Journal of Global Optimization **26:** 25–42, 2003.

Optimizing Supply Shortage Decisions in Base Stock Distribution Operations

JOSEPH GEUNES[1,*] and AMY Z. ZENG[2]
[1] *Department of Industrial and Systems Engineering, University of Florida, 303 Weil Hall, PO Box 116595, Gainesville, FL 32611, USA. *Corresponding author. (e-mail: geunes@ise.ufl.edu)*
[2] *Department of Management, Worcester Polytechnic Institute, Worcester, MA 01609, USA. (e-mail: azeng@wpi.edu)*

Abstract. This paper addresses policies and agreements between suppliers and customers for handling supply shortages in base-stock systems under uncertain demand. We investigate the impacts that backlogging and expediting decisions have on inventory and transportation costs in these systems and develop a model for deciding whether a supplier should completely backlog, completely expedite, or employ some combination of backlogging and expediting shortages. Our results indicate that practical cases exist where some combination of both expediting and backlogging supply shortages outperforms either completely expediting or backlogging all shortages. Including transportation costs in our model provides incentive to employ 'hybrid' policies that partially expedite and partially backlog excess demands within a given period. Our model demonstrates how inventory policy decisions directly impact transportation costs and provides a heuristic approach for jointly minimizing expected inventory and transportation costs.

Key words: Distribution, Logistics, Supply chain management

1. Introduction

Distribution operations form the backbone of supply chain material flows, and efficient management of these operations is, therefore, a key to enabling good supply chain performance. Distribution Center (DC) operations provide a natural focal point for studying mechanisms to improve cost performance in supply chains. Distributors of consumer goods typically maintain inventories and provide delivery service for the customers they serve. Inventory and transportation-related costs comprise the bulk of costs for these distributors, and effectively managing these costs, therefore, contributes significantly to the chain's ability to provide products at competitive prices. Since many of today's retail chains possess considerable power in the distribution chain, they can often dictate delivery terms to less powerful distributors and suppliers. The ability for distributors and suppliers to meet strict delivery requirements while controlling logistics costs is a significant factor in determining the chain's long-term success.

Traditional models of distribution operations, with several exceptions that we note later, have considered the transportation and inventory decision problems independently. This paper takes an integrated modeling approach for inventory and transportation decisions in base-stock systems under periodic and stationary, stochastic demand. Specifically, we consider how inventory shortage policies and agreements affect both inventory and transportation costs in these systems. By base-stock systems, we mean systems that use a periodic-review inventory control system and employ a policy that brings inventory position up to some target level (a decision variable) at the beginning of each planning period. Such systems are not uncommon in practice, since a base-stock policy constitutes an optimal policy structure for periodic review inventory systems where product procurement costs are proportional to the volume purchased, and inventory holding and shortage costs are proportional to the inventory on-hand and backordered, respectively, at the end of a period (see Karlin, 1958). Although many distribution contexts require a fixed order or delivery cost each time an order is placed (hence violating the assumption of linear procurement costs and resulting in the sub-optimality of base-stock policy structures), in many cases these order costs are either small due to electronic ordering capabilities, or they are constant due to a fixed weekly delivery schedule. In such contexts these fixed costs are, therefore, outside the control of the decision maker.

When a customer places demands on a supplier that exceed the supplier's stock levels one of three outcomes typically occurs:

(1) The supplier can expedite the product from an external source (such as another supplier or another distribution center) for immediate delivery to the customer,

(2) The customer can agree to wait another period for demand satisfaction, resulting in a backorder, or

(3) If the supplier finds it too costly to expedite the product and the customer refuses to wait an additional period, a lost sale results.

We demonstrate the important impacts these shortage policy decisions and agreements have on transportation costs, and emphasize the need for customers and suppliers to work together to achieve lower system cost performance. These choices involve an important tradeoff: choosing to backlog demand may in certain contexts increase shortage costs, such as loss-of-goodwill and administration expenses; however, as we will show, this decision can decrease expected transportation costs. So, even for cases where the unit backlogging cost exceeds the unit expediting cost, it may be more economical to backlog shortages due to savings in transportation costs. The supplier may be able to use these potential savings to negotiate delivery terms with customers by offering product discounts.

This work builds on prior work by Geunes and Zeng (2001) that shows the impacts backlogging decisions have on transportation costs in base-stock distribution operations. Their work shows that a policy that backlogs supply shortages can substantially reduce expected transportation costs (and overall distribution costs) when compared to a policy that expedites shortages. This cost reduction does not simply result from a lower cost per unit of backlogging (although this may be a contributing factor if the unit expediting cost exceeds the unit backlogging cost), but from a reduction in the variability of transportation capacity requirements and, hence, transportation costs.

Geunes and Zeng (2001) propose a method for determining per-unit product discounts based on the cost savings due to backlogging, which may induce a customer to allow the supplier to backlog shortages. Their model and methodology, however, assume a strict choice between backlogging and expediting for all supply shortages in every period. We consider a hybrid approach that allows for partial expediting *and* backlogging of supply shortages within a period. This approach generalizes their model and, as our computational results show, can lead to lower cost performance than policies that completely expedite all supply shortages, even for a variety of cases where the unit backlogging cost exceeds the unit expediting cost. Our model also generates insights for managing supply shortages, and for negotiating delivery agreements between a supplier and customer.

This paper is organized as follows. Section 2 contains our model, its relation to the literature, and some interesting properties and insights derived from the model. Section 2.1 reviews prior literature on relevant integrated inventory-transportation models in supply chains. Section 2.2 describes our combined inventory and transportation model with partial backlogging and expediting in base-stock systems under periodic review. Section 2.3 formulates the decision problem and establishes optimal policies for the problem in the absence of transportation costs. Section 3 studies the system under normally distributed demand and presents computational results that demonstrate the ability for our model to lead to lower cost solutions than under complete backlogging or complete expediting, while Section 4 presents concluding remarks.

2. Base stock system model with partial backlogging

This section presents a model of a base-stock distribution operation that allows partial backlogging and partial expediting of supply shortages. This model considers both inventory- and transportation-related costs at the DC. We first review related literature on combined inventory and transportation models under uncertain demand, in addition to past stochastic inventory models that allow partial backlogging.

2.1. RELATED LITERATURE

In one of the earliest models to combine inventory and transportation costs in a stochastic setting, Federgruen and Zipkin (1984) considered a one-warehouse, multiple-retailer system with random retailer demands in a single-period framework. Their model determines vehicle routes and inventory allocation and allows for portions of retailer orders to be filled (and the rest backordered) if insufficient truck capacity exists. Yano and Gerchak (1989) considered a two-stage system in which a manufacturing plant supplies an assembly plant with Just-in-Time (JIT) shipments of a high-volume part. They did not allow for backordering; instead they allowed for emergency shipments of parts when insufficient vehicle capacity exists. Their model determines the order-up-to point (base-stock level) for the part, the time between successive deliveries, and the number of vehicles contracted for deliveries between the supplier and assembly plant. The model minimizes the sum of expected inventory costs, contracted shipment costs, and emergency shipment costs. Ernst and Pyke (1993) built on the work of Yano and Gerchak (1989) by including the manufacturer's (or, in their case, the warehouse's) expected inventory costs plus per unit shipping costs (as opposed to a per truck shipping cost only). Henig et al. (1997) considered a combined inventory and transportation problem for a base-stock system in which the order cost equals zero for quantities below some contracted value of truck capacity, R. Their model applies a premium of k (dollars per unit) for quantities exceeding R, and they develop an optimal inventory replenishment policy structure that contains two base-stock levels, S_1 and S_2 (with $S_2 \geqslant S_1$). The optimal policy orders up to S_1 or S_2, or orders a fixed quantity of R, depending on the inventory position before ordering.

Qu et al. (1999) presented a continuous-time model of a multi-item system in which a central warehouse collects items from a set of suppliers. Their model considers distance-related routing plus stopover transportation costs, in addition to inventory ordering, holding, and backlogging costs, and minimizes average cost per unit time under a so-called modified periodic policy, in which the replenishment interval must be a multiple of some base planning cycle length. Their solution methodology decomposes the problem into separate inventory and transportation submodels.

Each of the stochastic models that combine inventory and transportation decisions we have summarized assumes that supply shortages are either completely backlogged or completely expedited. To our knowledge, no combined inventory and transportation model exists that allows for partial backlogging and partial expediting, while considering the impacts these decisions have on transportation costs, as we consider in this paper. Several papers consider inventory systems that allow for a combination of partial backlogging and partial lost sales when demand exceeds stock, including Thowsen (1975), Rosenberg (1979), Mak (1987), Padmanabhan and Vrat (1990, 1995), Abad (1996), and Vasudha Warrier and Shah (1999). While these articles address additional extensions, such as optimal pricing and demand perishability, they do not consider combining backlogging and expediting, and they include only inventory-related costs.

2.2. SYSTEM MODEL UNDER PARTIAL BACKLOGGING AND EXPEDITING

This section presents our model assumptions and develops expressions for average cost per period over an infinite horizon under these assumptions. We consider a distribution center (DC) that periodically delivers items to a customer or a set of customers using a fleet of trucks. We assume that the DC follows a fixed delivery schedule, i.e., the time between successive deliveries is fixed. Our experience has shown that many such fixed delivery schedules occur in practice where, for example, a DC makes weekly trips to customers to deliver goods and receives weekly deliveries from its supplier. For clarity of presentation, we consider a single customer, although this single customer may represent a concentrated cluster of customers in a geographical region aggregated into a 'super-customer'. We do not, therefore, consider vehicle routing costs; rather we focus on vehicle fleet operating costs, including the fixed costs of owning and maintaining vehicles plus transportation costs that depend on the volume shipped in each period. If the DC places orders to its supplier electronically and receives shipments from its supplier each week, then in many cases the DC need not consider an explicit fixed ordering cost for inventory replenishments. We assume that the lead time for replenishing DC inventory is significantly less than the period length and that the DC's supplier has infinite capacity, i.e., when the DC assesses its inventory level at the end of a week and places an order with its supplier, delivery is made in full at the beginning of the following week.

The sequence of events in a period occurs as follows. The DC receives a replenishment delivery from its supplier at the beginning of the period that brings the DC's inventory position up to some base-stock level (a decision variable). The DC's customer then places an order with the DC. If the customer's order quantity does not exceed the DC's stock level, immediate delivery of the entire order is made to the customer. If the DC does not have enough stock to meet the customer's demand, the DC can receive an immediately expedited order from its supplier or it can backorder the excess demand for delivery in the following period. Alternatively, the DC can choose to expedite a portion of the demand in excess of stock and to backlog the remainder until the following period. If the amount that the DC must ship in the current period does not exceed the DC's total vehicle capacity, then the DC uses its own vehicle fleet to transport the entire shipment. Any amount that the DC must ship exceeding the DC's fleet capacity is shipped using an outside LTL (less-than-truckload) carrier.

Our model assumes that the DC and customer agree upon a maximum amount of demand in any period that the DC will deliver in that period, which we call the *maximum current demand shipped* (MCDS), and that this amount is the same in all periods. The DC further guarantees delivery of any backlogged demand in the period immediately following the shortage. This implies that if the customer's demand in a period exceeds DC supply, the DC expedites the difference between the order quantity and its stock level, up to the MCDS. The total shipment in a period then equals the minimum between the current period's demand and the MCDS, plus the prior period's outstanding backlog. We note that the assumption of a fixed MCDS for all periods restricts us to a subset of all possible partial backlogging and expediting decision rules. The class of all possible partial backlogging and expediting decision rules includes policies that allow the MCDS to vary in different periods in addition to policies that impose no MCDS at all. A fixed value of MCDS, however, allows us to demonstrate that partial backlogging and expediting policies exist that outperform strictly backlogging or expediting (with respect to expected inventory plus transportation costs), even for cases where the unit backlogging cost exceeds the unit expediting cost.

2.2.1. *Transportation cost model*

The transportation cost model we present next is a special case of that developed by Ernst and Pyke (1993). Since transportation capacity is a long-term decision that is not easily changed from period to period, our approach minimizes the expected transportation cost as a function of the fleet capacity, the decision variable, and the probability distribution of shipping quantity in a period, which is a function of both the customer demand and, as we discuss later, the shortage policy used. Let T denote a variable for the DC's in-house truck fleet capacity. The transportation costs in a period include a fixed cost per shipment (which is a function of the fleet capacity), $g(T)$, a cost per unit shipped via in-house truck capacity, K_R, and a cost per unit shipped via LTL capacity, K_C (note that this cost per unit for using an LTL carrier could also represent the cost of arranging an additional 'overtime' shipment using in-house truck capacity). The cost per unit for using an outside carrier (or an overtime shipment) is typically much greater than that for using regular, in-house capacity, and we therefore assume $K_C > K_R$ throughout our analysis. The fixed cost per shipment function, $g(T)$, typically includes elements such as leasing cost, driver and fuel costs, and the costs of loading and unloading the vehicle(s).

Daganzo (1991) characterizes $g(T)$ as a subadditive and increasing step function which is often approximated by a linear or concave function. We adopt a linear approximation for $g(T)$, using the approximation $g(T) \cong K_{RL}T$. Let Q_s denote a random variable for the quantity shipped to the customer in a period and let $f_q(Q_s)$ and $\mathrm{F}_q(Q_s)$ denote the pdf and cdf of the same. (We address the form of this distribution later in this section, since it is directly impacted by our inventory decisions.) The expected transportation-related costs per period then become

$$K(T) = (K_{RL} + K_R)T - K_R \int_0^T (T - Q_s) f_q(Q_s) \mathrm{d}Q_s + \quad (1)$$
$$K_C \int_T^\infty (Q_S - T) f_q(Q_s) \mathrm{d}Q_s,$$

with the minimizing value of T satisfying the equation

$$F_q\left(T^*\right) = [(K_C - (K_{RL} + K_R)]/(K_C - K_R). \quad (2)$$

If we require that T takes one of a discrete set of values (which is typical of truck capacity), we can simply round the value of T^* to the nearest higher or lower discrete value (whichever of these gives the lower expected cost) due to the convexity of the expected transportation cost equation. Note that if the condition $K_C > K_{RL} + K_R$ is not met, then it is optimal for the DC to use the excess carrier exclusively and $T^* = 0$.

2.2.2. *Inventory cost model*

We assume that the DC's customer has a known probability distribution of demand in every period and that successive demands from the customer are independent and identically distributed (iid). Let x denote a random variable for customer demand in a period with mean μ and standard deviation σ, and let $f(x)$ and F(x) denote the pdf and cdf of single-period demand. Let c denote the cost per unit of product procured from the DC's supplier and let h denote the inventory holding cost per unit of inventory remaining at the DC at the end of a period, where h is proportional to c. Let p denote the cost per unit of backlogged demand at the DC (this cost typically includes administrative and loss-of-goodwill costs) and let e' denote the incremental cost (above the procurement cost) of immediately expediting items from the supplier. Let S_1 denote a decision variable for the base-stock level held at the DC and let S_2 denote the MCDS, also a decision variable. S_2 equals S_1 plus the maximum amount the DC plans to expedite from its supplier in a period (any demand in excess of S_2 is backlogged). We can alternatively view S_2 as the *backlog threshold level*. The average DC inventory procurement, holding, and shortage costs per period over an infinite horizon, $G(S_1, S_2)$ can be written as

$$\begin{aligned} G(S_1, S_2) \; = \; & c\mu + h\int_0^{S_1}(S_1 - x)f(x)\mathrm{d}x + e'\int_{S_1}^{S_2}(x - S_1)f(x)\mathrm{d}x \\ & + e'(S_2 - S_1)(1 - F(S_2)) + p\int_{S_2}^{\infty}(x - S_2)f(x)\mathrm{d}x \end{aligned} \qquad (3)$$

Note that in this system no demand is ever lost to the DC, since excess demand over the MCDS is backlogged. Note also that complete expediting is a special case of Equation (3) with $S_2 = \infty$, and complete backlogging is the special case where $S_1 = S_2$.

2.2.3. *Impacts of stock levels on transportation costs*

We next consider how inventory stock level decisions influence transportation costs. If all demand in excess of DC stock is immediately expedited in each period (i.e., $S_2 = \infty$ in Equation (3)), then the shipment quantity, Q_s, will exactly equal the demand, x, and we have $f_q(Q_s) = f(x)$ in Equation (1). Expediting, therefore, allows the decisions on transportation capacity and inventory stock level to be made independently from each other, since the stock levels do not affect transportation costs, nor does transportation capacity affect inventory costs. Under our assumptions then, we only need to separately minimize two convex functions (1) and (3), each of a single variable, and the optimal solutions are easily found.

If, on the other hand, all demand in excess of DC stock is completely backlogged in every period ($S_1 = S_2 = S$ in Equation (3)), Geunes and Zeng (2001) show that the inventory and transportation problems are not separable. In this case the shipment quantity in a period, Q_s, becomes a function of the stock level, S, and the expected transportation cost must be written as a function of not only T, but also S. If we let x_t denote a random variable for demand in period t, then the amount backlogged from the period immediately preceding period t equals $(x_{t-1} - S)^+$, i.e., the amount by which last period's demand exceeded the stock level. The amount shipped to meet current demand will equal the minimum between x_t and S. The total amount shipped in period t, which we denote by $Q_{s,t}$, will equal the sum of these quantities, i.e., $Q_{s,t} = (x_{t-1} - S)^+ + \text{Min}\{x_t, S\}$. Since $(x_{t-1} - S)^+$ depends only on S and x_{t-1}, and $\text{Min}\{x_t, S\}$ depends on S and x_t, and x_{t-1} and x_t are independent, we compute the variance of shipments under backlogging, σ_{bl}^2, using $\sigma_{bl}^2 = \text{Var}[Q_{s,t}] = \text{Var}[(x_{t-1} - S)^+] + \text{Var}[\text{Min}\{x_t, S\}]$. Letting $n(S) = \int_S^\infty (x - S) f(x) \text{d}x$ denote the expected number of units short in a period given a stock level of S, we can show the quantity shipped per period has mean μ and variance

$$\sigma_{bl}^2 = \sigma^2 - 2n(S)\{n(S) + (S - \mu)\}; \tag{4}$$

(see Geunes and Zeng, 2001). Equation (4) reveals an important relationship between backlogging and transportation costs in distribution systems: backlogging excess demand *decreases* the variance of weekly shipment quantities when compared to expediting and, therefore, results in lower expected transportation costs. This decrease in variance results from the DC's ability to *pool* demand from successive periods within a common shipment. Geunes and Zeng (2001) further show that under normal customer demand, the distribution of shipment quantity takes a shape very close to that of a normal distribution (with an additional point mass concentrated at the point S) when the DC's base stock level exceeds the average period demand by more than one standard deviation (which is highly likely in most distribution systems for consumer goods). We can show that Equation (4) also gives the variance of shipment quantities under the partial expediting and partial backlogging scheme we described, with the MCDS, S_2, replacing S in Equation (4), i.e.,

$$\sigma_{pbl}^2 = \sigma^2 - 2n(S_2)\{n(S_2) + (S_2 - \mu)\}, \tag{5}$$

where σ_{pbl}^2 denotes the variance of shipment quantities under partial backlogging with partial expediting up to the MCDS. Observe that for $S_2 > \mu$, Equation (5) implies that the variance of shipment quantities is less than the variance of shipments under complete expediting of shortages. Geunes and Zeng (2001) show that complete backlogging of shortages leads to lower expected transportation costs than complete expediting and can lead to lower total expected system costs, even for cases in which the unit backlogging cost exceeds the unit expediting cost, due to transportation cost savings. Equation (5) indicates that a partial backlogging policy also leads to lower expected transportation costs than a complete expediting policy.

The questions we seek to answer then, are whether or not solutions exist for the partial expediting and backlogging case with lower expected system costs (transportation plus inventory costs) than both the complete expediting and complete backlogging cases and, if so, do the savings from these hybrid policies justify their use in practice? Before answering these questions, we derive theoretical results that support the findings of our computational tests by first considering properties of the inventory cost function alone, and next considering the combined transportation and inventory cost function.

2.3. PROPERTIES OF THE INVENTORY COST FUNCTION

Under our assumptions of a fixed value of truck capacity, a fixed MCDS, and iid demands in all periods, we can show that a stationary base stock level, S_1, exists for minimizing the average cost per period over an infinite horizon. Combining this with our results from the previous section implies that the optimization problem the distributor faces for minimizing average inventory plus transportation costs per period over an infinite horizon under partial backlogging and expediting can be stated as follows:

$$\text{Minimize} \quad G(S_1, S_2) + K(T, S_2) \tag{6}$$

$$\text{Subject to:} \quad S_2 \geqslant S_1 \geqslant 0, \tag{7}$$

where we now write transportation cost, $K(T, S_2)$, as a function of both transportation capacity and the MCDS, based on our results from the preceding section. To better understand the above problem, we first consider the system in the absence of transportation costs. That is, we consider the problem of minimizing $G(S_1, S_2)$ over the region $S_2 \geqslant S_1 \geqslant 0$. The following lemma provides a condition that guarantees the convexity of $G(S_1, S_2)$.

LEMMA 1. *If $p \geqslant e'$ then $G(S_1, S_2)$ is convex in S_1 and S_2.*

Proof. We need to show that the diagonal elements and the determinant of the Hessian matrix are nonnegative for all values of S_1 and S_2. We can show that $\partial^2 G/\partial S_1^2 = (h + e')f(S_1)$, $\partial^2 G/\partial S_2^2 = (p - e')f(S_2)$, and $\partial^2 G/\partial S_1 \partial S_2 = 0$, which implies that the diagonal elements and the determinant of the Hessian matrix are nonnegative whenever $p \geqslant e'$. □

Lemma 1 implies that the following proposition holds when $p \geqslant e'$:

PROPOSITION 1. *If $p \geqslant e'$, then a solution exists that minimizes $G(S_1, S_2)$ using complete expediting of shortages. The stock levels $S_1 = F^{-1}(e'/(e' + h))$ and $S_2 = \infty$ minimize $G(S_1, S_2)$ in this case.*

Proof. Since $G(S_1, S_2)$ is convex when $p \geqslant e'$ we need to find a stationary point to find a global minimum. Equating $\partial G/\partial S_1$ and $\partial G/\partial S_2$ to zero produces the result. □

Proposition 1 confirms the intuition that a policy of fully expediting shortages is optimal when the unit cost of backlogging is greater than or equal to the unit cost of expediting. The following lemma shows that whenever $e' > p$, if we are given any solution $(S_1, S_2) = (S_1', S_2')$, with $S_2' > S_1'$, then the solution $S_1 = S_2 = S_1'$ always results in an expected cost less than or equal to $G(S_1', S_2')$.

LEMMA 2. *Suppose $e' > p$ and consider any solution (S_1', S_2') with $S_2' > S_1'$. The solution $S_1 = S_2 = S_1'$ has an expected cost less than or equal to the solution (S_1', S_2'), i.e., $G(S_1', S_2') \geqslant G(S_1', S_1')$.*

Proof. We can show that $G(S_1', S_2') - G(S_1', S_1') = (e' - p)(n(S_1') - n(S_2'))$, which is greater than or equal to zero for any $S_2' > S_1'$. □

Lemma 2 implies that we can drop the expediting cost term from Equation (3), and we are left with the case of complete backlogging, in which case the optimal solution is given by $S_1 = S_2 = F^{-1}(p/(p + h))$, which is consistent with existing results (see Nahmias, 2001).

3. Base-stock system under normally distributed demand

This section considers the base-stock system we have described under normally distributed customer demand and derives properties associated with the total expected system costs (note that although we assume normally distributed demand, we also assume a negligible probability of negative demand). Section 3.1 presents a convex heuristic approximation for the transportation cost Equation (1) as a function of the transportation capacity and the MCDS, which we use in our solution procedure. Section 3.2 then provides results from a set of computational tests based on the approximate expected total system cost equation we derive.

3.1. STOCK LEVEL DECISIONS IN THE PRESENCE OF TRANSPORTATION COSTS

Under normally distributed customer demand, the values of S_1 and S_2 can be substituted by their standardized values, $k_1 = (S_1 - \mu)/\sigma$ and $k_2 = (S_2 - \mu)/\sigma$, respectively. Letting L(k) denote the standard unit normal loss integral, i.e., $\mathrm{L}(k) = \int_k^{\infty}(u - k)\phi(u)\mathrm{d}u$ (where $\phi(u)$ denotes the pdf of the standard unit normal distribution), then the inventory cost Equation (3) under normally distributed customer demand becomes

$$G(k_1, k_2) = c\mu + h\sigma(k_1 + \mathrm{L}(k_1)) + e'\sigma(\mathrm{L}(k_1) - \mathrm{L}(k_2)) + p\sigma\mathrm{L}(k_2), \qquad (8)$$

Methods for finding a minimizing solution for (8) were described in the previous section. We next consider the problem with the addition of the expected transportation cost, $K(T, k_2)$, which includes the additional truck capacity decision variable, T, and results in the nonlinear program given by (6) and (7). Equation (1) gives the expected transportation cost as a function of the transportation capacity, T, and the distribution of shipments, $f_q(Q_s)$. We noted in the previous section that Q_s is a random variable with mean μ (the mean period demand) and variance $\sigma_{pbl}^2 = \sigma^2 - 2n(S_2)\{n(S_2) + (S_2 - \mu)\}$, and that Q_s is approximately normally distributed when customer demand is normally distributed. Using a normal approximation for $f_q(Q_s)$ and letting $n_q(T) = \int_T^\infty (Q_s - T) f_q(Q_s)\mathrm{d}Q$, we can rewrite Equation (1) as

$$K(T, S_2) = K_{RL}T + K_R\mu + (K_C - K_R)n_q(T), \tag{9}$$

where the dependence of expected transportation cost on S_2 is reflected through the variance of the distribution of Q_s. If we let $k_T = (T - \mu)/\sigma_{pbl}$, then under normal customer demand Equation (9) becomes

$$K(T, k_2) = K_{RL}T + K_R\mu + (K_C - K_R)\sigma_{pbl}\mathrm{L}(k_T). \tag{10}$$

Equation (10) encodes an extremely complex relationship between k_2 (the normalized MCDS), the truck capacity, T, and expected transportation costs that has not led to tractable mathematical analysis of the properties of the cost function. We therefore develop a heuristic approximation for $K(T, k_2)$ that allows us to obtain good solutions.

Assuming a normal distribution of shipment quantities, we can rewrite Equation (5) in the form $\sigma_{pbl} = \sigma\sqrt{1 - 2[\mathrm{L}(k_2)]^2 - 2\mathrm{L}(k_2)k_2} \equiv \theta\sigma$, and we therefore have $n_q(T) = \theta\sigma\mathrm{L}(k_T)$. Geunes and Zeng (2001) show that, under normal demand, the relationship between k_2 and θ can be closely approximated by a linear function over the range of k_2=[0, 3], i.e., $\theta \cong \hat{\theta} = c_1 + c_2k_2$, where c_1 and c_2 are positive constants determined by performing a linear fit of θ against k_2. They further show that the unit normal loss integral can be effectively (heuristically) approximated by an exponential function over this region and use the approximation $\mathrm{L}(k_T) \cong c_3\exp(-c_4k_T) \equiv \hat{\mathrm{L}}(k_T)$, where c_3 and c_4 are positive constants determined through an exponential fit for the loss function. Approximating k_T using $k_T \cong (T-\mu)/(\sigma(c_1+c_2k_2))$ we arrive at the following heuristic approximation for $n_q(T)$, which is also a function of k_2:

$$n_q(T, k_2) = \sigma_{pbl}L(k_T) \cong (c_1 + c_2k_2)\sigma c_3 \exp\left(-\frac{c_4(T - \mu)}{\sigma(c_1 + c_2k_2)}\right). \tag{11}$$

Our approximation for the expected transportation cost equation then becomes

$$\begin{aligned}\hat{K}(T, k_2) \;=\; & K_{RL}T + K_R\mu + (K_C - K_R)\\ & \times(c_1 + c_2k_2)\sigma c_3 \exp\left(-\frac{c_4(T - \mu)}{\sigma(c_1 + c_2k_2)}\right).\end{aligned} \tag{12}$$

Geunes and Zeng (2001) show that $\hat{K}(T, k_2)$ is a convex function of T and k_2 and a strictly increasing function of k_2 for $T \geqslant \mu$ and $k_2 \geqslant 0$.

Let $H(K_1, k_2, T)$ denote the sum of expected single-period inventory costs (Equation 8) and single-period transportation costs (Equation 10) and let $\hat{H}(k_1, k_2, T)$ denote the approximation for $H(k_1, k_2, T)$ obtained by replacing $K(T, k_2)$ with its approximation, $\hat{K}(T, k_2)$ given in Equation (12). The optimization problem for minimizing approximate expected total system cost per period is given by

$$\text{[ETSC]} \quad \text{Minimize} \quad \hat{H}(k_1, k_2, T) = G(k_1, k_2) + \hat{K}(T, K_2) \tag{13}$$

$$\text{Subject to:} \quad 0 \leqslant k_1 \leqslant k_2 \leqslant 3, T \in \{iT_0, (i+1)T_0, ..., mT_0\}, \tag{14}$$

where T_0 denotes the standard truck capacity, i is the smallest integer such that $iT_0 \geqslant \mu$, and m is the smallest integer such that $mT_0 \geqslant \mu + 3\sigma$ (we confine ourselves to integer multiples of some base truck capacity size, T_0, at least as great as the mean period demand and not exceeding three standard deviations above the mean period demand).

Since $G(k_1, k_2)$ is convex when $p \geqslant e'$, and $\hat{K}(T, k_2)$ is convex in T and k_2, this implies that $\hat{H}(k_1, k_2, T)$ is convex when $p \geqslant e'$. We are primarily interested in cases in which the expediting cost per unit is less than that for backlogging (i.e., $p > e'$), so we can ascertain whether practical cases exist that call for at least some level of backlogging (if not complete backlogging) due to the transportation cost savings backlogging can induce. Our computational tests therefore focus on cases in which $p > e'$. (Since $p \leqslant e'$ favors complete backlogging with respect to both inventory and transportation costs, we have not considered this case in our computational tests.) To provide more compact notation in our subsequent analysis, we let $\hat{\theta}$ denote our linear approximation for θ, i.e., $\hat{\theta} = c_1 + c_2 k_2$, and let $\hat{k}_T$ denote our approximation for k_T, i.e., $\hat{k}_T = (T - \mu)/(\sigma\hat{\theta})$. Using this notation, we can derive the following first-order optimality conditions for [ETSC]:

$$\Phi(k_1^*) = e'/(e' + h) \tag{15}$$

$$\Phi(k_2^*) = \frac{(p - e') - (K_C - K_R)c_2 c_3 \exp\big(-c_4\hat{k}_T\big)\big(1 + c_4\hat{k}_T\big)}{(p - e')} \tag{16}$$

$$T^* = \mu - \big(\sigma\hat{\theta}/c_4\big)\ln\left(\frac{K_R L}{(K_C - K_R)c_3 c_4}\right) \tag{17}$$

Since [ETSC] is a convex program, if we can find a feasible solution to [ETSC] satisfying the above conditions, then this solution is optimal. We first make some observations about these first-order conditions. It is interesting to note that the optimal value of k_1 is the same as that for the complete expediting case, which implies that the stock level is only a function of the relative values of the unit holding and expediting costs when $p > e'$. Note also that in Equation (17) T^* is a function of k_2 while in Equation (16) $\Phi(k_2^*)$ is a function of *T*. We therefore require an iterative procedure to find the optimal continuous value of T^*. However, since we confine ourselves to a limited set of discrete values of *T*, this enables us to find the value of k_2 that solves (16) for each value of *T* that we consider. Equation (16) does not, however, provide a closed-form solution for k_2 and so we must implement a search procedure to find whether a solution exists that satisfies this equation (in our computational tests, described in the next section, we were always able to find a solution that approximately satisfied Equations (15–17)). Equation (16) indicates that cases can exist where the optimal value of k_2 is finite, i.e., we do not fully expedite even when $p>e'$ in all cases. The computational test results presented in the following section bear this out. In fact we have found several instances for which the optimal solution *completely* backlogs all shortages, even though $p>e'$. For these cases, we obviously have equality holding between the right hand sides of Equations (15) and (16) in the optimal solution, i.e., $k_1^* = k_2^*$.

3.2. COMPUTATIONAL EXPERIMENTS

The preceding sections provide analytical results regarding the performance of our two options for handling excess demands (complete expediting versus partial expediting and backlogging) when the unit backlogging cost exceeds the unit expediting cost. This section provides a number of test examples that confirm our analytical results and provide additional managerial insights. We use the minimum expected inventory plus transportation cost per period (after subtracting the constant expected procurement cost per period, $c\mu$) under the partial expediting and backordering (PEB) approach, which we denote by TC_{PEB}^*, as a basis for comparing our results. We refer to policies that use partial expediting and backlogging as *hybrid policies*. Note that we use our heuristic approximation methods for $n_q(T)$ strictly to determine the appropriate values of k_1^* (normalized base-stock level) and k_2^* (normalized MCDS). Given k_1^* and k_2^*, we then explicitly calculate the value of $n_q(T)$ using only the normal approximation for Q_s. That is, we eliminate the linear approximation for σ_{pbl} and the exponential approximation for $\mathrm{L}(k_T)$ and calculate these quantities directly (assuming normality of shipment quantity) when computing TC_{PEB}^* to obtain a more accurate estimate of the optimal cost of a hybrid policy (under our hybrid policy structure assumptions). Letting $H_N(k_1, k_2, T)$ denote the value of $H(k_1, k_2, T)$ using the normal approximation for the distribution of Q_s as the only source of approximation, then we have $TC_{\text{PEB}}^* = H_N(k_1^*, k_2^*, T^*)$, where k_1^*, k_2^*, and T^* solve formulation [ETSC]. If we let TC_{CE}^* denote the optimal total costs under complete expediting (CE), we can use the following formula to calculate the percentage cost savings from using a hybrid policy:

$$\omega = \left[\left(TC_{CE}^*/TC_{PEB}^*\right) - 1\right] \times 100 \tag{18}$$

We chose the following parameters as base values for our numerical study:

Table 1. Computational test parameters

	Scenario[a]		
	I	II	III
Unit backlog cost, p	0.1	0.5	1
Procurement cost, c	{1,2,3,4,5}	{1,2,3,4,5}	{1,2,3,4,5}
Expediting cost, $K_C = e'$	{0.03, 0.06, 0.095}	{0.1, 0.25, 0.49}	{0.33, 0.67, 0.95}

[a]Every combination of parameters was tested within each scenario.

Table 2. Comparison of partial backordering (PEB) with complete expediting (CE) scenario I: $p = 0.1$

c	$e' = 0.03$[a]			$e' = 0.06$			$e' = 0.095$		
	k_1^*	k_2^*	ω	k_1^*	k_2^*	ω	k_1^*	k_2^*	ω
1	1.1	∞	0%	1.4	2.8	0.001%	*1.5*	*1.5*	*1.23%*
2	0.7	∞	0%	1.1	2.8	0.001%	*1.2*	*1.2*	*1.78%*
3	0.4	∞	0%	0.8	2.8	0.001%	*1*	*1*	*2.08%*
4	0.3	∞	0%	0.7	2.8	0.001%	*0.9*	*0.9*	*2.28%*
5	0.1	∞	0%	0.5	2.8	0.001%	*0.8*	*0.8*	*2.32%*

[a]Unit overflow carrier cost, K_C, is set equal to unit expediting cost, e', in all test cases.

h	h'	T_0	μ	σ	K_{RL}	K_R
\$0.005c	$= K_C$	40 000 lbs	100 000 lbs	30 000 lbs	\$0.007817	\$0.01

We have based this choice of parameters on weekly cost figures provided in an unrelated computational study of the distribution system of a Fortune 500 manufacturer (see Geunes, 1999). To provide a broad scope of test results, we varied the procurement cost per unit, c, the cost per unit for shipping via common carrier, K_C, and the cost per unit of backlogged demand, p, as shown in Table 1. Since the unit holding cost, h, is proportional to the unit procurement cost c, this implies that varying c is, therefore, equivalent to varying the unit holding cost. For each of the cost combinations tested, we calculated the minimum total costs of the two policies (hybrid policy and complete expediting) and the associated percentage savings of the hybrid policy. We have adopted a tabular format to present the numerical results, which we discuss later in this section.

Due to the convexity of the expected total cost Equation (13), the values of the two policy parameters, (k_1, k_2), that optimize the problem given in (13) and (14) could easily be found by a simple search of discrete points (in multiples of 0.1) over the space $0 \leqslant k_1 \leqslant k_2 \leqslant 3$ (although we realize that discretizing the points in this space results in a slight loss of accuracy). This procedure is a simple routine

Table 3. Comparison of partial backordering (PEB) with complete expediting (CE) scenario II: $p = 0.5$

c	$e' = 0.10$[a]			$e' = 0.25$			$e' = 0.49$		
	k_1^*	k_2^*	ω	k_1^*	k_2^*	ω	k_1^*	k_2^*	ω
1	1.6	∞	0%	2	∞	0%	*2*	*2*	*1.18%*
2	0.3	∞	0%	1.7	∞	0%	*1.7*	*1.7*	*1.97%*
3	1	∞	0%	0.5	∞	0%	*1.6*	*1.6*	*2.61%*
4	0.9	∞	0%	1.4	∞	0%	*1.5*	*1.5*	*3.10%*
5	0.7	∞	0%	1.2	∞	0%	*1.4*	*1.4*	*3.51%*

[a]Unit overflow carrier cost, K_C, is set equal to unit expediting cost, e', in all test cases.

Table 4. Comparison of partial backordering (PEB) with complete expediting (CE) scenario III: $p = 1.0$

c	$e' = 0.33$[a]			$e' = 0.67$			$e' = 0.95$		
	k_1^*	k_2^*	ω	k_1^*	k_2^*	ω	k_1^*	k_2^*	ω
1	2.1	∞	0%	*2.3*	*2.3*	*0.23%*	*2.2*	*2.2*	*0.78%*
2	1.8	∞	0%	*2.1*	*2.1*	*0.33%*	*2*	*2*	*1.42%*
3	1.6	∞	0%	*1.9*	*1.9*	*0.35%*	*1.9*	*1.9*	*1.90%*
4	1.5	∞	0%	1.8	2.8	0.33%	*1.8*	*1.8*	*2.32%*
5	1.4	∞	0%	1.7	2.8	0.32%	*1.7*	*1.7*	*2.70%*

[a]Unit overflow carrier cost, K_C, is set equal to unit expediting cost, e', in all test cases.

and the searching time is short. For each test problem, we found a pair of optimal solution values, (k_1^*, k_2^*) (within the set of discretized points for k_1 and k_2), which enables us to compare the performance of the two controlling policies in terms of the minimum average cost per period of holding inventory, expediting, backlogging, and transportation over an infinite horizon. Note that our search procedure invariably produced results that approximately satisfied first-order conditions (15) and (16).

Tables 2–4, categorized according to the three values of p, provide valuable insights into the relationships among the policies under consideration. Most importantly, these results identify cases where fully backlogging outperforms a policy that allows any expediting, even when the unit backlogging cost exceeds the unit expediting cost (these cases are italicized in the tables; note that in these cases the unit backlogging and expediting costs are, however, very close in magnitude). We also observe that some degree of backlogging was preferred in 25 of the 45 test cases. These results confirm that the magnitude of transportation cost savings obtained by backlogging outweighs the increased backlogging cost in the majority of test cases. As we would expect, when the unit expediting cost is significantly smaller than the unit backlogging cost, we employ little or no backlogging.

Further analysis of Tables 2–4 shows that increasing the unit product value, c (and equivalently the unit holding cost, h), provides incentive to increase expediting but not backlogging (except in the complete backlogging cases). In the partial backlogging case, changing the degree of backlogging has no effect on holding costs, and so changing the unit product value should not affect backlogging decisions. This is consistent with Equations (15) and (16), which indicate that the backlogging threshold level (or MCDS) level has no dependence on holding costs. Although the average percentage cost savings due to hybrid policies is somewhat small, several cases did exceed 2.5%, which can provide meaningful savings in many distribution contexts. We also point out that these savings are sensitive to the cost and demand parameters chosen (making a relative increase in transportation cost parameters will favor higher cost savings).

To gauge the accuracy of our approximation for $H(k_1, k_2, T)$ (the average inventory plus transportation cost per period; see Equation 13), we compared $\hat{H}(k_1^*, k_2^*, T^*)$ (which incorporates our approximation scheme for σ_{pbl} and $\mathrm{L}(k_T)$) to the value of $H_N(k_1^*, k_2^*, T^*)$ (which includes the normality assumption for Q_s as the only source of approximation) in each of our 45 test cases. To avoid overstating the accuracy of the approximation, we subtracted the expected procurement cost per period (i.e., $c\mu$, a constant) from both $H_N(k_1^*, k_2^*, T^*)$ and $\hat{H}(k_1^*, k_2^*, T^*)$. We computed the *approximation error* as the absolute value of the difference between $H_N(k_1^*, k_2^*, T^*)$ and $\hat{H}(k_1^*, k_2^*, T^*)$ taken as a percentage of $H_N(k_1^*, k_2^*, T^*)$ (after subtracting $c\mu$ from both cost terms). Our 45 test cases produced an average approximation error of 1.37%, with a range between 0.02 and 3.21%. For certain parameters the cost approximation is quite good, although this error tends to degrade as the expected cost of exceeding truck capacity per period forms a greater proportion of total costs. This is confirmed by the observation that the average approximation error for Scenario III (2.07%) exceeded that for Scenario II (1.71%), which exceeded that for Scenario I (0.33%). Note, however, that Tables 2–4 indicate that despite this approximation error, we were still able to identify good values of k_1^* and k_2^* that led to expected total cost improvements by using hybrid policies (the figures in Tables 2–4 use $H_N(k_1^*, k_2^*, T^*)$ for comparison to the complete expediting case, i.e., the costs used to calculate the % savings reported in these tables do not contain this approximation error).

An important message conveyed by the numerical illustrations indicates that including transportation costs tends to favor hybrid policies due to the transportation cost savings induced by backlogging. These analyses confirm our prior analytical results and provide a tool for a distributor to use in negotiating with customers regarding their shipping schedules when excess demands occur. The basic tradeoff the distributor and customer face is the choice between a high fill rate (proportion of demands met immediately from stock), which can be achieved through expediting, versus a lower fill rate (using backlogging) that can result in lower total inventory and transportation costs. Our model allows the distributor to capture the costs of these tradeoffs and to present the customer a variety of fill rate choices at different product delivery prices. For example, the distributor can offer quantity discounts to induce customers to wait one period to fulfill their partial backorders (see Geunes and Zeng, 2001, for an illustration of how to set discount levels based on expected cost savings), resulting in potential savings for both the distributor and customer.

4. Concluding remarks

Firms that provide inventory and delivery service for customers with varying demands confront the complex problem of determining how to efficiently manage inventory stock levels and transportation capacity. The effects that policies for managing supply shortages have on transportation costs has received little attention in prior literature. This paper attempts to partially fill this gap by providing a modeling and (heuristic) optimization approach for determining the best combination of inventory stock levels and transportation capacity when the distributor has both expediting and backlogging options for dealing with supply shortages.

We studied two control policies, namely complete expediting and partial expediting and backlogging, and developed a model that considers both inventory-related and transportation costs. This model enables us to analyze the impacts inventory policy decisions have on expected transportation costs, and to obtain solution procedures for setting stock levels that jointly minimize an approximate expected transportation and inventory cost equation. Our numerical examples under normally distributed demand showed that including transportation costs in the model favors a hybrid usage of partially expediting and backlogging shortages. We showed that these hybrid policies provide greater value under scenarios in which the unit backlogging and expediting costs are close in magnitude. These cost savings can support delivery policy negotiations between a supplier and customers for setting strategies to cope with excess demand. This work provides further evidence of the importance of supplier-customer coordination in supply chains.

Our work has focused on the costs borne by the DC only. A more complete *systems* approach would directly capture the economic impacts that these DC stock level decisions have on the customer (retailer) as well. By doing this, we could determine the appropriate decisions from the view of the entire system. However, this work focused on the insights obtained regarding the value hybrid supply shortage policies provide in managing DC operations, and particularly their impact on expected transportation costs incurred by a DC. We briefly touched on the potential for the DC to provide its customer economic incentives, such as product discounts, as a method for sharing the economic gains from employing hybrid policies. Extending the model to a systems framework that includes customer costs directly in the model provides a promising direction for further research.

Acknowledgement

The authors would like to thank two anonymous reviewers, whose guidance in the review process has helped significantly in the development of this paper.

References

Abad, P.L. (1996), Optimal pricing and lot-sizing under conditions of perishability and partial backordering, *Management Science* 42(8): 1093–1104.

Daganzo, C.F. (1991), *Logistics Systems Analysis*, Springer, Berlin.

Ernst, R. and Pyke, D.F. (1993), Optimal base stock policies and truck capacity in a two-echelon system, *Naval Research Logistics* 40: 879–904.

Federgruen, A. and Zipkin, P.H. (1984), A combined vehicle routing and inventory allocation problem, *Operations Research* 32: 1019–1037.

Geunes, J. (1999), *Models to Optimize Multi-Stage Linkages in Manufacturing and Distribution Operations*, Ph.D. Thesis, Penn State University, University Park, PA.

Geunes, J. and Zeng, A.Z. (2001), Impacts of inventory shortage policies on transportation requirements in two-stage distribution systems, *European Journal of Operational Research*, 129(2): 299–310.

Henig, M., Gerchak, Y., Ernst, R. and Pyke, D.F. (1997), An inventory model embedded in designing a supply contract, *Management Science* 43(2): 184–189.

Karlin, S. (1958), Optimal inventory policy for the arrow-Harris-Marschak dynamic model, in Arrow, K., Karlin, S. and Scarf, H. (eds.), *Studies in the Mathematical Theory of Inventory and Production*, Stanford University Press, Stanford CA.

Mak, K.L. (1987), Determining optimal production-inventory control policies for an inventory system with partial backlogging, *Computers & Operations Research* 14(4): 299–304.

Nahmias, S. (2001), *Production and Operations Analysis*, 4th ed., Irwin McGraw-Hill, Chicago, IL.

Padmanabhan, G. and Vrat, P. (1990), Inventory model with a mixture of backorders and lost sales, *International J. of Systems Science* 21(8): 1721–1726.

Padmanabhan, G. and Vrat, P. (1995), EOQ model for perishable items under stock dependent selling rate, *European J. of Operational Research* 86(2): 281–292.

Qu, W.W., Bookbinder, J.H. and Iyogun, P. (1999), An integrated inventory-transportation system with modified periodic policy for multiple Products, *European J. of Operational Research* 115: 254–269.

Rosenberg, D. (1979), New analysis of a lot-size model with partial backlogging, *Naval Research Logistics* 26(2): 349–353.

Thowsen, G.T. (1975), Dynamic, nonstationary inventory problem for a price/quantity setting firm, *Naval Research Logistics* 22(3): 461–476.

Vasudha Warrier, T. and Shah, N.H. (1999), Lot-size model with partial backlogging when the amount received is uncertain for deteriorating items, *International J. of Systems Science* 30(2): 205–210.

Yano, C.A. and Gerchak, Y. (1989), Transportation contracts and safety stocks for just-in-time deliveries, *J. of Manufacturing and Operations Management* 2: 314-330.