Theory and methodology

Impacts of inventory shortage policies on transportation requirements in two-stage distribution systems

Joseph Geunes a,*, Amy Z. Zeng b

a Department of Industrial and Systems Engineering, University of Florida, 303 Weil Hall, P.O. Box 116595, Gainesville, FL 32611, USA
b Department of Management, Worcester Polytechnic Institute, Worcester, MA 01609, USA

Received 29 February 2000

Abstract

This paper investigates the impacts inventory shortage policies have on transportation costs in base-stock distribution systems under uncertain demand. The model proposed demonstrates how backlogging arrangements can serve to decrease the variability of transportation capacity requirements, and hence the magnitude of transportation costs, when compared with policies that expedite demand shortages. The model shows how inventory policy decisions directly impact expected transportation costs and provides a new method for setting stock levels that jointly minimizes inventory and transportation costs. The model and solution method provide insights into the relationship between inventory decisions and transportation costs and can serve to support delivery policy negotiations between a supplier and customer that must choose between expediting and backlogging demand shortages. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Distribution; Logistics; Supply chain management

1. Introduction

This paper investigates the relationship between inventory shortage policies and the variability of shipment volumes in two-stage serial distribution systems controlled by periodic review. The main result of this paper demonstrates that a policy of backlogging excess demand can serve as an effective mechanism to decrease the variability (and hence costs) of transportation requirements. These variance-damping effects due to backlogging are quantified in this paper and a new approach is provided for setting base-stock levels in distribution systems where customers may receive financial incentives to wait an additional period for demand satisfaction.

Ernst and Pyke (1993) considered a two-echelon system in which the supplier and customer follow base-stock policies with a common review frequency. They showed that if the supplier immediately expedites demand occurring when stock on-hand equals zero, then inventory
stocking and transportation capacity decisions decouple and can be made independently. Under an expediting arrangement, the distribution of demand for transportation capacity equals that of product demand. This paper considers the case where the supplier backlogs excess demand (instead of expediting) and characterizes the relationship between the stock level and expected transportation costs. Since base-stock level has a direct effect on transportation costs, stocking and transportation decisions cannot be separated. A method is provided for setting stock levels that accounts for the impact backlogging has on transportation costs. Additional work on joint inventory and transportation decisions in base-stock systems can be found in Henig et al. (1997), Yano (1992), and Yano and Gerchak (1989). Past literature does not, however, consider the relationship between shortage policies and transportation costs as in this paper. For a recent summary of work on inventory-routing models, see Qu et al. (1999).

Traditional approaches (such as that of Ernst and Pyke (1993)) set base-stock levels for periodic review systems (with no fixed order cost) based on the tradeoff between unit overstocking and understocking costs. For example, consider a single product that experiences independent, identically distributed demands in successive periods and let \( x \) denote a random variable for demand in a period (let \( f(x) \) denote the pdf of demand in each period and \( F(x) \) the cdf). Suppose that the product costs \( c \) per unit, sells for \( P \), and costs \( h \) to hold in inventory for each period (charged against the end of period on-hand inventory). If it costs \( p \) per unit in loss of goodwill and bookkeeping expenses when demand is backlogged, then the stock level that minimizes the expected cost per period over an infinite horizon, \( S^* \), can be shown to be equal (Nahmias, 1997) to

\[
F(S^*) = p/(p + h).
\]

Suppose on the other hand that instead of backlogging, demands occurring when stock on hand equals zero are met by an expedited shipment that comes at a cost of \( e \) per unit expedited. If \( e \) includes loss of goodwill in addition to the expediting cost (which also includes the product cost, \( c \)), then \( p \) can simply be replaced by \( e - c \) in the above equation to obtain \( S^* \). Note that in both the backlogging and expediting cases all demand is eventually satisfied, implying that revenue is constant and minimizing cost also maximizes profit.

Under an expediting arrangement, all demand is satisfied in every period before the end of the period. If the stockholder also pays for delivery to a customer site (as a distribution center must often do), decisions regarding how goods are transported can be made independently from stock level decisions. Such transportation decisions typically include how much in-house truck fleet capacity to maintain for transporting goods and how to deal with periods in which demand exceeds in-house truck capacity.

In contrast to existing work integrating inventory and transportation decisions in base-stock systems, suppose that customers allow backlogging of unsatisfied demand and that the supplier receives replenishments from a source with unlimited capacity at the beginning of each period. The replenishment quantity received by the supplier equals the amount needed to bring inventory position to the base-stock level after observing the prior period’s demand (the lead time for supplier replenishments is therefore negligible). The customer then places an order with the supplier, who immediately ships the amount backlogged from the prior period, plus as much of the current order as possible. Under this backlogging arrangement, stock level and transportation capacity decisions no longer decouple. The key idea is that the stock level partially determines the distribution of shipment quantity per period to the customer, because the stock level effectively forces a maximum shipping quantity for satisfying current demands.

The goal of this study is to capture the economic impacts of backlogging in distribution systems and to determine stock levels that minimize inventory plus transportation costs. This approach generates insights into the implications of providing customers with a choice between expediting and backlogging in distribution systems. This choice involves an important tradeoff: Choosing to
backlog demand may in some cases increase shortage costs, such as loss-of-goodwill and administration expenses; however, it will be shown that this decision also decreases expected transportation costs. So, even for cases where unit backlogging cost exceeds unit expediting cost, it may be more economical to backlog supply shortages due to savings in transportation costs. The supplier can use these savings to negotiate delivery policies with the customer by offering product discounts. Transportation models have traditionally been decoupled from stock level decisions. It will be shown how backlogging does not allow this decoupling and requires jointly setting stock levels and transportation capacity. Finally, a method for setting stock and transportation capacity levels is presented, and numerical examples demonstrate the potential magnitude of cost savings.

2. Modeling the effects of backlogging on shipping requirements

2.1. Transportation cost model

This section considers a transportation cost model for a distributor that delivers goods to a single customer, a special case of the model by Ernst and Pyke (1993). Let \( T \) denote a decision variable for in-house transportation fleet capacity and assume availability of an outside (third party, less-than-truckload (LTL)) carrier. The costs for in-house truck capacity include a fixed cost per shipment, \( g(T) \), expressed as a function of truck capacity, and a variable charge of \( KR \) per unit shipped. The function \( g(T) \) incorporates allocated lease cost, a cost per truck-mile (independent of contents), and a driver cost. Daganzo (1991) characterizes \( g(T) \) as a subadditive and increasing step function which is often approximated by a linear or concave function. Assume a charge of \( KC \) per unit shipped via common carrier per period and assume throughout that \( KC > KR \), since LTL shipments typically require a much greater cost per unit than shipping via regular capacity. Letting \( Q_s \) denote a random variable for the shipping quantity in a period, then the expected total transportation cost per period, \( K(T) \), is

\[
K(T) = g(T) + KR \int_0^T Q_s f_q(Q_s) \, dQ_s +KR T \int_T^\infty f_q(Q_s) \, dQ_s +KC \int_T^\infty (Q_s - T) f_q(Q_s) \, dQ_s, \tag{2}
\]

where \( f_q(Q_s) \) denotes the probability distribution of single-period shipments required by the distributor. If the distributor must deliver all of the customer’s demand in every period (as occurs under an expediting arrangement), then \( f_q(Q_s) \) is exactly the same as the demand distribution seen by the customer. Under a linear approximation for \( g(T) \) (i.e., \( g(T) = KR LT \), where \( KR \) is a scalar), Eq. (2) takes the familiar form of the newsvendor equation given by

\[
K(T) = (KR + KR)T - KR \int_0^T (T - Q_s) f_q(Q_s) \, dQ_s +KC \int_T^\infty (Q_s - T) f_q(Q_s) \, dQ_s, \tag{3}
\]

with the minimizing value of \( T \) satisfying the equation

\[
F_q(T^*) = (K_C - (KR + KR))/(K_C - KR). \tag{4}
\]

If \( T \) can only assume a discrete set of values, the value of \( T^* \) can be rounded to the nearest higher or lower discrete value due to the convexity of the expected cost equation.

2.2. Variance damping impacts of backlogging

This section develops an expression for the variance of shipment quantity per period under backlogging and shows its impact on transportation costs. Assume that single-period demand has mean \( \mu \) and standard deviation \( \sigma \). The supplier’s expected inventory holding and shortage costs per period, \( G(S) \), are given by

\[
G(S) = c\mu + h \int_0^S (S - x) f(x) \, dx + p \int_S^\infty (x - S) f(x) \, dx. \tag{5}
\]
In period $t$, the amount backlogged from the prior period equals $(x_{t-1} - S)^+$, i.e., the amount by which last period’s demand $x_{t-1}$ exceeded the stock levels $S$. The amount shipped to meet current demand will equal the minimum between $x_t$ and $S$. The expected amount shipped in a period is then

$$E[Q_s] = E[(x_{t-1} - S)^+] + \text{Min}\{x_t, S\}. \quad (6)$$

It is straightforward to show that $E[Q_s]$ equals the mean demand per period, $\mu$. Since the first term in Eq. (6) depends only on $S$ and $x_{t-1}$, and the second term depends on $S$ and $x_t$, these terms are independent, and we compute the variance of shipments under backlogging, $\sigma^2_b$, using

$$\sigma^2_b = \text{Var}[Q_s] = \text{Var}[(x_{t-1} - S)^+] + \text{Var}[\text{Min}\{x_t, S\}] \quad (7)$$

Letting $n(S) = \int_S^{\infty} (x - S)f(x) \, dx$, i.e., the expected number of units short in a period with a stock level of $S$ we can show that (see Appendix A)

$$\sigma^2_b = \sigma^2 - 2n(S)n(S) + (S - \mu). \quad (8)$$

Eq. (8) shows the variance ‘damping’ effect backlogging has on the shipping quantity. For any $S > \mu$ (which is the case in most real world systems), $\sigma^2_b$ will clearly be strictly less than $\sigma^2$, the demand variance, which results in lower variability of shipment quantities than under an expediting arrangement. Under a normal demand distribution a stronger result can be shown, namely that $\sigma^2_b < \sigma^2$ regardless of the value of $S$, for finite values of $S$. This result implies that backlogging excess demand never amplifies the variance of shipments $\sigma^2_b$ over that of demand $\sigma^2$. Proposition 1 below shows that the maximum amount of variance damping available from backlogging is the same for any normal distribution and does not depend on specific distribution parameters. Let %VR denote the difference between the demand variance and the variance of shipment quantities, taken as a percentage of the demand variance, i.e.,

$$\%\text{VR} = \left(\frac{(\sigma^2 - \sigma^2_b)}{\sigma^2}\right) \times 100\%.$$  

Proposition 1 shows that setting the base-stock level equal to the mean demand per period maximizes %VR under normally distributed demand.

**Proposition 1.** Under normally distributed demand, the maximum value of percentage variance damping %VR is the same regardless of distribution parameters and equals approximately 68.2%.

**Proof.** See Appendix B.

**Proposition 2.** Under normally distributed demand, the maximum %VR occurs at $S = \mu$ (or, equivalently, at $k_S = 0$, where $k_S = (S - \mu)/\sigma$).

**Proof.** See Appendix C.

This result is actually stronger than the proposition states, since at $S = \mu$ we have a stationary point regardless of the demand distribution. $S = \mu$ therefore maximizes the percentage reduction in variance from backlogging for any distribution for which %VR is a pseudo-concave function of $S$ (which, as indicated in the proof in Appendix C, holds for the normal distribution). Proposition 2 also implies that setting the base-stock level equal to the mean demand gives the minimum variance of shipment quantities available under a backlogging policy. These results show that the stock level set by the distributor directly affects the variability (and hence the cost) of transportation requirements through the distribution of shipments, or $f_q(Q_c)$ in Eq. (3). This implies that expected transportation costs are a function of the stock level, and the inventory and transportation models cannot be separated as in Ernst and Pyke (1993) model. Although the mean and variance of $f_q(Q_c)$ are used, the shape of this distribution may be difficult to characterize, as illustrated below.

The quantity shipped to the customer, $Q_c$, consists of two components, the amount shipped to satisfy demand in the current period plus the backlog remaining from the prior period. Let $Q_c$ and $Q_b$ denote the amount shipped to satisfy current demand and the amount shipped to satisfy the prior backlog. Note that $Q_c = \text{Min}\{x, S\}$ and $Q_b = [x - S]^+$, where $x$ is a random variable for single-period demand with the time subscript omitted for convenience. If demand is normally distributed, then for a fixed stock value, $S$, both $Q_c$ and $Q_b$ have distributions consisting of a truncated normal with a probability mass concentrated at
the points $S$ and 0, respectively (this mass equals $F(S)$ for the distribution of $Q_b$ and $F(S)$ for the distribution of $Q_e$). Fig. 1 illustrates the shape of the distributions for $Q_e$ and $Q_b$.

The distribution of the quantity shipped per period, $f_Q(Q)$, is determined by the convolution of $f(Q_e)$ and $f(Q_b)$ and does not, therefore, take the form of any standard distribution (and hence does not lead to tractable mathematical analysis). To obtain further insight into the shape of this distribution, a system with normally distributed demand containing a mean, $\mu$, equal to 1000 and standard deviation of demand, $\sigma$, equal to 300 is simulated over 2000 time periods. This system follows the policy of stocking $S$ units at the beginning of each period and allocating the minimum between $S$ and the current period demand to satisfy demand. Backlogged demand is shipped in the period following the shortage. Fig. 2 illustrates the shape of the distribution of $Q_s$ for three levels of $S$.

![Fig. 1. Distributions of the components of shipment quantity.](image1)

The results of this simulation provide interesting insights regarding the nature of the distribution of shipping quantities, $f_Q(Q)$. As expected, the distribution takes an approximately normal shape, with a mass concentrated at the value $S$. This is not surprising because in periods where demand exceeds $S$ and no prior backlog exists, exactly $S$ units are shipped. As the value of $S$ increases, the mass concentrated at that point decreases because the likelihood of demand in a period exceeding $S$ decreases. For $S = 1900$ ($k = 3$, not shown in the figure), the distribution of shipments looks exactly like the demand distribution.

The results of the simulation confirm our prior analytical results showing that the variance of shipments decreases as $S$ approaches $\mu$. Table 1 below compares the percentage decrease in variance of shipment quantities in the simulation to analytical results suggested by Eq. (8) for various levels of safety factor, $k$. Note that $k_S$ is the standardized value of $S$, i.e., $k_S = (S - \mu)/\sigma$, often referred to as the safety factor, where safety stock, $ss$, defined as the expected amount on-hand when a replenishment arrives, is given by the equation $ss = k_S\sigma$.

![Fig. 2. Relative frequencies of shipping quantities in simulation experiment.](image2)

The results of the simulation combined with our analytical results suggest that, except for the discrete probability mass concentrated at the stock level $S$, the distribution of $Q_e$ (under normal customer demand) can be effectively approximated by a normal distribution with mean $\mu$ and standard deviation $\sigma_b$. Under a mild condition on the distribution of shipment quantities as a function of $S$ (which is satisfied by the normal distribution, among many others), Proposition 3 shows that for a fixed value of transportation capacity, setting $S$ equal to the average single-period demand

<table>
<thead>
<tr>
<th>$k_S$</th>
<th>%VR Analytical</th>
<th>%VR Simulated</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>0.2292</td>
<td>0.1376</td>
</tr>
<tr>
<td>-2</td>
<td>3.4108</td>
<td>3.8301</td>
</tr>
<tr>
<td>-1</td>
<td>18.0514</td>
<td>19.1112</td>
</tr>
<tr>
<td>0</td>
<td>31.8310</td>
<td>31.0976</td>
</tr>
<tr>
<td>1</td>
<td>18.0514</td>
<td>17.6707</td>
</tr>
<tr>
<td>2</td>
<td>3.4108</td>
<td>3.8301</td>
</tr>
<tr>
<td>3</td>
<td>0.2292</td>
<td>0.1376</td>
</tr>
</tbody>
</table>
minimizes expected transportation costs (as given by Eq. (3)). This requires the definition of the expected number of units exceeding in-house truck capacity, \( n_q(T) \), given by

\[
n_q(T) = \int_T^\infty (Q_i - T)f_q(Q_i)\,dQ_i.
\]

**Proposition 3.** Assume the distributor has a fixed truck fleet capacity exceeding average customer demand per period. Then, for any distribution of shipments, \( f_q(Q_i) \), in which the expected number of units exceeding truck capacity \( T \) decreases as the variance of shipments decreases (as occurs, for example, under a normal distribution with a fixed mean), setting the stock level equal to the mean demand minimizes expected transportation costs given by Eq. (3).

**Proof.** We can rewrite Eq. (3) as

\[
K(T, S) = K_R T + K_R \mu + (K_C - K_R)\int_T^\infty (Q_i - T)f_q(Q_i)\,dQ_i,
\]

\[
= K_R T + K_R \mu + (K_C - K_R)n_q(T).
\]

Proposition 2 implies that setting \( S = \mu \) minimizes \( \sigma_S^2 \), which equals \( \text{Var}[Q_i] \). Under the conditions of the proposition, the minimum value of \( \text{Var}[Q_i] \) also minimizes \( n_q(T) \), which in turn minimizes \( K(T, S) \) in Eq. (9) for a fixed value of \( T \). \( \square \)

### 3. A Heuristic method for setting stock levels

Recall that total expected cost for the distributor equals the sum of expected inventory and transportation costs. The stock level that minimizes expected inventory cost equals \( S^* \) as given by Eq. (1), while a stock level of \( S = \mu \) minimizes transportation costs under a distribution of shipments as described in Proposition 3. It is easily shown that expected inventory cost per period as given in Eq. (5) is a convex function of \( S \). However, transportation costs cannot be shown to be convex in \( S \) using the cost equation given in Eq. (9) due to the complex functional form of \( n_q(T) \), which leads us to use the convex approximation for \( n_q(T) \) described in the next section.

#### 3.1. Approximation scheme for \( n_q(T) \)

Assume that the distribution of shipment quantities can be approximated by a normal distribution. The results of the simulation study in Section 2.2 and Fig. 1 indicate that the normal distribution provides a close approximation for the distribution \( f_q(Q_i) \). The loss of accuracy due to this approximation decreases as the stock level \( S \) (or equivalently, \( k_S \)) increases, and approaches zero as \( S \) approaches infinity. Under a normal distribution, we can rewrite \( n_q(T) \) as \( \sigma_b L(k_T) \), where \( \sigma_b \) is given in Eq. (8), \( k_T \) is the standardized value of truck capacity, i.e., \( k_T = (T - \mu)/\sigma_b \), and \( L(k_T) \) is the standard normal loss function (see Nahmias, 1997). Then \( \sigma_b \) can be approximated as a linear function of \( k_S \) and \( L(k_T) \) as an exponential function of \( k_T \).

Eq. (8) can be rewritten as

\[
\sigma_b = \sigma\sqrt{1 - 2[L(k_S)]^2 - 2L(k_S)k_S} = \theta \sigma,
\]

where we let \( \theta \) denote the quantity

\[
\sqrt{1 - 2[L(k_S)]^2 - 2L(k_S)k_S}.
\]

Plotting \( \theta \) against values of \( k_S \) between zero and 2.5 (see Fig. 3) shows that a linear relationship provides an excellent fit, and using

![Fig. 3. Linear approximation for \( \theta \).](image-url)
optimal in-house fleet capacity that exceeded the logging case (our approximate model), resulted in expediting case (an exact model) and the backcomputational tests, which included testing the methods for approximating expected costs of similar to widely used and highly accurate heuristic ranges of the safety factor, higher service levels, or equivalently for smaller systems that operate under high service levels or that have due to the complexity of this term. For systems retail distribution operations), this approximated using in-house truck capacity (which is typical of an outside carrier cost that far exceeds the cost of proceeding in-house truck capacity per period, approximated is the expected number of units expected cost equation of both T and kS.

The only term in the expected cost equation approximated is the expected number of units exceeding in-house truck capacity per period, nq(T), due to the complexity of this term. For systems that operate under high service levels or that have an outside carrier cost that far exceeds the cost of using in-house truck capacity (which is typical of retail distribution operations), this approximated term will be a small component of total cost. Figs. 3 and 4 also indicate that both approximation methods increase in accuracy for systems with higher service levels, or equivalently for smaller ranges of the safety factor, k. This approach is similar to widely used and highly accurate heuristic methods for approximating expected costs of (Q, R) policies in continuous review inventory systems (see Hadley and Whitin, 1963). Finally, computational tests, which included testing the expediting case (an exact model) and the backlogging case (our approximate model), resulted in optimal in-house fleet capacity that exceeded the mean period demand by at least two standard deviations. This reinforces the idea that the expected cost of shipping via common carrier will be a relatively small component of total expected cost, with the approximation scheme generating a small loss of accuracy.

To derive properties of the expected total cost function using the approximations \( \theta \) and \( L(k_T) \) for \( \theta \) and \( L(k_T) \), additional notation is needed for a more compact presentation. Letting \( \beta = (T - \mu)/\sigma \), we can write \( k_T = \beta/\theta \) (note that \( \beta > 0 \)). The expected number of units exceeding truck capacity, \( n_q(T) \), can then be written as

\[
n_q(T) = \sigma_L L(k_T) \approx c_5 \sigma \hat{e}_1, \tag{10}
\]

where \( \hat{e}_1 = \exp(-c_4 \beta/\hat{\theta}) \). The expected single-period cost equation under backlogging can then be written as a function of \( k_S \) and \( T \) using the approximations for \( \theta \) and \( L(k_T) \) as

\[
TC(T, k_S) = [c_1 + h k_S \sigma + (p + h) \sigma L(k_S)] \\
+ [K_R T + K_S \mu + (K_C - K_R) c_3 \sigma \hat{e}_1].
\tag{11}
\]

The first term in Eq. (11) represents expected inventory holding and shortage costs while the second represents expected transportation costs. For a fixed value of \( T \), the first- and second-order derivatives of the second term in the right-hand side of (11) with respect to \( k_S \) (using approximations) yield

\[
\frac{\partial K(k_S)}{\partial k_S} = (K_C - K_R)(c_2 \sigma c_3 \hat{e}_1(1 + c_4 \beta/\hat{\theta}) > 0,
\]

and

\[
\frac{\partial^2 K(k_S)}{\partial k_S^2} = (K_C - K_R)(c_2 \sigma c_3 \beta)^2 \hat{e}_1 \sigma \hat{e}_1 > 0,
\]

which implies that, under these approximations, transportation cost is a strictly convex function of the safety factor, \( k_S \), and is increasing over the range of interest (\( k_S = [0, 2.5] \)).

The approximations also lead to the conclusion that \( k_S = 0 \) minimizes expected transportation cost, \( K(k_S) \), over the range \( k_S \geq 0 \). The following proposition shows the joint convexity of \( TC(T, k_S) \) in \( T \) and \( k_S \) and motivates the subsequent heuristic solution procedure.
Proposition 4. Under normally distributed demand, unique values of $T$ and $k_S$ exist that minimize expected single-period holding, shortage, and transportation costs given by Eq. (11) over the region $T, k_S \geq 0$.

Proof. See Appendix D.

3.2. Heuristic solution procedure

Under a fixed fleet capacity, $T$, finding a $k_S$ that satisfies the first-order condition of $TC(k_S)$ involves a cdf and an exponential function of $k_S$, and a search method to find the optimal $k_S$, which we denote by $k_S^T$. The search begins with $k_S^0$, then increases $k_S$ as long as the total cost function decreases, and stops when the total cost begins to increase.

The approach for minimizing Eq. (11) in $S$ and $T$ assumes available truck capacity in discrete multiples of a base truck capacity, $T_0$ at least as great as the average period demand. The procedure defines multiples of $T_0$ between $mT_0$ and $(m + k)T_0$, where $m$ is the smallest non-negative integer such that $mT_0 \geq \mu$, and $k$ is the smallest non-negative integer such that $(m + k)T_0 \geq \mu + 3\sigma$. The retained solution gives the minimum value of $TC(T, k_S^*(T))$ among all $T \in \{mT_0, (m + 1)T_0, \ldots, (m + k)T_0\}$. This algorithm can be easily implemented on a spreadsheet; furthermore, since it is likely that $0 \leq k < 2.5$, the searching technique will quickly identify the optimal solution with respect to the approximate cost equation (11).

4. Computational results

Denote minimum total cost under backlogging and expediting as $TC_b^*$ and $TC_e^*$, respectively, where $TC_b^*$ is obtained by minimizing (11) and $TC_e^*$ is obtained by minimizing the sum of (3) and (5) when $p$ is replaced by $e' = e - c$. The percentage cost savings due to backlogging, $\omega$, is then calculated using the following formula:

$$\omega = \left[1 - \frac{TC_b^*}{TC_e^*}\right] \times 100\%.$$

Based on experience with a large USA distributor, the following parameters are chosen as a basis for the numerical study, where costs are per pound and per week for time-sensitive costs such as $h, p,$ and $K_{RL}$, and $P$ denotes the price per pound paid by the customer to the distributor:

$$c \quad P \quad h \quad p \quad e'$$

<table>
<thead>
<tr>
<th></th>
<th>0.5</th>
<th>2c</th>
<th>0.005c</th>
<th>0.01c</th>
<th>$K_C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>100,000</td>
<td>$\sigma$</td>
<td>$K_{RL}$</td>
<td>$K_R$</td>
<td>$K_C$</td>
</tr>
<tr>
<td></td>
<td>30,000</td>
<td>0.007817</td>
<td>0.01</td>
<td>0.50</td>
<td></td>
</tr>
</tbody>
</table>

Assume a standard truck capacity, $T_0$, of 40,000 lbs and that the distributor operates on a 50-week year. The cost per lb, $c$, is incremented by 10% of its base value up to $1/lb and the cost per lb for shipping via common carrier, $K_C$, is incremented in steps of $0.50 up to $5/lb, resulting in $10 \times 5 (= 50)$ combinations. The 50 problems solved and the percentage cost savings are plotted in the three-dimensional surface shown in Fig. 5. The potential percentage savings are always positive (even for the case in which the unit backlogging cost equals the unit expediting cost) and in general significant, taking a value of 3.6% for our base case ($c = \$0.50/lb, K_C = \$0.50/lb$) and exceeding 7% in several cases. In addition, for a given cost per lb for using a common carrier, $K_C$ (or equivalently, the unit incremental expediting cost, $e'$), the percentage cost savings reaches its maximum within the chosen range of unit product cost, $c$. Fig. 5 also illustrates that cost savings become less sensitive to changes in $c$ as $K_C$ increases, while at high val-

![Fig. 5. Percentage cost savings resulting from backlogging policy.](image-url)
ues of unit cost, savings become extremely sensitive to changes in $K_C$.

Therefore the distributor can enjoy significant savings by employing a backlogging policy, assuming the customer is willing to wait for demand satisfaction. To persuade the customer to follow this backlogging policy, some form of incentive such as a product discounting scheme can be an effective strategy to compensate for potential customer loss. The distributor’s expected profit per period is determined by the difference between total revenue and expected cost per period, and can be written as $\text{Profit}_e = P\mu - TC_e^c$ if unmet demand is expedited, and $\text{Profit}_b = P\mu - TC_b^c$ if unmet demand is backlogged. The reduction in total cost in the backlogging case will allow the distributor to provide a discount on the selling price, $P$, to induce the customer to follow the backlogging policy. Let this discounted sales price be written as $P(1 - d^\ast)$, where $d^\ast$ denotes the discount rate. Thus, the maximum discount rate the distributor can offer should satisfy the condition that maximum $\text{Profit}_e = \text{Profit}_b$, that is,

$$P\mu - TC_e^c = P(1 - d_{\max})\mu - TC_b^c \quad \text{or}$$

$$d_{\max} = \frac{(TC_e^c - TC_b^c)/P\mu}. \quad (14)$$

Maximum possible discount rates for each problem are reported in Table 2. The highest allowable discount rate increases with the common carrier’s charge per lb, $K_C$, but decreases with the product cost per lb, $c$. This is not surprising since $c^\ast = K_C$ and as this value increases, the distributor can enjoy greater benefit from a backlogging policy and should, therefore, offer a higher discount rate. On the other hand, as $c$ increases, the distributor’s profit margin decreases (for a fixed value of $P$), and lower discount rates should be offered.

5. Concluding remarks

This paper studied a two-stage base-stock distribution system in which a distributor holds inventory and delivers products to a customer. Expediting and backlogging policies are considered to deal with supply shortages. In contrast to existing work, decisions on transportation capacity and base-stock levels are not made independently. A backlogging policy leads to reduced variability in transportation requirements and hence transportation cost. Total costs under the two policies are compared through a set of numerical examples using a new heuristic method for jointly setting stock levels and transportation capacity. The numerical examples demonstrate that the reduction in total cost resulting from a backlogging policy can be significant and can, therefore, serve as an effective delivery negotiation tool between a distributor and customer. The results also show that, when deciding between backlogging and expediting supply shortages, it is insufficient to simply consider the difference between the unit costs of expediting and backlogging and that additional savings in the form of reduced transportation costs should be included in the analysis.

### Appendix A

This appendix shows that

$$\text{Var}[Q] = \sigma_b^2 = \sigma^2 - 2n(S)(n(S) + (S - \mu)).$$

Note that

$$\text{Var}[Q] = \text{Var}[\langle x_r - S \rangle^\ast] + \text{Var}[\text{Min}\{x_r, S\}],$$

<table>
<thead>
<tr>
<th>$c$</th>
<th>$K_C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.06</td>
</tr>
<tr>
<td>1.0</td>
<td>0.07</td>
</tr>
<tr>
<td>1.5</td>
<td>0.08</td>
</tr>
<tr>
<td>2.0</td>
<td>0.09</td>
</tr>
<tr>
<td>2.5</td>
<td>0.10</td>
</tr>
<tr>
<td>3.0</td>
<td>0.11</td>
</tr>
<tr>
<td>3.5</td>
<td>0.12</td>
</tr>
<tr>
<td>4.0</td>
<td>0.13</td>
</tr>
<tr>
<td>4.5</td>
<td>0.14</td>
</tr>
<tr>
<td>5.0</td>
<td>0.15</td>
</tr>
</tbody>
</table>
\[ E[(x - S)^+] = n(S), \]

and
\[ E[\text{Min}\{x, S\}] = \mu - n(S). \]

Using the equation for variance
\[(\text{Var}[x] = E[|x - E[x]|^2])\]
we obtain
\[ \text{Var}[Q_n] = \sigma_n^2 \]
\[ = \int_{-\infty}^{\infty} (x - \mu + n(S))^2 f(x) \, dx \]
\[ + \int_{-\infty}^{S} (S - \mu + n(S))^2 f(x) \, dx \]
\[ + \int_{-\infty}^{S} (0 - n(S))^2 f(x) \, dx \]
\[ + \int_{S}^{\infty} (x - S - n(S))^2 f(x) \, dx. \]

By squaring the terms in parenthesis and simplifying we get
\[ \sigma_n^2 = \int_{-\infty}^{\infty} x^2 f(x) \, dx + \mu^2 + 2n^2 - 2\mu n \]
\[ + (2n - 2\mu) \int_{-\infty}^{S} xf(x) \, dx \]
\[ + \int_{-\infty}^{S} (2S^2 - 2\mu S + 4nS - 2xS - 2xn) f(x) \, dx, \]

where \( n = n(S) \) to simplify the notation. The above equation is equivalent to
\[ \sigma_n^2 = \int_{-\infty}^{\infty} x^2 f(x) \, dx + \mu^2 + 2n^2 - 2\mu n \]
\[ + (2n - 2\mu) \left( \mu - \int_{-\infty}^{\infty} xf(x) \, dx \right) \]
\[ + \int_{S}^{\infty} (2S^2 - 2\mu S + 4nS - 2xS - 2xn) f(x) \, dx. \]

Noting that \( \sigma^2 = \int_{-\infty}^{\infty} x^2 f(x) \, dx - \mu^2 \), we can simplify the above to
\[ \sigma_n^2 = \sigma^2 + 2n^2 - (4n - 2\mu + 2S) \int_{S}^{\infty} xf(x) \, dx \]
\[ + (2S^2 - 2\mu S + 4nS) \hat{F}(S), \]

where \( \hat{F}(\cdot) \) denotes the complementary CDF of demand. Next note that \( n(S) = \int_{S}^{\infty} xf(x) \, dx - S \hat{F}(S), \)

allowing the above equation to be rewritten as
\[ \sigma_n^2 = \sigma^2 + 2n^2 - (4n - 2\mu + 2S) \int_{S}^{\infty} xf(x) \, dx \]
\[ + (2S^2 - 2\mu S + 4nS) \hat{F}(S) \]
\[ + (4n - 2\mu + 2S) S \hat{F}(S) \]
\[ - (4n - 2\mu + 2S) S \hat{F}(S), \]

which simplifies to
\[ \sigma_n^2 = \sigma^2 + 2n^2 - (4n - 2\mu + 2S) \]
\[ \times \left[ \int_{S}^{\infty} xf(x) \, dx - S \hat{F}(S) \right] \]
\[ = \sigma^2 + 2n^2 - (4n - 2\mu + 2S)n \]
\[ = \sigma^2 - 2n^2 - 2n(S - \mu) \]
\[ = \sigma^2 - 2n(S) \{n(S) + (S - \mu)\}, \]

completing the proof. \( \square \)

**Appendix B**

**Proposition 1.** Under normally distributed demand, the maximum value of percentage variance damping, \( \%\text{VR} \), is the same regardless of the parameters of the distribution and is equal to approximately 68.2%.

**Proof.** Define \( k_S \) using the equation \( k_S = (S - \mu)/\sigma \) and let \( \phi(u) \) denote the pdf of the standard unit normal distribution. Nahmias (1997) shows that, for a normal distribution, \( n(S) = \sigma L(k_S) \), where \( L(k) = \int_{-\infty}^{\infty} (u - k) \phi(u) \, du \) is the standard normal loss function. From the definition of \( \%\text{VR} \) we have
\[ \%\text{VR} = \left( 1 - \frac{2n(S)\{n(S) + S - \mu\}}{\sigma^2} \right) \times 100\% \]
\[ = (1 - 2L(k_S)\{L(k_S) + k_S\}) \times 100\%. \quad (B.1) \]

Eq. (B.1) shows that, under the normal demand distribution assumption, the maximum percentage variance reduction is a function of the safety factor, \( k_S \) and does not depend on the parameters of
the normal demand distribution (although the corresponding stock level, $S$, depends on these parameters). For a normal distribution, $\%VR$ is maximized at $k_S = 0$ at which point Eq. (B.1) gives $\%VR = 68.2\%$. □

**Appendix C**

**Proposition 2.** Under normally distributed demand, the maximum $\%VR$ occurs at $S = \mu(k_S = 0)$.

**Proof.** Maximizing $\%VR$ is equivalent to maximizing $g(S) = \sigma^2 - \sigma^2 S$, since $\sigma^2$ is constant. This is an unconstrained maximization of the pseudo-concave function, $g(S)$. If a stationary point exists, it maximizes $g(S)$ due to the pseudo-concavity of $g(S)$ (see Bazaraa et al., 1993). Therefore

$$
\frac{dg(S)}{dS} = -4n(S)\hat{F}(S) + 2n(S) - 2(S - \mu)\hat{F}(S),
$$

where $\hat{F}(\cdot)$ denotes the complementary cumulative distribution function of the demand distribution. Observing that $\hat{F}(\mu) = 0.5$, it follows that $\frac{dg(S)}{dS} = 0$ at $S = \mu$. □

**Appendix D**

**Proposition 4.** Under normally distributed demand, unique values of $T$ and $k_S$ exist that minimize the sum of expected single-period inventory holding, shortage, and transportation costs as given by Eq. (11) over the region $T$, $k_S \geq 0$.

**Proof.** Eq. (11) can be shown to be jointly strictly convex in $T$ and $k_S$.

$$
TC(T, k_S) = [c \mu + h_k \sigma + (p + h)\sigma L(k_S)] + [K_R T + K_R \mu + (K_C - K_R)\sigma c_3 e_1].
$$

(11)

Note that $e_1 = \exp(-c_4 \beta / \hat{\theta})$, $\beta = (T - \mu)/\sigma$, and $\hat{\theta} = c_1 + c_2 k_S$. The following can be derived:

$$
\frac{\partial TC}{\partial T} = K_R L - (K_C - K_R) c_3 c_4 e_1,
$$

$$
\frac{\partial^2 TC}{\partial T^2} = (K_C - K_R) c_3 c_4^2 e_1 / \sigma \hat{\theta},
$$

$$
\frac{\partial TC}{\partial k_S} = h \sigma + (p + h)\sigma (\Phi(k_S) - 1) + (K_C - K_R) \sigma c_2 c_3 e_1 (1 + c_4 \beta / \hat{\theta}),
$$

when

$$
\frac{\partial^2 TC}{\partial k_S^2} = (p + h)\sigma \phi(k_S) + (K_C - K_R) \sigma c_3 (c_2 c_4 \beta)^2 e_1 / \hat{\theta}^3,
$$

and

$$
\frac{\partial^2 TC}{\partial k_S \partial T} = \frac{\partial^2 TC}{\partial T \partial k_S} = -(K_C - K_R) \sigma c_2 c_3 c_4^2 \beta e_1 / \hat{\theta}^2.
$$

Proving strict convexity requires the following conditions to hold:

$$
\frac{\partial^2 TC}{\partial T^2} > 0, \frac{\partial^2 TC}{\partial k_S^2} > 0
$$

and

$$
\frac{\partial^2 TC}{\partial T^2} \times \frac{\partial^2 TC}{\partial k_S^2} - \left[\frac{\partial^2 TC}{\partial T \partial k_S}\right]^2 > 0
$$

(see Bazaraa et al., 1993). The first two of these follow from the form written above since each of the corresponding terms is positive in each equation. Note that

$$
\frac{\partial^2 TC}{\partial T^2} \times \frac{\partial^2 TC}{\partial k_S^2} - \left[\frac{\partial^2 TC}{\partial T \partial k_S}\right]^2
$$

$$
= (K_C - K_R) c_3 c_4^2 e_1 (p + h)\phi(k_S) / \hat{\theta},
$$

which is always greater than zero and strict convexity follows. The fact that $TC(0,0)$ is finite and $TC(T,k_S)$ approaches infinity as $T$ and $k_S$ go to infinity imply that a minimizing solution exists in the region $T$, $k_S \geq 0$ and the strict convexity of $TC(T,k_S)$ therefore implies the existence of a unique minimizing solution (see Bazaraa et al., 1993). □

**References**

