

The Stochastic Economic Lot Scheduling Problem: Cyclical Base-stock Policies with Idle Times

Awı Federgruen • Ziv Katalan

Graduate School of Business, Columbia University, New York, New York 10027
The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104

In this paper we discuss *stochastic* Economic Lot Scheduling Problems (ELSP), i.e., settings where several items need to be produced in a common facility with limited capacity, under significant uncertainty regarding demands, production times, setup times, or combinations thereof. We propose a class of production/inventory strategies for stochastic ELSPs and describe how a strategy which minimizes holding, backlogging, and setup costs within this class can be effectively determined and evaluated. The proposed class of strategies is simple but rich and effective: when the facility is assigned to a given item, production continues until either a specific target inventory level is reached or a specific production batch has been completed; the different items are produced in a given sequence or rotation cycle, possibly with idle times inserted between the completion of an item's production batch and the setup for the next item. An optimal strategy within the class can be determined, and all relevant performance measures can be evaluated in just a few CPU seconds, using a 486-based PC. We also derive a number of easily computable lower bounds for the optimal cost value and establish a comparison with deterministic ELSPs.

(Multi-item; Stochastic Inventory Systems; Single Capacitated Manufacturing Facility; Setup Times)

1. Introduction and Summary

We discuss *stochastic* Economic Lot Scheduling Problems (ELSP), i.e., settings where several items need to be produced in a common facility with limited capacity, under significant uncertainty regarding demands, production times, setup times, or combinations thereof. We propose a class of production/inventory strategies and describe how a strategy which minimizes holding, backlogging, and setup costs within this class can be effectively determined and evaluated. The proposed class of cyclical base-stock strategies is simple but rich and effective: when the facility is assigned to a given item, production continues until either a specific target inventory level is reached or a specific production batch has been completed; the different items are produced in a given sequence, possibly with idle times inserted be-

tween the completion of an item's production batch and the setup for the next item. An optimal strategy within the class can be determined and all relevant performance measures can be evaluated in just a few CPU seconds, using a 486-based PC. We also establish comparisons with several alternative production systems, in particular, deterministic ELSPs and settings where each item has a dedicated facility. These comparisons characterize the price that is to be paid for various sources of uncertainty or shared capacity units.

In spite of its importance as a natural and basic model for the interaction between related products, little is known about the *general* model. Deterministic versions have been addressed via mixed integer models, which in general are difficult to solve; see the recent survey by Salomon (1990) or the so-called deterministic ELSP in

which all items' demands are assumed to occur at a constant deterministic rate and all production and setup times are assumed to be deterministic as well. Even the deterministic ELSP is too hard to be solved: Hsu (1983) proved that the problem is NP-Complete. The rather sizable literature on this model (e.g., Maxwell 1964, Bomberger 1966, Elmaghraby 1978, Dobson 1987, Gallego 1991, and Zipkin 1991) confines itself to a number of simple classes of replenishment rules, e.g., the rotation cycle strategy: here, the facility follows a cyclical schedule in which the items are produced in a fixed but arbitrary sequence; each cycle consists of a production period (in which the quantity produced of each item equals its deterministically known demand during the cycle) possibly followed by a single idle period. Most importantly, the above deterministic models are ill-suited to represent significant sources of uncertainty with respect to the demand processes, production, and setup times.

We now briefly review the much more limited and more recent literature on *stochastic* ELSPs. Several papers deal with the special case where *no* setup costs or setup times are incurred when switching production from one item to the next. These papers focus on dynamic scheduling policies in which at any production completion epoch the decision as to which item to produce next, if any, is made on the basis of the complete system wide state vector, i.e., the vector of the inventories for all individual items. Optimal dynamic scheduling policies cannot be computed for all but the smallest size systems; moreover the structure tends to be highly complex prohibiting their implementation even if such strategies could be computed in a reasonable amount of time. See, however, Ha (1992) for a characterization of an optimal policy in the case of *two* products. All other papers deal therefore with the development of *heuristics*, e.g., Wein (1992), Veatch and Wein (1992), and Peña and Zipkin (1993).

In spite of concerted efforts to reduce setup times in Just-In-Time and related programs, these remain significant in most practical production settings. Yet, all standard stochastic inventory models are driven by setup *costs* and ignore setup *times*, even though, as pointed out by Karmarkar (1987), often no *explicit* setup costs are incurred, and the latter are used merely to represent the opportunity costs of setup times. Such setup costs

are therefore difficult to estimate and are subject to change over time. Moreover, instead of being fixed and exogenously given, they are in fact dependent on the specific production strategy employed. Finally, systems *with* setup times exhibit important *qualitative* differences as opposed to those without such setup times. On the other hand, there are many settings where incremental setup costs are incurred in addition to or instead of setup times. Duenyas and Van Oyen (1993) describe examples in the production of asphalt shingles. Our paper simultaneously considers general setup costs and setup times.

A few other papers deal with dynamic or semidynamic scheduling policies for systems *with* setup costs or times. Browne and Yechiali (1989a, b) and Duenyas and Van Oyen (1992) deal with the special case of systems without inventories. None of their (semi-) dynamic scheduling rules can be evaluated analytically. Zipkin (1986) and Karmarkar (1987) address models with setup times and costs; they consider settings in which the production facility fails to have timely information about finished goods inventories and restrict themselves to a class of strategies under which production batches are triggered by independent single item (r, q) -rules, and priority between batches is determined on a FIFO basis. Leachman and Gascon (1988), Leachman et al. (1991), and Bourland and Yano (1991, 1994) consider semidynamic adaptations of the deterministic rotation cycle policy. These heuristics can not be evaluated analytically. Graves (1980) proposes a method requiring the solution of $2^N - N - 1$ single item Markov Decision problems where N denotes the number of distinct items.

As observed by Sarkar and Zangwill (1989), the basic cyclical polling model (with exhaustive or gated service) represents the special control rule in our proposed class of strategies in which the base stock levels equal zero and no idle times are inserted; see Takagi (1986, 1990) for recent surveys.

The remainder of this paper is organized as follows. In §2 we describe the model, and in §3 the proposed class of production strategies and an efficient method for identifying and evaluating an optimal strategy within this class. In §4 we report on a numerical study in which the proposed production strategy and its cost value are compared with those arising in deterministic

versions, or in settings where each item is produced on a dedicated facility. We also comment on the sensitivity of various performance measures with respect to certain parameters.

2. The Basic Model

Consider a system with N distinct items, with demands generated by independent demand Poisson processes; λ_i is the demand rate of item i ($i = 1, \dots, N$) and $\lambda = \sum_{i=1}^N \lambda_i$. (See §3 for generalizations of the results in this paper to *compound* Poisson demand processes.) The N items are produced in a common facility which can produce at most one item at a time. Production times for individual units are independent; those of item i are identically distributed with cdf $S_i(\cdot)$, mean $s_i < \infty$ and k th moment $s_i^{(k)}$ ($k \geq 2$) ($i = 1, \dots, N$). A setup time with cdf $R_i(\cdot)$, first moment $r_i < \infty$ and k th moment $r_i^{(k)}$ ($k \geq 2$) is incurred whenever the facility *starts* producing item i after being idle or after producing some other item. Consecutive setup times are independent. (More generally, our results are easily extended to the case of sequence dependent setup times; see the discussion below.) The utilization rate for item i is $\rho_i = \lambda_i s_i$; that of the system equals $\rho = \sum_{i=1}^N \rho_i$. We assume the system is stable, i.e., $\rho < 1$. Unfilled demand is backlogged.

Three types of costs are incurred. Let

$h_i(x)(p_i(x))$ = the inventory carrying (backlogging) cost for item i per unit of time at which x units of item i are carried in stock (backlogged) ($i = 1, \dots, N$).

K_i = the setup cost incurred per setup of item i ($i = 1, \dots, N$).

The functions $h_i(\cdot)$ and $p_i(\cdot)$ are convex and nondecreasing. The objective is to minimize the long run costs per unit time. Often, one prefers to control stockouts via service level constraints, e.g., lower bounds on the items' fill rate. This variant of the model calls for a minor adjustment; see §3.

3. A Class of Production Strategies: An Efficient Optimization and Evaluation Method

Under a base-stock policy, the facility rotates between the items in a fixed sequence, without loss of generality

the permutation $(1, \dots, N)$. When turning to item i , either (a) the facility continues to produce this item until its inventory level is increased to a base-stock level b_i , or (b) the facility produces a batch the size of which equals the difference between a base-stock level b_i and the prevailing inventory level. We refer to alternative (a) ((b)) as the exhaustive (gated) case. We initially assume that the same type of service (i.e., exhaustive or gated) is provided to *all* items; see subsection 3.4 for a brief discussion of the case of mixed service. When terminating production for item $i-1$ (modulo N), a (deterministic) idle time Δ_i is inserted prior to setting up for item i . Insertion of idle times may be beneficial to reduce the frequency of setups and hence the average setup cost. More surprisingly and as shown by Sarkar and Zangwill (1991), insertion of idle times may even be beneficial in the absence of setup costs, in particular when some of the setup times are highly variable. (It is unknown as yet whether the same phenomenon can occur under fully optimal or more sophisticated dynamic policies. As demonstrated by Sarkar and Zangwill, the phenomenon *can* occur under the base-stock policies considered here.)

The choice of the above class of base-stock policies is motivated by the following considerations. First, base-stock policies are easy to implement and can effortlessly be monitored with minimal informational requirements. Indeed, many manufacturing companies have adopted some form of cyclical scheduling; see, e.g., Hall (1982), Schonberger (1987), Leachman and Gascon (1988), and Smith et al. (1993).

Second, base-stock policies are a natural generalization of the rotation cycle policies advocated for deterministic ELSPs where they are close to optimal under most reasonable parameter combinations; see, e.g., Jones and Inman (1989). Also, a base-stock policy governs each item's inventory via a variant of a (T, S) -rule. Under a *pure* (T, S) -rule, the item's inventory position is increased to a base-stock level S every T time units; under a base-stock policy, production of a given item is likewise continued until a fixed "target level" is reached, but the interreplenishment interval T is somewhat random. (T, S) rules are among the most widely used policies in *single* item systems; see, e.g., Silver and Peterson (1985). They are also effective in coordinating replenishments across items so as to exploit economies

of scale; see, e.g., Atkins and Iyogun (1988). Finally, an alternative generalization of the rotation cycle policies for deterministic ELSPs would prescribe a cyclical schedule with *fixed* production and idle time intervals of appropriately chosen lengths and hence *random* target levels S . However, no acceptable analytical method is known (or, as shown in Borst 1994, is even likely to exist) to evaluate a single such time-window policy, let alone to identify the best N -vector of production windows. Leachman and Gascon's (1988) heuristic can be viewed as a variant of the time-window policies.

One may consider certain *dynamic* adjustments of base-stock policies, e.g., (i) where an item is skipped when it is its turn to be produced but its inventory is still at its base-stock level, or (ii) where upon completion of an item the facility switches to one with the largest shortfall from its base-stock level (perhaps weighted by the holding cost rate), or (iii) the length of the inserted idle times is determined dynamically. No analytical evaluation method is available for any of these dynamic adjustments. Moreover, they do not need to result in improvements; see, e.g., Duenyas and Van Oyen (1992). Also, the event under which an item is skipped under policy (i) is rare under reasonably large utilization rates and/or cycle times, so that the performance measures of the static base-stock policy may be used as a good approximation for those of the dynamic policy (i), even if the latter is desired. Bertsimas and Xu (1993) show, for make-to-order systems, that policy (ii) is often inferior to a static policy.

Contrary to the case of deterministic ELSPs, the chosen permutation cycle has an impact on various performance measures, albeit very minor compared to that of the base-stock levels and idle times; see §4 for details. Thus, for all practical purposes, a single permutation can be chosen arbitrarily. In case setup times are sequence dependent, we suggest to choose the permutation which optimizes the Traveling Salesman Problem with the mean switchover time between items i and j as the distance between them.

3.1. An Optimization Method for Base-stock Policies

A base-stock policy is specified by an N -vector of base-stock levels $b = (b_1, \dots, b_N)$ and a vector of idle times $\Delta = (\Delta_1, \dots, \Delta_N)$. Note that under a given idle time

vector Δ the process of system-wide inventories is related to the queue size process in a cyclical polling system, i.e., a system with N stations at which customers arrive to be serviced by a common server. The server visits the stations in a fixed permutation and stays at a station until its queue is emptied out (the exhaustive case) or all customers present upon his arrival have been served, (the gated case). When the server switches between stations a station-specific switchover time is incurred.

For a given vector Δ , the "corresponding polling system" is specified as follows: each item is identified with a station. The arrival processes at the stations are given by the items' Poisson demand processes. Unit service times at a station are given by the unit production times of the corresponding item. The switchover times from station i to $i + 1$ (modulo N) are given by the setup times for item $i + 1$, each augmented by Δ_{i+1} .

For a given vector of base-stock levels b , let for all $i = 1, \dots, N$:

$IL_i(t)$ = the inventory level of item i at time t .

$L_i(t)$ = the queue size at station i in the corresponding polling system at time t .

The correspondence between the ELSP and the polling system implies:

$$IL_i(t) = b_i - L_i(t) \quad \text{for all } t \geq 0. \quad (1)$$

Since the system is stable ($\rho < 1$) it is easily verified to be regenerative, e.g., at epochs at which production of item 1 is terminated while the inventory levels of all items equal their respective base-stock levels (i.e., in the corresponding polling system, the system is empty). In particular, the processes $IL_i(t)$ and $L_i(t)$ converge to steady-state distributions IL_i and L_i ($i = 1, \dots, N$). Likewise, let

C_i = the steady-state cycle time i.e., the time between two consecutive epochs at which production of item i is started ($i = 1, \dots, N$).

The C_i -variables may have different *distributions* but their means coincide (see Takagi 1986).

$$EC_i = \bar{C}, \quad i = 1, \dots, N$$

$$\text{where } \bar{C} \stackrel{\text{def}}{=} \frac{\sum_{i=1}^N (r_i + \Delta_i)}{1 - \rho}. \quad (2)$$

In view of (1), the long run average cost under a given base-stock policy is thus given by:

$$\frac{\sum_{i=1}^N K_i}{\bar{C}} + \sum_{i=1}^N E\{h_i(IL_i^+) + p_i(IL_i^-)\}$$

$$= \frac{\sum_{i=1}^N K_i}{\bar{C}} + \sum_{i=1}^N E\{h_i([b_i - L_i]^+) + p_i([L_i - b_i]^+)\}, \quad (3)$$

where $x^+ = \max(x, 0)$ and $x^- = \max(-x, 0)$.

The advantage of this representation is that the distribution of the random variables $\{L_i; i = 1, \dots, N\}$ is independent of the vector of base-stock levels. We conclude:

PROPOSITION 1. Consider all base-stock policies with a given vector of idle times Δ . Let $\{L_i; i = 1, \dots, N\}$ denote the steady-state queue sizes in the corresponding polling system. For each item i ($i = 1, \dots, N$), one obtains the optimal base-stock level b_i^* by determining the unique minimum of the (single variable) convex function

$$\psi_i(x) = E\{h_i([x - L_i]^+) + p_i([L_i - x]^+)\}.$$

In other words, the optimal base-stock level b_i^* is obtained by solving a newsboy problem with L_i as the demand distribution ($i = 1, \dots, N$).

PROOF. The vector b has no impact on the average setup cost, i.e., the first term in (3) and the distribution of $\{L_i; i = 1, \dots, N\}$ is independent of b as well. (3) is thus separable in b ; in particular b_i^* is obtained by minimizing the single variable function $\psi_i(\cdot)$ which is convex (since $h_i(\cdot)$ and $p_i(\cdot)$ are convex and nondecreasing), and hence has a unique minimum. \square

REMARK 1. In case a minimum fill rate α is to be guaranteed for some $i = 1, \dots, N$, one sets b_i^* as the 100 α th percentile of the L_i -distribution.

It thus remains to be shown how for any given vector of idle times Δ the distributions $\{L_i; i = 1, \dots, N\}$ can be computed and in particular how an optimal vector Δ can effectively be determined. We address the second question first.

PROPOSITION 2. The long run average cost (3) depends on Δ only via $\Delta_{\text{tot}} = \sum_{i=1}^N \Delta_i$.

PROOF. The distribution of $\{L_i; i = 1, \dots, N\}$ and hence, the second and third terms in (3) depend on Δ only via $\Delta_{\text{tot}} = \sum_{i=1}^N \Delta_i$; see Theorems 1(a) and 3(a) in Federgruen and Katalan (1993). The same clearly holds for the first term in Equation (3), see Equation (2). \square

In other words, a single idle time period, inserted prior to the setup of any of the N items, can be used in any given cycle *without loss of optimality*. The search for optimal idle times thus reduces to that for a single scalar Δ_{tot} . An optimal base-stock policy is thus obtained by minimizing the single variable function $\Phi(\cdot)$ where

$\Phi(\Delta)$ = the minimum average cost when inserting a single idle time of length Δ in every cycle and employing an optimal corresponding vector of base-stock levels.

We now show how $\Phi(\Delta)$ can efficiently be evaluated. In view of (3) and proposition 1, the evaluation is straightforward given the distributions of $\{L_i; i = 1, \dots, N\}$. In the next subsection, we describe a fast method to determine the L_i -distributions for any initial value of Δ . An even faster procedure (in subsection 3.3) can be used for any subsequent values of Δ , as required when searching for an optimal value of Δ .

REMARK 2. Based on partial results in Katalan (1995, p. 144), we claim that nothing is gained by implementing random idle times.

3.2. Evaluation of the L_i -distributions

Federgruen and Katalan (1994) recently developed an efficient algorithm to compute the complete steady-state queue size distributions in polling systems. While approximate, the method is remarkably accurate as verified in an extensive simulation study. Here, we confine ourselves to a brief description of this method. See the appendix for a complete algorithmic description.

Fix $i = 1, \dots, N$. Let X_i be the steady-state number of customers at station i at a polling instant there, i.e., an instant when the server is ready to resume service and B_i the busy period at station i , i.e., the amount of time service is provided at station i during an arbitrary cycle. It follows from Fuhrmann and Cooper (1985) that L_i can be decomposed as the independent sum of two simpler components:

$$L_i = L_i' + L_i''. \quad (4)$$

Here L_i' is the steady-state number of customers at station i , if the server were exclusively assigned to this station, i.e., the steady-state queue size in an M/G/1 queue with arrival rate λ_i and service time distribution $S_i(\cdot)$; L_i'' is the number of customers at station i at an arbitrary tagged epoch within one of station i 's intervisit

periods I_i . (An intervisit period starts when the server completes serving the station and ends at the next epoch at which he is ready to resume service there.) The values of the pdf of L_i' can be computed via a simple linear recursion, see Tijms (1986). As to L_i'' , under exhaustive service, it is the number of Poisson arrivals with rate λ_i in an interval of time distributed as I_i' , the forward recurrence time of I_i . The distribution of I_i is approximated by a mixture of Erlang distributions chosen to match a prespecified number of moments (say m). The distribution of I_i' is thus approximated as a mixture of Erlangs itself and that of L_i'' as a mixture of negative binomials with parameters easily computed from the first m moments of I_i . But,

$$EI_i' = \lambda_i^{-r} E(X_i(X_i - 1) \cdots (X_i - r + 1)), \quad r \geq 1, \quad (5)$$

and the moments of X_i can be obtained efficiently via the method in Konheim et al. (1994).

Under gated service, an intervisit period starts with the customers that arrived to the station during the preceding busy period. L_i'' is thus the number of arrivals of item i in an interval of time \bar{I}_i ending at an arbitrary tagged epoch during an intervisit period and starting at the beginning of the preceding busy period. \bar{I}_i is again approximated by a mixture of Erlangs with parameters chosen to match a pre-specified number of moments. Close approximations for the moments are obtained from the corresponding moments of C_i and B_i ; these in turn are easily computed from those of X_i , again efficiently obtained via the method of Konheim et al. (1994).

The approximation method for the distribution of L_i is both extremely accurate and fast, even when using two moment approximations only. Among 1746 problem instances, with numbers of items varying from 5 to 50, and systematic variation from light (balanced) traffic to heavy (unbalanced) traffic and with a variety of distributional forms for the setup and production time distributions, Federgruen and Katalan observe a maximum pointwise absolute difference between the approximated cdf curves of the variables L_i ($i = 1, \dots, N$) and the exact cdf curves, estimated via high precision simulation, of no more than 0.018. The relative accuracy of the method is of the same order of magnitude, guaranteeing high precision even for rare events.

To assess how accurately the optimal base stock levels are determined, consider for an item i the most prevalent case with $h_i(x) = h_i x$ and $p_i(x) = p_i x$ for given constants $h_i, p_i > 0$. Proposition 1 shows that $b_i^* = L_i^{-1}(p_i/p_i + h_i)$ with L_i^{-1} the inverse of the cdf of L_i . Federgruen and Katalan show that the difference between the approximated and the exact values of L_i^{-1} is typically less than or equal to one for critical ratios $p_i(p_i + h_i)^{-1}$ varying from 0 to 0.999; differences of more than one unit arise only when the relative error is less than 3%.

The approximation method is also extremely fast. Even instances with 50 stations and high traffic intensity require only a few CPU seconds on a PC. The complexity of the method, for a given choice of Δ , is given by $O(N \max\{k^{*2}, N \log_p \epsilon\})$ to evaluate the first k^* pdf values of the queue sizes in a complete system, with ϵ the desired numerical precision.

3.3. Optimization of the Inserted Idle Time Δ

Having shown how $\Phi(\Delta)$ may efficiently be evaluated for a given value of Δ , we now show how it can be minimized efficiently over Δ . We first describe how after an *initial* evaluation of the function $\Phi(\cdot)$ an even faster procedure may be used for *subsequent* evaluations. Note first from equations (2) and (3) that the average setup cost component is an explicit hyperbolic function of Δ : $(\sum_{i=1}^N K_i)(1 - \rho) / (\sum_{i=1}^N r_i + \Delta)$. Next, addressing the remaining cost components, fix $i = 1, \dots, N$ and consider the effort to compute the holding and backlogging cost for item i under alternative values of Δ . The distribution of L_i' is independent of Δ ; i.e., it never needs to be reevaluated. As to L_i'' , recall that its evaluation starts with the computation of m moments of X_i . Thereafter only a few closed form expressions need to be evaluated to determine the parameters of the mixture of negative Binomials approximation of L_i'' , the convolution of this distribution with that of L_i' is computed, and holding and backlogging costs for item i are determined by the solution of a single newsboy problem; see proposition 1.

We now show that after the initial evaluation of $\Phi(\cdot)$, the required m moments of X_i are available as given polynomials of Δ with known coefficients, reducing their reevaluation to a trivial computation. Federgruen and Katalan (1993) show:

$$X_i(\Delta) = X_i(0) + Y_i(\Delta),$$

with $X_i(0)$ independent of $Y_i(\Delta)$, (6)

Here $Y_i(\Delta)$ is the number of events in Δ time units in a compound Poisson process with rate λ and compounding distribution Z_i , which is independent of Δ and whose first m moments $z_i^{(1)}, \dots, z_i^{(m)}$ ($m \geq 1$) are easily obtained in the process of computing the first m moments of $X_i(0)$ via the method of Konheim et al. (1994). In particular:

PROPOSITION 3. Fix $i = 1, \dots, N$. Let $\xi_i^r = E\{X_i(0)\}^r$ for $r = 1, \dots, m$.

(a)

$$E\{X_i(\Delta)\}^r = \sum_{l=0}^r \binom{r}{l} \xi_i^{r-l} \sum_{\sum_{i=1}^l a_i = l} \lambda^{(a_1 + \dots + a_l)} \prod_{k=1}^l \{z_i^{(k)}\}^{a_k} \Delta^{(a_1 + \dots + a_l)},$$

i.e., the r th moment of $X_i(\Delta)$ is a polynomial of degree r in Δ .

(b) In the exhaustive case, $E\{I_i(\Delta)\}^r$ is a polynomial of degree r in Δ and in the gated case, $E\{C_i(\Delta)\}^r$ is a polynomial of degree r in Δ .

PROOF. (a) It follows from $X_i(\Delta) = X_i(0) + Y_i(\Delta)$ that

$$E\{X_i(\Delta)\}^r = \sum_{l=0}^r \binom{r}{l} \xi_i^{r-l} E\{Y_i(\Delta)\}^l.$$

Note that

$$Y_i(\Delta) = \sum_{p=1}^M Z_{i,p} \Delta, \quad (7)$$

where $Z_{i,1}, \dots, Z_{i,M}$ are independent and distributed as Z_i . M has an independent Poisson distribution with rate $\lambda\Delta$. Thus, conditioning on M , we obtain:

$$E\{Y_i(\Delta)\}^l = E_M E \left[\left\{ \sum_{p=1}^n Z_{i,p} \right\}^l \middle| M = n \right].$$

Let $n^{(a)} \stackrel{\text{def}}{=} n \cdot (n-1) \cdots (n-a+1)$ for any integer $a \geq 1$. In any given term in the multinomial expansion of $\{\sum_{p=1}^n Z_{i,p}\}^l$ and for any $k = 1, \dots, l$ let a_k denote the number of distinct factors of the type $(Z_{i,p}^k)$ for some index $\leq p \leq n' \leq p \leq n$. For any given choice of numbers a_1, \dots, a_k note that $\sum_{k=1}^l k \cdot a_k = l$ and that there are $n^{(a_1 + \dots + a_l)}$ distinct terms in the multinomial expansion

with a_k as the number of factors that are k th powers of the same random variable, since any such term includes $(a_1 + \dots + a_l)$ distinct random variables and there are $n^{(a_1 + \dots + a_l)}$ (ordered) combinations of such variables. Moreover, each such term in the multinomial expansion, has expected value $\{\prod_{k=1}^l [z_i^{(k)}]^{a_k}\}$. But $EM^{(a_1 + \dots + a_l)} = (\lambda\Delta)^{(a_1 + \dots + a_l)}$ since the r th factorial moment of a Poisson variable equals the r th power of its mean.

(b) Immediate from part (a) and

$$EI_i^r = (EC_i^r) = EX_i(X_i - 1) \cdots (X_i - r + 1)$$

under exhaustive (gated) service; see Equation (5). \square

As far as an appropriate search method for the optimal idle time Δ^* is concerned, any standard method for minimizing a nonlinear function of a single variable can be used; see Rinnooy Kan and Timmer (1989). Moreover, in our numerical experience, the function $\Phi(\cdot)$ is always quasi-convex, i.e., it has a unique local minimum. This property permits the use of a bisection technique under which the number of evaluations of $\Phi(\cdot)$ is merely logarithmic in the length of the search interval. We conjecture that the function $\Phi(\cdot)$ is quasi-convex in general, based on the following partial characterization and the above numerical experience. In the remainder of this subsection we restrict ourselves to the case of exhaustive service.

PROPOSITION 4. Under exhaustive service, all moments of I_i^r , $i = 1, \dots, N$, are quasi-convex in Δ .

PROOF. We first show that the family $\{X_i(\Delta): \Delta \geq 0\}$ is stochastically increasing convex (SICX) in Δ , i.e., $Ef(X_i(\Delta))$ is a nondecreasing convex function of Δ for any nondecreasing convex function f . $X_i(0)$ is clearly (SICX) in Δ . Also, $\{Y_i(\Delta): \Delta \geq 0\}$ is (SICX) in view of Equation (7) and Theorem 6.8 in Shaked and Shanthikumar (1990). Thus, $X_i(\Delta) = X_i(0) + Y_i(\Delta)$ is (SICX) (see Theorem 5.6 in Shaked and Shanthikumar).

Now, fix an integer $r \geq 1$ and note that

$$f(x) = x(x-1)^+(x-2)^+ \cdots (x-r+1)^+$$

is nondecreasing and convex as the product of r nondecreasing, nonnegative, and convex functions. Thus, using (5),

$$\begin{aligned} EI_i^r &= \lambda_i^{-r} Ef(X_i(\Delta)) \\ &= \lambda_i^{-r} EX_i(\Delta)(X_i(\Delta) - 1) \cdots (X_i(\Delta) - r + 1) \end{aligned}$$

is convex in Δ . Next, $E(I_i^r)^r = EI_i^{r+1}/(r+1)EI_i$; see Tijms (1986, p. 6), and by Proposition 3(b) EI_i is linear in Δ , so that $E(I_i^r)^r$ is quasi-convex in Δ . (If $f(\cdot)$ is a convex function then $g(x) \stackrel{\text{def}}{=} f(x)/(ax+b)$ for given nonnegative constants a, b is quasi-convex, since for any $c > 0$, the set

$$A_c \stackrel{\text{def}}{=} \{x: g(x) \leq c\} = \{x: f(x) \leq axc + bc\}$$

is convex.) \square

Proposition 4 suggests that $\{I_i^r(\Delta): \Delta \geq 0\}$ may be stochastically convex, in which case $L_i^r(\Delta)$ and hence $L_i(\Delta)$ would be stochastically convex, implying that the long run average cost is convex in Δ for any given vector of base stock levels.

In summary, at the end of §3.2, we have demonstrated that a single value of the function $\Phi(\cdot)$, i.e., the minimum cost value for a *given* idle time, can be evaluated in just a few CPU seconds, even for systems with 50 items and high utilization rate. Since subsequent evaluations of $\Phi(\cdot)$ can be performed even faster, and only few evaluations are required, an overall optimal strategy within the class of base-stock strategies can be found with comparable effort.

3.4. Generalizations

Both the model and the class of base-stock policies are easily extended to allow for a number of important generalizations. First, consider the case where the demands are generated by *compound* rather than unit Poisson processes. One easily verifies, using Corollary 6 in Federgruen and Katalan (1993), that without loss of optimality, at most a *single* idle time needs to be inserted in each cycle. For a given idle time Δ and corresponding distributions $\{L_i; i = 1, \dots, N\}$ optimal base-stock levels can be computed as described in Proposition 1. The problem thus reduces once again to that of minimizing the function $\Phi(\cdot)$. The above method for the initial evaluation of $\Phi(\cdot)$ is easily modified; see Federgruen and Katalan (1994). Also, (6) continues to apply, see Corollary 6 in Federgruen and Katalan (1993), allowing for the same acceleration in subsequent evaluations of $\Phi(\cdot)$.

It is also possible to use mixed service, i.e., to provide a different *type* of service to different items, treating some items with exhaustive and some with gated service. Optimality of a single idle time per cycle, the above fast initial evaluation method of $\Phi(\cdot)$ and its enhance-

ment for subsequent evaluations, all continue to apply with minor modifications; see Federgruen and Katalan (1993, 1994).

4. Comparison with Alternative Lot Scheduling Systems and Numerical Study

The most common strategies for ELSPs are those derived for their deterministic version. A straightforward adaptation of the deterministic rotation cycle strategy to our stochastic setting, specifies a base-stock policy in which each item's base-stock level and the idle time per cycle are chosen as the maximum inventory level and idle time under the deterministic rotation cycle policy. In this section, we report on a numerical study which compares this heuristic with an optimal base-stock rule.

We also establish comparisons with several alternative production systems, in particular deterministic ELSPs and settings where each item has a dedicated facility. These comparisons characterize the price to be paid for various sources of uncertainty or shared capacity. We first evaluate the minimum average cost incurred if the facility could be dedicated exclusively to each item and if production could be initiated without setup time. This is clearly a lower bound for the *minimum* system-wide cost under any feasible policy. It decomposes into N stochastic single item problems for which an (s, S) -policy is optimal; see Federgruen and Zheng (1993). Under an (s, S) policy, production of the item is continued until a target level S is reached, and production is resumed when inventory drops to a level $s < S$. Federgruen and Zheng also show how the optimal (s, S) -parameters and the corresponding minimum average cost value $Z(s, S)$ can be computed with a fast optimization procedure.

The cost of an optimal deterministic rotation cycle is another lower bound for the minimum cost value among all *stochastic* base-stock (but not necessarily among all *feasible*) policies. This bound can be represented by a closed form expression. Consider, e.g., the most prevalent case where all inventory and backlogging costs are *linear*, i.e., $h_i(x) = h_i x$ and $p_i(x) = p_i x$ for given constants $h_i, p_i > 0$ ($i = 1, \dots, N$). If a rotation cycle of length T is employed, one easily verifies that the average holding and backlogging cost of item i is

given by $\lambda_i T(1 - \rho_i)h_i p_i / 2(h_i + p_i)$ and the total average cost by

$$\frac{\sum_{i=1}^N K_i}{T} + T \sum_{i=1}^N \frac{\lambda_i(1 - \rho_i)h_i p_i}{2(h_i + p_i)}. \quad (8)$$

To be feasible, we must have $T \geq \sum_{i=1}^N r_i + \rho T$ or $T \geq \sum_{i=1}^N r_i / (1 - \rho)$ to allow for $\sum_{i=1}^N r_i$ in setup time and ρT in production time. (The lower bound for T equals \bar{C} when $\Delta = 0$; see (2).) The minimum cost value among all rotation cycles is the minimum of (8) subject to the bound on T :

$$Z_D = \frac{\sum_{i=1}^N K_i}{T^*} + T^* \sum_{i=1}^N \frac{\lambda_i(1 - \rho_i)h_i p_i}{2(h_i + p_i)} \quad \text{where}$$

$$T^* = \max \left\{ \sqrt{\frac{2 \sum_{i=1}^N K_i}{\sum_{i=1}^N \frac{\lambda_i(1 - \rho_i)h_i p_i}{2(h_i + p_i)}}}, \frac{\sum_{i=1}^N r_i}{(1 - \rho)} \right\}. \quad (9)$$

In our numerical study we have evaluated 360 problem instances which are partitioned into ten sets of 36 instances each. All instances have five items, all production times are exponential with mean one, and all setup times are Erlangs with five phases and mean one. All items share the same cost parameters h , p , and K . Sets 1–3 represent our *base* category with settings with relatively high traffic ($\rho = 0.8$) and significant *imbalance* in the workload associated with the different items, i.e., $\rho_{\max}/\rho_{\min} = 16$ where $\rho_{\max} = \max\{\rho_i\}$ and $\rho_{\min} = \min\{\rho_i\}$. Sets 4–6 represent settings with the same total utilization rate $\rho = 0.8$ but with relatively *balanced* workloads ($\rho_{\max}/\rho_{\min} = 1.6$). With sets 7–9 we return to the base category, merely changing the total utilization rate from $\rho = 0.8$ to $\rho = 0.5$. Finally, in set 10 we systematically vary the total utilization rate from $\rho = 0.15$ to $\rho = 0.9$, under a given set of relative utilization rates for the different items reflecting *moderate* imbalance ($\rho_{\max}/\rho_{\min} = 7$). In sets 1, 4, and 7 we systematically vary the values of h and p for a fixed value of K ; likewise, in sets 2, 5, and 8 (3, 4, and 9) we vary the values of h and K (p and K), fixing the value of $p(h)$. Each of the three cost parameters is systematically chosen from a list of six possible values:

$$h = \{1, 5, 10, 15, 20, 25\}; \quad p = \{5, 25, 50, 75, 100, 125\};$$

$$K = \{20, 200, 400, 600, 800, 1000\}.$$

(Each possible set consists of 36 instances since two parameters are varied in each.) Finally, in set 10 we fix h and p and vary ρ and K , choosing once again six distinct values for each.

The following performance measures have been evaluated. (All these measures refer to the case of exhaustive service, except for Z_S^{gated} defined below.)

∂b = the maximum difference across all items between the base-stock level in the optimal (stochastic) base-stock rule and the base-stock level in the optimal rotation cycle in the deterministic version of the ELSP.

b_s = the base-stock level in the optimal stochastic base-stock rule for which the maximum difference Δb is achieved.

C_s = the expected cycle length in the optimal stochastic base-stock rule (identical for all items; see (2)).

C_D = the length of the optimal deterministic rotation cycle.

Δ_s = the inserted idle time in the optimal stochastic base-stock rule.

Δ_D = the idle time in the optimal deterministic rotation cycle.

Z_s = the cost value of the optimal stochastic base-stock rule under exhaustive service.

Z_S^{gated} = the cost value of the optimal stochastic base-stock rule under gated service.

Z_{SD} = the cost value of the stochastic base-stock rule obtained by adapting the optimal rotation cycle to the stochastic setting, as explained above.

Finally Z_D and $Z(s, S)$ are the two lower bounds described above.

The specific values of these performance measures for all 360 instances can be found in Katalan (1995). Here, we display (in Table 1) the results for one of the sets, set 1. The results imply a number of important conclusions: first, to appropriately manage the risk due to uncertain demands, production, and setup times, an optimal base-stock policy employs significantly less idle time, i.e., significantly shorter cycles than the optimal rotation cycle in the deterministic version. In other words, reducing the uncertainty in the system results not only in a significant cost reduction but allows for a larger fraction of the available capacity to be reserved for new or unanticipated activities. The value C_s/C_D can be as large as 1.47; that of Δ_D/Δ_s can be as large as 4.4.

Table 1 Inventory Cost versus Backlogging Cost (Set 1: Unbalanced system ($\rho = 0.8$, $K = 500$))

N	h	p	∂b	b_s	C_s	C_D	Δ_s	Δ_D	Z_s	Z_s^{gated}	Z_{SD}	Z_D	$Z(s, S)$
1	1	5	2	22	97.2	101.6	14.4	15.3	67.2	71	69.2	49.2	46.1
2	1	25	6	28	80.8	94.6	11.2	13.9	89.7	94.7	132.2	52.9	49.9
3	1	50	9	31	79.5	93.7	10.9	13.7	98.9	104	209	53.4	51.8
4	1	75	9	31	72	93.4	9.4	13.7	103.8	109.4	285.6	53.6	53.1
5	1	100	10	32	70.6	93.2	9.1	13.6	107.4	112.9	362.1	53.6	53.9
6	1	125	11	33	71.3	93.1	9.3	13.6	110.1	115.6	438.3	53.7	54.6
7	5	5	0	0	57.3	58.6	6.5	6.7	109.7	119.6	109.8	85.2	79.8
8	5	25	2	11	37.4	45.4	2.5	4.1	185.8	199.6	202.2	110	102.7
9	5	50	4	13	36	43.5	2.2	3.7	220	235.5	295.2	114.9	109.1
10	5	75	4	14	35.6	42.8	2.1	3.6	240.6	255.7	370.1	116.7	113.8
11	5	100	5	15	35.6	42.5	2.1	3.5	255.3	270.1	452.6	117.6	118
12	5	125	6	16	35.6	42.3	2.1	3.5	267	281.5	507.6	118.2	121.8
13	10	5	-1	0	47	50.8	4.4	5.2	123.8	137.6	124.1	98.4	92.3
14	10	25	1	7	30	34.7	1	1.9	238.1	258.8	244.3	144.1	136
15	10	50	3	9	27.5	32.1	0.5	1.4	295	319	347.1	155.6	144.8
16	10	75	3	10	26.9	31.2	0.4	1.2	332.1	356.5	424.1	160.1	154.1
17	10	100	3	10	25	30.8	0	1.2	358.6	386.1	509.8	162.5	160.2
18	10	125	4	11	25	30.5	0	1.1	377.4	406	564.2	164	165.5
19	15	5	-1	0	47.1	47.9	4.4	4.6	128.9	146	131	104.4	97.7
20	15	25	0	0	27	30.3	0.4	1.1	267.6	294.4	271.2	165.1	154.9
21	15	50	2	7	25	27.3	0	0.5	345.6	378.8	378	183.1	167.5
22	15	75	3	8	25	26.2	0	0.2	393.8	430.2	478.5	190.6	180.7
23	15	100	4	9	25	25.7	0	0.1	431.5	469.8	549.5	194.7	190.2
24	15	125	5	10	25	25.3	0	0.1	461.9	501	636	197.3	197.2
25	20	5	0	0	46	46.4	4.2	4.3	132.3	151	132.3	107.8	100.5
26	20	25	1	5	27.4	27.8	0.5	0.6	287.1	318.6	289.8	179.7	167.8
27	20	50	2	6	25	25	0	0	382.3	423.5	404.1	203.8	185.2
28	20	75	2	7	25	25	0	0	446.1	491.4	500.8	214.7	197.4
29	20	100	3	8	25	25	0	0	491.8	540.3	587.8	221.1	211.9
30	20	125	4	9	25	25	0	0	529.7	580.9	685.9	225.3	222.1
31	25	5	0	1	45.8	45.4	4.2	4.1	134.5	154.3	134.7	110	102.7
32	25	25	1	4	25	26.2	0	0.2	302.6	338.8	305.8	190.6	178.3
33	25	50	2	6	25	25	0	0	414.6	460.9	421.8	221.1	200.3
34	25	75	2	7	25	25	0	0	488.3	543	521.3	236.2	212.6
35	25	100	3	8	25	25	0	0	545.7	602.8	627.5	245.3	230.6
36	25	125	3	8	25	25	0	0	589.7	650.4	709.8	251.4	241.6

Second, comparing the Z_s and Z_D measures, one concludes that a very significant price is paid for the uncertainty in the demand processes, production and setup times. The ratio Z_s/Z_D is sometimes as large as 4; this ratio increases significantly in all problem instances as the cost of backlogging (p) is increased. In other words, the cost of unreliable production or setup times or that of variable demands is particularly large when a high level of service is required. The cost ratio does

not vary significantly with h , the cost rate of carrying inventories; it decreases significantly as K , the setup cost, is increased. This is due to the fact that with larger values of K , the average setup cost component becomes more dominant and while it is larger than the average setup cost in the deterministic version, due to the use of less idle times and smaller cycle lengths, the increase in this component is less significant than that of the carrying and backlogging cost component.

Figure 1 Deterministic versus Stochastic Base-stock Policies

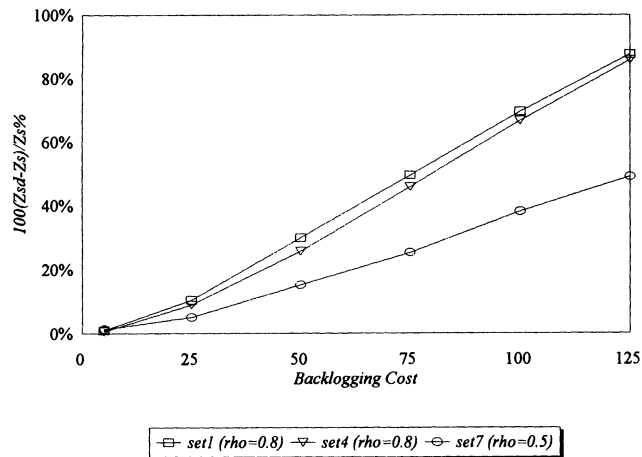
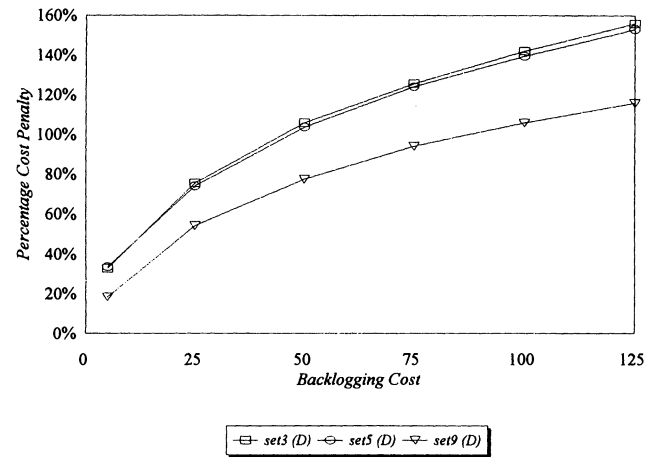


Figure 3 Cost of Variability



Another important observation is that the base-stock level rule obtained from a straightforward adaptation of the deterministic ELSP version can perform rather poorly. This is apparent when comparing Z_s with Z_{SD} . The ratio Z_{SD}/Z_s can be as large as 3.98. Like the ratio Z_s/Z_D discussed above, the latter ratio tends to increase significantly with p . It is rather insensitive to variations in h , but tends to increase significantly as K decreases. (In this case, monotonicity with respect to p and K sometimes fails to apply.)

Similarly, significant differences can be observed between the base-stock levels employed by an optimal stochastic base-stock rule and those obtained by adapting

the optimal rotation cycle in the deterministic version. The ratio $\partial b/b_s$ can be as large as 1 and as small as -1 .

Finally, the above lower bound $Z(s, S)$ can be used to assess the increase in operating cost, which is due to the fact that the different items compete for the availability of the same facility rather than having access to a dedicated facility. This comparison is therefore useful in capacity studies. Similar to the ratio Z_s/Z_D , the ratio $Z_s/Z(s, S)$ tends to increase with p , decreases with K and is rather invariant to changes in h .

The above observations regarding the comparison of Z_s with Z_D and Z_s with Z_{SD} hold across the board under light, moderate, and heavy (total) utilization

Figure 2 Deterministic versus Stochastic Base-stock Policies

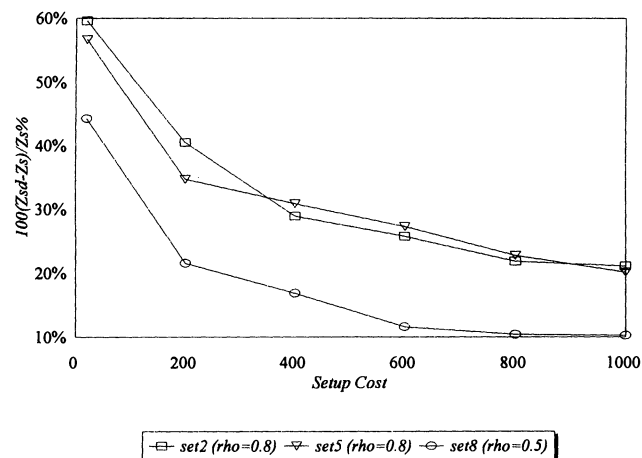
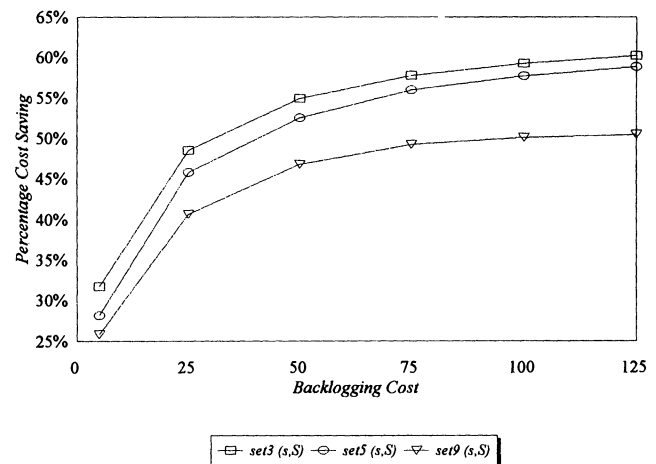


Figure 4 Savings from Having a Dedicated Facility for Each Product



rates. The ratios of these cost measures fail to be monotone in ρ .

We have observed that exhaustive service policies outperform gated service policies in *all* 360 instances; however, the gap between the two policies tends to become smaller as higher levels of service (p/h ratios) are required. This leads us to conjecture that exhaustive service is preferred as long as the cost structure is identical for all items and as long as the required service level is not extremely high. Additional support for this conjecture is provided by Theorem 3 in Levy et al. (1990), showing that the expected total amount of work in the system is almost surely smaller under exhaustive service than under gated service. However, under non-identical cost structures, gated service may outperform exhaustive service more readily. To verify this, we have repeated the 36 instances in set 2, merely scaling the holding and backlogging cost rates for the fast moving items (1, 4) down by a factor of 100, while leaving those of the remaining three items unaltered. Gated service outperforms exhaustive service, sometimes by as much as 4.2%, whenever $p_i/h_i \geq 5$, i.e., whenever a significantly high service level is required. An example in Katalan (1995) also exhibits that a *mixed* policy may both be better than or worse than *each* of the pure alternatives.

Figure 1 displays for sets 1, 4, and 7 the percentage cost increase incurred when using the deterministic policy in the stochastic environment, i.e., $100(Z_{SD} - Z_S)/Z_S$, as a function of the backlogging cost, taking averages over the six considered holding cost parameters. Similarly, Figure 2 displays for sets 2, 5, and 8 the same percentage cost increase as a function of the setup costs. Figures 3 and 4 exhibit for sets 3, 6, and 9 the relative discrepancies between Z_S and the two considered alternatives Z_D and $Z(s, S)$, respectively, i.e., $100(Z_S - Z_D)/Z_D$ and $100(Z_S - Z(s, S))/Z_S$ as a function of the backlogging costs, taking averages over the six considered setup costs.

As observed in §3, the specific permutation in which the items are produced in each cycle has some impact on the cost performance, contrary to the deterministic case where the cost of any rotation cycle is invariant to the chosen permutation; see (9). However, the cost differences are very small; changing the permutation may reduce the cost value by a few percentage points only,

even for settings with a high utilization rate and large imbalance between the items' workloads. Thus for all practical purposes, an arbitrary permutation can be selected. To illustrate these conclusions, we have chosen the ninth instance of problem set 1 (with $\rho = 0.8$ and large imbalance; $h = 5$, $p = 50$, $K = 500$). $C_{\max}/C_{\min} = 1.0039$, $\Delta_{S,\max}/\Delta_{S,\min} = 1.0124$, $Z_{S,\max}/Z_{S,\min} = 1.0023$ and $Z_{SD,\max}/Z_{SD,\min} = 1.0055$ where the subscript max (min) refers to the maximum (minimum) value across all $4! = 24$ permutations.

Appendix

In this appendix we give an algorithmic description of the procedure described in §3 to determine the distribution of the L -variables and associated optimal base-stock levels $\{b_i^*; i = 1, \dots, N\}$ for a given value of Δ . We confine ourselves to the case where exhaustive service is provided and describe the procedure to calculate the distribution of L_1 and b_1^* ; those for the other items require straightforward adaptation. Let

$$\pi_1(j) = \Pr[L_1 = j], \quad \pi_1'(j) = \Pr[L_1' = j],$$

$$\pi_1''(j) = \Pr[L_1'' = j], \quad j = 0, 1.$$

Algorithm

Step 1. (Computing variance of X_1 , using Konheim et al. (1994))

Initialize: $\text{sum_beta_2} := 0$; $\text{alpha}[i] := \text{alpha_2}[i] := 0$ ($2 \leq i \leq N$);

$\text{alpha}[1] := \text{alpha_2}[1] := \text{sum_alpha} := \text{sum_alpha_2} := \lambda_1$;

$\theta[i] := s_i/(1 - \rho_i)$; $\theta_2[i] := s_i^{(2)}/(1 - \rho_i)^3$ ($1 \leq i \leq N$);

for $c := 0$ to ∞ do begin

for $i := N$ to 1 by -1 do begin

$\text{sum_beta_2} := \text{sum_beta_2} + (r_i^{(2)} - r_i^2)(\text{sum_alpha})^2 + r_i$

sum_alpha_2 ;

$\text{sum_alpha} := \text{sum_alpha} - \text{alpha}[i]$;

$\text{sum_alpha_2} := \text{sum_alpha_2} - \text{alpha_2}[i]$;

$\text{alpha}[i] := \lambda_i \theta[i] \text{sum_alpha}$;

$\text{alpha_2}[i] := \lambda_i (\theta_2[i] (\text{sum_alpha})^2 + \theta[i] \text{sum_alpha_2})$;

$\text{sum_alpha} := \text{sum_alpha} + \text{alpha}[i]$;

$\text{sum_alpha_2} := \text{sum_alpha_2} + \text{alpha_2}[i]$;

end

end

$\text{Var}(X_1) = \text{sum_beta_2}$;

Step 2. (Computing first two moments of I_1)

$E(I_1) := (1 - \rho_1) \sum_i r_i / (1 - \rho)$; $E(I_1^2) := \lambda_1^{-2} (\text{Var}(X_1) + (EX_1)^2 - EX_1)$;

Step 3. (Fitting mixture of Erlangs to I_1 , using Tijms 1986)

$cv^2 := E(I_1^2) / (E(I_1))^2 - 1$;

if $cv^2 > 1$ then (fit via hyperexponential) begin

$p := (1 + \sqrt{((cv^2 - 1)/(cv^2 + 1))})/2$;

$\mu_1 := 2p/EI_1$; $\mu_2 := 2(1 - p)/EI_1$;

end

else (fit via mixture of two Erlangs with common shape parameter μ and phases $k, k-1$)
begin
 $k := 2$; while $cv^2 < k^{-1}$ do $k := k + 1$;
 $p := (kcv^2 - \text{sqrt}(k(1 + cv^2) - k^2cv^2))/(1 + cv^2)$; $\mu := (k - p)/EL_1$;
end

Step 4. (Computing b_1^*)

Initialize: $f := p_1/(p_1 + h_1)$; $j := 0$;
 $\pi_1'(0) = 1 - \rho_1$;
 if $cv^2 > 1$ then begin
 $q_1 := \lambda_1/(\lambda_1 + \mu_1)$; $q_2 := \lambda_1/(\lambda_1 + \mu_2)$; $pr := p/(p + (\mu_1/\mu_2)(1 - p))$;
 $\text{sum1} := (1 - q_1)$; $\text{sum2} := (1 - q_2)$;
 $\pi_1''(0) = pr \text{sum1} + (1 - pr)\text{sum2}$;;
 end
 else begin
 $q := \lambda_1/(\lambda_1 + \mu)$;
 $\text{sum1} := \text{sum2} := \text{sum3} := 1 - q$;
 for $c := 2$ to $k - 1$ do begin
 $\text{sum2} := \text{sum2}(1 - q)$;
 $\text{sum3} := \text{sum3} + \text{sum2}$;
 end
 $\text{sum2} := \text{sum2}(1 - q)$;
 $\pi_1'(0) := \text{sum2}(1 - p)/k + \text{sum3}(p + k - 1)/(k(k - 1))$;
 end
 $\Pi_1 := \pi_1(0) := \pi_1'(0)\pi_1''(0)$;
 while $\Pi_1 < f$ do begin
 $A[j] := \int_0^\infty (1 - S(t))e^{-\lambda_1 t}(\lambda_1 t)^j/j! dt$; $j := j + 1$;
 (computable in closed form for deterministic, mixture of Erlangs, Uniform etc.)
 $\pi_1'(j) := (\lambda_1 \pi_1'(0)A[j - 1] + \sum_{l=1}^{j-1} \lambda_1 \pi_1'(l)A[j - l])/(1 - \lambda_1 A[0])$;
 if $cv^2 > 1$ then begin
 $\text{sum1} := \text{sum1} q_1$; $\text{sum2} := \text{sum2} q_2$;
 $\pi_1''(j) := pr \text{sum1} + (1 - pr)\text{sum2}$;
 end
 else begin
 $\text{sum1} := \text{sum1} q$;
 $\text{sum2} := \text{sum3} := \text{sum1}$;
 for $c := 2$ to $k - 1$ do begin
 $\text{sum2} := \text{sum2}(1 - q)(j + c - 1)/(c - 1)$;
 $\text{sum3} := \text{sum3} + \text{sum2}$;
 end
 $\text{sum2} := \text{sum2}(1 - q)(j + k - 1)/(k - 1)$;
 $\pi_1'(0) := \text{sum2}(1 - p)/k + \text{sum3}(p + k - 1)/(k(k - 1))$;
 end
 $\pi_1(j) := \sum_{l=0}^j \pi_1'(l)\pi_1''(j - l)$;
 $\Pi_1 := \Pi_1 + \pi_1(j)$;
 end
 $b_1^* := j$;

References

- Atkins, D. and P. Iyogun, "Periodic versus 'Can-Order' Policies for Coordinated Multi-item Inventory Systems," *Management Sci.*, 34 (1988), 791-796.
- Bertsimas, D. and H. Xu, "Optimization of Polling Systems and Dynamic Vehicle Routing Problems on Networks," Working Paper, Sloan School of Management and Operations Research Center, MIT, Cambridge, MA, 1993.
- Bomberger, E. E., "A Dynamic Programming Approach to a Lot Size Scheduling Problem," *Management Sci.*, 12, 11 (1966), 778-784.
- Borst, S., *Polling Systems*, Ph.D. Dissertation, University of Tilburg, The Netherlands, 1994.
- Bourland, K. E. and C. A. Yano, "Modeling Random Demand when the Exposure Period is a Decision Variable," Working Paper No. 270, The Amos Tuck School of Business Administration, Dartmouth College, Hanover, NH, 1991.
- and —, "Use of Capacity Slack in the Economic Lot Scheduling Problem with Random Demand," *Management Sci.*, 40 (1994), 1690-1704.
- Browne, S. and U. Yechiali, "Dynamic Routing in Polling Systems," in M. Bonatti (Ed.), *Teletraffic Science for New Cost-Effective Systems, Networks and Services*, ITC-12, Elsevier Science Publishers, North Holland, Amsterdam, 1989a.
- and —, "Dynamic Priority Rules for Cyclic-type Queues," *Advances in Applied Probability*, 21, 2 (1989b), 432-450.
- Dobson, G., "The Economic Lot Scheduling Problem: Achieving Feasibility Using Time-varying Lot Sizes," *Oper. Res.*, 35, 5 (1987), 764-771.
- Duenyas, I. and M. P. Van Oyen, "Heuristic Scheduling of Parallel Heterogeneous Queues with Setups," Technical Report 92-60, Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI, 1992.
- and —, "Stochastic Scheduling of Parallel Queues with Setup Costs," Technical Report 93-09, Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI, 1993.
- Elmaghraby, S., "The Economic Lot Scheduling Problem (ELSP): Review and Extensions," *Management Sci.*, 24, 6 (1978), 587-598.
- Federgruen, A. and Z. Katalan, "The Impact of Setup Times on the Performance of Multi-class Service and Production Systems," to appear in *Oper. Res.*
- and —, "Approximating Queue Size and Waiting Time Distributions in General Polling Systems," *Queueing Systems*, 18 (1994), 353-386.
- and Y. S. Zheng, "Optimal Control Policies for Stochastic Inventory Systems with Endogenous Supply," *Probability in the Engineering and Informational Sciences*, 7 (1993), 257-272.
- Fuhrmann, S. W. and R. B. Cooper, "Stochastic Decompositions in the M/G/1 Queue with Generalized Vacations," *Oper. Res.*, 33, 5 (1985), 1117-1129.
- Gallego, G., "Scheduling the Production of Several Items with Random Demands in a Single Facility," *Management Sci.*, 36, 12 (1990), 1579-1592.
- Graves, S. C., "The Multi-product Production Cycling Problem," *AIEE Transactions*, 12 (1980), 233-240.
- Ha, A., "Optimal Scheduling and Stock Rationing Policies in Make-to-stock Production Systems," Ph.D. Dissertation, Stanford University, Stanford, CA, 1992.
- Hall, R. W., *Kawasaki, USA*, American Production and Inventory Control Society, Falls Church, VA, 1982.

- Hsu, W., "On the General Feasibility Test of Scheduling Lot Sizes for Several Products on One Machine," *Management Sci.*, 29, 1 (1983), 93–105.
- Jones, P. and R. Inman, "Where Is the ELSP Easy?," *IIE Transactions*, 21 (1989), 11–20.
- Karmarkar, U. S., "Lot Sizes, Lead Times and In-process Inventories," *Management Sci.*, 33 (1987), 409–418.
- Katalan, Z., "Production and Service Management Under Setup Times and Uncertainties," Ph.D. Dissertation, Columbia University, New York, 1995.
- Konheim, A. G., H. Levy, and M. M. Srinivasan, "Descendant Set: An Efficient Approach for the Analysis of Polling Systems," *IEEE Transactions on Communications*, 42, 2/3/4 (1994), 1245–1253.
- Maxwell, W., "The Scheduling of Economic Lot Sizes," *Naval Res. Logistics Quarterly*, 11 (1964), 89–124.
- Leachman, R. C. and A. Gascon, "A Heuristic Scheduling Policy for Multi-item, Single-machine Production Systems with Time-varying Stochastic Demands," *Management Sci.*, 34, 3 (1988), 377–390.
- , Z. K. Xiong, A. Gascon, and K. Park, "Note: An Improvement to the Dynamic Cycle Lengths Heuristic for Scheduling the Multi-item, Single Machine," *Management Sci.*, 37, 9 (1991), 1201–1205.
- Levy, H., M. Sidi, and O. J. Boxma, "Dominance Relations in Polling Systems," *Queueing Systems*, 6 (1990), 155–172.
- Peña, A. and P. Zipkin, "Dynamic Scheduling Rules for a Multiproduct Make-to-stock Queue," Working Paper, Graduate School of Business, Columbia University, New York, 1993.
- Rinnooy Kan, A. H. G. and G. T. Timmer, "Global Optimization," in G. L. Nemhauser et al. (Eds.), *Handbooks in OR & MS*, Volume 1, Elsevier Science Publishers, B.V., North-Holland, 1989.
- Salomon, M., "Deterministic Lotsizing Models for Production Planning," Ph.D. Dissertation, Erasmus University, The Netherlands, 1990.
- Sarkar, D. and W. I. Zangwill, "Expected Waiting Time for Nonsymmetric Cyclic Queueing Systems—Exact Results and Applications," *Management Sci.*, 35, 12 (1989), 1463–1474.
- and —, "Variance Effects in Cyclic Production Systems," *Management Sci.*, 37 (1991), 444–453.
- Schonberger, R., *World Class Manufacturing Case Book*, The Free Press, New York, 1987.
- Shaked, M. and J. G. Shanthikumar, "Parametric Stochastic Convexity and Concavity of Stochastic Processes," *Annals of the Institute of Statistical Mathematics*, 42 (1990), 509–531.
- Silver, E. and R. Peterson, *Decision Systems for Inventory Management and Production Planning*, second edition, John Wiley & Sons, New York, 1985.
- Smith, W., J. Ingraham, and D. Rurak, "Time and Pull: a DuPont Case History," *Target*, 9, 1 (1993).
- Takagi, H., *Analysis of Polling Systems*, MIT Press, Cambridge, MA, 1986.
- , "Queueing Analysis of Polling Models: An Update," in H. Takagi (Ed.), *Stochastic Analysis of Computer and Communication Systems*, North-Holland, Amsterdam, 1990.
- Tijms, H. C., *Stochastic Modeling and Analysis: A Computational Approach*, John Wiley & Sons, UK, 1986.
- Veatch, M. and L. Wein, "Scheduling a Make-to-stock Queue: Index Policies and Hedging Points," Working Paper, MIT, Cambridge, MA, 1992.
- Wein, L., "Dynamic Scheduling of a Multi-class Make-to-stock Queue," *Oper. Res.*, 40 (1992), 724–735.
- Zipkin, P., "Models for Design and Control of Stochastic, Multi-item Batch Production Systems," *Oper. Res.*, 34, 1 (1986), 91–104.
- , "Computing Optimal Lot Sizes in the Economic Lot Scheduling Problem," *Oper. Res.*, 39, 1 (1991), 56–63.

Accepted by Luk Van Wassenhove; received February 4, 1994. This paper has been with the authors 4 months for 2 revisions.