ELSEVIER

Operations Research Letters 29 (2001) 187–192

operations research letters

www.elsevier.com/locate/dsw

Minimizing holding and ordering costs subject to a bound on backorders is as easy as solving a single backorder cost model

Tamer Boyaci[a,*], Guillermo Gallego[b]

[a]Faculty of Management, McGill University, 1001 Sherbrooke St. West, Montreal, Quebec, Canada H3A 1G5,
[b]Department of IE&OR, Columbia University, 500 West 120th Street, Rm 331, New York, NY, 10027, USA

Received 1 April 2000; received in revised form 1 January 2001; accepted 1 April 2001

Abstract

Minimizing average ordering and holding cost subject to a constraint on expected backorders has required to iteratively solve the backorder cost model with different backorder penalty rates until the constraint is satisfied. Here we present a direct approach, that is as simple as solving a single backorder cost model.

Keywords: Stochastic inventory; Service measures; Expected backorders; Waiting times

1. Introduction and motivation

The traditional backorder cost (BC) model minimizes the sum of long-run average ordering, inventory holding, and backorder costs per unit time. Both inventory holding and backorder costs are traditionally charged at constant rates, proportional to the number of inventory and backorders, i.e. ($/*unit*/*time*). In practice, the backorder costs charged by the BC model are rarely out-of-pocket and therefore are difficult to specify. The use of proportional backorder penalty costs is justified by dualizing a constraint on the (steady-state) expected number of backorders in a model whose objective is to minimize the average ordering and holding costs only, see Hadley and Whitin [8]. As a consequence, there is a one-to-one correspondence between the tolerance for expected number of backorders (or through Little's Law for expected waiting time of demands) [1] and the imputed penalty cost rate of the BC model. Finding the imputed penalty cost rate has required, until now, a search that involves solving, at each step, a BC model for a specific penalty rate. The penalty cost rate is updated at each iteration until the one corresponding to the backorder tolerance is found. We note that BC models can be used in a similar manner for other service measures such as the fill-rate, but in this case they would be suboptimal, see Boyaci [1].

* Corresponding author.

E-mail addresses: boyaci@management.mcgill.ca (T. Boyaci), ggallego@ieor.columbia.edu (G. Gallego).

[1] The waiting of demands corresponds to the waiting time of orders only if orders are for one unit at a time.

0167-6377/01/$ - see front matter
PII: S0167-6377(01)00095-5

In this letter, we show that the problem of minimizing average ordering and holding costs subject to a constraint on the expected number of backorders is actually as easy as solving a *single* BC model. We present the basic qualitative properties and optimality conditions for this model in the context of (Q,r) inventory systems. We also develop a simple algorithm to compute an optimal (Q,r) pair. As a bonus, we obtain the imputed penalty cost rate and the corresponding fill-rate (proportion of demand filled immediately from stock).

A brief review of the vast literature on (Q,r) and related inventory management models is presented in Section 2. Relevant notation and the model formulation is contained in Section 3. Optimality properties and the algorithm are presented in Section 4. Section 5 discusses computation of the imputed penalty cost rate from the optimal solution and Section 6 contains a numerical example.

2. Literature review

Initial treatment of the BC (Q,r) model dates back to late 1950s (see Whitin [23]). An exact formulation under the assumptions of Poisson demand and positive reorder levels was developed and analyzed by Galliher et al. [7] and by Hadley and Whitin [8]. Since then, substantial research has been done on (Q,r) and (s,S) policies. Various authors have analyzed the operating characteristics and the objective functions of (Q,r) and (s,S) policies under different system settings: Browne and Zipkin [2], Sahin [14], Serfozo and Stidham [17], Simon [19], Sivazlian [20], Zipkin [26], among others. In spite of these efforts, streamlined results on the optimality conditions for the (Q,r) model have been obtained only recently by Zheng [25]. An efficient algorithm for computing an optimal (Q,r) pair is developed for discrete demands in Federgruen and Zheng [5].

The literature on service constrained single-stage inventory systems consists mostly of approximate models for the fill-rate service measure. Models for (Q,r) policies include Hadley and Whitin [8], Nahmias [9,10], Silver and Wilson [18], Yano [24], and for (s,S) policies, Cohen et al. [4], Schneider [15], Schneider and Ringuest [16], Tijms and Groenevelt [22], among others. Platt et al. [11] provide a good review of this literature. For periodic review (Q,r) policies, Tempelmeier [21] studies the problem of finding optimal reorder point r for a given (and large) order quantity Q, under an expected waiting time service measure. More recently, Rosling [12] establishes conditions to guarantee quasiconvexity of the Lagrangian cost functions that arise from dualizing constraints on various service measures, and presents an algorithm in [13]. We show in this letter that when expected number of backorders (or waiting times of demands) is the desired service, dealing with the service constraint directly results in a significantly simpler algorithm.

3. Preliminaries

For a given continuous review (Q,r) policy, let $c(Q,r)$ denote the long-run average ordering and inventory holding cost, and let $B(Q,r)$ denote the expected number of backorders. Then the model can be formulated as

$$\min_{Q,r} \quad c(Q,r) \tag{1}$$

$$\text{s.t.} \quad B(Q,r) \leqslant \eta, \tag{2}$$

where η is the predetermined maximum tolerance for backorders.

Let λ denote the demand rate (the average demand per unit time) and D the sum of the demands that occur over the fixed leadtime interval $(t, t+L]$ in steady state. We assume that $D \geqslant 0$. Let F be the cdf of D. Define $x_+ = \max\{x, 0\}$. Let

$$G(y) = hE[y-D]_+ \tag{3}$$

denote the rate at which expected inventory holding cost accumulates at time $t+L$ when the inventory position at time t is y and the holding cost rate is h. Let K denote the fixed cost of ordering. Our analysis is based on the average cost (rate) function $c(Q,r)$ given by

$$c(Q,r) = \frac{K\lambda + \int_r^{r+Q} G(y)\,\mathrm{d}y}{Q}. \tag{4}$$

Eq. (4) is the long-run average cost of operating under a (Q,r) policy when the distribution of the inventory position in steady state is uniformly distributed on the interval $(r, r+Q]$ and is independent of the leadtime demand D. These conditions are met if cumulative demand rate is of the form $\int_0^t \lambda(x(s))\,\mathrm{d}s$, where $x(\cdot)$ is a time homogeneous Markov process and $\lambda(\cdot)$

is a sufficiently smooth (but not constant) function. See Browne and Zipkin [2], and Serfozo and Stidham [17] for a detailed account of when these conditions hold. Zipkin [26] provides conditions under which (4) remains valid for stochastic leadtimes. As shown in Chao and Gallego [3], (4) is also valid for a large class of single-item stochastic *production* systems where items are manufactured rather than ordered. In that case, however, D is not the leadtime demand. It is worth noting that the average cost function (4) implicitly assumes that leadtime demand D is continuous, but it can also be viewed as an approximation to the expected average cost when D is discrete and demands are for one unit at a time, e.g., a Markov modulated Poisson process.

For a given (Q,r) policy, the expected number of backorders $B(Q,r)$ is given as

$$B(Q,r)=\frac{1}{Q}\int_r^{r+Q} E[D-y]_+\,\mathrm{d}y. \tag{5}$$

4. Analysis and algorithm

Our analysis is largely inspired by Zheng [25]. The two-dimensional *service-constrained* minimization problem $\min_{Q,r} c(Q,r)$ can be carried out sequentially: $\min_Q[\min_r c(Q,r)]$. Let $r(Q)$ denote an optimal r for fixed Q. Let

$$\begin{aligned}S(Q,r)&=\int_r^{r+Q} E[D-y]_+\,\mathrm{d}y-\eta Q\\&=Q(B(Q,r)-\eta).\end{aligned}$$

Notice that for fixed Q, $c(Q,r)$ is increasing and $B(Q,r)$ is decreasing in r. Therefore $r(Q)$ is defined by the identity

$$S(Q,r)\equiv 0. \tag{6}$$

Let $r'(Q)$ denote the derivative of $r(Q)$.

Lemma 1. *$r(Q)$ is decreasing and $r(Q)+Q$ is increasing in Q. Specifically,*

$$-\tfrac{1}{2}\leqslant r'(Q)\leqslant 0.$$

Proof. From the implicit function theorem,

$$\begin{aligned}r'(Q)&=-\frac{\partial S(Q,r)/\partial Q}{\partial S(Q,r)/\partial r}\\&=-\frac{E[D-(r+Q)]_+-\eta}{E[D-(r+Q)]_+-E[D-r]_+}\leqslant 0.\end{aligned} \tag{7}$$

On the other hand, since $E[D-y]_+$ is decreasing convex,

$$\begin{aligned}\eta&=\frac{1}{Q}\int_r^{r+Q}E[D-y]_+\,\mathrm{d}y\\&\leqslant\frac{E[D-r]_++E[D-(r+Q)]_+}{2},\end{aligned}$$

which implies that $r'(Q)\geqslant-\frac{1}{2}$. □

Let $c(Q)=c(Q,r(Q))$. Since $[y-D]_+=y-D+[D-y]_+$, it follows from the definition of $G(y)$ that

$$\begin{aligned}&\frac{1}{Q}\int_r^{r+Q}G(y)\,\mathrm{d}y\\&\quad=h\frac{1}{Q}\int_r^{r+Q}(y-\mu)\,\mathrm{d}y\\&\qquad+h\frac{1}{Q}\int_r^{r+Q}E[D-y]_+\,\mathrm{d}y\\&\quad=h\left(r+\frac{Q}{2}-\mu\right)+hB(Q,r).\end{aligned}$$

Consequently from (4) and (6) we have

$$c(Q)=\frac{K\lambda}{Q}+h\left(r(Q)+\frac{Q}{2}-\mu\right)+h\eta. \tag{8}$$

Lemma 2. *$c(Q)$ is convex.*

Proof. We will show that $r(Q)$ is convex, which implies that $c(Q)$ is convex. For notational convenience we will drop the arguments of the functions $r(Q)$ and $r'(Q)$. Let

$$g(Q)=\frac{1+r'}{r'}=\frac{\eta-E[D-r]_+}{\eta-E[D-(r+Q)]_+}$$

and notice that $g'=-r''/(r')^2$. Convexity of r follows if $g'\leqslant 0$. This is true since $E[D-y]_+$ is decreasing convex, and from Lemma 1, r is decreasing and $(r+Q)$ is increasing in Q. □

Lemmas 1 and 2 show that the model has essentially the same qualitative properties as the original BC model, see [25]. Let $c'(Q)$ denote the derivative of $c(Q)$ and let $Q_d=\sqrt{2K\lambda/h}$ denote the EOQ. Notice that the first order condition (FOC) implies that

$$Q^* = \sqrt{\frac{2K\lambda}{h(1+2r'(Q^*))}} = \frac{Q_d}{\sqrt{1+2r'(Q^*)}} \geqslant Q_d. \tag{9}$$

The following iterative-substitution procedure can be used to compute the optimal policy (Q^*, r^*) and the optimal cost c^*, for a predetermined desired precision level ε.

Algorithm to compute the optimal (Q, r) policy

1. Set $n = 1$, $Q_n = Q_d$.
2. Set $r_n = r(Q_n) = \min\{r\colon S(Q_n, r) = 0\}$.
3. Compute Q_{n+1}:

$$Q_{n+1} = Q_d \sqrt{\frac{1}{2r'(Q_n)+1}} = Q_d \sqrt{\frac{E[D-r_n]_+ - E[D-(r_n+Q_n)]_+}{E[D-r_n]_+ + E[D-(r_n+Q_n)]_+ - 2\eta}}.$$

IF $|Q_{n+1} - Q_n| \leqslant \varepsilon$, GOTO Step 4; ELSE set $n = n+1$ and GOTO Step 2.

4. $(Q^*, r^*) = (Q_n, r_n)$ and $c^* = K\lambda/Q_n + h(r_n + Q_n/2 - \mu) + h\eta$.

Notice that in computing r_n at Step 2 of the algorithm, a complete search is not required, since by virtue of Lemma 1,

$$r_n - r_{n-1} \geqslant -\tfrac{1}{2}(Q_n - Q_{n-1}).$$

The convergence of the algorithm is proved in the following proposition.

Proposition 1.

$$\lim_{n\to\infty} |Q_{n+1} - Q_n| \to 0$$

and

$$\lim_{n\to\infty} Q_n \to Q^*.$$

Proof. From the convexity of $c(Q)$ and the fact that $c(Q) \to \infty$ as $Q \to \infty$, and as $Q \downarrow 0$, there exists a finite global minimizer Q^*. Using the FOC for $c(Q)$ it can be shown that the sequence $\{Q_{2k-1}\}_{k=1}^{k=\infty} \leqslant Q^*$ is monotonically increasing, whereas the sequence $\{Q_{2k}\}_{k=1}^{k=\infty} \geqslant Q^*$ is monotonically decreasing. This implies $|Q_{n+1} - Q_n|$ is monotonically decreasing, so the first result follows. When $Q_{n+1} = Q_n$, the FOC is satisfied, proving the second result. □

5. Imputed backorder penalty rate

Here we show that it is possible to easily obtain the imputed backorder penalty cost as well as the fill-rate at the end of the algorithm. To see this, recall that the traditional BC model minimizes (4) with $G(y)$ replaced by $G^{\mathrm{T}}(y) = h(y-D)_+ + pE[D-y]_+$, where p is the linear backorder penalty rate. It is shown in Zheng [25] that for fixed Q, the optimal reorder level $r^{\mathrm{T}}(Q)$ for the BC model is selected such that

$$G^{\mathrm{T}}(r^{\mathrm{T}}(Q)) = G^{\mathrm{T}}(r^{\mathrm{T}}(Q) + Q).$$

As shown in Gallego [6], this condition is equivalent to

$$\frac{1}{Q}\int_{r^{\mathrm{T}}(Q)}^{r^{\mathrm{T}}(Q)+Q} P(D > y)\,\mathrm{d}y = \frac{1}{Q}\{E[D - r^{\mathrm{T}}(Q)]_+ - E[D - (r^{\mathrm{T}}(Q)+Q)]_+\} = \frac{h}{h+p}. \tag{10}$$

Rewriting $r'(Q^*)$ given by (7) using (10), substituting in the FOC (9), and solving for the imputed backorder cost results in

$$p^* = h\left\{\left(\frac{Q_d^2 + (Q^*)^2}{2Q^*(E[D-r^*]_+ - \eta)}\right) - 1\right\}. \tag{11}$$

This implies that the fill-rate is given by

$$\beta(Q^*, r^*) = \frac{1}{Q^*}\int_{r^*}^{r^*+Q^*} P(D \leqslant y)\,\mathrm{d}y = \frac{p^*}{p^*+h} = 1 - \frac{2Q^*(E[D-r^*]_+ - \eta)}{Q_d^2 + (Q^*)^2}. \tag{12}$$

To our knowledge, this is the first formula relating the tolerance for backorders to the fill-rate.

Table 1
Numerical results

Iteration (n)	$\mu = 10$			$\mu = 100$		
	Q_n	r_n	c_n	Q_n	r_n	c_n
1	7.071	7.116	51.875	22.361	124.313	476.733
2	11.730	5.840	48.364	43.734	117.550	461.330
3	9.757	6.335	47.762	32.897	120.710	457.579
4	10.335	6.185	47.708	36.836	119.502	457.070
5	10.138	6.235	47.702	35.157	120.008	456.978
6	10.202	6.219	47.702	35.832	119.803	456.963
7	10.181	6.224	47.702	35.553	119.888	456.960
8	10.188	6.222	47.702	35.667	119.853	456.959
9	10.185	6.223	47.702	35.620	119.867	456.959
10	**10.186**	**6.223**	**47.702**	35.639	119.862	456.960
11				35.632	119.864	456.960
12				35.635	119.863	456.959
13				35.633	119.863	456.959
14				**35.634**	**119.863**	**456.959**
Imputed penalty	$p^* = 16.495$			$p^* = 112.082$		
Fill-rate	62.26%			91.81%		

6. Numerical example

The purpose of this section is to illustrate the results of our model and to demonstrate the performance of the algorithm presented in Section 4. The following parameters are used in the example: $\{L = 1, K = 25, h = 10\}$. The leadtime demand is assumed to be normally distributed with mean μ and standard deviation σ. Two distributions are tested: $(\mu, \sigma) \in \{(10, 2.5), (100, 25)\}$. The tolerance limit for backorders is set as $\eta = 1$, corresponding to 10% (resp. 1%) of mean leadtime demand in the case $\mu = 10$, (resp. $\mu = 100$). Table 1 reports (Q_n, r_n, c_n) as obtained from every iteration of the algorithm. The iterations are stopped when the precision level $\varepsilon = 0.001$ is achieved. The imputed penalty rate p^* and the fill-rate is also reported. As seen from the results in Table 1, the order quantities Q_n generated by algorithm converge to Q^* quite fast, usually after several iterations. Notice also that the costs c_n converge typically faster than Q_n. Hence the algorithm can be terminated earlier, if desired. It is also interesting to note that using the EOQ, as suggested in textbooks, can result in significant additional (ordering and holding) costs: 8.75% for the case $\mu = 10$, and 4.33% for the case $\mu = 100$.

References

[1] T. Boyaci, Supply chain coordination and service level management, Ph.D. Thesis, Department of IEOR, Columbia University, New York, NY, 1998.

[2] S. Browne, P. Zipkin, Inventory models with continuous, stochastic demands, Ann. Appl. Probab. 1 (1991) 419–435.

[3] X. Chao, G. Gallego, Stochastic production systems with finite capacity: bounds, heuristics, and algorithms, Working paper, Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027, 1999.

[4] M.A. Cohen, P. Kleindorfer, H. Lee, Service constrained (s, S) inventory systems with priority demand classes and lost sales, Management Sci. 34 (1988) 482–499.

[5] A. Federgruen, Y.S. Zheng, An efficient algorithm for computing an optimal (r, Q) policy in continuous review stochastic inventory systems, Oper. Res. 40 (1992) 808–813.

[6] G. Gallego, New bounds and heuristics for (Q, r) policies, Management Sci. 44 (1998) 219–233.

[7] H. Galliher, P. Morse, M. Simmond, Dynamics of two classes of continuous-review inventory systems, Oper. Res. 7 (1959) 362–384.

[8] G. Hadley, T. Whitin, Analysis of Inventory Systems, Prentice-Hall, Englewood Cliffs, NJ, 1963.

[9] S. Nahmias, On the equivalence of three approximate continuous review inventory models, Nav. Res. Logist. Quart. 23 (1976) 31–36.
[10] S. Nahmias, Production and Operations Analysis, Irwin, Homewood, IL, 1993.
[11] D.E. Platt, L.W. Robinson, R.B. Fruend, Tractable (Q, r) heuristic models for constrained service levels, Management Sci. 43 (1997) 951–965.
[12] K. Rosling, Applicable cost rate functions for single-item inventory control, Working paper, Department of Industrial Engineering, Lund University, Lund, Sweden, 1997.
[13] K. Rosling, The textbook algorithm for single-item inventory optimization, Working paper, Department of Industrial Engineering, Lund University, Lund, Sweden, 1997.
[14] I. Sahin, On the objective function behavior in (s, S) inventory models, Oper. Res. 30 (1982) 709–725.
[15] H. Schneider, Methods for determining the re-order point of an (s, S) ordering policy when a service level is specified, J. Oper. Res. Soc. 29 (1978) 1181–1193.
[16] H. Schneider, J.L. Ringuest, Power approximation for computing (s, S) policies using service level, Management Sci. 36 (1990) 822–834.
[17] R. Serfozo, S. Stidham, Semi-stationary clearing processes, Stochastic Process Appl. 6 (1982) 165–178.
[18] E.A. Silver, T.G. Wilson, Cost penalties of simplified procedures for selecting reorder points and order quantities, American Production and Inventory Control Society International Conference Proceedings, 1975, pp. 219–233.
[19] R.M. Simon, The uniform distribution of inventory position for continuous review (s, Q) policies, Rand Corporation Paper P3938, Santa Monica, CA, 1968.
[20] B. Sivazlian, A continuous review (s, S) inventory system with arbitrary interarrival distribution between unit demand, Oper. Res. 22 (1974) 65–71.
[21] H. Tempelmeier, Inventory control using a service constraint on the expected customer order waiting time, European J. Oper. Res. 19 (1985) 313–323.
[22] H.C. Tijms, H. Groenevelt, Simple approximations for the reorder point in periodic and continuous review (s, S) inventory systems with service level constraints, European J. Oper. Res. 17 (1984) 175–190.
[23] T.M. Whitin, The Theory of Inventory Management, Revised Edition, Princeton University Press, Princeton, NJ, 1957.
[24] C.A. Yano, New algorithms for (Q, r) systems with complete backordering using a fill-rate criterion, Nav. Res. Logist. Quart. 32 (1985) 675–688.
[25] Y.S. Zheng, On properties of stochastic inventory systems, Management Sci. 38 (1992) 87–103.
[26] P. Zipkin, Stochastic lead times in continuous-time inventory models, Nav. Res. Logist. Quart. 33 (1986) 763–774.