Note that the supply chain’s revenue is completely determined by the demand rates.) Now fix the demand rates (thus retail prices) at those found in the (truly) optimal pricing/replenishment solution, and consider the remaining problem of choosing a replenishment strategy to minimize the system-wide inventory costs (i.e., setup and inventory holding costs). This problem is exactly the one considered in Roundy (1985). Let $LB$ be a lower bound on the system-wide inventory costs of any feasible policy. As mentioned, one can find an optimal power-of-two policy so that the corresponding system-wide inventory costs, denoted by cost', is within 6% of $LB$ (assuming the base period is fixed), i.e., cost' $\leq$ 1.06LB. Because cost' $\leq$ LB, 1.06cost' $\geq$ cost'.

Let profit' be the resulting system-wide profits. Because the demand rates are fixed at those found in the truly optimal solution, the total revenue for this system is still revenue'. Thus $\text{profit}^0 = \text{revenue}^* - \text{cost}^*$. Note that this sequential solution (fix demand rates first, then optimize within the class of power-of-two policies) is only a feasible solution to the integrated pricing/replenishment problem under the power-of-two restriction. Thus profit' $\leq$ profit'. The above observations lead to the following inequalities:

$$\frac{\text{profit}^0}{\text{profit}^*} = 1 - \frac{\text{cost}^* - \text{cost}'}{\text{profit}^*} \geq 1 - 0.06 \frac{\text{cost}^*}{\text{profit}^*}.$$
Therefore, if the supply chain has a relatively large profit margin (corresponding to a relatively large value of profit/cost), we can expect the profit gap caused by the power-of-two restriction to be quite small. Conversely, if the profit margin is small, the profit gap can be large. Nevertheless, power-of-two policies represent the state-of-the-art solution for the kind of replenishment problem considered here.

Recall from Roundy (1985) that under a power-of-two policy the long-run average (system-wide) setup and holding costs are given by

\[
\frac{K_0}{T_0} + \sum_{i=1}^{N} \left[ \frac{K_i}{T_i} + \frac{1}{2} h_0 d_i(p_i) \max \{T_0, T_i\} + \frac{1}{2} h_i d_i(p_i) T_i \right].
\]

(2)

Treating \( \{d_i, i = 1, \ldots, N\} \) as the decision variables, along with the intervals \( \{T_i, i = 0, \ldots, N\} \), we obtain the following expression for the system-wide profits \( \Pi \):

\[
\Pi = -\frac{K_0}{T_0} + \sum_{i=1}^{N} G_i(d_i, T_i, T_0),
\]

(3)

where

\[
G_i(d_i, T_i, T_0) = (p_i d_i - c_0 - c_i) d_i + \Psi(d_i) - \frac{K_i}{T_i} - \frac{1}{2} h_0 d_i \max \{T_0, T_i\} - \frac{1}{2} h_i d_i T_i.
\]

The problem thus reduces to determining a power-of-two vector \( T \) satisfying Equation (1) and a vector of demand rates \( d \) achieving the maximum in Equation (3). Let \( (T_0^*, \ldots, T_N^*) \) and \( (d_1^*, \ldots, d_N^*) \) be the optimal solution, and \( \Pi^* \) the maximum channel profits.

Below, we develop an efficient algorithm to compute an optimal pair \((T, d)\) maximizing channel-wide profits. We first need the following preliminary results. For any fixed vector of demand rates \( d \), define for \( i = 0, \ldots, N \)

\[
\hat{T}_i(d) = \text{the replenishment interval for facility } i \text{ which achieves the unconstrained minimum of Equation (2), i.e., without consideration of the constraints in Equation (1).} \text{(Roundy 1985 establishes the existence of a unique minimum of Equation (2) for any given } d.)
\]

\[
T_i^*(d) = \text{the smallest replenishment interval for facility } i \text{ which achieves the minimum of Equation (2) subject to the power-of-two constraints (1).}
\]

It follows from Roundy (1985) that \( T_i^*(d) \) is obtained from \( \hat{T}_i(d) \) by rounding the latter to the power-of-two multiple of \( T_i \) that is closest in the relative sense. More formally, let \( m = \lceil \log_2(\hat{T}_i(d)/T_i) \rceil \), where \( \lceil y \rceil \) is the largest integer less than or equal to \( y \). Then, by the definition of \( m, 2^n T_L \leq \hat{T}_i(d) < 2^{n+1} T_L \), and

\[
T_i^*(d) = \begin{cases} 
2^n T_L & \text{if } \hat{T}_i(d) \leq \sqrt{2^{2^n} T_L}, \\
2^{n+1} T_L & \text{otherwise.}
\end{cases}
\]

In the sequel, "rounding" refers to the above procedure.

Similarly, for any fixed vector of power-of-two intervals \( T = (T_0, T_1, \ldots, T_N) \), let \( d_i^*(T) = \text{the demand rate for retailer } i \) which achieves the maximum of Equation (3), \( i = 1, \ldots, N \).

The following lemma identifies some useful monotonicity properties. We say a function \( f(x) \) is decreasing in the \( N \)-dimensional vector \( x \) if \( f(x') \geq f(x'') \) for all \( x' \) and \( x'' \) with \( x_i' \leq x_i'' \) for \( i = 1, \ldots, N \).

**Lemma 1.** (a) \( T_i^*(d) \) is decreasing in \( d \) for \( i = 0, \ldots, N \).

(b) \( d_i^*(T) \) is decreasing in \( T \) for \( i = 1, \ldots, N \).

**Proof.** (a) Observe first that the rounding procedure transforming \( \hat{T} \) to \( T^* \) is order-preserving, i.e., if \( \hat{T}_i(d^1) \leq \hat{T}_i(d^2) \) for some demand rate vectors \( d^1 \) and \( d^2 \) and some \( i = 0, \ldots, N \) then \( T_i^*(d^1) \leq T_i^*(d^2) \). It thus suffices to prove that \( \hat{T}(d) \) is decreasing in \( d \).

We first derive an equivalent formulation for Equation (2). Introduce auxiliary variables \( T_{0i} = \max \{T_0, T_i\}, i = 1, \ldots, N \). The problem of minimizing Equation (2) is thus equivalent to the following convex program:

\[
(\text{P}) \min_{T_i, T_{0i}} \sum_{i=0}^{N} \frac{K_i}{T_i} + \frac{1}{2} \sum_{i=0}^{N} \frac{1}{2} h_0 d_i T_{0i} + \frac{1}{2} \sum_{i=1}^{N} h_i d_i T_i,
\]

s.t. \( T_{0i} \geq T_0, \quad i = 1, \ldots, N \) \hspace{1cm} (5)

\( T_{0i} \geq T_i, \quad i = 1, \ldots, N \) \hspace{1cm} (6)

(Because for any \( i = 1, \ldots, N \) the objective (5) is increasing in \( T_{0i} \) in an optimal solution to (P) at least one of the two constraints is binding, i.e., \( T_{0i} = \max \{T_0, T_i\} \).

Next, substitute the \( T_i \) and \( T_{0i} \) variables by \( x_i = T_i^{-1} \) and \( x_{0i} = T_{0i}^{-1} \).

\[
(\text{P}) \min_{x_i, x_{0i}} f(x, d) = \sum_{i=0}^{N} K_i x_i + \sum_{i=1}^{N} \frac{1}{2} h_0 d_i / x_{0i}
\]

\[+ \sum_{i=1}^{N} \frac{1}{2} h_i d_i / x_i, \]

s.t. \( x_{0i} \leq x_i, \quad i = 1, \ldots, N \) \hspace{1cm} (9)

\( x_{0i} \leq x_{0i}, \quad i = 1, \ldots, N \). \hspace{1cm} (10)

Let:

\[ S = \{x = (x_0, x_1, \ldots, x_N, x_{01}, \ldots, x_{0N}) \in R^{2N+1} : x \geq 0 \}
\]

satisfies (9) and (10)),

\[ D = \{d \in R^N : d \geq 0\} \]

\( x_i^* = \text{the vector } x \in S \text{ which achieves the minimum in Equation (8). (Because } \hat{T}(d) \text{ is the unique minimizer of (2) for fixed } d, x_i^* \text{ is uniquely determined as well.)} \)

Note that \( D \) is a partially ordered set. Also, \( S \) is a lattice, since it is a partially ordered set and it can be shown that if \( x, y \in S \) then \( \min \{x, y\} \in S \) and \( \max \{x, y\} \in S \). \( \min \{x, y\} \) is the vector of component-wise minima and
max\{x, y\} the vector of component-wise maxima.) We verify that min\{x, y\} ∈ S. (The proof that max\{x, y\} ∈ S is analogous.) Because x ∈ S, for all i = 1, . . . , N, x0 ≥ x0 ≥ max\{x0, y0\} and y ∈ S, y0 ≥ y0 ≥ max\{x0, y0\}. Thus, min\{x0, y0\} ≥ min\{x0, y0\}, verifying Equation (9). The proof that min\{x, y\} satisfies Equation (10) as well as analogous.

The fact that x0 is increasing in d ∈ D and hence Lemma 1 follows from Theorem 6 in Topkis (1978) by verifying that:

(i) \( f(x, d) \) is submodular in x ∈ S for each d ∈ D
(ii) \( f(x, d) \) has antitone differences in (x, d) on S × D.

Part (i) is immediate because f(x, d) is in fact a modular, i.e., a separable function in the x-variables. To verify (ii), we need to show that for all x1 ≤ x2 and d1 ≤ d2:

\[
 f(x2, d2) - f(x1, d2) ≤ f(x2, d1) - f(x1, d1).
\]

Note that the corresponding inequality holds term by term in Equation (8).

(b) Note first that for fixed T, the profit function in Equation (3) is separable in d, i = 1, . . . , N. More specifically, di is chosen to maximize

\[
(p_i(d_i) - c_i) di - \Psi(d_i) - \left( \frac{1}{2} h_i \max\{T_0, T_i\} + \frac{1}{2} h_i T_i \right) d_i,
\]

which can be written as

\[
 r_i(d_i) - \tilde{c}_i d_i,
\]

where \( r_i(d_i) = (p_i(d_i) - c_i) d_i - \Psi(d_i) \) and \( \tilde{c}_i = \frac{1}{2} h_i \max\{T_0, T_i\} + \frac{1}{2} h_i T_i \), which is clearly increasing in T. Therefore, it suffices to show that the value of d that maximizes the value of Equation (11) is decreasing in \( \tilde{c}_i \). Take any \( \tilde{c}_i < \tilde{c}_i \). Let \( d_i \) and \( \tilde{d}_i \) be the corresponding optimal values of di. Note that for any d > \( \tilde{d}_i \), we have

\[
 r_i(d_i) - \tilde{c}_i \tilde{d}_i \geq r_i(d) - \tilde{c}_i d_i
\]

which is equivalent to

\[
 r_i(d) - \tilde{c}_i d_i > r_i(d) - \tilde{c}_i \tilde{d}_i.
\]

Thus \( \tilde{d}_i \) ≤ \( d_i \).

The above result is quite intuitive. We can interpret \( \tilde{c}_i \) as the "imputed" marginal cost for retailer i. As the marginal cost increases, the retailer tends to charge a higher price and sell less.

We now derive bounds on \{T∗(d), i = 0, . . . , N\} that apply to all achievable demand vectors d. Let \( d_i^{\max} = d_i(p_i^{\max}) \) and \( d_i^{\min} = d_i(p_i^{\min}) \) for i = 1, . . . , N. Define

\[
 \tau_i = \sqrt{\frac{2K_i}{h_i d_i^{\max}}} \quad {\text{and}} \quad \tau_i = \sqrt{\frac{2K_i}{h_i d_i^{\min}}} \quad {\text{for}} \quad i = 1, . . . , N,
\]

and

\[
 \tau_0 = \min\{\tau_i, i = 1, . . . , N\} \quad {\text{and}} \quad \tau_0 = \max\{\tau_i, i = 1, . . . , N; z\},
\]

where \( z = \sqrt{\frac{2K_0}{h_0 \sum_{i=1}^{N} d_i^{\min}}} \). For i = 0, 1, . . . , N, let \( T_i \) and \( \tilde{T}_i \) be the rounded values of \( \tau_i \) and \( \tau_i \), respectively.

**Lemma 2.** \( \tilde{T}_i \leq T_i(d) \leq \tilde{T}_i, i = 0, 1, . . . , N \).

**Proof.** Fix the demand vector d. Thus the revenue is fixed, and profit maximization reduces to minimizing the cost function in Equation (2). Take any i = 1, . . . , N. Because

\[
 \sqrt{\frac{2K_i}{h_i d_i}} \leq \tilde{T}_i(d) \leq \sqrt{\frac{2K_i}{h_i d_i}}.
\]

(see, e.g., Roundy 1985) and \( d_i^{\min} \leq d_i \leq d_i^{\max} \), we have \( \tau_i \leq \tilde{T}_i(d) \leq \tau_i \). The lemma follows because rounding (to a power-of-two number) is order-preserving. We next prove the lemma for i = 0. Because \( \tilde{T}_0(d) \geq \min_{i=1, \ldots, N} \tilde{T}_i(d) \geq \tau_0 \) (see, e.g., Roundy 1985 for the first inequality), the lower bound on \( T_0(d) \) follows. To prove the upper bound, suppose \( T_i = T_i(d) \) for i = 1, . . . , N. For \( T_0 \geq \max\{\tilde{T}_i, i = 1, . . . , N\} \) which implies \( T_0 = T_0 \) for i = 1, . . . , N, the only part in Equation (2) that depends on \( T_0 \) is

\[
 \frac{K_0}{2} + \frac{1}{2} h_0 \sum_{i=1}^{N} d_i T_0.
\]

which is convex and its unconstrained minimum is x = \( \sqrt{\frac{2K_0}{h_0 \sum_{i=1}^{N} d_i^{\min}}} \). If z ≤ \( \max\{\tilde{T}_i, i = 1, . . . , N\} \) then Equation (12) is increasing in \( T_0 \) over the range \( T_0 \geq \max\{\tilde{T}_i, i = 1, . . . , N\} \), in which case \( T_0 \) is the only part in Equation (2) that depends on \( T_0 \).

We are now ready to present an algorithm for solving the joint optimization problem. Recall from Equation (3) that for a given \( T_0 \), the channel-wide profit function is separable in \( (d_i, T_i) \) for i = 1, . . . , N. The algorithm therefore has two loops. The first loop iterates over all values of \( T_0 \) within the bounds identified in Lemma 2. Within this loop, for each retailer i and each value of \( T_i \) in \[ \tilde{T}_i, \tilde{T}_i \], we optimize the profit function \( G_i(d_i, T_i, T_0) \) over \( d_i \). Let \( n_i \) be the number of power-of-two values between \( \tilde{T}_i \) and \( \tilde{T}_i, i = 0, 1, . . . , N \). Initially, the optimization over \( d_i \) is over the range \( [d_i^{\min}, d_i^{\max}] \). For each retailer i, we keep track of \( n_i \) upper bounds on \( d_i \), \( d_{i,j} \) j = 1, . . . , \( n_i \), where \( d_{i,j} \) applies when \( T_i \) is set at its jth smallest value. By Lemma 1’s monotonicity property, we have \( d_{i,j} \geq d_{i,j+1} \) for all j = 1, . . . , \( n_i - 1 \). By the same monotonicity property, any upper bound \( d_{i,j} \) which applies for a given value of \( T_0 \), continues to apply when \( T_0 \) is replaced by a larger value, in particular as \( T_0 \) is replaced by 2\( T_0 \) in the outer loop. As a result, when executing the (j + 1)st iteration of the inner loop for retailer i, we set \( d_{i,j+1} \) equal to \( d_{i,j} \) or the current value of \( d_{i,j+1} \), whichever is smaller.
ALGORITHM I.

Initialize: \( \Pi^* := 0; \)

For \( i := 1, \ldots, N \) do \( d_i := d_i^{\text{min}}; \)

For \( i := 1, \ldots, N \) and \( j := 1, \ldots, n_i \) do \( \tilde{d}_{ij} := d_i^{\text{max}}, \)

\( T_0 := \tilde{\tau}_0. \)

Step 0: \( \Pi := -K_0 / T_0^6; \)

Step 1: For \( i := 1, \ldots, N \) do begin

\( j := 1; T_i := \tilde{\tau}_i; \Pi := -\infty; \)

Iterative Step:

Determine the smallest maximum \( x \) of \( G_i(x, T_i, T_0) \) on the interval \([\tilde{d}_{ij}, \tilde{d}_{ij}]; \)

\( \tilde{d}_{ij} := x; \)

If \( G_i(x, T_i, T_0) > \Pi \) then begin

\( T_0^6 := T_i; \)

\( d_i^6 := x; \)

\( \Pi := G_i(x, T_i, T_0); \)

end;

If \( j < n_i \) then begin

\( \tilde{d}_{ij+1} := \min \{ \tilde{d}_{ij}, \tilde{d}_{ij}; \}

\( j := j + 1; T_i := 2T_i; \) and go to Iterative Step;

end;

\( \Pi := \Pi + \Pi_i, \)

end;

If \( \Pi > \Pi^* \) then begin

\( \Pi^* := \Pi; \)

\( T_0^6 := T_0^6; \)

For \( i := 1, \ldots, N \) do \( T_i^* := T_0^6 \) and \( d_i^* := d_i^6; \)

end;

\( T_0 := 2T_0^6; \)

If \( T_0 \leq \tilde{\tau}_0 \) then go to Step 0.

The computational effort of Algorithm I consists primarily of the multiple optimizations of the single-variable function \( G_i(x, T_i, T_0) \) over \( x \). These functions are often strictly concave, e.g., in the absence of account management costs \( (\Psi(\cdot) = 0) \) or when the function \( \Psi(\cdot) \) is of the fixed-plus-linear type. In this case, maximizing \( G_i \) is entirely straightforward, e.g., by bisection or by finding the unique root of its derivative, while a more general single-variable optimization technique is required in case \( G_i \) fails to be concave. As is the case with all optimization problems involving general, nonlinear functions, we refer to the single-variable maximization procedure optimizing \( G_i \) as an "oracle" and we assess the complexity of the algorithm in terms of the number of oracle calls required. Note that the outer loop in Algorithm I is traversed \( n_0 \) times and the inner loop \( \sum_{i=1}^{N} n_i \) times. Each execution of the inner loop involves a single call to the oracle and a constant number of multiplications and comparisons.

The overall effort is thus given by \( n_0 \sum_{i=1}^{N} n_i \) oracle calls and \( O(n_0 \sum_{i=1}^{N} n_i) \) elementary operations. The expression \( n_0 \sum_{i=1}^{N} n_i \) fails to be polynomial for completely arbitrary parameters. However, as has been frequently observed for other algorithms involving power-of-two policies (see, e.g., Federgruen and Zheng 1995), it would be exceedingly rare to have an instance where more than (say) eight distinct power-of-two values need to be investigated for any of the facilities, because otherwise, the largest considered replenishment interval would be more than 500 times the smallest considered interval. Thus, for all practical purposes, the complexity of Algorithm I can be conservatively bounded by \( 64N \) calls to the oracle and \( O(N) \) elementary operations. This bound does not incorporate the very significant shortcuts achieved by the updating of the upper bounds \( \{d_{ij}\} \) described above.

5. THE SILO SOLUTION

The silo solution assumes that the supplier and the retailers are part of a single firm, where pricing and replenishment decisions are delegated to separate departments. More specifically, the marketing department first determines the demand rates (or the retail prices) to maximize the firm’s gross profits \( \sum_{i=1}^{N} (p_i(d_i) - c_i - c_i) d_i \). Let \( d_i^* \) be the optimal demand rate for retailer \( i, i = 1, \ldots, N \). Given these demand rates, the production/distribution department then determines the replenishment intervals to minimize the firm’s logistics costs. This problem is identical to the problem Roundy (1985) has solved; thus one can use his \( O(N \log N) \) algorithm to identify an optimal power-of-two policy. (An alternative implementation by Queyranne 1987 results in a linear time algorithm.) Alternatively, one can use a modified version of Algorithm I. First, redefine

\[ \tau_i^* = \sqrt{\frac{2K_i}{h_i d_i^*}} \quad \text{and} \quad \tau_i = \sqrt{\frac{2K_i}{h_i d_i^*}} \quad i = 1, \ldots, N, \]

and \( z = \sqrt{2K_0/(h_0 \sum_{i=1}^{N} d_i^*)} \). The rounded values of \( \tau_i^* \) and \( \tau_i \) are again the lower and upper bounds on \( T_i^*(d_i^*), i = 0, 1, \ldots, N, \) where \( d_i^* = (d_1^*, \ldots, d_N^*). \) With these updated bounds, one can use Algorithm I to find the silo solution by skipping the step optimizing over the demand rates. Therefore, this algorithm searches over all possible values of the replenishment intervals. Because no oracle calls are required for the silo solution, the complexity of this modified version of Algorithm I is \( O(n_0 \sum_{i=1}^{N} n_i) \).

As argued above, in most practical instances \( n_i \leq 8 \) for \( i = 0, 1, \ldots, N, \) resulting in a linear time algorithm.

6. THE STACKELBERG GAME

In this section, we assume that the supplier and the retailers are independent firms, each of which maximizes its own profits. The supplier, in anticipation of the retailers’ reactions, chooses a constant wholesale price and a replenishment strategy for itself with the objective of maximizing its own profits. The retailers take the wholesale price
as given and each of them maximizes its individual profits by charging an optimal retail price and following an optimal replenishment strategy. In other words, the players in the distribution channel play a Stackelberg game, with the supplier as the leader, and the retailers as followers. This *modus operandi* represents many traditional distribution channels in which the channel members fail to coordinate their decisions. Nevertheless, no solution procedure has been proposed to solve this decentralized system. The objective of this section is to develop an efficient algorithm for the Stackelberg game. In addition to providing the optimal policy parameters for such decentralized systems, the algorithm enables us to quantify the relative benefits of the centralized solution.

Every decentralized supply chain requires an upfront specification of a contract, i.e., a set of ground rules for the commercial interactions between the different parties involved. Such a contract may involve the specification of a pricing rule, the commitment to deliver in whole or in part (possibly within a specified leadtime), return policies, restrictions on the times at which orders may be placed and delivered (e.g., daily, every Tuesday or Friday, etc.), among others. In our model, the contract consists of the following provisions. We assume, without loss of generality, that the system is without inventory at time 0. The supplier commits himself to satisfy all retailer orders in their entirety without backlogging and to deliver them with a fixed leadtime, which is without loss of generality normalized to be zero. The supplier is responsible for all (fixed and variable) costs associated with its orders, the holding costs incurred for its own inventories as well as the account management costs. The retailers are responsible for all costs incurred for orders they place with the supplier and all inventories they carry at their sites. (Later, we will show that the same analysis goes through if the supplier incurs a fixed order-processing cost for each retailer order.) Finally, all channel members place their replenishment orders at epochs chosen from the discrete set of power-of-two values \(\{2^m T_0 : m = \ldots, -2, -1, 0, 1, 2, \ldots\}\). This restriction results in considerable simplifications and efficiencies for the supplier at minimal expense to the retailers. As discussed before and known since Brown (1959), the latter’s inventory related costs increase by at most 6%. On the other hand, if the retailers were able to choose their replenishment intervals without any restriction, these would be set according to the EOQ formula and the resulting order stream for the supplier would be highly nonstationary and in general without a periodically repeating pattern. This rather intricate order pattern from the retailers would certainly represent a difficult managerial task for the supplier. Moreover, no satisfactory solution exists for the resulting inventory problem for the supplier and the supplier’s costs are significantly larger than if orders come in according to power-of-two patterns. It can therefore be expected that the supply-chain-wide profits in the Stackelberg game with the power-of-two restriction are larger than those in a version of the game without such restrictions. If the restriction to

power-of-two intervals is difficult to enforce, the supplier may offer an annual rebate to the retailers equal to the modest increase in their inventory and setup costs resulting from the interval restriction. (Such rebates are most easily computed and clearly do not affect the supply-chain-wide profits arising from the Stackelberg game.)

We now develop an efficient algorithm for solving the above Stackelberg game. We begin by considering retailer \(i\)’s problem. Let \(w\) be the wholesale price and \(T_0\) the supplier’s replenishment interval. Given these decisions, retailer \(i\) sets his own retail price \(p_i\) and replenishment interval \(T_i\) to maximize his own profit. Therefore, retailer \(i\)’s problem can be formulated as

\[
\max_{d_i, T_i \mid w} \pi_i(d_i, T_i \mid w) \overset{\text{def}}{=} (p_i(d_i) - c_i - w)d_i - \frac{K_i}{T_i} - \frac{1}{2}h_i d_i T_i,
\]

\(i = 1, \ldots, N,\) (13)

where \(T_i\) only takes on power-of-two values and \(d_i\) is confined to the finite interval \([d_i^{\text{min}}, d_i^{\text{max}}]\). (As before, we use \(d_i\) as the decision variable. The retail price is again uniquely determined through the inverse demand function.) Note that the profit measure in Equation (13) depends only on a single parameter specified by the supplier, i.e., the (constant) wholesale price \(w\). In particular, the profit measure is independent of \(T_0\).

We adopt the convention that when given a choice between two pairs of decision variables \((d_i, T_i)\) with different \(T_i\)-but identical profit values, the retailer will always choose the pair with the lower \(T_i\) value. Note that for any fixed \(T_i\) value, the corresponding optimal \(d_i\) value is uniquely determined since the profit function \(\pi_i\) is strictly concave in \(d_i\). Furthermore, this optimal \(d_i\) value is strictly decreasing in \(T_i\), as follows from the first-order condition. Therefore, if there are multiple optimal solutions to the retailer’s problem, they must have distinct \(d_i\) and \(T_i\) values. Thus, given the above convention, let \((d_i(w), T_i(w))\) be the (uniquely) chosen optimal solution in Equation (13).

Just like for any given \(T_i\), there is a unique corresponding optimal \(d_i\) value, the converse holds as well, given the above convention. For any fixed \(d_i\), the chosen optimal corresponding value of \(T_i\) is obtained by rounding the unconstrained minimum point \(\sqrt{2K_i/h_i d_i}\) of the EOQ cost function

\[
\min_{T_i} \frac{K_i}{T_i} + \frac{1}{2}h_i d_i T_i
\]

to the power-of-two value that is closest in the following relative sense. Let \(t_m = 2^m T_0, m\) integer. The optimal value of \(T_i\) equals \(t_m\) if and only if

\[
\frac{t_m}{\sqrt{2}} \leq \sqrt{\frac{2K_i}{h_i d_i}} \leq \frac{t_m}{\sqrt{2}} \quad \text{or} \quad d_i^m \leq d_i < d_i^{m+1},
\]

(14)

where \(d_i^n = K_i/h_i t_i^n\) for any integer \(n\). Because \(d_i\) is confined to a finite interval \([d_i^{\text{min}}, d_i^{\text{max}}]\), there are only a finite number of \(T_i\) values that need to be considered. Let \(n_i\) be the number of such \(T_i\) values.
Lemma 3. For \(i = 1, \ldots, N\), we have (i) \(d_i(w)\) is decreasing in \(w\), (ii) \(T_i(w)\) is an increasing step function of \(w\) with finitely many steps, and (iii) \(T_i(w)\) is still an increasing step function of \(w\) if \(T_i\) is confined to any finite subset of power-of-two values.

Proof. The rounding procedure in Equation (14) implies that the optimal value of \(T_i\) is uniquely determined by \(d_i\) and is decreasing in \(d_i\). Moreover, as noted above, there are only a finite number of \(T_i\) values that are potentially optimal and thus the function \(T_i(w)\) takes on finitely many values. Therefore, (i) implies (ii).

To show (i), take any \(w' < w\). Note that
\[
\pi_i(d_i, T_i|w') - \pi_i(d_i, T_i|w) = (w' - w)d_i.
\]
\[
\pi(d_i', T_i'|w') - \pi_i(d_i', T_i'|w) = (w' - w)d_i
\]
\[
= \pi_i(d_i', T_i'|w'),
\]
where the inequality follows because \(\pi_i(d_i', T_i'|w') \geq \pi_i(d_i, T_i'|w')\) by the definition of \(d_i'\) and \(T_i'\), and \(d_i > d_i'\). Therefore, \(d_i(w') \leq d_i\).

Now suppose \(T_i\) is confined to a finite subset of power-of-two values. It is clear that in this case, the optimal value of \(T_i\) is still uniquely determined by \(d_i\) and decreases as \(d_i\) increases. (iii) thus follows from (i).

Because \(T_i(w)\) is an increasing step function of \(w\) with finitely many steps (Lemma 3), there exist a finite number of breakpoints for the wholesale price so that \(T_i(w)\) remains constant between any two consecutive breakpoints. We now show how these breakpoints can be determined efficiently. First, note that we can restrict \(w\) to a finite interval \((a, b)\) with \(a = c_0\) and \(b = \max\{p_i^{\text{max}}, i = 1, \ldots, N\}\). Let \(d_i(w, T_i)\) be the optimal demand rate for retailer \(i\) as a function of \(w\) and \(T_i\). Let \(\pi_i(w|T_i)\) denote \(\pi_i(d(w, T_i), T_i|w)\). (Clearly, \(\pi_i(w|T_i)\) is a decreasing function of \(w\) for any given \(T_i\)). Now take any pair of reorder intervals \(T_i'\) and \(T_i''\) with \(T_i' < T_i''\) and suppose the retailer chooses between these two intervals. By Lemma 3 (iii), one of the following three cases prevails: (1) the retailer chooses \(T_i'\) for all \(w \in (a, b)\); (2) the retailer chooses \(T_i''\) for all \(w \in (a, b)\); and (3) there exists a point \(w_0\) such that \(T_i'\) is chosen to the left of \(w_0\) and \(T_i''\) is chosen to the right of \(w_0\). Thus in all three cases, there exists a unique cross point \(\hat{w}\) such that at this point, the retailer's choice switches permanently from \(T_i'\) to \(T_i''\): in case (1), \(\hat{w} = b\); in case (2), \(\hat{w} = a\); in case (3), \(\hat{w} = w_0\). Note that \(\hat{w}\) can be determined easily by bissection, starting with the evaluation of the sign of \(\pi_i(w|T_i') - \pi_i(w|T_i'')\) for \(w = a\) and \(w = b\).

Recall that there are \(n_i\) distinct power-of-two values of \(T_i\) that are potentially optimal. We construct the step function \(T_i(w)\) by an iterative procedure. In the first iteration, we consider only the two smallest feasible power-of-two values, in the \(k\)th iteration the \((k + 1)\)th smallest values until reaching the \(n_i - 1\)st iteration in which all feasible power-of-two values are considered. By Lemma 3 (iii), if only the \(k\) smallest values of \(T_i\) are considered, the function \(T_i(w)\) is still an increasing step function and can therefore be characterized by a possibly empty list of breakpoints \(W_i = \{w_{i(1)}, \ldots, w_{i(k)}\}\) with \(0 \leq L < k\) and an associated list of indices \(J_i = \{j_{i(1)}, \ldots, j_{i(k)}\}\) such that for all \(l = 1, \ldots, L, T_i(w)\) equals the \(j_{i(l)}\)th smallest power-of-two value on the interval \((w_{i(l-1)}, w_{i(l)})\) (with the convention \(w_{i(0)} = a\)), while \(T_i(w)\) equals the \(k\)th smallest power-of-two value on \((w_{i(L)}, b)\). To simplify the exposition below, we write \(T_i = k\) when \(T_i\) equals the \(k\)th smallest power-of-two value. After the first iteration, both lists contain one element with \(w_{i(1)}\) the cross point of the profit functions associated with the two smallest power-of-two values. Assume now that the \(k - 1\)th iteration has been completed. We describe how the two lists can be updated efficiently in the \(k\)th iteration. First, compute the cross point \(\hat{w}\) of the profit functions of \(T_i = k\) and \(T_i = k + 1\) and compare this with \(w_{i(L)}\). Case I: \(w > w_{i(L)}\); set \(L = L + 1\), add \(w_{i(L)} = \hat{w}\) at the back of list \(W_i\), and add \(k\) at the back of \(J_i\). By the cross point definition, \(T_i = k + 1\) dominates \(T_i = k\) for \(w > \hat{w}\) while the list \(W_i\) indicates that \(T_i = k\) dominates all smaller power-of-two values for \(w > \hat{w} > w_{i(L)}\). Thus \(T_i = k + 1\) dominates the \(k\) smallest values for \(w > \hat{w}\), while one of the \(k\) smallest values dominates for \(w \leq \hat{w}\). Indeed, for all \(l = 1, \ldots, L, T_i = j_{i(l)}\) continues to dominate on the interval \((w_{i(l-1)}, w_{i(l)})\). Case II: \(w \leq w_{i(L)}\). In this case, \(T_i = k\) is dominated by one of the other values throughout the interval \((a, b)\) and can therefore be eliminated from consideration. Thus, delete \(w_{i(L)}\) and \(j_{i(L)}\) from the lists. The situation is now equivalent to that of iteration \(J_i\), assuming all values \(j_{i(L)} + 1, \ldots, k\) are eliminated from consideration. Thus, as before, compute the cross point \(\hat{w}\) of the profit functions \(T_i = j_{i(L)}\) and \(T_i = k + 1\). Let \(L := L - 1\), and compare the cross point with the new \(w_{i(L)}\) value. Repeat this process, eliminating the elements from the back of the lists \(W_i\) and \(J_i\) and reducing the value of \(L\) by one until either Case I emerges or the lists are emptied out. After the \(n_i\)th iteration, the lists \(W_i\) and \(J_i\), both of length \(L \leq n_i - 1\), fully characterize the step function \(T_i(w)\) for \(a < w < b\).

Note that the above algorithm bears similarity to other list-based procedures used to solve dynamic programs of special structure, see e.g., Aggarwal et al. (1987), Hirshberg and Larmore (1987), Wilber (1988), Miller and Myers (1988), Galil and Giancarlo (1989), and Federgruen and Tzur (1991).

Let \(\bigcup_{i=1}^{\infty} W_i = \{w_0, w_1, \ldots, w_m\}; a = w_0 < w_1 < \cdots < w_m < w_{m+1} = b\) for some integer \(M \leq \sum_{i=1}^{N} (n_i - 1)\) denote the complete list of all of the retailers' breakpoints. For all \(w \in [w_m, w_{m+1}]\), the optimal replenishment intervals for the retailers remain fixed, \(m = 0, \ldots, M\). Let \(T_i^m\) be retailer \(i\)'s optimal replenishment interval when \(w \in [w_m, w_{m+1}]\), for \(m = 0, \ldots, M\) and \(i = 1, \ldots, N\). Note that on the
Equation (16) when w is restricted to the mth interval, [\( w_m, w_{m+1} \)]. Note that
\[
\pi_0(w, T_0) = \sum_{i=1}^{N} \left( \left( w - c_i - \frac{1}{2} h_0 [T_0 - T_i^w]^+ \right) d_i(w) - \Psi(d_i(w)) \right) \frac{K_0}{T_0}, \quad w_m \leq w \leq w_{m+1},
\]
We next show that only a limited number of power-of-two values need to be considered for \( T_0 \). Define
\[
S^m = \min \{ T_i^m, i = 1, \ldots, N \},
\]
\[
U^m = \max \{ T_i^w, i = 1, \ldots, N \},
\]
\[
T^m = \text{smallest power-of-two value greater than}
\]
or equal to \( \sqrt{2K_0 / (h_0 \sum_{i=1}^{N} d_i(w_{m+1}))} \).

**Lemma 5.** \( S^m \leq T^m \leq \max\{T^m, U^m\}, m = 0, \ldots, M \).

**Proof.** The lower bound follows because \( \pi_0(w, T_0) \) is increasing in \( T_0 \) for \( T_0 < S^m \). To prove the upper bound, first note that for \( T_0 > \max\{T^m, U^m\} \)
\[
\pi_0(w, T_0) = \sum_{i=1}^{N} \left( (w - c_i)d_i(w) - \Psi(d_i(w)) + \frac{1}{2} h_0 d_i(w) T_i^w \right)
\]
\[- \left[ \frac{1}{2} h_0 \sum_{i=1}^{N} d_i(w) T_0^+ \frac{K_0}{T_0} \right].
\]
Also note that \( \sqrt{2K_0 / h_0 \sum_{i=1}^{N} d_i(w) T_0^+} \left[ \frac{K_0}{T_0} \right] \) is the minimum point of \( \sqrt{2K_0 / h_0 \sum_{i=1}^{N} d_i(w) T_0^+} \) and that for any \( w \in [w_m, w_{m+1}] \),
\[
\sqrt{2K_0 / h_0 \sum_{i=1}^{N} d_i(w_{m+1})} \leq \frac{2K_0}{h_0 \sum_{i=1}^{N} d_i(w_{m+1})} \leq T^m,
\]
where the first inequality follows from Lemma 3, \( d_i(w) \geq d_i(w_{m+1}) \) for all \( w \in [w_m, w_{m+1}] \). Therefore, \( \pi_0(w, T_0) \) is decreasing in \( T_0 \) over the range \( T_0 > \max\{T^m, U^m\} \). The upper bound follows.

The supplier's problem on the interval \([w_m, w_{m+1}], m = 0, \ldots, M\), can be solved by fixing \( T_0 \) at each of the power-of-two values between \( S^m \) and \( \max\{T^m, U^m\} \) and maximizing \( \pi_0(w, T_0) \) over \( w \). Later, we will show that for an important special case, \( \pi_0(w, T_0) \) is concave in \( w \) over \([w_m, w_{m+1}]\), allowing for a simple maximization.

We are now ready to present an algorithm for solving the Stackelberg game. The first step of the algorithm is to determine the breakpoints \( a = w_0, w_1, \ldots, w_M, w_{M+1} = b \) and the optimal reorder intervals \( T_i^w \) for all retailers \( i \) and all intervals \([w_m, w_{m+1}]\). The second step involves two loops. The outer loop iterates through the intervals \([w_m, w_{m+1}], m = 0, \ldots, M\). For each of these intervals, the inner loop investigates each of a finite number of distinct power-of-two values of \( T_0 \). For a given interval \([w_m, w_{m+1}]\) and a given value of \( T_0 \), we optimize \( \pi_0(w, T_0) \) over \( w \in [w_m, w_{m+1}] \). If \( \pi_0 \) is concave in \( w \), this optimization is
straightforward, e.g., by bisection. (Later, we will show that this is indeed the case when the demand functions are linear.) Otherwise, if \( \pi_0 \) fails to be concave in \( w \), then a more general single-variable optimization technique is required. When these two loops are completed, we have an optimal solution to the Stackelberg game.

**Algorithm II.**

**Step 1.** (Determine \( \{w_{m+1}\}_{n=0}^m \) with \( w_0 = a \) and \( w_{M+1} = b \) and \( T^m_i, i = 1, \ldots, N, m = 0, 1, \ldots, M \))

For \( i = 1, \ldots, N \) do

begin

Set \( w_{(1)} \) as the cross point of \( \pi_i(\cdot|T_i = 1) \) and \( \pi_i(\cdot|T_i = 2) \);

\( j_{(1)} := 1; W_i := \{w_{(1)}\}; J_i := \{j_{(1)}\}; L := 1; \hat{k} := 2; \)

For \( k = 2, \ldots, n_i - 1 \) do

begin

Set \( \hat{w} \) as the cross point of \( \pi_i(\cdot|T_i = \hat{k}) \)

and \( \pi_i(\cdot|T_i = k + 1) \);

While \( \hat{w} \leq w_{(k)} \) do

begin

\( \hat{k} := j_{(k)}; \) Eliminate \( j_{(k)} \) and \( w_{(k)} \) from \( J_i \) and \( W_i \);

Set \( \hat{w} \) as the cross point of \( \pi_i(\cdot|T_i = \hat{k}) \)

and \( \pi_i(\cdot|T_i = k + 1) \);

\( L := L - 1; \)

end;

\( L := L + 1; j_{(L)} := \hat{k}; w_{(L)} := \hat{w}; \) Add \( j_{(L)} \)

to \( J_i \), and \( w_{(L)} \) to \( W_i \);

\( \hat{k} := k + 1; \)

end;

end;

Merge the lists \( \{W_i, i = 1, \ldots, N\} \) into a single list \( W \).

**Step 2.** \( \Pi^0_0 := 0 \);

For \( m := 0, \ldots, M \) do

begin

For each power-of-two value of \( T_0 \) between \( S^m \) and \( \max\{T^m_i, U^m_i\} \) do

begin

Determine the maximum \( x \) of \( \pi_0(x, T_0) \) on the

interval \( [w_{m}, w_{m+1}] \);

If \( \pi_0(x, T_0) \geq \Pi^0_0 \) then

\( \Pi^0_0 := \pi_0(x, T_0); w^0 := x; T_0^0 := T_0; \)

For \( i = 1, \ldots, N \) do

\( d_i^m := d_i(x); T_i^m := T^m_i; \)

\( \Pi^0_i := \pi_i(d_i^m, T_i^m|x); \)

end;

end;

end;

We will measure the complexity of Algorithm II in terms of (i) elementary operations, (ii) the number of times a cross point needs to be computed of a pair of profit functions, and (iii) the number of times the continuous, single-variable function \( \pi_0(x, T_0) \) is optimized on one of the intervals \( [w_m, w_{m+1}] \). Recall that a cross point can be determined with simple bisection evaluating the sign of a difference function \( \pi_i(\cdot|T_i) - \pi_i(\cdot|T_j) \) for some pair of values \( T_i, T_j \). Fix \( i = 1, \ldots, N \). The effort to construct the lists \( W_i \) and \( J_i \) consists of \( n_i - 1 \) iterations. In each iteration, there is one cross point calculation followed by the addition of an element to each of the lists \( W_i \) and \( J_i \) and an increase of \( L \) (Case I) and possibly several cross point calculations resulting in eliminations of elements and a reduction of \( L \) (Case II). However, the total number of times Case II occurs across all \( (n_i - 1) \) iterations is at most \( n_i - 1 \) itself, since each such case is uniquely associated with one of the distinct power-of-two values. The entire effort to construct the lists \( W_i \) and \( J_i \) for a given \( i = 1, \ldots, N \) thus consists of \( O(n_i) \) elementary operations and at most \( 2(n_i - 1) \) cross point calculations. To construct the complete set of breakpoints across all \( i = 1, \ldots, N \), therefore involves \( O\left(\sum_{i=1}^N n_i\right) \) elementary operations and at most \( 2\sum_{i=1}^N (n_i - 1) \) cross point calculations. An additional \( O(N \log N + \sum_{i=1}^N n_i) \) operations is needed to merge the lists \( \{W_i, i = 1, \ldots, N\} \) into a single list.

To assess the complexity associated with the second step of Algorithm II, let \( n_0 \) be the number of power-of-two values between \( \min\{S^m : m = 0, \ldots, M\} \) and \( \max\{T^m_i, U^m_i : m = 0, \ldots, M\} \). Thus the outer loop of the second step of Algorithm II is traversed \( M \leq \sum_{i=1}^N (n_i - 1) \) times and the inner loop at most \( n_0 \) times for each value of \( m \). Each time this inner loop is executed, we maximize \( \pi_0(x, T_0) \) (a nonlinear, continuous, single-variable, closed-form function) over a finite interval of \( x \). Therefore, the second step of Algorithm II involves doing this maximization at most \( n_0 \sum_{i=1}^N (n_i - 1) \) times.

The overall effort for the algorithm thus consists of \( O(N \log N + \sum_{i=1}^N n_i) \) elementary operations, \( 2\sum_{i=1}^N (n_i - 1) \) cross-point calculations, and at most \( n_0 \sum_{i=1}^N (n_i - 1) \) maxima of a continuous, closed-form, single-variable function. As mentioned, practical problems have \( n_i \leq 8 (i = 0, 1, \ldots, N) \), resulting in an algorithm that requires at most \( 14N \) cross-point calculations, \( 56N \) single-variable function maximizations, as well as \( O(N \log N) \) elementary operations.

**Remark 1.** Suppose for each order by retailer \( i, i = 1, \ldots, N \), the supplier incurs a fixed order-processing cost \( K_i^i \) and the retailer incurs a fixed setup cost \( K_i^r \). In this case, the problem facing retailer \( i \) in the Stackelberg game is still characterized by Equation (13) with \( K_i \) replaced by \( K_i^r \). Clearly, this problem has the same structure as before. Therefore, it is still true that there exist breakpoints of the wholesale price \( \{w_m\}_{m=0}^{M+1} \), such that for all \( w \in [w_m, w_{m+1}] \), \( m = 0, \ldots, M \), the optimal replenishment intervals for the retailers remain fixed, i.e., \( T_i(w) = T_i^m, i = 1, \ldots, N \). Note that the supplier’s objective function is obtained by subtracting

\[
\sum_{i=1}^N K_i^r T_i(w)
\]
7. NUMERICAL EXAMPLES

This section presents numerical examples that were used to compare the centralized solution, the silo solution, and the Stackelberg solution.

We have evaluated two sets of examples, one with identical retailers and the other with nonidentical retailers. In all examples, we assumed $\Psi(d) = f + ed$ and $p_i(d_i) = a_i - b_i d_i$ for some positive constants $f, e, a_i$ and $b_i$, $i = 1, \ldots, N$. The base case for the first set of examples has the following parameters: $N = 10$, $K_0 = 500$, $h_0 = 5$, $c_0 = 10$, $f = 10$, $e = 1$, and $K_i = 10$, $h_i = 1$, $c_i = 1$, $a_i = 100$, $b_i = 20$ for $i = 1, \ldots, N$. The base case for the second set is the same except $c_i = 1 + i/10$, and $b_i = \delta(20 + i/10)$, $i = 1, \ldots, N$, with $\delta = 1$. All the examples in a set were obtained from the base case of the set by varying one parameter at a time.

Both Algorithm I and Algorithm II are very efficient, solving each instance with $N = 10$ retailers in a fraction of a second (less than 0.02 secs) on a PC with a Pentium 450-MHz processor. Because the computational effort is effectively linear in $N$ (under the above described assumptions), this implies that even instances with several hundred retailers can be solved in approximately one CPU second. Figures 1 and 2 summarize the results, where the percentage profit decrease is relative to the maximum profit.
channel profits (the joint optimum). Based on the results, we make the following observations.

- The system-wide profits decrease significantly in the Stackelberg game. In some cases, the supplier cannot ensure itself of a positive profit, even though a vertically integrated system (or an appropriately decentralized one, see Chen et al. 2001) results in healthy profits. In such settings, the supplier, in spite of being the price leader, is forced to withdraw from the market.

- Sequential (as opposed to joint) decision making can be quite costly, especially when the number of retailers is small or the price elasticity is large (a smaller value of \( a_i \) means larger price elasticity at any price level), or the slope of the inverse demand function \( (b_i) \) is large. (Because the demand function is \( d_i = (a_i - p_i)/b_i \), a larger value of \( b_i \) means a smaller market at any given price.) In fact, we have identified instances where sequential decision making results in negative profits even though the maximum channel profits are positive. (In Figures 1 and 2, negative channel-wide profits are replaced by zero profits because in such cases the channel members withdraw from the market.) Note that the percentage profit decrease for small values of \( b_i \) (or \( \delta \)) is consistent with an observation by Boyaci and Gallego (1997), who show that the relative profit decrease due to sequential decision making disappears as \( b_i \to 0 \).

- The economics literature (starting with Spengler 1950) has observed that wholesale prices arising in Stackelberg games exceed the marginal costs in a vertically integrated channel, resulting in higher retail prices and lower sales volumes. This phenomenon is usually referred to as double marginalization. We conjecture that this phenomenon is guaranteed to arise in our Stackelberg solution as well. While it is intrinsically difficult to prove this conjecture analytically, all our numerical results confirm the hypothesis. We observed in several instances that the sales volumes of the retailers in the Stackelberg solution are reduced by as much as 50% as compared to the centralized solution.

- It is often claimed that supply chain integration results in shorter replenishment cycle times, regarded by many as
a benefit in and of itself. For example, Eckstut and Tang’s (1996) recent survey of supply chain management practices in the pharmaceutical industry, observed that companies with explicit strategies to improve their supply chains have shorter cycle times both in manufacturing and replenishment as compared to others. Our results indicate that the mere integration or coordination of the decision making processes may fail to have the desired effect, by itself. As an example, in the base case with identical retailers, the retailers’ replenishment intervals are 50% shorter in the Stackelberg solution as compared to the centralized solution. On the other hand, shorter cycle times are likely to arise if a coordinated or integrated decision making framework in the supply chain is exploited to achieve reductions in setup costs and setup times. Continuing with the same example, supply chain coordination must be accompanied by a reduction of the fixed costs associated with retailer orders ($K_I$) of at least 25% to achieve shorter cycle times.

REFERENCES


—, —, M. Tzur. 1991. A simple forward algorithm to solve general dynamic lot sizing models with $n$ periods in $O(n \log n)$ or $O(n)$ Time. *Management Sci.* 37(8) 909–925.


—. 1998. Coordination of pricing and multi-period production across multiple constant priced goods. Working Paper, Case Western Reserve University, Cleveland, OH.


____. 1985. 98% effective integer-ratio lot-sizing for one
warehouse multi-retailer systems. Management Sci. 31
1416–1430.
Econom. 58 347–352.
Thomas, J. 1970. Price-production decisions with deterministic

Wagner, H., T. Whitin. 1958. Dynamic version of the economic
lot size model. Management Sci. 5 89–96.
Whitin, T. 1955. Inventory control and price theory. Management
Sci. 61–68.
Wilber, R. 1988. The concave least weight subsequence problem
Sci. 46 375–388.