Continuous review inventory models where random lead time depends on lot size and reserved capacity

Metin Çakanyıldırım¹, James H. Bookbinder, Yigal Gerchak*

Department of Management Sciences, University of Waterloo, Waterloo, Ont., Canada N2L 3G1

Received 6 April 1998; accepted 2 August 1999

Abstract

The processing time of large orders is, in many industries, longer than that of small orders. This renders supply lead times in such settings to be increasing in the order size. Yet that pattern is not reflected in existing inventory control models, especially those allowing for random lead times. This work aims at rectifying the situation. Our setting is an order-quantity/reorder-point model with backordering, where the shortage penalty is incurred per unit per unit time. The processing time of each unit is random; the processing time of a lot is correlated with its size. For the case where lead time is proportional to the lot size, we obtain a closed-form solution. That is, unlike the classical \((Q,r)\) model (where lead time is independent of lot size), no iterations are required here. We also analyze a case where the processing time exhibits economies of scale in the lot size. Finally, we consider a situation where a customer can secure shorter processing times by reserving capacity at the supplier’s manufacturing facility. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Inventory control; Continuous review; Random lead time

1. Introduction

The goal of this work is to explore the implications, in a continuous review inventory model, of random lead times being contingent on lot size. It is motivated by the rather obvious observation that manufacturing larger lots often requires longer time. While waiting time and machine setup time are usually independent of lot size, the actual processing-time portion of lead time could well depend on the lot size in many industries which manufacture discrete parts or products. Yet existing inventory control literature (e.g. [1–3]) does not explicitly incorporate lot-size dependency into the way it models lead times. We aim at rectifying this situation.

Some indirect implications of replenishment policies for supply lead times have been investigated before. In references on multiple sourcing when lead times are random (e.g. [4]), order splitting among several suppliers causes one of the orders to tend to arrive earlier than had there been a single supplier, thereby reducing effective lead time. Selecting suppliers by the length and/or variability of their lead times has been explored by Gerchak
and Parlar [5], Choi [6] and Bookbinder and Çakanyıldırım [7]. But neither strand of the literature allowed a given supplier's lead time to be dependent on (i.e. increasing in) the lot size.

Our setting is a \((Q,r)\) order-quantity/reorder-point model with backordering, where the shortage penalty is incurred per unit per unit time. Demand rate is assumed constant, while the lead-time is random. In line with most researchers, we assume that immediately after the arrival of an order the installation stock will always exceed the reorder level, so at most one order will be outstanding at any time. The processing time of a single-unit order is a random variable, \(T\); the processing time of a lot of \(Q\) units is assumed to be a multiple of \(T, Q^\theta T\), where the value of the positive parameter \(\theta\) indicates the extent of economies of scale and/or learning effects, if any, in production speed. We focus specifically on the values \(\theta = 1\) (no economies of scale) and \(\theta = 1/2\) (positive economies).

For the case \(\theta = 1\), in Section 3 we prove joint convexity in \((Q,r)\) and obtain closed-form expressions for the decision variables. Thus, as opposed to the classical lot-size-independent lead-time model (which might be viewed as corresponding to \(\theta = 0\) in our setting), our “linear” model does not require an iterative procedure for obtaining the solution. The case \(\theta = 1/2\) is more complicated to analyze; there we establish joint convexity within a certain range. An example (Section 4) is then used to show that the joint-convexity range may indeed be large enough to guarantee global optimality of the solution to the necessary conditions. That solution, however, must be obtained in an iterative fashion, which we illustrate. Numerical examples are presented in Section 5, comparing solutions for \(\theta = 1\) and \(1/2\) and conducting some sensitivity analyses with respect to the cost parameters.

In the previously described models, the time needed to produce a single item was assumed to be an exogenous random variable. Our final model relaxes this assumption in Section 6 via the use of reserved capacity. The idea of using supply contracts which “reserve” capacity at the supplier for a particular customer, thereby ensuring timely supply, appears to be growing in popularity, and has recently been explored by several researchers, including Karmarkar et al. [8], Silver and Jain [9], Jain and Silver [10] and Anupindi and Bassok [11]. Our model assumes that the capacity available for a particular order is the sum of the reserved production rate and an additional random rate (which will depend on the supplier’s situation at the moment). A single item’s processing time is then assumed to be the reciprocal of that sum of capacities. Random capacity was previously envisioned as restricting output quantity [12–14]. We, on the other hand, view it as a factor in determining the lead time.

For any level of reserved capacity, one can find the best values of \(Q\) and \(r\) using the methods outlined earlier. A search for the optimal capacity to reserve is then conducted. We outline a procedure, but leave the detailed investigation of this model for future research.

2. The model and its properties

Production costs are not explicitly included, since all demand is eventually met. An inventory holding cost of \(h\) per unit per unit time is assumed, and a fixed cost of \(K\) is incurred for each order placed. Also, a shortage cost of \(\tilde{n}\) per unit per unit time is incurred when a demand is backlogged. For reasons that will become clear, let us for the moment denote by \(\hat{Q}\) and \(\hat{r}\) the order quantity and reorder point.

Let \(g\) be the probability density of lead time. A typical realization of installation stock is depicted in Fig. 1. From it, by conditioning on whether or not a stockout occurs, the expected holding and
shortage costs per cycle can be computed as
\[
\left(\frac{h}{2D}\right) \int_{x=0}^{r/D} \left[(\hat{r} - Dx + \hat{Q})^2 - (\hat{r} - Dx)^2\right] g(x) \, dx
\]
\[
\quad + \left(\frac{h}{2D}\right) \int_{x=r/D}^{x=\hat{r}/D} (\hat{r} - Dx + \hat{Q})^2 g(x) \, dx
\]
\[
\quad + \left(\frac{\hat{r}}{2D}\right) \int_{x=\hat{r}/D}^{x=\infty} (Dx - \hat{r})^2 g(x) \, dx
\]
= \left(\frac{h}{2D}\right) \int_{x=0}^{r/D} (\hat{r} - Dx)^2 g(x) \, dx
\]
\[
\quad + \left(\frac{\hat{r}}{2D}\right) \int_{x=\hat{r}/D}^{x=\infty} (Dx - \hat{r})^2 g(x) \, dx
\]
\[
\quad + \left(\frac{h}{2D}\right) \int_{x=0}^{\infty} (\hat{r} - Dx + \hat{Q})^2 g(x) \, dx.
\]

We shall now scale the constant demand rate to unity; the decision variables \(\hat{Q}\) and \(\hat{r}\) will then be expressed in units of time rather than quantity. For illustration, suppose that \(D = 10\) packets per day, \(\hat{Q} = 50\) packets, \(\hat{r} = 20\) packets. We may scale demand to unity by visualizing a single day’s demand (10 packets) as one quantity-unit. Let \(Q\) and \(r\) denote \(\hat{Q}\) and \(\hat{r}\), respectively, after scaling, then \(Q = 5\) days’ demand and \(r = 2\) days’ demand. In general, \(Q\) will be \(\hat{Q}/D\) day’s demand and \(r\) will be \(\hat{r}/D\) days’ demand. Thus the dimensions of \(Q\) and \(r\) are time, not quantity any more.

To maintain accuracy in the costs, we have to scale them as well. Letting \(h\) and \(\pi\) be the costs after scaling, we have \(\hat{h} = Dh\) and \(\hat{\pi} = Dr\). The scaled holding and shortage costs per cycle are then
\[
- h \int_{x=0}^{r} (r - x)^2 g(x) \, dx/2
\]
\[
+ \pi \int_{x=r}^{\infty} (x - r)^2 g(x) \, dx/2
\]
\[
+ h \int_{x=0}^{\infty} (r - x + Q)^2 g(x) \, dx/2.
\]

Thus by the renewal reward theorem (e.g. [15]), the expected cost per unit time is
\[
E\{C(Q, r)\} = \left\{K - h \int_{x=0}^{r} (r - x)^2 g(x) \, dx/2
\]
\[
+ \pi \int_{x=r}^{\infty} (x - r)^2 g(x) \, dx/2
\]
\[
+ h \int_{x=0}^{\infty} (r - x + Q)^2 g(x) \, dx/2\right\} / Q
\]
\[
= K/Q + (\pi + h) \int_{x=r}^{\infty} (x - r)^2 g(x) \, dx/2Q
\]
\[
+ h \int_{x=0}^{\infty} (r - x + Q/2) g(x) \, dx. \tag{2}
\]

The random lead time \(L(Q)\) for processing a lot of size \(Q\) is assumed to equal
\[
L(Q) = Q^\theta T, \quad \theta \geq 0,
\]
where \(T\) is the processing time of a single-unit lot. Clearly, \(L(Q)\) is stochastically increasing in \(Q\).

Let \(G\) denote the distribution function of lead time. \(G\) and \(g\) can respectively be expressed in terms of the distribution \(F\) and probability density function \(f\) of the processing time \(T\) of a single-unit lot, as follows:
\[
G_Q(x) = P\{L(Q) \leq x\} = P(Q^\theta T \leq x)
\]
\[
= P(T \leq x/Q^\theta) = F(x/Q^\theta)
\]
and therefore
\[
g_Q(x) = f(x/Q^\theta)/Q^\theta.
\]

Substituting in terms of the distribution of \(T\), we have
\[
E\{C(Q, r)\}
\]
\[
= K/Q + (\pi + h) \int_{x=r}^{\infty} (x - r)^2 f(x/Q^\theta) \, dx/2Q^{\theta+1}
\]
\[
+ h \int_{x=0}^{\infty} (r - x + Q/2) f(x/Q^\theta) \, dx/Q^\theta. \tag{3}
\]

This completes our formulation of the problem.

Before discussing convexity properties and solution algorithms, let us make the following remarks.

In modeling the lead time as \(L(Q) = Q^\theta T\), we have
neglected the portion \( t_0 \) of lead time that is independent of \( Q \). The total lead time is thus \( t_0 + L(Q) \), where \( t_0 \) includes allowances for materials handling, waiting and setup. Although \( t_0 \) will often be the major component of total lead time, there are many circumstances where \( t_0 \) can be approximated as constant, independent of the random variable \( T \).

Bookbinder and Çakanyildirim [7] then show that \( m(x) \), the probability density function of lead time, accounting for the lead-time density \( g(x) \) pertinent to only the portion \( L(Q) \). They found therefore that the optimal \( Q \) was unaffected by the shift, while \( r \) was simply increased by \( t_0 \). This is why, in the present paper, we have concerned ourselves with just the \( L(Q) \)-segment of lead time.

Returning now to finding the optimal pair \((Q, r)\) for objective function (3), one seeks critical points of the expected-cost function over \( \{(Q, r) \in \mathbb{R}^2 : Q \geq 0, r \geq 0\} \). Such a search will converge to the global minimum if joint convexity of (3) can be proven. First, however, we shall study convexity of that portion of \( E[C(Q, r)] \) corresponding to the expected number of units backordered. It is worth mentioning that Zipkin [16] proved convexity of this function for a lot-size-independent lead time.

Letting \( y = x/Q^\theta \) in (3), the expected number of units backordered is given by

\[
B(Q, r) = \int_{y=r/Q^\theta}^{\infty} (yQ^\theta - r)^2 f(y) \, dy/2Q
= \int_{y=r/Q^\theta}^{\infty} Q^{2\theta-1}(y - r/Q^\theta)^2 f(y) \, dy/2.
\]

**Proposition 1.** For \( \theta = 1 \), the expected number of units backordered is jointly convex in \((Q, r)\).

**Proof.** Taking second partial derivatives of the function \( B(Q, r) \), the Hessian matrix is found to be

\[
H = \begin{pmatrix}
F(r/Q) - rF(r/Q)/Q & -r(1 - F(r/Q))/Q^2 \\
-r(1-F(r/Q))/Q^2 & 1 - F(r/Q)/Q
\end{pmatrix},
\]

(5)

Since the term \( 1 - F(r/Q) \) is non-negative, and the determinant of the Hessian matrix is zero, the Hessian is positive semidefinite (e.g. [17]). Hence, the expected-units-backordered cost function is jointly convex. □

To help us analyze the case \( \theta = 1/2 \), we let \( \delta = r/Q^\theta \), so

\[
B(Q, \delta) = Q^{2\theta-1} \int_{y=\delta}^{\infty} (y - \delta)^2 f(y) \, dy/2.
\]

**Proposition 2.** If \( \theta = 1/2 \), the expected number of units backordered is jointly convex in \((Q, \delta)\).

**Proof.** For that value of \( \theta \), \( B(Q, \delta) \) reduces to

\[
B(\delta) = \int_{y=\delta}^{\infty} (y - \delta)^2 f(y) \, dy/2.
\]

Thus convexity reduces to showing that the second derivative of (7) with respect to \( \delta \) is positive, which is immediate. □

We have so far proved convexity of the expected number of units backordered. To consider convexity of the entire objective function, let \( \mu = E(T) \). The objective (3) takes the following form when \( \theta = 1 \):

\[
E[C(Q, r)] = K/Q + (\pi + h)B(Q, r)
+ h(r + Q/2 - \mu Q).
\]

Using Proposition 1, it is easy to show:

**Corollary 1.** For \( \theta = 1 \), the expected-cost function (8) is jointly convex in \((Q, r)\).

By similar arguments, the objective function for \( \theta = 1/2 \) turns out to be

\[
E[C(Q, \delta)] = K/Q + (\pi + h)B(\delta)
+ h(\delta Q^{1/2} + Q/2 - \mu Q^{1/2}).
\]

Looking at (9), term by term, in conjunction with Proposition 2, reveals that all terms are always convex in \( \delta \), and are convex in \( Q \) if \( \delta \leq \mu \). Hence we have
Corollary 2. For $\theta = 1/2$, the expected-cost function is convex in $\delta$, and, if $\delta \leq \mu$, it is also convex in $Q$.

We now examine the Hessian of total expected cost for $\theta = 1/2$. After some algebra, that Hessian turns out to equal

$$
\begin{pmatrix}
(\pi + h)(1 - F(\delta)) & h/2Q^{1/2} \\
2K/Q^3 + (\mu - \delta)h/4Q^{3/2} & h/2Q^{1/2}
\end{pmatrix}.
$$

(10)

Without substituting the optimality conditions, it is not easy to show positive definiteness of (10). We therefore defer this discussion to Section 4, where we obtain the critical points for concave processing time. Let us first study the necessary conditions when $\theta = 1$.

3. Critical points for linearly varying processing time

Since we established the convexity of expected costs, critical points of the objective function (8) will be global minimizers. Proposition 3 characterizes those points through first-order optimality conditions for $Q$ and $r$. The proof follows directly by differentiating the function, plus some algebra.

Proposition 3. The optimality conditions are

$$
\int_{x=r/Q}^{\infty} (x - r/Q)f(x) \, dx = h/(\pi + h)
$$

(11)

and

$$
-K/Q^2 + (\pi + h) \int_{x=r/Q}^{\infty} (x - r/Q)^2 f(x) \, dx/2
+ hr/Q + h/2 - h\mu = 0.
$$

(12)

At first glance, both (11) and (12) appear to be functions of $Q$ and $r$, and iterating between these two conditions would thus be unavoidable. However, an important observation eliminates the need for iterative solutions. Let us refer to $\delta = r/Q$ as the “single-unit reorder point”, since $\delta$ is the optimal reorder level when the lot size is one. The optimality conditions become

$$
\int_{x=\delta}^{\infty} (x - \delta)f(x) \, dx = h/(\pi + h),
$$

(13)

$$
Q^2 = K \left\{ \left( (\pi + h) \int_{x=\delta}^{\infty} (x - \delta)^2 f(x) \, dx/2
+ h\delta + h/2 - h\mu \right) \right\}.
$$

(14)

Then, minimization of expected costs can be achieved via the following algorithm:

1. Solve (13) for the single-unit reorder point, $\delta$.
2. Calculate the lot size, $Q$, by solution of (14).
3. The reorder point is $r = Q\delta$.

In contrast to solving for the optimal values of $Q$ and $r$ in the classical $(Q,r)$ model (e.g. [18]), corresponding to $\theta = 0$ in our general model, we obtained here (when $\theta = 1$) a one-pass procedure (not requiring iterations) to find the decision variables.

It is easy to see that the single-unit reorder point $\delta$ increases with the shortage cost $\pi$ and decreases with the holding cost $h$, and that the optimal lot size $Q$ increases in the setup cost $K$.

Let us illustrate our findings by specializing to exponential single-unit processing time.

Example 1. Suppose

$$
f(t) = \lambda \exp(-\lambda t), \quad t \geq 0.
$$

(15)

Note that $E(T)$ must be less than one, i.e. $\lambda > 1$; otherwise, in the long run, the production rate will not suffice to meet demand. For this density (15), the expected-cost expression becomes

$$
E[C(Q,r)] = K/Q + (\pi + h)Q(\exp(-\lambda r/Q))/\lambda^2
+ h(r + Q/2 - Q/\lambda).
$$

(16)

We thus obtain the following solution:

$$
\delta = -\left\lceil \ln[h\lambda/(\pi + h)]/\lambda \right\rceil,
$$

(17)

$$
Q^2 = 2K/h(2\delta + 1).
$$

(18)

Eqs. (17) and (18) reveal that $Q$ decreases in $\pi$, whereas $r(Q\delta)$ increases in $\pi$. 
4. Critical points for concave processing time
\( (\theta = 1/2) \)

Concave processing time appears to be more plausible in practice than linear processing time. Having already discussed convexity and established the necessary conditions in Corollary 2, we shall next investigate the critical points of (9). Proposition 4 will directly provide the optimality conditions for \( Q \) and \( \delta \). (\( \delta \) is still the “single-unit reorder point”, since \( \delta = r/Q^{1/2} \) remains the optimal reorder level when \( Q = 1 \).)

**Proposition 4.** The optimality conditions are
\[
\int_{y=\delta}^{\infty} (y-\delta)f(y)\,dy = hQ^{1/2}/(\pi + h), \tag{19}
\]
\[
Q^2 + (\delta - \mu)Q^{3/2} - 2K/h = 0. \tag{20}
\]

**Proof.** Eqs. (19) and (20) follow by differentiating (9) with respect to \( \delta \) and \( Q \), respectively, and setting each equal to zero. \( \square \)

Returning to the issue of the joint convexity of (9), our first result is

**Proposition 5.** The second derivative of the expected-cost function with respect to \( Q \) is positive where \( Q \) is critical, i.e. where (20) is satisfied.

**Proof.** From (10), this second derivative equals
\[
2K/Q^3 + (\mu - \delta)h/4Q^{3/2}. \tag{21}
\]

It follows that (21) is positive if and only if
\[
8K/h + (\mu - \delta)Q^{3/2} \tag{22}
\]
is positive. This is clearly so for any \( Q \) satisfying (20). \( \square \)

Corollary 2 implies that the expected costs are convex for \( \delta \) in a certain range, whereas Proposition 5 says that the expected costs are convex at critical \( Q \) values over all values of \( \delta \). We now assert the joint convexity of the expected costs.

**Proposition 6.** If \( \delta \leq F^{-1}(\pi/(\pi + h)) \), the Hessian (10) is positive definite at critical points \( Q \).

**Proof.** We need to show only that the determinant of (10) is positive at critical values of \( Q \) and over \( \delta \) satisfying the above condition. That determinant equals
\[
(\pi + h)(1 - F(\delta))(\mu - \delta)h/4Q^{3/2} + 2K/Q^3 - h^2/4Q^2
\]
\[
= [(\pi + h)(1 - F(\delta))(\mu - \delta)hQ^{3/2} + 8K]
\]
\[
- h^2Q^{21}/4Q^3. \tag{23}
\]

Next substitute \( Q^2h - 2K \) for \( (\mu - \delta)hQ^{3/2} \), since \( Q \) satisfies (20). After some algebra, we find that (23) is positive if and only if
\[
\pi - (\pi + h)F(\delta) + 6Kh/(6K + Q^2h) > 0, \tag{24}
\]
which is always true for the given range of \( \delta \). \( \square \)

We shall henceforth denote by \( A \) the region \( \{(Q, \delta) \in \mathbb{R}^2 : Q > 0, 0 < \delta \leq F^{-1}(\pi/(\pi + h))\} \).

Proposition 6 is important in the sense that it guarantees joint convexity at a critical point \( Q \) for any \( \delta \) in \( A \). Solution of (19) and (20) thus yields a local minimum if \( (Q, \delta) \in A \). Recall, however, that a differentiable function, convex at all its critical points in a region, can have at most one minimizer there. Thus if \( (Q, \delta) \in A \) for some critical \( Q \) and \( \delta \), then this point is a global minimizer in \( A \). So the larger the range \( A \), the more likely it is that a solution of (19) and (20) will yield the real minimizer — the global one. As our examples will reveal, that range is large enough for reasonable problem parameters.

Unlike the optimality condition (13), Eq. (19) does depend on \( Q \) hence we cannot avoid an iterative solution scheme. We apply coordinate descent (e.g. [17]) to obtain the critical points of (9). Our method is outlined below.

0. Initialize \( Q = EOQ \).
1. Given \( Q \), solve for \( \delta \) in (19).
2. With that \( \delta \), obtain \( Q \) via (20).
3. Return to Step 1 unless \( Q \) does not differ much from its previous value.
4. Find the reorder point from \( r = Q^{1/2} \delta \).
It is known that all coordinate-descent algorithms, minimizing a function with unique minimum along any coordinate direction, converge [17]. By Corollary 2, the expected costs are such a function for $\delta \leq \mu$. Therefore, over the region $\{Q > 0, 0 < \delta \leq \mu\}$, our algorithm converges to the pair $(Q, \delta)$ which simultaneously solves (19) and (20). Outside that region, convergence cannot be guaranteed.

Another issue is whether solving (19) and (20) indeed yields minimizers, as opposed to maximizers or saddle points. It is clear from Corollary 2 that the expected-cost function cannot be concave. Critical points are thus not maximizers, although they may be saddle points if $(Q, \delta) \notin A$. Therefore, to establish convexity outside the region $A$, it is necessary to evaluate the Hessian. To illustrate, let us again specialize to exponential single-unit processing time, as in Example 1.

**Example 2.** The expected costs will now equal

$$E[C(Q,r)] = K/Q + (\pi + h)\{\exp(-\lambda \delta)\}/\lambda^2 + h(\delta Q^{1/2} + Q/2 - Q^{1/2}/\lambda).$$

(25)

The optimality equations given in (19) and (20) respectively reduce to

$$\delta = -\{\ln(h\lambda Q^{1/2}/(\pi + h))\}/\lambda,$$

(26)

$$Q^2 + (\delta - 1/\lambda)Q^{3/2} - 2K/h = 0.$$  

(27)

We now eliminate $\delta$ from (27) by inserting (26) to obtain

$$Q^2 - \{\ln(h\lambda Q^{1/2}/(\pi + h)) + Q^{3/2}\}/\lambda - 2K/h = 0.$$  

(28)

Eq. (28) can be solved numerically for $Q$. For example, if $\lambda = 2$, $h = 10$, $\pi = 40$ and $K = 50$, we obtain $Q = 3.513$. The corresponding $\delta$ turns out to equal 0.144. (Note that since $\delta < \mu = 1/2$, an iterative coordinate descent would have converged here).

For the exponential density, the sufficient condition of Proposition 6 becomes

$$\exp(-\lambda \delta) \geq h/(\pi + h).$$

(29)

From (26), it is clear that (29) holds as long as $\lambda Q^{1/2} \geq 1$. Hence if processing time is exponential, there is a large range of $\delta$ for which expected costs will be jointly convex.

### 5. Numerical examples

In this section, we again assume an exponential distribution for the time $T$ to produce one unit, thereby extending Examples 1 and 2. We shall examine changes in the optimal lot size $Q$ and reorder point $r$ as the cost parameters vary, for both the linear and concave models.

#### 5.1. $\theta = 1$

Tables 1–3 illustrate results when lead time depends *linearly* on $Q$.

The reader may have wondered whether the dependence of lead time on lot size is a small effect which can be neglected. We now show with an example that this need not be so.
Consider two inventory managers, A and B, both with a cost structure $K = 40$, $h = 10$ and $\pi = 30$. In each case, the true relationship is that lead time depends linearly on $Q$ and the single-unit lead time is exponential with mean $\mu = 2/3$. Manager A recognizes the dependences of lead time on lot size. She thus finds, from the first row of Table 1, that $Q = 1.86$ and $r = 1.22$, for a (minimum) cost/unit time of $\$42.97$.

Manager B, however, employs the classical $(Q, r)$ model; he overlooks the relationship $L(Q)$. Suppose, in the first instance, B takes the lead-time for the entire lot as exponentially distributed with mean $2/3$. He will then obtain $Q = 3.34$ and $r = 0$, incurring a cost per unit time of $\$66.53$, $54\%$ above that of A.

Now suppose that Manager A tells B that she calculated the optimal lot size as 1.86, which made the expected lot lead time be 1.24. If Manager B takes lead time as exponential, but still independent of lot size, with mean 1.24, he will end up with $(Q = 4.33, r = 0.17)$ and an inventory cost of $\$76.30$ per unit time, $77\%$ higher than what A incurs. Comparing his cost performance to A’s, B may think that he wrongly anticipated short lead times. Hence, B may assume a stochastically longer lead time, say with expected value of 2.28 (twice the whole lot lead time A experiences). Then B will find $(Q = 6.24, r = 1.15)$, incurring $\$91.70$ per unit time, now $113\%$ more than A’s cost.

It is worth noting that B, failing to realize the dependence of lead time on lot size, tended to choose larger lot sizes, hence faced longer lead times than he would have thought. Not appreciating the lot-size dependence also caused B to incur excess costs.

Suppose now that manager C employs the classical $(Q, r)$ model (with per-unit-time shortage penalty), but the lead-time distribution used in each step of the iterations is the one corresponding to the current value of $Q$. (Iterations would begin with $Q = EOQ$, perhaps as in the model allowing shortages.) Thus C recognizes the correct form of lead time’s dependence on the lot size, but does so within optimality equations derived from a model which ignores such dependence.\(^2\)

If lead-time distribution is exponential, then the expected number of units backordered is

$$n(r) = e^{-\lambda r / \lambda}.$$  

The optimality conditions become

$$Q = \sqrt{2(K + \pi e^{-\lambda r / \lambda})} / h, \quad e^{-\lambda r / \lambda} = Q h / \pi.$$  

Thus

$$Q = 1/\lambda + \sqrt{1/\lambda^2 + 2K/h}$$

and

$$r = -\ln[h(1 + \sqrt{1 + 2\lambda^2 K/h}/\pi)]/\lambda.$$  

Since from the above

$$\lambda = Q/(Q^2/2 - K/h),$$

that will be the (reciprocal of the) mean lead-time for a particular $Q$.

We looked at some numerical examples and the costs turned out to be very high relative to the optimal costs. We feel, however, that this conclusion may be rather meaningless, since the models are not really comparable; the lead-time distribution, which is here consequential, cannot be made identical with an externally hypothesized distribution of $T$ in $QT$.

5.2. $\theta = 1/2$

Tables 4–6 show how the optimal $Q$ and $r$ respond to changes in cost parameters when the lead

\(^2\) We wish to thank Ton de Kok and Søren Glud Johansen for proposing this idea.
Table 4
Minimum expected costs and optimal decision variables with variation in \( K \), the setup cost per order, when \( \theta = 1/2, h = 10, n = 30 \)

<table>
<thead>
<tr>
<th>( K )</th>
<th>( \lambda = 1.5 )</th>
<th>( \lambda = 2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( Q )</td>
<td>( r )</td>
</tr>
<tr>
<td>40</td>
<td>3.21</td>
<td>0.47</td>
</tr>
<tr>
<td>50</td>
<td>3.61</td>
<td>0.43</td>
</tr>
<tr>
<td>60</td>
<td>3.97</td>
<td>0.39</td>
</tr>
</tbody>
</table>

*The optimal (unrounded) \( r \) equals zero.*

Table 5
Values of objective function (25) (\( \theta = 1/2 \)) and global minimizers. Sensitivity analysis with respect to the shortage cost \( \pi \) per unit per unit time, for \( h = 10 \) and \( K = 60 \)

<table>
<thead>
<tr>
<th>( \pi )</th>
<th>( \lambda = 1.5 )</th>
<th>( \lambda = 2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( Q )</td>
<td>( r )</td>
</tr>
<tr>
<td>30</td>
<td>3.97</td>
<td>0.39</td>
</tr>
<tr>
<td>45</td>
<td>3.67</td>
<td>0.83</td>
</tr>
<tr>
<td>60</td>
<td>3.52</td>
<td>1.14</td>
</tr>
</tbody>
</table>

*The optimal (unrounded) \( r \) equals zero.*

Table 6
Minimum expected costs (25) and optimal \( (Q, r) \) as we vary \( h \), the holding cost per unit per unit time. \( K = 60, n = 60 \)

<table>
<thead>
<tr>
<th>( h )</th>
<th>( \lambda = 1.5 )</th>
<th>( \lambda = 2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( Q )</td>
<td>( r )</td>
</tr>
<tr>
<td>10</td>
<td>3.52</td>
<td>1.14</td>
</tr>
<tr>
<td>15</td>
<td>3.04</td>
<td>0.75</td>
</tr>
<tr>
<td>20</td>
<td>2.76</td>
<td>0.52</td>
</tr>
</tbody>
</table>

For both cases (linear and concave), the changes in optimal values of the decision variables can be summarized as follows:

(a) \( Q \) increases in \( K \), and decreases in \( \pi \) and \( h \).
(b) \( r \) increases in \( \pi \), but decreases in \( h \).

These observations are intuitive [cf. comparative statics for the classical \((Q, r)\) model with time-independent shortage penalty in [19]].

The impact of the setup cost \( K \) on the reorder point \( r \) is, however, interesting. In the linear-lead-time case, \( r \) increases in \( K \) (Table 1), whereas for concave lead time \( r \) decreases in \( K \). This could be explained as follows. Eq. (13) reveals that in the linear model, the optimal single-unit reorder point \( d \) is actually independent of the setup cost \( K \). Since the optimal \( r \) is a product of \( d \) and \( Q \), Eq. (19), so when \( K \) rises, \( Q \) increases and \( d \) decreases. Thus the direction of change in \( r = Q^{1/2}d \) is not clear. Apparently, in our examples, the decline in \( d \) was great enough to compensate for the rise in the lot size \( Q \), causing the reorder point \( r \) to go down.

Another way to explain the preceding is that, when \( L(Q) = Q^{1/2}T \), increasing \( Q \) economizes on setups and reduces the average lead time per unit. When the average lead time per unit decreases, it is plausible that the reorder point also does.

One could interpret the parameter \( \lambda \) as a measure of capacity. It is noteworthy that by augmenting \( \lambda \) from 1.5 to 2 in each of Tables 1-6, the total costs decrease in every row. Therefore, it is quite reasonable to look for ways to increase capacity permanently, even at some expense.

6. Lot-size-dependent lead time under reserved production rate

Until now, we assumed that the stochastic lead time \( T \) to produce one unit was independent of the decision variables; in other words, it was exogenous. In this section, that duration will be made an endogenous random variable by establishing a link between the reserved production rate and lead time. Basically, we are motivated by the observation that larger reserved capacity should result in shorter processing time. Securing a reserved capacity at a cost is thus a reasonable way of reducing the processing time of one’s orders.
We envision the lead time to be inversely related to capacity. The capacity allocated to a particular order consists of an amount \( x \) dedicated (contracted-for) to that customer, and a random portion \( C \), which depends on congestion at the moment and commitments to other customers. Thus

\[
L(Q) = Q/(x + C).
\]

Suppose \( C \) has distribution function \( H \). It follows that the lead-time distribution, which we still denote by \( G \), depends on that of \( C \) via

\[
G_Q(x) = P\{Q/(x + C) \leq x\} = P(C \geq Q/x - x) = H(Q/x - x). \tag{30}
\]

To obtain the expected inventory costs for a given \( x \), begin by putting (2) in the form

\[
E\{C(Q, r)\} = \frac{K}{Q} + (\pi + h)\int_{x=r}^{\infty} (x - r)G(x)\,dx/Q
\]

\[+ \frac{hQ}{2} + hr - h\int_{x=0}^{\infty} G(x)\,dx. \tag{31}\]

Let us henceforth restrict attention to the case \( \theta = 1 \). If one substitutes (30) in the expected costs, keeping in mind that lead time cannot exceed \( Q/x \), we get

\[
E\{C(Q, r)\} = \frac{K}{Q}
\]

\[+ (\pi + h)\int_{x=r}^{Q/x} (x - r)H(Q/x - x)\,dx/Q
\]

\[+ \frac{hQ}{2} + hr - h\int_{x=0}^{\infty} H(Q/x - x)\,dx. \tag{32}\]

Due to the upper bound on lead time, the reorder point satisfies \( r < Q/x \). Otherwise, the holding costs could be reduced by lowering the reorder level without altering other costs.

To complete the expected-cost expression, we must include the costs to reserve a production capacity \( x \). A customer may have to compensate the manufacturer for dedicating this capacity to its exclusive use. Such an expense arises because the manufacturer distorts its production process, perhaps incurs overtime or hires extra workers, or pays a tardiness penalty to some other customers.

The cost to reserve a production rate can be charged to the customer in one of the two ways: either on the basis of cost per unit quantity or cost per unit time (cf. [5]). However, since demand was scaled to be one unit per time period, it does not formally matter how we model the cost to reserve the rate; expected-cost expressions will be the same in both cases.

Although the upper level of reserved production could, in principle, be the manufacturer’s total production rate \( R \), a customer would be charged excessively to bring its reserved rate so high. Reservation costs will thus be convex; a possible shape \( \phi(x) \), to reserve the production rate \( x \), is in Fig. 2.

If \( \phi(x) \) is added to the previous expected-cost expression (32), with the reserved production rate \( x \) a third decision variable, the objective function becomes

\[
E\{C(Q, r, x)\} = \frac{K}{Q} + (\pi + h)
\]

\[\int_{x=r}^{Q/x} (x - r)H(Q/x - x)\,dx/Q
\]

\[+ \frac{hQ}{2} + hr - h\int_{x=0}^{\infty} H(Q/x - x)\,dx + \phi(x). \tag{33}\]

This is to be minimized over \( Q, r \) and \( x \). Observe that, for a fixed reserved capacity \( x \), Eq. (33) reduces to (3) with \( \theta = 1 \). Thus, all convexity and first-order optimality results are still valid, i.e. (33) is jointly convex in \( Q \) and \( r \), and closed-form expressions are available for the minimizers. What we need then is a mechanism to find the \( x \) which minimizes (33) for fixed \( Q \) and \( r \); we shall again employ a coordinate-descent-based algorithm. By differentiating (33) with respect to \( x \), we obtain the following equation whose solutions are critical values of \( x \):

\[
(\pi + h)\int_{y=0}^{Q/x-x} H(y)\left(-\frac{3Q}{(y + x)^2} + \frac{2r}{(y + x)^3}\right)\,dy
\]

\[+ 2Qh\int_{y=0}^{\infty} \frac{H(y)}{(y + x)^3}\,dy + \phi'(x) = 0. \tag{34}\]
Therefore, we propose in Fig. 3 an algorithm to find optimal values of the decision variables.

That algorithm starts by finding $Q$ and $\delta = r/Q$ for fixed $z$. This is Step 1, and we have already given the closed-form results for that in Section 3. Step 2 is basically solution of Eq. (34) when $Q$ and $r$ are given. Step 1 is extremely simple, so Step 2 (solving a nonlinear equation) dictates the level of effort. We note that one has to check for joint convexity by evaluating numerically the Hessian at critical points, before concluding that the result is really a minimizer.

7. Conclusions

This paper modeled, and explored the implications of, dependence of random lead time on lot size and production capacity. We discussed mainly two cases: lead time linear and concave in lot size. Joint convexity of expected costs was established in the decision variables $Q$ and $r$. First-order conditions for $Q$ and $r$ were laid out, and (for $\theta = 1$) closed-form results for optimal decision variables were obtained from those conditions. Numerical examples and sensitivity analyses on cost parameters were provided for both models. A comparison of the way linear and concave lead-time models respond to cost parameters was presented.

Reserving some of a manufacturer’s production capacity was then examined. Orders were assumed to be allocated the sum of the reserved and random capacity. The reserved capacity, $z$, then became a decision variable, in addition to $Q$ and $r$. Since for fixed reserved capacity this model reduces to the one discussed earlier in the paper, many results carried over automatically. Indeed, the difficult part was solving the nonlinear optimality equation for $z$, given $Q$ and $r$. Some of this was left for future work.

The models presented here might be extended in the following ways. Our results pertaining to a backordering situation could serve as the basis for analysis of the lost-sales case. Another possibility is to work with discrete lead-time, which might arise if a certain capacity were allocated to an order every period, and this order could be considered complete only at the end of a period. Çakanyıldırım [20] contains some results regarding the lost-sales case and discrete lead-time.

In this paper, we tacitly assumed that reserved capacity does not affect the random portion of capacity. This assumption is valid as long as reserved capacity is small by comparison. If such is not true, the formulation should be modified to reflect the dependence of the random capacity on reserved capacity.

References


