OPTIMAL LOT SIZES WITH GEOMETRIC PRODUCTION YIELD AND RIGID DEMAND

SHOSHANA ANILY, AVRAHAM BEJA, and AMIT MENDEL

Faculty of Management, Tel Aviv University, Tel Aviv, Israel
anily@post.tau.ac.il • abeja@post.tau.ac.il • amit_12@netvision.net.il

(Received September 1996; revisions received December 1998, March 2000, September 2000; accepted November 2000)

A job shop has to deliver a given number of custom-order items. Production is performed in "lots" that require costly setup, and output quality is stochastic. The size of each lot must be set before output quality is observed. We study production processes whose yield is distributed according to a generalized truncated geometric distribution: Production can randomly go "out of control," in which case all subsequent output of that lot is defective. The "generalization" allows both hazard rates and marginal production costs to vary as production progresses. Our results characterize the optimal lot sizes. In particular, we show that for small demands the optimal lot size equals the outstanding demand, and for larger demands it is less than the outstanding demand, with all lot sizes being uniformly bounded. For sufficiently large demands, we identify conditions under which the optimal lots are precisely those that minimize the ratio of production cost to the expected number of good items. A tighter characterization is given for the standard case, where hazard rates and marginal costs are constant.

1. INTRODUCTION

Technological and economic considerations often dictate production in "lots" involving some set-up cost. This applies not only to batch processing: When production is sequential, one can consider the set of items to be manufactured prior to the next inspection and realignment of the machine as a "production lot." The key issue is that the lot size must be determined before output quality can be ascertained. Small lot sizes result in lower risk of poor quality but more frequent set-ups. Large lot sizes may result in a waste of producing too many defective items or good items in excess of the demand. The later factor is especially crucial for custom-order demand, in which nondeliverable good items have little, if any, salvage value.

This paper considers the problem of optimal lot sizes for custom orders. The demand to be satisfied is "rigid"—one must deliver a given number of items that meet some prespecified quality standards. If the number of good items on hand is less than the outstanding demand, an additional production lot must be initiated, and so forth until the demand is fully satisfied. It is assumed that defective items and good items in excess of the required quantity have the same salvage value. The production planning problem, then, is to determine the optimal lot size for each possible quantity of outstanding demand. The objective is to minimize the expected total cost of satisfying the demand. Our study addresses this problem for processes where the production yield obeys the truncated geometric probability distribution. One way of interpreting this distribution is the following. Suppose that the process can be either "in control," in which case the items produced are of acceptable quality, or "out of control," in which case the items produced are defective. As time goes on, an "in control" process may deteriorate, go out of control, and throughout the remainder of the run produce only items that are of substandard quality. At the end of the run, the output is inspected (and the process is realigned if a new run is needed). Monden (1983) discusses such a system in a Toyota factory, where an automated high-speed punch press is fed by lots of 50 or 100 units and may randomly go out of control before the end of the run.

1.1. A Brief Review of Earlier Literature

The problem of determining optimal lot sizes to satisfy a rigid demand has commanded considerable theoretical analysis from as far back as the 1950s (e.g., Llewellyn 1959, and Levitan 1960). The models in the literature typically include set-up costs and linear production costs. Early studies of the problem were devoted almost exclusively to processes with binomial yield distribution, and they primarily suggested heuristic computational approximations. Because part of the output was expected to come out defective, the optimal lots were always taken to be larger than the outstanding demand to allow for possible rejects, and the problem was then accordingly dubbed the "reject allowance" problem. White (1965) gives a dynamic programming formulation of this problem with general yield distribution but gives no insight on the structure of optimal policies. In particular, researchers continued to look for structural properties of optimal policies for production processes with binomial yield distribution. Implicit in this case is the assumption that occurrences of failures during a run are independent and identically distributed. Beja (1977) considered processes where failures are independent but not necessarily of equal probability. He studied a family of processes with "constant marginal efficiency of production,"

Subject classifications: Production planning: single machine, lot sizes. Dynamic programming.
Area of review: Logistics and Supply Chain Operations.

0030-364X/02/5003-0424 $05.00
1526-5463 electronic ISSN

meaning that the expected production cost per *good* item is fixed throughout the run (as is the case with binomial yield and linear cost). He showed that there is an optimal policy in which the optimal lot size is strictly increasing in the outstanding demand (and hence the optimal lot is never less than the outstanding demand, justifying the notion of "reject allowance"). For the family of processes studied, Beja also gave additional structural properties that greatly reduce the required computation in the search for optimal policy.

More recent studies of the problem allow for failures that are not necessarily independent. Anily (1995) shows results similar to those of Beja (1977) for the case of discrete uniform yield distribution with linear cost, a process that does not exhibit stochastic independence but does have constant marginal efficiency of production. Grosfeld-Nir and Gerchak (1996) consider some properties of the solutions for general yield distributions. They prove that when yield is geometrically distributed and production costs are linear, the optimal lot size never exceeds the outstanding demand. Zhang and Guu (1998) also studied the geometric distribution. In particular, they prove that the optimal lot size never increases by more than one unit in response to a unit increase in the demand. In addition, they prove that the optimal lot sizes are uniformly bounded by a function of the cost parameters and the failure rate. Grosfeld-Nir (1995) and Grosfeld-Nir and Robinson (1995) consider lot sizing problems with rigid demands in multistage production systems and various yield distributions (but they do not consider the geometric distribution).

Deteriorating production processes in EOQ-type problems were investigated by Porteus (1986, 1990), Rosenblatt and Lee (1986a, 1986b), and Lee and Rosenblatt (1987). These papers have modeled the time at which the production process goes out of control, by a geometric probability distribution in the discrete case, or by an exponential probability distribution in the continuous case.

For a more extensive review of the literature on optimal lot sizing with random yields, the reader is referred to Yano and Lee (1995).

1.2. An Overview of this Paper

The geometric yield distribution studied in this paper represents an extreme degree of stochastic dependence—the conditional probability of success for the nth item in a production run is positive if all previous items are good, and zero if there has been an earlier failure (e.g., a failure occurs when a machine tool gets out of its prescribed setting, and as a result all subsequent items in that run are defective). This model is thus in some sense diametrically different from the earlier models of the rigid demand problem, with stochastic independence and binomial yields. Not surprisingly, the resulting optimal policies are accordingly also diametrically different.

The model of the production process is presented in §2. Where possible, we generalize the familiar standard geometric distribution to allow for possibly varying hazard rates and marginal costs, in a way analogous to Beja's generalization of processes with stochastic independence. The results of the model typically depend on various assumptions regarding the yield distribution and/or the cost structure.

The presentation of our results starts with a sequence of fairly straightforward observations. It is shown that under fairly general conditions, the geometric yield model exhibits the following properties: (1) The optimal lot size never exceeds the outstanding demand; (2) total cost is *strictly* increasing in the outstanding demand; (3) the optimal lot sizes are uniformly bounded from above, if, in addition, the marginal production costs are nondecreasing, i.e., the marginal production cost of the $i+1$st item in the run is at least as large as that of the ith item, then we further obtain that (4) a unit increase in the outstanding demand never induces more than a unit increase in the optimal lot size (but it may induce a decrease in the optimal lot size; and (5) there is a critical value L such that the optimal lot size is equal to the outstanding demand if and only if the outstanding demand does not exceed L.

Section 3 deals with the optimal lot sizes for large demands. We prove that the optimal lot sizes for sufficiently large outstanding demands converge to a set of at most two consecutive integers under some assumptions about the marginal costs, and when the yield distribution is the interrupted geometric. This set consists of all integer values that maximize the expected number of good items produced per dollar invested in initiating the run (equivalently minimize the ratio of production cost to the expected number of good items in the run). Usually, this set consists of a single integer. In such a case, the optimal policy for large demands is to produce a lot of this size until the outstanding demand becomes sufficiently small. If the set consists of two integers, then the optimal lot size for a sufficiently large demand is either of these two integers. This convergence is of special importance because recursive computation of optimal lot sizes by dynamic programming gets progressively cumbersome as demands get larger.

Section 4 gives a tighter characterization of the optimal policy for the standard case of constant hazard rates and linear costs. We identify the largest value of the outstanding demand for which it is optimal to produce exactly the demanded quantity, and we also show that this value is the largest lot size consistent with optimality over all conceivable outstanding demands. For the reader's convenience, the more lengthy proofs of §§3 and 4 are deferred to appendixes. Finally, §5 offers a few concluding remarks on possible extensions of the results of this paper.

2. THE MODEL

A production process is conceptually defined by its yield distribution and cost function. In this paper, we study discrete production processes (lot sizes and yields are nonnegative integers) where the potential deterioration in output quality is stochastic and persistent, i.e., if an item happens to be defective then all subsequent items in the same run are also defective. Such processes are usually modeled in the literature by the *truncated geometric probability distribution:* Letting G_n denote the number of good items in a run of n items, the yield distribution is given by

$$P\{G_n \geqslant k\} = q^k \quad \text{for } k = 0, 1, 2, \ldots, n$$
$$= 0 \quad \text{for } k > n,$$

where $0 < q < 1$ is the quality parameter of the production process and $(1-q)$ is its constant *hazard rate*; i.e., the probability that an item is defective, given that all previous items in that run are of good quality.

Some of our results allow for possible variations in the hazard rates as production progresses. This gives rise to a *generalized truncated geometric distribution*, defined by

$$P\{G_n \geqslant k\} = Q_k \quad \text{for } k = 0, 1, \ldots, n$$
$$= 0 \quad \text{for } k > n,$$

where $Q_0 \equiv 1$.

The technological characteristics of the process are specified here by a nonincreasing parameter sequence, $1 \geqslant Q_k > Q_{k+1} > 0$. These characteristics are equivalently defined by the parameters $\{0 < q_k < 1,\ k = 1, 2, \ldots\}$, where for $k = 1, 2, \ldots$ $q_k = Q_k/Q_{k-1} < 1$, so that $Q_k = q_1 q_2 \ldots q_k$. As before, the complement value $(1-q_k)$ is the probability that the kth item in a production run fails, given that all previous $k-1$ items were good, so that the hazard rate now depends on k. The salient feature of the generalized geometric distribution is the property that, for $k \leqslant n$, $P\{G_n \geqslant k\}$ is independent of n.

The cost of initiating a run of size n, denoted $C(n)$, consists of a set-up cost $\alpha \geqslant 0$, independent of the size of the run, and variable costs, which are specified by the sequence of marginal costs $\beta_i > 0$, viz.,

$$C(n) = \alpha + \sum_{i=1}^{n} \beta_i,$$

i.e., β_i is the marginal cost of producing the ith unit in a production run.

Nondecreasing marginal costs and hazard rates reflect the effects of wear and tear in sequential processing, or of overcrowding a limited space in batch processing; nonincreasing marginal costs and hazard rates reflect economies of scale (particularly learning effects in sequential processing). Our results depend on various assumptions regarding the structure of technological parameters and/or the cost parameters (e.g., nondecreasing or constant marginal costs, bounded or constant hazard rates, etc.). Clearly, stronger assumptions allow for stronger results (our tightest characterization of the optimal policy relates to the standard case of constant hazard rates and linear costs).

As explained in the introduction, we deal here with production processes for custom orders. For such processes, it is reasonable to assume that salvage values, if any, are the same for defective items and for good items in excess of the required quantity, and that these are small, relative to production costs. More specifically, we assume that the common salvage value is strictly less than $\beta = \inf_{i \geqslant 1} \beta_i$ (otherwise, it may sometimes pay to produce items just for scrap). Consequently, without loss of generality, we can assume that the salvage value is zero: Suppose that the salvage value, say S, is strictly positive. Then, for each of the items produced, except for the D good items that meet the demand, the revenue from the salvage value can be credited against the production cost. Because D is constant and $S < \beta_i$, we can define an equivalent problem in which the salvage value is $S' = 0$, by redefining the marginal costs for all i, as $\beta_i' = \beta_i - S$.

The objective of minimizing the expected cost required to fulfill an outstanding rigid demand is investigated through three interrelated functions:

$V(D) =$ The minimal expected cost to fulfill an order for D items (i.e., the expected cost when an optimal policy is followed throughout).

$U(n, D) =$ The expected cost incurred in fulfilling a rigid demand for D items if a run of size n is initiated once, and an optimal policy is followed thereafter.

$W(n, D) =$ The expected cost incurred in fulfilling a rigid demand for D items if a run of size n is initiated whenever the outstanding demand is equal to D units, and an optimal policy is used whenever the outstanding demand is less than D.

Given that salvage values are taken to be zero, as discussed above, we can simplify the notation by extending the definition of V to negative integers as well. Thus, we let

$$V(D) = 0 \qquad \text{for } D = 0, -1, -2, \ldots$$

The three functions satisfy the following relationships:

$$U(n,D) = C(n) + \sum_{k=0}^{n} P\{G_n = k\} V(D-k),$$

$$W(n,D) = \left\{ C(n) + \sum_{k=1}^{n} P\{G_n = k\} V(D-k) \right\} \Big/ P\{G_n > 0\},$$

$$V(D) = \min_{n=1,2,\ldots} U(n,D).$$

Also, let

$$N(D) = \arg\min_{n=1,2,\ldots} U(n, D).$$

Note (from the above) that $P\{G_n > 0\}\{W(n,D) - V(D)\} = U(n,D) - V(D)$, thus $\text{sign}[W(n,D) - V(D)] = \text{sign}[U(n,D) - V(D)]$, hence also (as expected from the definitions),

$$V(D) = \min_{n=1,2,\ldots} W(n,D) \quad \text{and}$$

$$N(D) = \arg\min_{n=1,2,\ldots} W(n, D).$$

For the case of the generalized geometric yield, substituting for $C(n)$ and $P\{G_n = k\}$ and simplifying, we get

$$U(n,D) = V(D)+\alpha+\sum_{i=1}^{n}\beta_i - \sum_{k=1}^{n} Q_k[V(D+1-k)-V(D-k)], \tag{1}$$

$$U(n+1,D)-U(n,D) = \beta_{n+1} - Q_{n+1}[V(D-n)-V(D-n-1)]. \tag{2}$$

Inspection of the above equations gives rise to a number of straightforward results, which because of their simplicity are stated as "observations." Their proofs are given in Appendix A.

OBSERVATION 1. For all D, Max $N(D) \leqslant D$.

Furthermore, for all $n \geqslant D$,

$$U(n+1,D)-U(n,D) = C(n+1)-C(n) = \beta_{n+1}.$$

The first part of Observation 1 was stated by Grosfeld-Nir and Gerchak (1996) for the special case of linear production costs and standard geometric yield, and by Zhang and Guu (1998) for linear production costs when the yield distribution satisfies certain conditions that cover the generalized geometric distribution.

Observation 1 implies that $N(1) = \{1\}$, and from Equation (1), we get

$$V(1) = \frac{\alpha+\beta_1}{q_1}.$$

OBSERVATION 2. For all D, $V(D) > V(D-1)$. It should be noted that while $V(D) \geqslant V(D-1)$ is universal and obvious, $V(D) > V(D-1)$ is not universally true and depends on the yield distribution, hence it cannot be taken for granted. For example, if $P\{G_1 = 0\} = 1$ and $P\{G_n \geqslant 2\} = 1$ for $n = 2, 3, \ldots$, then $V(2) = V(1)$. Our proof (see Appendix A) uses the fact that the generalized geometric distribution satisfies $P\{G_{n+1} = k\} = P\{G_n = k\}$ for all $k < n$, and in fact our proof establishes the tighter bound $V(D)-V(D-1) \geqslant \max_{i\in N(D)} \beta_i/q_1$. (The stated version of the observation was given preference over the stronger alternative both for its simplicity and also to emphasize its being nontrivial.)

OBSERVATION 3. (a) Suppose marginal costs are bounded away from zero, i.e., $\beta_i \geqslant \beta > 0$ for all $i = 1, 2, \ldots$, and suppose further that $Q_i \to 0$ as $i \to \infty$. Then,

for all D, $\max N(D) \leqslant M$, where

$$M = \max\{i = 1, 2, \ldots : \beta_i \leqslant Q_i V(1)\}.$$

(b) If, in addition, the hazard rates are bounded away from zero, i.e., $q_i \leqslant q < 1$ for all $i = 1, 2, \ldots$, then

for all D, we have that $\max N(D) \leqslant \theta$,

where $\theta = \frac{\log V_1 - \log \beta}{\log(1/q)}$, i.e., $q^\theta V(1) = \beta$.

OBSERVATION 4. Suppose that marginal costs are nondecreasing, i.e., $\beta_i \leqslant \beta_{i+1}$ for all i. Then,

for $D \geqslant 2$, $\max N(D) \leqslant \min N(D-1)+1$.

Zhang and Guu (1998) have stated Observation 4 (in a slightly weaker version) for the special case of standard geometric yield and linear costs.

OBSERVATION 5. Suppose that marginal costs are nondecreasing, i.e., $\beta_i \leqslant \beta_{i+1}$ for all i. Then,

(i) $D \in N(D)$ implies $N(d) = \{d\}$ for $d \leqslant D-1$,
(ii) $D \notin N(D)$ implies $d \notin N(d)$ for $d \geqslant D+1$.

REMARKS. Observations 1, 3, and 4 give upper bounds on the optimal lot size. The first of these is in sharp contrast with the familiar results regarding other yield distributions, as noted in the introductory section. Yet, this observation is quite intuitive and it holds for general marginal costs and general hazard rates. With a geometric yield distribution, output beyond the Dth item in a production run can be of good quality only if all the first D items in that run are of good quality, in which case the excess output is not needed. Indeed, the economic implication of extending the production run by one more item is exactly the added production cost for this marginal item, without any benefit to compensate for this extra cost.

Observation 3 gives a *uniform* upper bound on the optimal lot sizes for all possible values of the outstanding demand: No matter how large this demand is, it is always optimal to produce in lots of size not exceeding M. The stated conditions for the existence of the bound are fairly general: Changes in the marginal costs need not be monotonic, and the condition $Q_i \to 0$ only requires that for sufficiently large lots, eventual failure becomes almost a certainty. Tighter results are given in later sections for more restricted contexts: (i) For large demands, we specify in §3 conditions under which a given lot size, typically much lower than M, is not only an asymptotic bound but in fact an optimal lot size; (ii) for the case of constant hazard rates and linear costs, we identify in §4 the lowest possible uniform upper bound on all $N(D)$, i.e., the largest lot size which is actually optimal for *some* outstanding demand.

Observations 4 and 5 apply to processes with nondecreasing marginal costs (with no restriction on hazard rate patterns). Observation 4 gives a "comparative" upper bound on the optimal lot size, relative to the optimal lots for smaller demands. Note that the observation does not restrict downward changes in the optimal lot size. In extensive numerical experiments, we have observed that drastic reductions in the optimal lot size (in response to increases in the outstanding demand) are typical occurrences at demand levels, where it is first optimal to produce less than the demanded quantity.

Observation 5 states conditions under which the set of demand values for which it is optimal to produce exactly the demanded amount is contiguous. This range of values always starts at $D=1$, where $N(1)=\{1\}$ is dictated by the nonoptimality of ever producing more than the outstanding demand. For the special case of constant hazard rates and linear costs, the demand value at which this range ends is fully specified in §4.

The restriction of Observations 4 and 5 to processes with nondecreasing marginal costs is not a limitation of our proof, but rather a genuine reflection of the nature of the optimal policy: When $\beta_D > \beta_{D+1}$, it can indeed be the case that $N(D)=\{D-1\}$ and $N(D+1)=\{D+1\}$ (for example, if $\alpha=0.9$ $\beta_1=\beta_2=1.0$ $\beta_3=0.1$ and $q_1=q_2=q_3=0.5$, then $N(2)=\{1\}$ and $N(3)=\{3\}$). It is worth noting that when marginal costs and hazard rates are nondecreasing, the only incentive for increasing lot sizes is to save costly set-ups; otherwise, there may be additional incentives as well (such as an attempt to take advantage of potential economies of scale), and idiosyncrasies in the marginal cost and hazard rate patterns can induce idiosyncrasies in the optimal lot size pattern.

3. OPTIMAL LOT-SIZE FOR LARGE DEMANDS

In the previous section, we described some properties of optimal lot-sizes for finite demands under mild conditions for the hazard rates and the marginal costs. These properties enable us to somewhat reduce the search for the optimal lot-size when solving the recursive formula. Nevertheless, the computation of the optimal lot-sizes for a given outstanding demand $D \geqslant 2$ requires the solution of $V(d)$ and the sets $N(d)$ for $d=1,\ldots,D-1$. A question that is of particular interest is whether the sets $N(d)$ exhibit certain convergence patterns when the outstanding demand d increases. Such convergence would imply two benefits: From an implementation viewpoint, it would imply that the optimal lot-size is independent of the remaining demand as long as the outstanding demand is sufficiently large, i.e., the same lot-size can be used repetitively until the remaining demand becomes small enough. Computationally, such convergence would imply that the problem of finding the optimal lot-size for any demand can be solved by applying the recursive formula up to the first value of d where convergence is guaranteed. In this section, we investigate the conditions that ensure a convergence of the sets $N(d)$ for $d \geqslant 1$. We prove that $\lim_{d\to\infty} N(d)$, exists under mild conditions for the marginal costs and when hazard rates are constant.

Indeed, under the conditions stated in Theorem 1, we prove that for sufficiently large demands it is optimal to use lot sizes that solve a related problem—minimizing the ratio of the production cost of a run to the expected number of good items produced in that run. We introduce this related problem in §3.1, and analyze some of its more relevant mathematical features. The asymptotic behavior of the optimal policy and of the minimum cost function is then studied in §3.2.

3.1. A Related Problem

The following concepts and notation are used in our statement of the related problem. The expected number of good items in a run of size n is denoted $E[G_n]$, i.e.,

$$E[G_n]=\sum_{k=1}^{\infty} k\,P\{G_n=k\}=\sum_{k=1}^{\infty} P\{G_n\geqslant k\}=\sum_{k=1}^{n} Q_k.$$

We study the function

$$f(n)=\frac{C(n)}{E[G_n]}=\frac{\alpha+\sum_{k=1}^{n}\beta_k}{\sum_{k=1}^{n}Q_k}.$$

Intuitively, one can think of $f(n)$ as "the average cost per good item in a run of size n." Also, let us define

$$\varphi=\min_n f(n), \tag{3}$$

$$N_0=\arg\min_n f(n). \tag{4}$$

Very mild conditions suffice to guarantee that the minimum is attained, so that φ exists and N_0 is nonempty. In particular, it is sufficient (but not necessary) that $C(n)$ is unbounded as $n\to\infty$ and $q_i \leqslant q < 1$ for all i.

Lemma 1 states conditions under which $f(n)$ is minimized on at most two adjacent integers, strictly decreasing for lower values of n and strictly increasing for higher values of n. The basic argument of the proof can be intuitively summarized as follows: When marginal costs are nondecreasing the marginal contribution of each subsequent item in the lot to the numerator of f is higher than that of previous items, and (since hazard rates are positive) its marginal contribution to the denominator (i.e., the probability that the item is good) is lower than that of previous items; thus if $f(n)\geqslant f(n-1)$ then $f(n+1)>f(n)$.

LEMMA 1. *Suppose that $\varphi=\min_n f(n)$ exists, and that marginal costs are nondecreasing, i.e., $\beta_i \leqslant \beta_{i+1}$ for all i. Then there exists an integer μ that fully characterizes N_0 as follows*:

$$\mu=\min\left\{m:\ \frac{\beta_{m+1}}{Q_{m+1}}\geqslant\frac{\alpha+\sum_{i=1}^{m}\beta_i}{\sum_{i=1}^{m}Q_i}\right\},$$

$$N_0=\{\mu\}\qquad \text{if } \frac{\beta_{\mu+1}}{Q_{\mu+1}}>\frac{\alpha+\sum_{i=1}^{\mu}\beta_i}{\sum_{i=1}^{\mu}Q_i}$$

$$=\{\mu,\mu+1\}\qquad \text{if } \frac{\beta_{\mu+1}}{Q_{\mu+1}}=\frac{\alpha+\sum_{i=1}^{\mu}\beta_i}{\sum_{i=1}^{\mu}Q_i}.$$

PROOF. The proof is based on the following basic property: Given four positive real numbers a_1, b_1, a_2, and b_2 with $\frac{a_1}{b_1}<\frac{a_2}{b_2}$, then $\frac{a_1}{b_1}<\frac{a_1+a_2}{b_1+b_2}<\frac{a_2}{b_2}$. In addition, if $\frac{a_1}{b_1}=\frac{a_2}{b_2}$, then $\frac{a_1+a_2}{b_1+b_2}=\frac{a_1}{b_1}$. In our context, $\beta_i\leqslant\beta_{i+1}$ and $Q_i>Q_{i+1}$ imply that $\frac{\beta_i}{Q_i}<\frac{\beta_{i+1}}{Q_{i+1}}$ for $i\geqslant 1$. Consider now the following sequence of ratios, $\frac{\alpha+\beta_1}{Q_1}, \frac{\beta_2}{Q_2}, \frac{\beta_3}{Q_3},\ldots$, i.e., the numerator of the ith ratio is β_i except for $i=1$, where the numerator is $\alpha+\beta_1$, and the denominator of the ith ratio is Q_i for $i\geqslant 1$. If $\frac{\alpha+\beta_1}{Q_1}<\frac{\beta_2}{Q_2}$ then $f(n)$ is strictly increasing in n

and therefore $N_0 = \{1\}$. If $\frac{\alpha+\beta_1}{Q_1} = \frac{\beta_2}{Q_2}$, then $f(1) = f(2)$ and $f(n+1) > f(n)$ for $n \geqslant 2$, implying $N_0 = \{1, 2\}$. Otherwise, $\frac{\alpha+\beta_1}{Q_1} > \frac{\beta_2}{Q_2}$, and we have $f(2) < f(1)$. Moreover, as long as $\frac{\beta_{n+1}}{Q_{n+1}} < f(n)$, then $f(n+1) < f(n)$. So $f(n)$ is strictly decreasing up to $\mu = \min\{m: \frac{\beta_{m+1}}{Q_{m+1}} \geqslant \frac{\alpha+\sum_{i=1}^{m}\beta_i}{\sum_{i=1}^{m} Q_i}\}$. The existence of μ is guaranteed because otherwise the minimum of f is not attained, contradicting the theorem's assumptions. If $\frac{\beta_{\mu+1}}{Q_{\mu+1}} > \frac{\alpha+\sum_{i=1}^{\mu}\beta_i}{\sum_{i=1}^{\mu} Q_i}$, then $f(n)$ is strictly increasing for $n \geqslant \mu$, thus $N_0 = \{\mu\}$. Otherwise, if $\frac{\beta_{\mu+1}}{Q_{\mu+1}} = \frac{\alpha+\sum_{i=1}^{\mu}\beta_i}{\sum_{i=1}^{\mu} Q_i}$ then $f(\mu) = f(\mu+1)$ and $f(n)$ is strictly increasing for $n \geqslant \mu+1$, thus $N_0 = \{\mu, \mu+1\}$. □

3.2. Asymptotic Behaviour of $N(D)$ and $V(D)$ for Constant Hazard Rates

We now investigate a class of production processes for which (i) the optimal lot-sizes for large demands coincide with those that solve the related problem presented in the previous subsection, and (ii) the marginal cost of satisfying a unit increase in the demand converges to (but need not coincide with) the minimum cost φ of the related problem. This class of production processes is characterized by constant and positive hazard rates and a cost function $C(n)$ that is unbounded.

We first state the main result of this section, Theorem 1, applying the definitions of f, φ, N_0, and μ, given in the previous subsection, as well as the definitions and notation of §2.

THEOREM 1. *Suppose that $q_i = q < 1$ for all i, and that $C(n)$ is unbounded. Let δ and H be defined by*

$$\delta = \min\{[f(n) - \varphi]\sum_{k=1}^{n} q^k : n \notin N_0\},$$

$$H = \frac{\log[V(1) - \varphi] - \log \delta}{\log(1/q)} \qquad \text{if } \varphi < V(1)$$

$$= 0 \qquad \text{if } \varphi = V(1).$$

Then

(a) *For all $D > H$, $N(D) \subseteq N_0$.*

(b) *If, in addition, marginal costs are nondecreasing, i.e., $\beta_i \leqslant \beta_{i+1}$ for all i, then for all $D > H$, $N(D) = N_0$,*

(c) *If $\varphi = V(1)$, then for all D, $N(D) = N_0$, and it is always optimal to produce one unit at a time, i.e., $1 \in N(D)$.*

The chain of arguments leading to the proof of Theorem 1 is broken down into a number of steps. In the first step, we rewrite the equation for $U(n, D)$, and identify a summation term, to be denoted $S(n, D)$, that plays a pivotal role in the subsequent analysis. In the second step, we prove that $S(n, D)$ tends to vanish for large D, a convergence that is stated as Lemma 2(a). This convergence property is a key element in the chain of arguments. It also implies part (b) of the lemma, where we prove that the growth rate of $V(d)$ converges to φ. Finally, we present the concluding arguments that prove the main theorem.

As noted above, in the first step, when φ exists, we can rewrite Equation (1) as

$$\begin{aligned} U(n, D) &= V(D) + \alpha + \sum_{i=1}^{n}\beta_i - \sum_{k=1}^{n} q^k[V(D+1-k) - V(D-k)] \\ &= V(D) + [f(n) - \varphi]\sum_{k=1}^{n} q^k - \sum_{k=1}^{n} q^k[V(D+1-k) - V(D-k) - \varphi]. \end{aligned}$$

For convenience, we denote the last term on the right-hand side of the above equation by $S(n, D)$, i.e., for $D = 1, 2, \ldots$, and $n = 1, 2, \ldots, D$, define

$$S(n, D) = \sum_{k=1}^{n} q^k[V(D+1-k) - V(D-k) - \varphi].$$

$S(n, D)$ can be viewed as a measure, exponentially weighted, of the differences between φ and the growth rates of the sequence $V(d)$, for d ranging from $D - n$ to D. Equation (1) then becomes

$$U(n, D) = V(D) + [f(n) - \varphi]\sum_{k=1}^{n} q^k - S(n, D). \qquad (5)$$

In the second step, we present the next lemma, whose proof is deferred to Appendix B. In its first part we show that $S(n, D)$ converges to zero as D becomes large. This result is central to the proof of Theorem 1. In its second part we show that the growth rate of $V(d)$ converges to φ.

LEMMA 2. *Suppose hazard rates are constant, i.e., $q_i = q < 1$ for all i, and costs are such that φ exists. Then,*

(a) *for $D = 1, 2, \ldots$, and $n = 1, \ldots, D$,*

$$|S(n, D)| \leqslant [V(1) - \varphi]q^D.$$

(b) *$V(D+1) - V(D) \to \varphi$ as $D \to \infty$.*

PROOF. See Appendix B.

CONCLUSION OF PROOF FOR THEOREM 1. Note that the assumptions of the theorem imply that for all D, $N(D)$ is nonempty, that φ exists, and that δ exists and is positive. $V(1) = f(1)$ implies $\varphi \leqslant V(1)$. We note that the statement of part (c) for the case $\varphi = V(1)$ is stronger than the statements in parts (a) and (b) for the general case. Therefore, we need only to prove parts (a) and (b) for the case $\varphi < V(1)$.

PART (a). We note from the definition of H that

$$(1/q)^H = \frac{V(1) - \varphi}{\delta},$$

$$[V(1) - \varphi]q^H = \delta.$$

Under the postulated conditions, Lemma 2 holds; hence,

$$\text{for } D > H,\ |S(n, D)| \leqslant [V(1) - \varphi]q^D < [V(1) - \varphi]q^H = \delta.$$

Thus, for $D > H$, $S(n, D) < \delta$ for all $n = 1, \ldots, D$,

and if $n \notin N_0$, $[f(n) - \varphi]\sum_{k=1}^{n} q^k \geqslant \delta$ (from the definition of δ).

Using these two inequalities, we have from Equation (5) for $D > H$, and $n \notin N_0$,

$$U(n, D) = V(D) + [f(n) - \varphi]\sum_{k=1}^{n} q^k - S(n, D) \geqslant V(D) + \delta - S(n, D) > V(D).$$

This implies that for $D > H$, $n \notin N(D)$ whenever $n \notin N_0$; thus $N(D) \subseteq N_0$, as claimed.

PART (b). Under the postulated conditions, both Lemma 1 and Part (a) above apply.

(b.1) $N(D)$ is nonempty, hence if $|N_0| = 1$ (and by Lemma 1 $\frac{\beta_{\mu+1}}{Q_{\mu+1}} > \frac{\alpha+\sum_{i=1}^{\mu}\beta_i}{\sum_{i=1}^{\mu}Q_i}$), then $N(D) = N_0$ follows immediately from part (a).

(b.2) Alternatively, suppose that $|N_0| \neq 1$. Then from Lemma 1, it follows that $N_0 = \{\mu, \mu+1\}$, with $\frac{\beta_{\mu+1}}{Q_{\mu+1}} = \frac{\alpha+\sum_{i=1}^{\mu}\beta_i}{\sum_{i=1}^{\mu}Q_i}$. To allow for possible perturbations in α, we set the above values as α^1 and μ^1, i.e., $\frac{\beta_{\mu^1+1}}{Q_{\mu^1+1}} = \frac{\alpha^1+\sum_{i=1}^{\mu^1}\beta_i}{\sum_{i=1}^{\mu^1}Q_i}$. We also define α^2 and α^0 as those values of α that satisfy the following equalities, respectively: $\frac{\beta_{\mu^1+2}}{Q_{\mu^1+2}} = \frac{\alpha^2+\sum_{i=1}^{\mu^1+1}\beta_i}{\sum_{i=1}^{\mu^1+1}Q_i}$ and $\frac{\beta_{\mu^1}}{Q_{\mu^1}} = \frac{\alpha^0+\sum_{i=1}^{\mu^1-1}\beta_i}{\sum_{i=1}^{\mu^1-1}Q_i}$. The fact that the sequence $\frac{\beta_i}{Q_i}$ for $i \geqslant 1$ is increasing in i implies that $\alpha^0 < \alpha^1 < \alpha^2$. We now want to see how N_0, $W(n, D)$, and $V(D)$ change with α around α^1. From Lemma 1 it follows that

for $\alpha^1 < \alpha < \alpha^2$, $N_0 = \{\mu^1 + 1\}$,

for $\alpha = \alpha^1$, $N_0 = \{\mu^1, \mu^1 + 1\}$,

for $\alpha^0 < \alpha < \alpha^1$, $N_0 = \{\mu^1\}$.

It follows from (b.1) above that for α in the range $\alpha^1 < \alpha < \alpha^2$, we have $N(D) = \{\mu^1 + 1\}$, hence $V(D) = W(\mu^1 + 1, D)$, and for α in the range $\alpha^0 < \alpha < \alpha^1$, we have $N(D) = \{\mu^1\}$, hence $V(D) = W(\mu^1, D)$.

But inspection of the definitions of the functions $V(D)$ and $W(n, D)$ shows (recursively on $D = 12, \ldots$) that these functions are continuous (indeed piecewise linear) in α. Left continuity implies that at $\alpha = \alpha^1$, $V(D) = W(\mu^1, D)$, hence $\mu^1 \in N(D)$, and right continuity implies that at $\alpha = \alpha^1$, $V(D) = W(\mu^1 + 1, D)$, hence $\mu^1 + 1 \in N(D)$. Thus, by Lemma 1, $N_0 \subseteq N(D)$, and with Part (a) above, this gives $N(D) = N_0$.

PART (c). From Lemma 2, Part (a) it follows that in this case $S(n, D) = 0$ for all n and D, and then it is clear from Equation (5) that $U(n, D) = V(D)$ if and only if $f(n) = \varphi$, i.e., $n \in N_0$. Next, note that $\varphi = V(1) = f(1)$ implies $1 \in N_0$, and hence $1 \in N(D)$. □

4. FURTHER RESULTS FOR STANDARD GEOMETRIC YIELD AND LINEAR COST

All previous literature on optimal lot-sizing for these processes has been restricted to a special "standard" case, where both hazard rates and marginal costs were assumed constant. In this section, we give a tighter characterization of the optimal policy for the standard case, i.e., $q_i = q$, where $0 < q < 1$ and $\beta_i = \beta$ for all i. Without loss of generality, we henceforth simplify the exposition by letting $\beta = 1$.

In general terms, our earlier results indicated the following. When the yield is geometrically distributed, it is never optimal to produce more than the outstanding demand. For small demands, it is optimal to produce exactly the demanded quantity (Observation 5). This holds up to some critical level, say L, beyond which the optimal lot sizes are always strictly less than the outstanding demand. In Theorem 2 below, we precisely identify the critical value L for the standard case. Also, as we have noted in §2, all optimal lot sizes are uniformly bounded from above, and we have identified possible values for such bounds in Observation 3. Theorem 3 below states that in the standard case the critical value L is in fact the lowest upper bound on all optimal lot sizes, i.e., a lot of size L is optimal for an outstanding demand of size L, and lots larger than L are *never* optimal, whatever the outstanding demand may be. For the reader's convenience, we defer the detailed proofs of these theorems to Appendix C, and in the text we give only a brief exposition of the underlying arguments.

The critical value L is specified indirectly by a number of intermediate parameters, defined as follows:

$$\xi = \frac{-\alpha + \sqrt{\alpha^2 + 4\alpha/\ln(1/q)}}{2}$$

[i.e., ξ solves the equation $\xi(\alpha + \xi)\ln(1/q) = \alpha$].

Let $\lfloor \xi \rfloor$ be the largest integer not exceeding ξ, and define

$$\kappa = \lfloor \xi \rfloor \quad \text{if} \quad q \leqslant \frac{\lfloor \xi \rfloor}{(\lfloor \xi \rfloor + 1)} \frac{(\alpha + \lfloor \xi \rfloor + 1)}{(\alpha + \lfloor \xi \rfloor)},$$
$$= \lfloor \xi \rfloor + 1 \quad \text{otherwise.}$$

Note that κ is a strictly positive integer because $\kappa = 0$ implies $\lfloor \xi \rfloor = 0$ and therefore $q \leqslant 0$, contradicting the fact that $0 < q < 1$. Also, recall from §2 Observation 3 part (b) that for all $D \geqslant 1$, $\max N(D) \leqslant \theta$. Replacing β by 1 yields the following expression for θ:

$$\theta = \frac{\log V(1) - \log \beta}{\log(1/q)} = \frac{\log \frac{1+\alpha}{q}}{\log \frac{1}{q}},$$

thus $\theta > 1$.

Define

$$L_1 = \begin{cases} \theta & \text{if } \kappa = 1, \\ \kappa + \dfrac{\log \frac{\alpha+\kappa}{\kappa}}{\log \frac{1}{q}} & \text{if } 1 < \kappa \leqslant \dfrac{q}{1-q}\alpha, \\ \dfrac{q}{1-q}\alpha + 1 & \text{if } \kappa > \dfrac{q}{1-q}\alpha \text{ and } \kappa > 1. \end{cases}$$

$$L = \lfloor L_1 \rfloor.$$

THEOREM 2. *For all D, $D \in N(D)$ if and only if $D \leqslant L$.*

The proof is presented in Appendix C. Here we outline the basic arguments of the proof, and also explain the origins of the parameters ξ and κ that define L.

PROOF OUTLINE. For $1 < \theta < 2$ we prove the theorem directly by using definitions of θ and L. The proof for $\theta \geqslant 2$ is by induction on D: Assuming inductively that $d \in N(d)$ for $d = 1, \dots, D$ we show that $D+1 \in N(D+1)$ if and only if $D+1 \leqslant L$.

For this purpose, we define the function $y(x) = xq^x/(\alpha + x)$ on $x \geqslant 0$, and we prove that under the inductive assumption, $D+1 \in N(D+1)$ if and only if $q^{D+1} \geqslant \max\{y(k)\colon k = 1, \dots, D\}$. Thus, equivalently, we prove the theorem by showing that under the inductive assumption $[q^{D+1} \geqslant \max\{y(k)\colon k = 1, \dots, D\}$ if and only if $D+1 \leqslant L]$. Hence, we study the function $y(x) = xq^x/(\alpha+x)$ on $x \geqslant 0$ and show that it is unimodal with a peak at $x = \xi$. κ is the integer maximizer of $y(x)$, i.e., $y(\kappa) = \max\{y(k)\colon k = 1, 2, \dots\}$. As noted above, $\kappa \geqslant 1$. Finally, we show that L as defined above coincides with the largest integer $D+1$ satisfying $q^{D+1} \geqslant \max\{y(k)\colon k = 1, \dots, D\}$ which depends on how large κ is (hence the three different expressions in the definition of L_1).

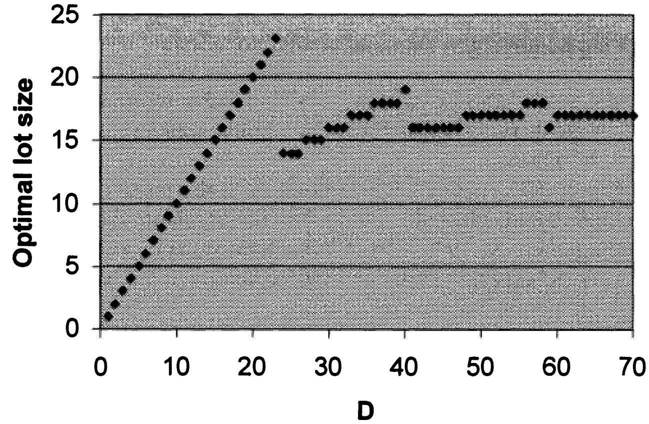
THEOREM 3. $\max\{k \in N(D), D = 1, 2, \dots\} = L$.

A typical pattern of how the optimal lot sizes change with the outstanding demand is given in Figure 1. The pattern exhibits clearly the properties of optimal policies noted in Theorem 3 and in the observations of §2.

Because the optimal lot-sizes and the critical value L are specified rather indirectly, it is not easy to see exactly how they depend on the system's hazard rates and production costs. Intuitively, the optimal lots (and the critical value L) should be smaller when (a) output quality is lower (not worthwhile to produce in large lots when many of the items are likely to turn out defective), and (b) set-up costs are lower. If output quality and/or set-up costs are sufficiently low, it may be best to avoid nontrivial lots altogether, and always produce only one item at a time. For the standard case studied in this section, we can give a simple precise specification of the conditions for which the production of a single unit is optimal for all demand levels.

COROLLARY 1. $N(D) = \{1\}$ for all $D = 1, 2, \dots$, (i.e., $L = 1$), if and only if $q < \frac{1}{\alpha+1}$.

Figure 1. Optimal lots for $q = 0.95$, $C(n) = 10 + n$. For these parameters, $\mu = 17$ and convergence of $N(D)$ to $\{17\}$ occurs as soon as $D \geqslant 60$.



PROOF. The condition $q < \frac{1}{\alpha+1}$ implies $\theta < 2$ [recall $q^{\theta-1}(\alpha+1) = 1$], hence by Observation 3(b), $N(D) = \{1\}$ for all D. On the other hand, if $N(2) = \{1\}$ then $U(2,2) - U(1,2) > 0$, and from Equation (2), $U(2,2) - U(1,2) = 1 - q^2 \frac{\alpha+1}{q}$, hence $q < \frac{1}{\alpha+1}$. □

5. CONCLUDING COMMENTS ON POSSIBLE EXTENSIONS

We conclude with a few comments concerning potential extensions of the results presented here. The results of this paper can be extended in two different directions. One possible direction for future research is to identify additional interesting properties of the optimal policies for the classes of production processes studied in this paper. In particular, consider Figure 1, which depicts the optimal lot sizes for a range of outstanding demands. Beyond the properties of the optimal policy stated in the previous sections, Figure 1 exhibits the following noteworthy property. When a unit increase in the outstanding demand induces a drop in the optimal lot size from, say, n to $m < n$, then the optimal lot sizes for higher outstanding demands are all between m and n. This property recurred consistently in extensive numerical investigations that we conducted. In the absence of formal proof, we state this as a conjecture for future investigation.

CONJECTURE. In the standard case,

if $n \in N(D)$, $m \in N(D+1)$ and $m < n$,
then for all $d > D$ $k \in N(d)$ implies $m \leqslant k \leqslant n$.

Another potential direction for future research is to investigate the robustness of the properties of the optimal policy studied in this paper with respect to the assumed characteristics of the production process. For example, the existence of a limit to $N(D)$ as D gets large, as discussed in §3, seems to apply also to other production processes that do not comply with the conditions of Theorem 1. It would be interesting to characterize the class of production processes for which a similar property holds.

APPENDIX

The appendices can be found at the *Operations Research* Home Page ⟨http://or.pubs.informs.org/pages/collect.html⟩.

REFERENCES

Anily, S. 1995. Single-machine lot sizing with uniform yields and rigid demands: robustness of the optimal solutions. *IIE Trans.* **27** 625–633.

Beja, A. 1977. Optimal reject allowance with constant marginal production efficiency. *Naval Res. Logist. Quart.* **24** 21–33.

Grosfeld-Nir, A. 1995. Single bottleneck systems with proportional expected yields and rigid demand. *Eur. J. Oper. Res.* **80** 297–307.

Grosfeld-Nir, A., Y. Gerchak. 1996. Production to order with random yields: single stage multiple lot sizing. *IIE Trans.* **28** 669–676.

Grosfeld-Nir, A., L. W. Robinson. 1995. Production to order on a two machine line with random yields and rigid demand. *Eur. J. Oper. Res.* **80** 264–276.

Lee, H. L., M. J. Rosenblatt. 1987. Simultaneous determination of production cycle and inspection schedules in production systems. *Management Sci.* **33** 1125–1136.

Levitan, R. E. 1960. The optimum reject allowance problem. *Management Sci.* **6** 172–186.

Llewellyn, R. W. 1959. Order sizes for job lot manufacturing. *J. Industrial Engrg.* **10** 176–180.

Monden, Y. 1983. *Toyota Production System*. Industrial Engineering and Management Press, Norcross, GA.

Porteus, E. L. 1986. Optimal lot sizing, process quality improvement and setup cost reduction. *Oper. Res.* **34** 137–144.

——. 1990. Impact of inspection delay on process and inspection lot sizing. *Management Sci.* **36** 999–1007.

Rosenblatt, M. J., H. L. Lee. 1986a. A comparative study of continuous and periodic inspection policies in deteriorating production systems. *IIE Trans.* **18** 2–9.

——, ——. 1986b. Economic production cycles with imperfect production processes. *IIE Trans.* **18** 48–55.

White, D. J. 1965. Dynamic programming and systems of uncertain duration. *Management Sci.* **12** 37–67.

Yano, C. A., H. Lee. 1995. Lot sizing with random yields: a review. *Oper. Res.* **43** 311–334.

Zhang, A. X., S. Guu. 1998. The multiple lot sizing problem with rigid demand and interrupted geometric yield. *IIE Trans.* **30** 427–431.