DETERMINING PRODUCTION SCHEDULES UNDER BASE-STOCK POLICIES IN SINGLE FACILITY MULTI-ITEM PRODUCTION SYSTEMS

AWI FEDERGRUEN
Columbia University, New York, New York

ZIV KATALAN
University of Pennsylvania, Philadelphia, Pennsylvania

(Received January 1996; revisions received September 1996; accepted September 1996)

In this paper we address periodic base-stock policies for stochastic economic lot scheduling problems. These represent manufacturing settings in which multiple items compete for the availability of a common capacity source, in the presence of setup times and/or costs, incurred when switching between items, and in the presence of uncertainty regarding demand patterns, production, and setup times. Under periodic base-stock policies, items are produced according to a given periodic item-sequence. This paper derives effective heuristics for the design of a periodic item-sequence minimizing system-wide costs. This sequence is constructed based on desirable production frequencies for the items, obtained as the solution of lower bound mathematical programs. An extensive numerical study gauges the quality of the proposed heuristics.

In this paper we consider manufacturing settings in which multiple, say N, items compete for the availability of a common capacity source, i.e., production facility, in the presence of *setup times* incurred when switching between items. In addition to a setup *time*, switching may involve an explicit out of pocket setup *cost* as well. We focus in particular on settings where uncertainty prevails regarding the demand patterns, production times as well as the setup times. Such settings are often referred to as Stochastic Economic Lot Scheduling Problems (SELSP). A variety of strategy classes have been proposed to govern these systems effectively; these can be partitioned into (i) *static* and (ii) *dynamic* strategies. Dynamic strategies determine at any point in time which of the items, if any, is to be produced in the facility on the basis of the *complete* state of the system (which includes the inventory levels of all items and the most recent assignment of the facility). Static policies, on the other hand, use only state information that pertains to the item currently being produced.

One important class of static strategies are the so-called periodic base-stock policies: 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 periodic sequence, possibly with idle times inserted between the completion of an item's production batch and the setup for the next item. Other examples of static policies include variations in which the sequence of items produced by the facility is determined by a state-independent random process (see, e.g., Kleinrock and Levy 1988) as well as *time-window* policies in which the assignment of the facility is completely determined by the time of day, the day of the week, etc.

The exact optimal policy under most common performance measures or cost criteria is often of the dynamic type. Such criteria include the minimum average system-wide setup-, holding-, and backlogging costs or the minimum of average setup and holding costs under specific service level (e.g., fill rate) constraints. Unfortunately, identification of an optimal dynamic policy requires the solution of a dynamic program with an N-dimensional state space; this is feasible only when $N = 2$ or 3. Moreover, the structure of an optimal policy tends to be highly complex, prohibiting their implementation even if such strategies could be computed in a reasonable amount of time. See, e.g., Ha (1992) and Qui and Loulou (1995) for a characterization of an optimal policy in the special case of two products, with zero and positive setup times and costs, respectively. Thirdly, coordination with other related activities (e.g., procurement of raw materials, input components, externally performed setups) is often difficult under dynamic policies as opposed to some of the above classes of static policies. Indeed, under many of the latter (e.g., the periodic base-stock policies) predictability of the plant is greatly enhanced. As pointed out in Tayur (1994), increased predictability promotes more accurate due-date quotations and thus reduces the need for expediting and schedule changes and associated ad-hoc setups, effectively improving the production capacity.

Finally, there is now evidence that the restriction to static policies comes with only a *moderate* loss of optimality: Markowitz et al. (1995) recently proposed a class of *dynamic* cyclic policies as a dynamic alternative to the cyclic base stock policies. Under a dynamic cyclic policy, the facility continues to cycle through the products in a fixed sequence or permutation; however, when engaged in the

Subject classifications: Inventory/production: multi-item/echeon/stage. Inventory/production: stochastic. Production/scheduling: stochastic.
Area of review: Manufacturing Operations.

0030-364X/98/4606-0883 $05.00

production of a given item, the facility may switch to the next item or go idle as a function of the system wide vector of inventories, not just when reaching a given target level. The authors propose a specific rule within this class, based on heavy traffic approximations. They compare the performance of their dynamic cyclic rule with a minor static variant of the cyclic base-stock policies. Where possible, i.e., in the case of two product problems (only), they also provide a comparison with the optimal (dynamic) policy. See Section 4 for a summary of the comparison results. Moreover, Markowitz et al. (1995) only consider *cyclic* strategies. We shall demonstrate that significant improvements can be achieved by replacing *cyclic* by *general periodic* strategies, producing different items with different relative frequencies. The magnitude of these improvements is on average similar, but sometimes significantly larger than those achieved in going from static to dynamic cyclic rules.

To overcome part of the difficulties associated with fully dynamic policies several classes of so-called *semi-dynamic* policies have been proposed; see, e.g., Browne and Yechiali (1989a, b), Duenyas and Van Oyen (1995, 1996) for the special case of systems without inventories, and Zipkin (1986), Leachman and Gascon (1988), Leachman et al. (1991), and Bourland and Yano (1991, 1994) for the more general case of make-to-stock systems. (See Federgruen and Katalan 1995, 1996a for a more detailed review of these policy classes.) However, no accurate analytical methods are known to evaluate policies or to identify the optimal policy within any of these classes, except for standard dynamic programming methods which, as mentioned above, are practical only when $N = 2$ or $N = 3$. Indeed, to our knowledge, the only class of policies permitting such evaluation and optimization methods for general values of N is the above described class of (static) base-stock policies. Federgruen and Katalan (1995, 1996a) describe an efficient method to accurately evaluate any given base-stock policy and to identify optimal base-stock levels for a *given* periodic sequence. The latter confines itself to cyclic sequences in which the facility rotates among the items according to a specific permutation; Federgruen and Katalan (1995) deals with general periodic sequences and certain semi-dynamic extensions in which part of the items are made-to-order, with dynamic priority rules governing their insertion into the schedule for the make-to-stock items. Anupindi and Tayur (1998) recently developed a simulation based method to compute optimal base-stock levels for a *given*, *cyclic* sequence, under somewhat more general demand processes than those considered in Federgruen and Katalan (1995, 1996a).

In this paper we complement Federgruen and Katalan (1995, 1996a) by showing how an effective periodic item-sequence can be selected among all such sequences and this periodic schedule outperforms cyclical base-stock policies quite significantly. Identifying an effective periodic item sequence represents an exceedingly difficult combinatorial problem as there are $O(N!\binom{M}{N}N^{M-N})$ potential sequences of a *given* periodicity M ($\geq N$) alone. The same question has been addressed by Dobson (1987) and Zipkin (1991) in the context of deterministic ELSPs in which demands for all items occur at constant rates and the production and setup times are constants as well. Boxma et al. (1990, 1991, 1993) and Bertsimas and Xu (1993) develop heuristics for the determination of an optimal polling table in periodic polling systems with exhaustive and gated service. A polling system consists of several (say N) stations and a server who at any point in time can be assigned to at most one of them incurring setup times when switching between stations. Under a periodic polling system, the server travels between the stations according to a given periodic sequence described by the so-called polling table; under exhaustive (gated) service the server continues at a given station until its queue is emptied out (until all customers present upon his arrival to the station are served). Polling systems thus correspond with the special case of our model in which no positive inventories are allowed. Boxma et al. (1990, 1991) and Bertsimas and Xu (1993) focus on the objective of minimizing a weighted average of the steady-state waiting times at the different stations, while Boxma et al. (1993) are concerned with the weighted average of the steady-state workloads.

The remainder of this paper is organized as follows. In Section 1 we first introduce the general model and its notation as well as a brief review of the method to evaluate a specific periodic base-stock policy. We then give a brief outline of a three-phase approach to generate effective production tables. In Section 2 we address the question of determining optimal production *frequencies* of the different items (Phase I). (The production frequency of an item is specified as the frequency with which a new production batch for this item is initiated.) In Section 3 we generate, with these frequencies determined, a periodic item-sequence (or production table) (Phase III). Section 4 concludes the paper with an extensive numerical study, gauging the quality of the proposed strategies.

1. MODEL DESCRIPTION; A THREE-PHASE APPROACH TO GENERATE EFFECTIVE PRODUCTION TABLES

We consider a production system with N distinct items, demands for which are generated by independent Poisson processes, with λ_i the rate at which demand arises for item i ($i = 1, \ldots N$). Let $\lambda = \sum_{i=1}^{N} \lambda_i$. (The results in this paper easily carry over to the more general case where all demand processes are compound Poisson.) The N items are produced in a common facility that can produce only one unit of one of these items at a time. Production times for individual units are assumed to be independent; those of item i are identically distributed with cdf $S_i(\cdot)$ and mean $s_i < \infty$ ($i = 1, \ldots N$). A possibly random setup time with cdf $R_i(\cdot)$ and mean $r_i < \infty$ ($i = 1, \ldots, N$), is incurred when setting up to produce item i, i.e., whenever the facility starts producing item i after being idle or after producing some other item. Consecutive setup times are independent. The utilization rate for item i is $\rho_i = \lambda_i s_i$; that of the system equals $\rho = \sum_{i=1}^{N} \rho_i$ and we assume the system is stable, i.e., $\rho < 1$.

If a demand cannot be filled from existing inventory, it is backlogged. Three types of costs are incurred. Let:

h_i = the inventory carrying cost rate for item i per unit of time $(i = 1, \ldots, N)$;
p_i = the backlogging cost rate for item i per unit of time $(i = 1, \ldots, N)$;
K_i = the setup cost incurred per setup of item i $(i = 1, \ldots, N)$.

The objective is to minimize the long run total of average holding-, backlogging-, and setup costs.

A periodic base-stock policy is specified by:

(a) a vector of base-stock levels $b = (b_1, \ldots, b_N)$;
(b) a table $\underline{T} = \{T(j); j = 1, \ldots, M\}$ of length $M \geq N$; for all $j = 1, \ldots, M$ $T(j)$ denotes one of the items in $\{1, \ldots, N\}$; and
(c) a vector of inserted idle times $\underline{\Delta} = (\Delta_1, \ldots, \Delta_M)$.

At time 0, the facility starts to produce the first item listed in table $\underline{T}$, i.e., item $T(1)$ and continues its production until its inventory level is increased to a base-stock level $b_{T(1)}$ (the *exhaustive* case). Alternatively a batch is produced, the size of which is given by the difference between $b_{T(1)}$ and the starting inventory level (the *gated* case). The facility then switches to the second item in the table, $T(2)$, after a possible idle time Δ_2 and setup time $R_{T(2)}$. This protocol continues until the Mth production run for item $T(M)$. Thereafter the facility returns to the beginning of the table, producing its first item $T(1)$ after a possible idle time Δ_1 and setup time $R_{T(1)}$, and continuing the above protocol. (To facilitate the exposition, we confine ourselves to the case where the *exhaustive* production discipline is applied to every entry of the table. All our results are straightforwardly extended to the case where the gated discipline is applied to some or all of the table's entries.)

For a given table $\underline{T}$, vector of base-stock levels $\underline{b}$ and inserted idle times $\underline{\Delta}$, we define for all $i = 1, \ldots N$ the (steady-state) shortfall distribution L_i by:

$$L_i = b_i - IL_i, \tag{1}$$

where IL_i denotes the steady-state inventory level of item i.

The shortfall variables $\{L_i\}$ have the advantage of being independent of the base-stock levels. A method to compute the shortfall distribution is developed in Federgruen and Katalan (1994); it is based on a well-known decomposition result due to Fuhrmann and Cooper (1985). More specifically, fix $i = 1, \ldots, N$; a nonproduction epoch for item i refers to any epoch at which the system is not engaged in producing a unit for this item. Then, as shown in Federgruen and Katalan (1994),

$$L_i = L'_i + L''_i \tag{2}$$

with L'_i and L''_i independent of each other;
where

L'_i = the steady-state shortfall in a system exclusively dedicated to item i, i.e., the steady-state number of customers in an M/G/1 system with Poisson arrivals at rate λ_i and i.i.d. service times distributed as $S_i(\cdot)$,
L''_i = the shortfall distribution for item i at an arbitrary tagged nonproduction epoch for this item.

To further characterize L''_i, assume item i appears M_i times in the table $\underline{T}$ and let

$I_{i,j}[C_{i,j}]$ = the jth intervisit [cycle] time for item i $(j = 1, \ldots, M_i)$, i.e., the steady-state time interval between the start of the jth production run of item i and the termination [start] of the preceding production run of the same item.
$I_i[C_i]$ = the unconditional intervisit [cycle] time for item i, i.e., the mixture of the steady-state distributions $\{I_{i,j}: j = 1, \ldots, M_i\}[\{C_{i,j}: j = 1, \ldots, M_i\}]$ with mixing probabilities

$$\left\{\frac{EI_{i,j}}{\sum_{\ell=1}^{M_i} EI_{i,\ell}} : j = 1, \ldots, M_i\right\}$$

$$\left[\left\{\frac{EC_{i,j}}{\sum_{\ell=1}^{M_i} EC_{i,\ell}} : j = 1, \ldots, M_i\right\}\right].$$

Also, for any random variable U, let U^e denote an interval of time distributed as the equilibrium excess distribution of U, and let

$N_i(U)$ = the total demand for item i in an interval of time distributed like $U (i = 1, \ldots, N)$.

As shown in Federgruen and Katalan (1994, Proposition 2) L''_i is distributed as a mixture

$$\text{of distributions } \{N_i(I^e_{i,j}): j = 1, \ldots, M_i\} \tag{3}$$

with mixing probabilities

$$\left\{\frac{EI_{i,j}}{\sum_{\ell=1}^{M_i} EI_{i,\ell}} : j = 1, \ldots, M_i\right\}.$$

The method in Federgruen and Katalan (1994) is exact except for the fact that the intervisit times are approximated by numerically convenient phase type distributions, fitting *any* desired number of moments. The moments of these intervisit times are themselves computed within any required relative precision $\epsilon > 0$, via a recursive so-called descendant set method due to Konheim et al. (1994). To compute the first m (say) moments of the intervisit times via the descendant set method, it suffices to know the first m moments of the production and setup time distributions. The overall worst case complexity is $O(\max\{Nk^{*^2}, M^2 \log_\rho \epsilon\})$ where k^* denotes the largest shortfall level that needs to be evaluated. As demonstrated in Federgruen and Katalan (1994), for tables with up to 50 entries (say), the complete method requires only a few seconds on a PC and is remarkably accurate as verified in an extensive numerical study. Most recently Choudhury and Whitt (1994) proposed an alternative evaluation method based on numerical transform inversions. This method can be designed to provide exact evaluations of the shortfall distributions up to any prespecified precision, and it is asymptotically somewhat faster for fixed k^* as M tends to infinity. Its complexity is $O(M^{1+\alpha}k^*)$ with α typically in the range 0.6 to 0.8. The shortfall, and hence inventory level, distributions (see (1)) permit us to evaluate all system-wide costs.

Let $\bar{C}$ denote the expected length, in steady-state, of a *complete* cycle, i.e., the time between two consecutive returns to the beginning of the table $\underline{T}$. It is well known (see, e.g., Takagi 1986), that

$$\bar{C} = \sum_{j=1}^{M} (r_{T(j)} + \Delta_j)/(1 - \rho). \tag{4}$$

The long-run average cost under a given base-stock policy is thus given by:

$$\frac{\sum_{j=1}^{M} K_{T(j)}}{\bar{C}} + \sum_{i=1}^{N} \{h_i E(IL_i^+) + p_i E(IL_i^-)\}, \tag{5}$$

where $x^+ = \max(x, 0)$ and $x^- = \max(-x, 0)$. Substituting (1) into (5) we obtain:

$$Z^* = \frac{\sum_{j=1}^{M} K_{T(j)}}{\bar{C}} + \sum_{i=1}^{N} \{h_i E(b_i - L_i)^+ + p_i E(L_i - b_i)^+\}. \tag{6}$$

Note that the distribution of the random variables $\{L_i: i = 1, \ldots, N\}$ is *independent* of the vector of base-stock levels $\underline{b}$, but it of course does depend on the production table $\underline{T}$ and vector of idle times $\underline{\Delta}$. Thus, for a *given* table $\underline{T}$ and vector $\underline{\Delta}$, the optimal base-stock levels, b_i^* for item i ($i = 1, \ldots, N$), can be determined by solving a newsvendor problem with L_i as the demand distribution, i.e.,

$$b_i^* = \min\{k: \Pr[L_i \leqslant k] \geqslant p_i/(p_i + h_i)\}. \tag{7}$$

(See, e.g., Proposition 1 in Federgruen and Katalan 1996a.)

Our objective is therefore to characterize the relationships between the production table $\underline{T}$ and vector of idle times $\underline{\Delta}$ on the one hand, and the cost expression (6) on the other hand, so as to develop heuristics for an optimal choice of $\underline{T}$ and $\underline{\Delta}$.

To describe possible construction procedures for the production table we need the following additional notation. Let

$\underline{M_i}$ = the *absolute* frequency with which item i appears in the production table $\underline{T} = |\{j = 1, \ldots, M: T(j) = i\}|$ (note $\sum_{i=1}^{N} M_i = M$);

$m_i = M_i/M$ = the *relative* frequency of item i in the production table $\underline{T}$ ($\sum_{i=1}^{N} m_i = 1$);

F_i = the long-run average *absolute* frequency with which a setup for item i is initiated ($=M_i/\bar{C}$ by the theory of regenerative processes);

δ = the long-run average frequency with which the system is idle ($=\sum_{i=1}^{M} \Delta_i/\bar{C}$).

Note that

$$m_i = \frac{M_i}{\sum_{\ell=1}^{N} M_\ell} = \frac{M_i/\bar{C}}{\sum_{\ell=1}^{N} (M_\ell/\bar{C})} = \frac{F_i}{\sum_{\ell=1}^{N} F_\ell}, \quad i = 1, \ldots, N. \tag{8}$$

We propose the following three-phase approach in constructing an effective production table $\underline{T}$:

Phase (I) (*Relative Frequencies*): Generate frequency values $\{F_i: i = 1, \ldots, N\}$ and hence, by (8) relative frequencies $\{m_i: i = 1, \ldots, N\}$.

Phase (II) (*Table Size*): Select M, the length of the production table and *absolute* frequencies $\{M_i: i = 1, \ldots, N\}$ with relative frequencies $\{M_i/M: i = 1, \ldots, N\}$ closely matching the values $\{m_i: i = 1, \ldots, N\}$ in Phase (I).

Phase (III) (*Sequencing*): Select a table $\underline{T}$ in which the items appear with the frequencies $\{M_i: i = 1, \ldots, N\}$ as determined in Phase (II).

The same three-phase approach was initially proposed by Boxma et al. (1990, 1991, 1993) for the construction of a polling table in polling systems under which a *linear* combination of the *expected* waiting times (or workloads) at the different stations is minimized. Bertsimas and Xu (1993) deal with the same model and objective, however, under *sequence-dependent* setup times, i.e., the setup time for a new production run depends both on the new item being setup and the item last produced. Under sequence dependent setup times it is important to disaggregate the F_i- and m_i-quantities, i.e., to characterize the quantities $F_{ij}(m_{ij})$ which denote the frequency of switching from item i to item j per unit of time (in the table $\underline{T}$); $i, j = 1, \ldots, N$. (Clearly, $F_i = \sum_{j=1}^{N} F_{ij}$ and $m_i = \sum_{j=1}^{N} m_{ij}$ for all $i = 1, \ldots, N$.) Like Boxma et al. (1990, 1991, 1993), Bertsimas and Xu (1993) propose the above three-phase approach as well, merely replacing $\{F_i\}$ by $\{F_{ij}\}$ and $\{m_i\}$ by $\{m_{ij}\}$.

In Phase (I) we generate production frequencies from the solution of mathematical programming models of special structure which approximate the exact system-wide cost objective (6).

As far as Phase (II) is concerned, we propose as in Boxma et al. (1990, 1991, 1993) that M be incremented sequentially, starting with $M = N$ until the maximum rounding error incurred when rounding the numbers $\{Mm_i: i = 1, \ldots, N\}$ to their nearest integer values falls below a prespecified precision η, i.e., $M = \inf\{x \geqslant N: \min(\lceil xm_i \rceil - xm_i, xm_i - \lfloor xm_i \rfloor) < \eta$ for all $i = 1, \ldots, N\}$. As shown by Boxma et al. (1990, 1991, 1993) this phase appears to be the least crucial one in the heuristic.

As far as the sequencing Phase (III) is concerned, we propose that an *initial* sequence be constructed on the basis of the *Golden Ratio* heuristic proposed in Boxma et al. or a *machine scheduling* heuristic proposed by Dobson (1987) for *deterministic* Economic Lot Scheduling problems. The former attempts to space all items i with $M_i \geqslant 2$ equidistantly in the table $\underline{T}$, i.e., it attempts to equalize the *number of entries* between consecutive appearances of item i in the table. Dobson's machine scheduling heuristic, on the other hand, attempts to equalize for all items the *times* between their consecutive production runs. Construction of an initial sequence is potentially followed by iterative sequence improvement routines, described in detail below.

2. PHASE (I): DETERMINING OPTIMAL PRODUCTION FREQUENCIES

In this section we focus on Phase (I). It is clear that the *exact* system-wide objective (6) can not be determined on the basis of the values $\{F_i\}$ and δ alone. As in other problems of determining optimal priority schemes (see, e.g., Bertsimas and Niño-Mora 1993, Bertsimas et al. 1994, Bhattacharya et al. 1991, Coffman and Mitrani 1980, Federgruen and Groenevelt 1988a and b, Katehakis and Veinott 1987, and Shanthikumar and Yao 1992) as well as Boxma et al. (1990, 1991, 1993) and Bertsimas and Xu (1993), we develop a lower bound approximation for the exact objective which *is* determined by the vector $\underline{F}$ and δ, and for which a minimizing pair $(\underline{F}, \delta)$ *can* be obtained. For objectives which are *linear* and *increasing* in the vector of *mean* waiting times or *mean* queue sizes or shortfalls $\{EL_i: i = 1, \ldots, N\}$ (as is the case in Boxma et al. and Bertsimas and Xu), it suffices to obtain a lower bound for each of the mean shortfalls $\{EL_i: i = 1, \ldots, N\}$:

Lemma 1. *Fix* $i = 1, \ldots, N$.

$$EL_i \geqslant \lambda_i \left\{ \frac{\lambda_i E(S_i^2)}{2(1-\rho_i)} + s_i + \frac{1}{2}\frac{(1-\rho_i)}{F_i} \right\}.$$

Proof. It follows from (2) that $E(L_i) = E(L_i') + E(L_i'')$.

By the Pollaczek-Khintchine formula,

$$E(L_i') = \frac{\lambda_i^2 E(S_i^2)}{2(1-\rho_i)} + \rho_i, \tag{9}$$

and by (3):

$$E(L_i'') = E_J E(N_i(I_{iJ}^e) | I_{iJ}^e),$$

where J is a discrete random variable, with

$$\Pr[J = j] = \frac{E(I_{i,j})}{\sum_{\ell=1}^{M_i} E(I_{i,\ell})}.$$

Thus,

$$E(L_i'') = E_J[\lambda_i E(I_{iJ}^e)] = E_J\left[\lambda_i \frac{E(I_{iJ}^2)}{2E(I_{iJ})}\right]$$
$$\geqslant \frac{1}{2}\lambda_i E_J[E(I_{iJ})] = \frac{1}{2}\lambda_i E(I_i), \tag{10}$$

where the inequality follows from $E(I_{i,j}^2) \geqslant E^2(I_{i,j})$. Observe also that $F_i = 1/E(C_i)$, with $E(C_i) = (1-\rho_i)E(C_i) + \rho_i E(C_i) = E(I_i) + \rho_i E(C_i)$. ($\lambda_i C_i$ is the expected number of units of item i demanded in a cycle of length C_i and $s_i \lambda_i C_i = \rho_i C_i$ the expected amount of production time required to satisfy this demand.) It follows that

$$E(I_i) = (1-\rho_i)/F_i. \tag{11}$$

Lemma 1 now follows by substituting (11) into (10) and adding (9) to the resulting inequality. □

Remark 1. The proof of Lemma 1 is similar to that used in the derivation of Bertsimas and Xu's lower bound.

The objective (6) in our model depends on the *entire* distribution of the variables $\{L_i: i = 1, \ldots, N\}$, not just on their *means*. Our approach in developing a lower bound approximation is therefore to (i) develop lower bounds for a given number (say $m \geqslant 2$) of the moments of $\{L_i: i = 1, \ldots, N\}$, and (ii) to develop an approximation of (6) in terms of these m moments only. Here we confine ourselves to the case $m = 2$. We therefore start with a lower bound for the *variance* of the shortfall distributions $\{\mathrm{Var}(L_i): i = 1, \ldots, N\}$.

Lemma 2. *Fix* $i = 1, \ldots, N$.

$$(a)\ \mathrm{Var}(L_i) \geqslant V_i' + \frac{\lambda_i}{2}\frac{(1-\rho_i)}{F_i} + \frac{\lambda_i^2}{12}\frac{(1-\rho_i)^2}{F_i^2},$$

where

$$V_i' = \mathrm{Var}(L_i') = \frac{\lambda_i^4 E^2(S_i^2)}{4(1-\rho_i)^2} + \frac{\lambda_i^3(2E(S_i^3) - 6s_i E(S_i^2)) + 9\lambda_i^2 E(S_i^2)}{6(1-\rho_i)} + \rho_i(1-\rho_i). \tag{12}$$

$$(b)\ E(L_i^2) \geqslant \frac{\lambda_i^4 E^2(S_i^2)}{2(1-\rho_i)^2} + \frac{2\lambda_i^3 E(S_i^3) + 9\lambda_i^2 E(S_i^2)}{6(1-\rho_i)} + \rho_i + \frac{\lambda_i(1-\rho_i)(1+s_i+\lambda_i E(S_i^2))}{2F_i} + \frac{(1-\rho_i)^2(\lambda_i^2+3)}{12F_i^2}.$$

Proof. (a) It follows from (2) that $\mathrm{Var}(L_i) = \mathrm{Var}(L_i') + \mathrm{Var}(L_i'')$. We first show that $\mathrm{Var}(L_i') = V_i'$.

Let L_i^q denote the number of customers in queue in an $M/G/1$ system with arrival rate λ_i and service time distribution $S_i(\cdot)$. Note that

$$\Pr[L_i' = 0] + \Pr[L_i' = 1] = \Pr[L_i^q = 0]$$

so that

$$\Pr[L_i' = 1] = \Pr[L_i^q = 0] + \rho_i - 1,$$

while for $k \geqslant 2$, $\Pr[L_i' = k] = \Pr[L_i^q = k-1]$. This implies that

$$E(L_i'^2) = \sum_{k=1}^{\infty} k^2 \Pr[L_i' = k] = \Pr[L_i^g = 0] + \rho_i - 1 + \sum_{k=2}^{\infty} k^2 \Pr[L_i^g = k-1]$$
$$= \Pr[L_i^g = 0] + \rho_i - 1 + \sum_{k=1}^{\infty} (k+1)^2 \Pr[L_i^g = k] = \rho_i - 1 + \sum_{k=0}^{\infty} (k+1)^2 \Pr[L_i^g = k]$$
$$= \rho_i - 1 + E([L_i^g + 1]^2)$$
$$= \rho_i + E(L_i^g(L_i^g - 1)) + 3E(L_i^g).$$

It is well known that

$$E(L_i^g) = \frac{\lambda^2 E(S_i^2)}{2(1-\rho_i)}$$

and

$$E(L_i^g(L_i^g - 1)) = \frac{\lambda_i^3 E(S_i^3)}{3(1-\rho_i)} + 2E^2(L_i^g)$$

(see, e.g., (4.32) in Tijms 1986). Together with (9) we obtain (12).

It thus suffices to derive a lower bound for $\text{Var}(L_i'')$. Defining the discrete random variable J as in the proof of Lemma 1, we obtain from (3) that

$$\text{Var}(L_i'') = \text{Var}(N_i(I_{i,J}^e))$$
$$= E_J \text{Var}[N_i(I_{i,J}^e)|J = j] + \text{Var}_J E[N_i(I_{i,J}^e)|J = j]$$
$$\geqslant E_J\{E_{I_{i,J}^e} \text{Var}[N_i(I_{i,J}^e)|I_{i,J}^e, J = j)] + \text{Var}_{I_{i,J}^e} E[N_i(I_{i,J}^e)|I_{i,J}^e, J = j]\}$$
$$= E_J\left\{\frac{\lambda_i E(I_{i,J}^2)}{2E(I_{i,J})} + \lambda_i^2\left[\frac{E(I_{i,J}^3)}{3E(I_{i,J})} - \left[\frac{E(I_{i,J}^2)}{2E(I_{i,J})}\right]^2\right]\right\}$$
$$= E_J\left\{\frac{\lambda_i E(I_{i,J}^2)}{2E(I_{i,J})} + \lambda_i^2\left[\frac{4E(I_{i,J})E(I_{i,J}^3) - 3E^2(I_{i,J}^2)}{12E^2(I_{i,J})}\right]\right\}$$
$$\geqslant E_J\left\{\frac{\lambda_i}{2} E(I_{i,J}) + \left[\frac{\lambda_i^2 E(I_{i,J}^3)}{12E(I_{i,J})}\right]\right\}$$
$$\geqslant E_J\left\{\frac{\lambda_i}{2} E(I_{i,J}) + \frac{\lambda_i^2}{12} E(I_{i,J}^2)\right\}$$
$$= \frac{\lambda_i}{2} E(I_i) + \frac{\lambda_i^2}{12} E(I_i^2) \geqslant \frac{\lambda_i}{2} E(I_i) + \frac{\lambda_i^2}{12} E^2(I_i). \qquad (13)$$

To verify the second equality, note that $[N_i(I_{i,J}^e)|I_{i,J}^e, J = j]$ has a Poisson distribution whose mean and variance equal $\lambda_i I_{i,J}^e$, while $E_{I_{i,J}^e}(\lambda_i I_{i,J}^e) = \lambda_i E(I_{i,J}^e)$ and

$$\text{Var}_{I_{i,J}^e}(\lambda_i I_{i,J}^e) = \lambda_i^2[E(I_{i,J}^e)^2 - \{E(I_{i,J}^e)\}^2];$$

finally, recall that $E((U^e)^k) = E(U^{k+1})/(k+1)E(U)$ for any random variable U. The last three inequalities in (13) follow from Jensen's inequality and $E(I_{i,j})E(I_{i,j}^3) \geqslant E^2(I_{i,j}^2)$, an application of Hölder's inequality $E(XY) \leqslant E(X^p)^{1/p} E(Y^q)^{1/q}$ for any pair of nonnegative random variables X, Y and p, q with $1/p + 1/q = 1$. (Take $X = I_{i,j}^{3/2}$, $Y = I_{i,j}^{1/2}$ and $p = q = 2$.) Part (a) now follows by substituting (11) into (13) and adding the resulting expression to (12). Part (b) is immediate from part (a) and Lemma 1. □

Remark 2. Fix $i = 1, \ldots, N$. It follows from the proof of Lemma 2 that the bound for $\text{Var}(L_i)$ is accurate if (a) $\text{Var}_J E[N_i(I_{i,J}^e)|J = j]$ is small and (b) the intervisit time distributions $\{I_{i,j}: j = 1, \ldots, M\}$ have relatively low variability. As discussed in Section 3, equalization of the individual intervisit time distributions, and hence reduction of $\text{Var}_J E[N_i(I_{i,J}^e)|J = j]$ can be promoted by appropriate spacing of the items in the table $\underline{T}$ (Phase III).

We are now ready to derive a nonlinear program, with the quantities $\{F_i: i = 1, \ldots, N\}$ and δ as the primary decision variables, the value of which provides a lower bound for the exact objective (6). Our approach is to replace the true distributions of the L_i-variables by those of a set of variables $\{\hat{L}_i: i = 1, \ldots, N\}$ achieving a *minimal* value for the objective (6) among all distributions with the same first and second moments. A similar approach has been followed by Klincewicz and Whitt (1984) to obtain bounds for the mean queue length in a GI/M/1 queueing system as a function of the first two moments of the interarrival time distribution. Thus let

$$x_{ik} = \Pr[\hat{L}_i = k], \; k = 0, 1, \ldots$$

$$\text{(LB)} \quad \underline{Z} = \min \sum_{i=1}^{N} \sum_k \{p_i[k - b_i]^+ + h_i[b_i - k]^+\}x_{ik} + \sum_{i=1}^{N} K_i F_i, \qquad (14)$$

subject to

$$\sum_k x_{ik} = 1, \quad i = 1, \ldots, N, \qquad (15)$$

$$\sum_k k x_{ik} \geqslant \lambda_i\left\{\frac{\lambda_i E(S_i^2)}{2(1-\rho_i)} + s_i + \frac{1}{2}\frac{(1-\rho_i)}{F_i}\right\}, \quad i = 1, \ldots, N, \qquad (16)$$

$$\sum_k k^2 x_{ik} \geqslant \frac{\lambda_i^4 E^2(S_i^2)}{2(1-\rho_i)^2} + \frac{2\lambda_i^3 E(S_i^3) + 9\lambda_i^2 E(S_i^2)}{6(1-\rho_i)} + \rho_i + \frac{\lambda_i(1-\rho_i)(1 + s_i + \lambda_i E(S_i^2))}{2F_i} + \frac{(1-\rho_i)^2(\lambda_i^2 + 3)}{12F_i^2}, \quad i = 1, \ldots, N, \qquad (17)$$

$$\sum_{i=1}^{N} r_i F_i + \delta = 1 - \rho, \qquad (18)$$

$$\text{all } x_{ik}, b_i, F_i, \delta \geqslant 0. \qquad (19)$$

The objective (14) coincides with (6), with expectations written explicitly in terms of the minimizing probabilities $\{x_{ik}\}$. Equation (15) ensures that for all $i = 1, \ldots, N$, the quantities $\{x_{ik}\}$ correspond with a proper density function.

The inequalities (16) and (17) follow immediately from Lemmas 1 and 2(b), since $\Sigma_k kx_{ik} = E\hat{L}_i$ and $\Sigma_k k^2 x_{ik} = E\hat{L}_i^2$ for all $i = 1, \ldots, N$ while Equation (18) is immediate from the definition of $\{F_i\}$ and δ; the fraction of time available for all setups is clearly given by $1 - \rho - \delta$ and the fraction of time devoted to setups for item i is given by $r_i F_i$, $(i = 1, \ldots, N)$. We conclude:

Theorem 3. *The optimal value of the nonlinear program (LB) provides a lower bound for the best periodic base-stock policy, i.e.,* $\underline{Z} \leq Z^*$.

(LB) represents a fairly large nonlinear program (depending upon the level at which the distributions $\{x_{ik}\}$ are truncated) which fails to be convex. A natural solution approach employs Lagrangean Relaxation of the *single* constraint (18) since the relaxed problem decomposes into N separate nonlinear programs each with only three constraints (in addition to nonnegativity constraints for all variables)! Since (LB) is not convex, the Lagrangean dual may exhibit a duality gap, thus representing a further lower bound.

Remark 3. In many inventory models, analyses are performed on an item-by-item basis with setup cost parameters determined, in whole or in part, from estimates of the (marginal) time value of the associated setup times, usually an exceedingly difficult task. Observe that in the Lagrangean relaxation of (LB) the setup cost parameters $\{K_i: i = 1, \ldots, N\}$ are transformed to $\{\bar{K}_i = K_i + \lambda r_i: i = 1, \ldots, N\}$ with λ the Lagrangean multiplier. Using the values $\bar{K}_i = K_i + \lambda^* r_i$ with λ^*, the maximizing value of the Lagrangean multiplier in the Lagrangean dual, provides a rigorous foundation for the computation of such "allocated" setup cost parameters.

We observe that each of the minimizing distributions in the lower bound problem (LB) are concentrated on at most three points. (Let $(x^*, b^*, F^*, \delta^*)$ denote an optimal solution for (LB); fix $b = b^*$, $F = F^*$ and $\delta = \delta^*$ and note that for all $i = 1, \ldots, N\{x_{ik}^*: k = 0, 1, \ldots\}$ is an optimal solution of the resulting Linear Program with three constraints only.) The L_i-distributions which arise under periodic base-stock policies are much smoother; see, e.g., Federgruen and Katalan (1994). As in Klincewicz and Whitt (1984) a significantly improved approximation to the exact objective (6) can be expected when adding extra shape constraints to the constraint set (15)–(19). For example, we have observed that the shortfall distributions tend to be unimodal. Thus, for any given $i = 1, \ldots, N$, there is an integer $k_0(i)$ such that

$$x_{ik} \geq x_{ik-1} \quad \text{for } k \leq k_0(i) \text{ and } x_{ik} \leq x_{ik-1} \text{ for } k > k_0(i). \tag{20}$$

Likewise, we have observed that the shortfall distributions tend to be IFR (Increasing Failure Rate), which implies the constraints:

$$x_{ik} X_{ik-1} \geq x_{ik-1} X_{ik} \quad \text{where } X_{ik} = 1 - \sum_{l=0}^{k-1} x_{i\ell}, \qquad i = 1, \ldots, N,\ k = 1, 2, \ldots. \tag{21}$$

If the shortfall distributions are known to be log-concave, the following constraints are satisfied:

$$x_{ik}^2 \geq x_{ik-1} x_{ik+1}, \quad i = 1, \ldots, N,\ k = 1, 2, \ldots. \tag{22}$$

Thus, improved lower bounds can be obtained by adding one or more of the constraint sets (20)–(22) to the constraints (15)–(19), albeit at the expense of considerably enlarging each of the subproblems in the Lagrangean relaxation procedure. As a result these nonlinear programs may potentially be prohibitively large. We therefore seek to simplify the mathematical program (LB).

A Simplified Mathematical Program for Phase (I)

A vastly simplified mathematical program for the determination of optimal production frequencies (Phase (I)) would be available provided the optimal cost values of the newsvendor problems associated with (6) could be obtained as closed form expressions of the first m moments (say) of the L_i-distributions $(i = 1, \ldots, N)$. Recall L_i serves as the demand distribution in the corresponding newsvendor problem for item i, the minimal cost of which depends unfortunately on the *complete* distribution of the L_i-variable. However, several reasonable closed form approximations can be obtained as a function of L_i's mean and standard deviation only.

One such approximation is due to Scarf (1958), which surprisingly received little attention until its recent resurrection by Gallego (1998) and Gallego and Moon (1993). Fix $i = 1, \ldots, N$. Note that

$$\text{INV}_i = p_i E[L_i - b_i]^+ + h_i E[b_i - L_i]^+,$$

the long-run average holding and backlogging cost for item i, can be written as

$$\begin{aligned}\text{INV}_i &= (p_i + h_i) E[L_i - b_i]^+ + h_i[b_i - E(L_i)] \\ &\leq \text{INV}_i^S \stackrel{\text{def}}{=} \frac{1}{2}(p_i + h_i)\left[\sqrt{\text{Var}(L_i) + (b_i - E(L_i))^2} - (b_i - E(L_i))\right] \\ &\quad + h_i(b_i - E(L_i)),\end{aligned}$$

since

$$(L_i - b_i)^+ = \{|L_i - b_i| + (L_i - b_i)\}/2$$

and since

$$E|L_i - b_i| \leq \sqrt{E(L_i - b_i)^2} = \sqrt{\text{Var}(L_i) + (b_i - E(L_i))^2}$$

by the Cauchy-Schwarz inequality; see Lemma 1 in Gallego and Moon (1993). Observe that the approximation INV_i^S depends on the distribution of L_i only via its mean and standard deviation. Moreover, the minimum value of INV_i^S under the optimal base-stock level b_i^* can be obtained in closed form, as we show in Lemma 4 below.

An alternative approximation for INV_i, the cost of the newsvendor problem associated with item i, can be obtained by approximating the L_i-distribution by one in a class of two-parameter distributions $\mathcal{G}$, e.g., the Normal distributions or any other class in which every distribution $G(\cdot)$ with parameters ν and τ can be written as $G(t) = G^0((t - \nu)/\tau)$; here G^0 is a *unique* standard distribution. Such classes include (in addition to the Normals) (i) the delayed (or shifted) exponentials with ν the delay (or shift) and τ the standard deviation of the distribution, as well as (ii) all classes of delayed Gamma, Weibull, or Pareto distributions with a common shape parameter. Let INV_i^G denote the minimum cost of the newsvendor problem under the class of distribution $\mathcal{G}$. Also, let $G^{0-1}(\cdot)$ denote the inverse of $G^0(\cdot)$ and $\alpha^0(\cdot)$ the associated loss function, i.e., $\alpha^0(x) = \int_x^\infty (u - x)\, dG^0(u)$.

Lemma 4. *Fix* $i = 1, \ldots, N$. *Let* $\mu_i = E(L_i)$ *and* $\sigma_i^2 = \mathrm{Var}(L_i)$.

(*a*) $\inf_{b_i} \mathrm{INV}_i^S = \beta_i^S \sigma_i$ *where* $\beta_i^S = \sqrt{p_i h_i}$.
(*b*) $\inf_{b_i} \mathrm{INV}_i^G = \beta_i^G \tau_i + h_i(\nu_i - \mu_i)$ *where*

$$\beta_i^G = \left[(p_i + h_i)\alpha^0\left(G^{0-1}\left(\frac{p_i}{p_i + h_i}\right)\right) + h_i G^{0-1}\left(\frac{p_i}{p_i + h_i}\right)\right].$$

(*c*) *Let* $\Phi(\cdot)$ *denote the cdf of the standard normal and* $\alpha(\cdot)$ *the associated loss function. Let* $\mathcal{N}$ *denote the class of normal distributions. Then,* $\inf_{b_i} \mathrm{INV}_i^N = \beta_i^N \sigma_i$ *where*

$$\beta_i^N = \left[(p_i + h_i)\alpha\left(\Phi^{-1}\left(\frac{p_i}{p_i + h_i}\right)\right)\right.$$
$$\left. + h_i \Phi^{-1}\left(\frac{p_i}{p_i + h_i}\right)\right].$$

Proof. (a) The proof of this part is given in Gallego and Moon (1993). To keep the presentation self contained we give a brief argument here. Let $\bar{b}_i = b_i - \mu_i$. Note that

$$\mathrm{INV}_i^S = (p_i + h_i)\frac{\sqrt{\sigma_i^2 + \bar{b}_i^2} - \bar{b}_i}{2} + h_i \bar{b}_i \tag{23}$$

is convex in $\bar{b}_i$. The minimum value of $\bar{b}_i$ is thus obtained as the unique root $\bar{b}_i^*$ of the first derivative of INV_i^S:

$$\frac{1}{2}(p_i + h_i)\bar{b}_i(\sigma_i^2 + \bar{b}_i^2)^{-1/2} + \frac{(h_i - p_i)}{2} = 0,$$

so that

$$\bar{b}_i^* = \left[\left(\frac{p_i - h_i}{p_i + h_i}\right) \Big/ \sqrt{1 - \left(\frac{p_i - h_i}{p_i + h_i}\right)^2}\right]\sigma_i.$$

Part (a) thus follows by substituting this value of $\bar{b}_i$ into (23).

(b) Let $\bar{b}_i = b_i - \nu_i$. It is well known and easily verified that the optimal value of $\bar{b}_i$ is given by

$$G^{0-1}\left(\frac{p_i}{p_i + h_i}\right)\tau_i.$$

Also,

$$\begin{aligned}
\mathrm{INV}_i^S &= (p_i + h_i)E[L_i - \nu_i - \bar{b}_i]^+ \\
&\quad + h_i \bar{b}_i + h_i(\nu_i - \mu_i) \\
&= (p_i + h_i)\int_{\bar{b}_i}^{\infty} (u - \bar{b}_i)\, dG^0\left(\frac{u}{\tau_i}\right) \\
&\quad + h_i G^{0-1}\left(\frac{p_i}{p_i + h_i}\right)\tau_i + h_i(\nu_i - \mu_i) = (p_i + h_i)\tau_i \\
&\quad \cdot \int_{G^{0-1}(p_i/(p_i + h_i))}^{\infty} \left(u - G^{0-1}\left(\frac{p_i}{p_i + h_i}\right)\right) dG^0(u) \\
&\quad + h_i G^{0-1}\left(\frac{p_i}{p_i + h_i}\right)\tau_i + h_i(\nu_i - \mu_i) \\
&= \left[(p_i + h_i)\alpha^0\left(G^{0-1}\left(\frac{p_i}{p_i + h_i}\right)\right)\right. \\
&\quad \left. + h_i G^{0-1}\left(\frac{p_i}{p_i + h_i}\right)\right]\tau_i + h_i(\nu_i - \mu_i)
\end{aligned}$$

Part (c) is immediate from part (b). □

We thus conclude that both under the Scarf and under the Normal approximation, the minimal newsvendor cost associated with item i, can be obtained as a closed form expression proportional to σ_i, the standard deviation of L_i; for the other two parameter classes of distributions, $\inf_{b_i}$ INV_i^G is likewise expressed as a simple closed form function of μ_i and σ_i, the mean and standard deviation of L_i. The usefulness of the above classes of two-parameter distributions in terms of generating closed form expressions for minimum inventory costs was first observed in Zipkin (1982).

We henceforth confine ourselves to the Scarf and Normal approximations. Under these approximations and invoking Lemma 2, the lower bound mathematical program (LB) for the determination of the optimal production frequencies drastically simplifies to:

$$\text{(P)} \quad \min \sum_{i=1}^{N} \beta_i \psi_i(F_i) + \sum_{i=1}^{N} K_i F_i, \tag{24}$$

subject to

$$\sum_{i=1}^{N} r_i F_i \leqslant 1 - \rho, \tag{18'}$$

$$F_i \geqslant 0;\ i = 1, \ldots, N, \tag{25}$$

where

$$\psi_i(F_i) = \sqrt{V_i' + \frac{\lambda_i}{2}\frac{(1 - \rho_i)}{F_i} + \frac{\lambda_i^2}{12}\frac{(1 - \rho_i)^2}{F_i^2}}$$

and under the Scarf (Normal) approximation $\beta_i = \beta_i^S (= \beta_i^N)$ $(i = 1, \ldots, N)$. Exceedingly simple solution methods are available to solve this simple nonlinear program. This follows from Lemma 5.

Lemma 5. *The objective* (24) *of* (*P*) *is separable and convex in the decision variables* $\{F_i\}$.

Proof. Each of the ψ_i-functions is of the form

$$\psi(x) = \sqrt{\frac{A}{x^2} + \frac{B}{x} + C}$$

with A, B, $C \geq 0$. Taking the second derivative ψ'' of ψ, we obtain:

$$\begin{aligned}\psi''(x) &= \frac{1}{2}\left[\frac{A}{x^2} + \frac{B}{x} + C\right]^{-1/2}\left[\frac{6A}{x^4} + \frac{2B}{x^3}\right] \\ &\quad - \frac{1}{4}\left[\frac{A}{x^2} + \frac{B}{x} + C\right]^{-3/2}\left[\frac{-2A}{x^3} + \frac{B}{x^2}\right]^2 \\ &= \frac{1}{4}\left[\frac{A}{x^2} + \frac{B}{x} + C\right]^{-3/2} \\ &\quad \cdot\left\{2\left[\frac{A}{x^2} + \frac{B}{x} + C\right]\left[\frac{6A}{x^4} + \frac{2B}{x^3}\right] - \left[\frac{-2A}{x^3} + \frac{B}{x^2}\right]^2\right\} \\ &= \frac{1}{4}\left[\frac{A}{x^2} + \frac{B}{x} + C\right]^{-3/2} \\ &\quad \cdot\left\{\frac{8A^2}{x^6} + \frac{12AB}{x^5} + \frac{3B^2 + 12AC}{x^4} + \frac{4BC}{x^3}\right\} \geq 0. \end{aligned}$$

□

Observe also that the nonnegativity constraints (25) are never binding. To find the optimal solution for (P), we first solve the unconstrained problem by finding for all $i = 1, \ldots, N$, the unique root of the equation $\psi_i'(F_i) = -K_i$ (see Lemma 5). (As shown below this amounts to finding the unique root of a sixth-degree polynomial.) If the resulting solution satisfies (18′) it is of course optimal. Otherwise, (18′) is binding; let $\lambda > 0$ denote the (optimal) Lagrangean multiplier associated with (18′). For a given value of λ the optimal corresponding F_i-values are obtained by determining the unique root of the equation (see Lemma 5)

$$\begin{aligned}\psi_i'(F) &= \frac{1}{2}\left[\frac{A_i}{F^2} + \frac{B_i}{F} + C_i\right]^{-1/2}\left[\frac{-2A_i}{F^3} + \frac{B_i}{F^2}\right] \\ &= -\bar{K}_i \stackrel{\text{def}}{=} -(K_i + r_i\lambda) \end{aligned} \tag{26}$$

(the time adjusted setup cost; see the discussion above), where

$$A_i = \frac{\lambda_i^2(1-\rho_i)^2}{12}, \quad B_i = \frac{\lambda_i(1-\rho_i)}{2}, \quad \text{and} \quad C_i = V_i',$$

or, in terms of $x = F^{-1}$ the unique root of the sixth degree polynomial

$$\bar{K}_i^2[A_i x^2 + B_i x + C_i] = [2A_i x^3 + B_i x^2]^2.$$

Since, by Lemma 5, the functions $\psi_i'(\cdot)$ are increasing, we obtain that the root of (26) decreases with λ. The optimal value of λ can thus be found by simple bisection verifying for each trial value of λ whether the corresponding F_i values satisfy $\sum_{i=1}^N r_i F_i \geq 1 - \rho$ or $\sum_{i=1}^N r_i F_i \leq 1 - \rho$.

While the above method is simple, it involves repeated computations of (unique) roots of sixth-degree polynomials. As an alternative, we may wish to change variables to $y_i = r_i F_i$, obtaining the formulation:

$$(\mathrm{P}') \quad \min \sum_{i=1}^N \beta_i \psi_i\left(\frac{y_i}{r_i}\right) + \sum_{i=1}^N \frac{K_i}{r_i} y_i, \tag{24′}$$

subject to

$$\sum_{i=1}^N y_i \leq 1 - \rho, \tag{18″}$$

and restrict ourselves to y_i-values which are integer multiples of a given grid size d (e.g., $d = 1\%$ or $d = 1‰$). It is well known (see Gross 1956 and Fox 1958) that this discretized version can be solved by the greedy procedure; starting with $y_i^0 = d (i = 1, \ldots, N)$; we sequentially increment by one unit of d a component y_i for which this increment results in the largest positive reduction of the cost objective (24′) and terminate when no such value can be found or when $\sum_{i=1}^N y_i = 1 - \rho$.

The greedy procedure clearly requires at most $O(N + (1-\rho)d^{-1})$ elementary operations. Simple and *fully* polynomial time algorithms for (**P′**) are due to Galil and Megiddo (1979), Frederickson and Johnson (1982), and Groenevelt (1986).

We conclude with a few qualitative observations regarding the solution of (P). First, note that (18′) is always binding, i.e., $\delta = 0$, in the important special case where no *external* setup costs are incurred, i.e., $K_i = 0$ for all $i = 1, \ldots, N$. (In this case, the objective (24) is decreasing in all F_i-variables.) In other words, in the absence of external setup costs, it never pays to insert idle times. While intuitive, this result may fail to hold for the exact system-wide objective (6), as opposed to its approximation (24), albeit that such failures occur rarely, e.g., under extremely variable setup times. This phenomenon was first pointed out by Sarkar and Zangwill (1991) showing that mean waiting times can sometimes *decrease* as mean setup or idle times are *increased*. See also Zangwill (1992, 1994) for further discussion on this phenomenon.

Staying with the important special case of zero external setup costs, we also observe that the optimal vector of the frequencies $(\underline{F}, \delta)$ is quite insensitive to the specific approximation used in (P), i.e., the solution hardly depends on which of the above described approximations for the newsvendor costs $\{\mathrm{INV}_i\}$ is employed. This follows from the fact that in most practical settings, the p_i/h_i-ratios of the different items tend to be equal or close to equal; indeed, both the holding cost and backlogging cost rates are often set proportional to the unit cost values of the items. Observe that if all p_i/h_i-ratios are exactly identical, the β_i-coefficients in Lemma 4 obtained under alternative approximations differ from each other by a proportionality factor only, thus resulting in the same optimal solution for (P) (since all $K_i = 0$). Even more generally, the same argument reveals that the *relative* frequencies $\{F_i/F_j\colon i, j = 1, \ldots, N)$ are unaffected by the specific approximation used in specifying the β-coefficients in (24) as long as both the $\{p_i/h_i\}$-ratios and the $\{K_i/r_i\}$-ratios are (close to) identical. These observations regarding the robustness of the optimal vector of production frequencies instill further comfort in the use of the approximation approach leading up to problem (P).

Observe also that in problems (LB) and (P), the (approximated) cost objective depends on the vector of inserted idle times only via its sum $\Delta_{tot} = \sum_{j=1}^{M} \Delta_j = \delta\bar{C}$. Indeed, Federgruen and Katalan (1996a) proposition 2 shows for *cyclical* base-stock policies that this invariance result applies to the exact cost objective; in other words, it suffices, without loss of optimality, to insert a *single* idle time period and it is immaterial after *which* of the N items this idle time period is inserted. The invariance result fails to apply to the exact cost objective for more general periodic base-stock policies, even though partial invariance results continue to apply, see Federgruen and Katalan (1996b, corollary 5). In view of the characteristics of the approximating mathematical programs (LB) and (P) we recommend that a single idle time period of length $\delta^*\bar{C}$ be inserted at the end of the polling table (with δ^* the optimal value of δ), even when the optimal frequencies result in a noncyclical base-stock policy.

3. PHASE (III): THE ITEM-SEQUENCE IN THE PRODUCTION TABLE

With the table size M and the item frequencies $\{M_i\}$ determined in Phases (I) and (II), only the spacing of the items in the table T needs to be resolved. This is the objective of Phase (III) of our procedure. Several heuristics have been proposed for this problem in other related contexts, in particular, (i) the Golden Ratio heuristic and (ii) a machine (makespan) scheduling heuristic proposed for deterministic ELSPs. As mentioned in Section 1, the former attempts to equalize the *number* of entries between consecutive appearances of each item in the table while the latter attempts to equalize the *times* between consecutive production runs of the same item. One can think of this spacing problem as a scheduling problem with M jobs, the first M_1 of which are of type 1, the next M_2 of which are of type 2, et cetera.

More specifically, the Golden Ratio rule associates with job k an index $I_k = k\phi^{-1}$ modulo1 ($k = 1, \ldots, M$) where $\phi^{-1} = 0.618034 = (\sqrt{5} - 1)/2$ (the so-called golden ratio associated with the Fibonacci numbers), and sequences the jobs in ascending order of their index values. It appears that the Golden Ration rule was first proposed in Knuth (1973) where an extensive discussion of its properties can be found, see also Boxma et al. (1990, 1993) for a review of other spacing problems solved by this method.

It appears more appropriate to attempt to equalize the intervisit times $\{I_{i,j}: j = 1, \ldots, M_i\}$ associated with any given item i ($i = 1, \ldots, N$). This is the rationale underlying Dobson's (1987) makespan heuristic. Indeed, as observed in Remark 2, the lower bound for $\mathrm{Var}(L_i)$ in Lemma 2 and hence the lower bound in the problem (LB) and the approximate lower bound in (P) are all tightest if $\mathrm{Var}_J E[N_i(I^e_{i,J})|J = j] = 0$, i.e., if the intervisit times $\{I_{i,j}: j = 1, \ldots, M_i\}$ are identically distributed. Note that all M_i jobs associated with a given item i ($i = 1, \ldots, N$) are of the same expected duration $r_i + \rho_i\bar{C}/M_i$ if the intervisit times $\{I_{i,j}: j = 1, \ldots, M_i\}$ have identical means; here $\bar{C}$ is the expected cycle time given by (4) with $\sum_{j=1}^{M} \Delta_j = \delta^*$, the optimal value of δ obtained in Phase (I). Dobson's heuristic assumes that all $\{M_i\}$-frequencies have been rounded to power-of-two integers and conceives of the sequencing problem as a scheduling problem with M jobs, each with the above described duration, and with M^+ machines where $M^+ = \max_i \{M_i\}$. The ℓth machine ($\ell = 1, \ldots, M^+$) corresponds with the ℓth of M^+ subintervals of the full cycle $\bar{C}$ (each of which is of equal length $\bar{C}/M^+$). Starting with $t = 0, 1, \ldots, \log_2 M^+$ the rule schedules for all items with $M_i = M^+/2^t$ their $M^+/2^t$ corresponding jobs exclusively on *one* of the machine sets $\{kM^+/2^t + \ell : k = 0, 1, \ldots, M^+/2^t - 1\}$ ($\ell = 1, \ldots, 2^t$), where the specific machine set is chosen so as to heuristically minimize the machine makespan, e.g., by the LPT (Longest Processing Time first) procedure.

The requirement in the above heuristic that all table frequencies $\{M_i\}$ be rounded to power-of-two values represents one of its drawbacks. As an alternative capable of handling arbitrary M_i-values, we propose an iterative shortest path based heuristic. Starting with an arbitrary initial sequence (e.g., the sequence obtained by the Golden Ratio rule or the makespan scheduling procedure) we attempt, in the kth iteration to improve the positions of the M_k entries of item k (modulo N) as follows: as above, associate with each entry for item i a duration $r_i + \rho_i\bar{C}/M_i$ ($i = 1, \ldots, N$). Let π denote the current periodic sequence after removal of all but the first entry of item k. Now, insert the remaining $(M_k - 1)$ entries of item k so as to minimize the sum of the absolute deviations of the expected intervisit times from their target value $(1 - \rho_k)\bar{C}/M_k + r_k$, using the above durations of the jobs in computing the intervisit times. We replace the existing set of entries for item k by the newly obtained set only if the resulting table $\underline{T}$ evaluated by the above described method, is associated with a strictly lower cost value in (6).

The above positioning problem can be solved as follows: let p denote the first entry of item k and cut the path $\{p + 1, p + 2, \ldots, M, 1, \ldots, p\}$ in the sequence π into M_k links so as to minimize the sum of the absolute deviations of the arc lengths from the target value $(1 - \rho_k)\bar{C}/M_k + r_k$. The problem of finding the shortest path with M_k arcs from entry p back to entry p can be solved exactly in $O(M^2M_k)$ time; many faster highly effective heuristics can be used instead. This iterative repositioning procedure is terminated as soon as in a series of N consecutive iterations, none of the positions of the entries can be improved. Note that this procedure terminates after a finite number of iterations since the sequence of optimal cost values associated with the corresponding table $\underline{T}$ is monotonically decreasing, thus precluding any cycling phenomena.

4. NUMERICAL STUDY

We conclude this paper with a numerical study designed to test the quality of the proposed heuristics for determining

Table I
Ratio of the Largest and the Smallest ρ-values for ω-vectors and N

N	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$
3	1.60	5.00	20.00
5	1.67	7.00	33.00
7	2.00	5.00	29.00

production schedules. Recall that the optimal production schedule represents an exceedingly difficult combinatorial problem as there are $(N!\binom{M}{N}N^{M-N})$ potential sequences of a *given* periodicity M ($\geqslant N$) alone, and the optimal M value itself may be quite large, resulting in a large number of potential absolute frequency vectors, $\{M_i;\ \sum_{i=1}^{N} M_i = M,\ M_i \geqslant 1\}$. Therefore, we restrict ourselves to problem instances with at most seven items. As demonstrated above, the heuristics themselves can be used confidently for systems with scores or hundreds of items.

We have confined ourselves to instances with *zero* external setup costs for all items (i.e., $K_i = 0$ for all $i = 1, \ldots, N$). Recall that the expression for the average setup cost component in the objective function of (P′) (or (P)) is exact. Thus, the approximation of problem (P′) is relatively more accurate in the presence of setup costs. By demonstrating that our heuristics come close to identifying optimal periodic base-stock policies for settings *without* external setup costs, we thus are able to conclude that the same will apply (a fortiori) for general settings with *positive* setup costs.

In addition to addressing the more limited set of problem instances (18 in total) considered in the setup time problems of Markowitz et al. (1995), we have evaluated 312 additional problem instances which are partitioned into three sets; set 1 (2, 3) consists of 270 (36, 6) problem instances with $N = 3$ (5, 7); within each set, we vary the *total* utilization rate ρ; $\rho \in \{0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.90\}$ in set 1, $\rho \in \{0.25, 0.45, 0.65, 0.85\}$ in set 2 and $\rho \in \{0.25, 0.45\}$ in set 3. For a given value of ρ, the demand rates of the individual items are specified by a non-negative allocation vector $\omega = (\omega_1, \ldots, \omega_N)$ with $\sum_{i=1}^{N} \omega_i = 1$, i.e., $(\rho_1, \cdots, \rho_N) = (\omega_1\rho, \ldots, \omega_N\rho)$. Three distinct allocation vectors ω, are considered so as to cover relatively balanced and two types of unbalanced systems. The ratios of the largest and the smallest ρ-value are reported in Table I.

For each of the (ρ, ω) combinations, we generate 10 (3, 1) problem instances in set 1 (2, 3) as follows: for a given value of ρ, the traffic intensities of the individual stations are specified by a non-negative allocation vector $\omega = (\omega_1, \ldots, \omega_N)$ with $\sum_{i=1}^{N} \omega_i = 1$, i.e., $(\rho_1, \cdots, \rho_N) = (\omega_1\rho, \ldots, \omega_N\rho)$. For each problem instance, we generate combinations of production/setup time distributions as follows: all production and setup times are of one of the following four types: deterministic, exponential, uniform and Erlang. We assign to each of the random variables $\{S_i, R_i;\ i = 1, \cdots, N\}$ one of the four distributional forms *independently* and with equal probability ¼. Having determined that a specific service time or switching time variable, S_i or R_i ($i = 1, \cdots, N$) is of one of the four types, we select the parameter(s) associated with this distribution as follows: the mean s_i (r_i) is selected uniformly on the interval [0.25, 1.0] ([1.0, 2.5]). For Erlang distributions, we select the phase uniformly among the integers between 2 and 10; finally, for a uniform distribution, we select the lower bound of its support uniformly between 0 and the assigned mean. For all problem instances, the inventory holding (backlogging) cost rates, h_i (p_i), are selected uniformly on the interval [0.1, 10]([1, 100]) for all $i = 1, \ldots, N$.

For each problem instance, we first determine the relative frequency vector $\{F_i\}$ (Phase (I)) via the mathematical program (P′) under the Scarf and Normal approximations. The problem is solved via the greedy procedure with a grid size $d = 0.001$. The vector of absolute frequencies, $\{M_i\}$ (Phase (II)) is selected via the simple rounding procedure described in Section 1 with a prespecified precision, $\eta = 0.25$, and with an upper bound on the maximum table size, $M_{\max}$($M_{\max} = 30, 50, 70$ for $N = 3, 5, 7$, respectively). To differentiate between the impact of the frequency vector $\{M_i\}$ (Phases (I) and (II)) and the sequence selection (Phase (III)), we initially determine the production table $\underline{T}$ via the Golden Ratio procedure exclusively. We denote by T^S and T^N the solutions obtained by the Scarf and Normal approximations, respectively. For each of the two initial solutions, we perform a *limited* local search among at most $2N$ neighbors by changing *only* one of the M_i-values by $+1$ or -1 for each i ($i = 1, \ldots, N$) at a time. For each production table $\underline{T}$, we compute the system-wide long-run average cost via the evaluation methods described in Federgruen and Katalan (1996a) and select the one which achieves the minimum cost. We also conduct an *extensive* local search among all 3^N neighbors in the set $X_{i=1}^{N}\{M_i - 1, M_i, M_i + 1\}$ of each of the two initial solutions and again select the one which minimizes the overall long-run average cost. Let T^L and T^E denote the solution obtained by the limited and extensive local search procedures, respectively.

As mentioned above, determining the true optimal production table and its associated cost is impractical even for systems with only three items. Therefore, we select the best production table within a large neighborhood of our best heuristic solution T^E; for systems of size 3 (5, 7) we search among the 10^3 (5^5, 5^7) neighbors in the set $X_{i=1}^{N}\{1, \ldots, 10\}(X_{i=1}^{N}\{M_i - 2, M_i - 1, M_i, M_i + 1, M_i + 2\})$.

To assess the impact of the sequence selection, we apply a pairwise interchange heuristic to the solution T^L. In each iteration we evaluate the cost improvement associated with interchanging each of the pairs of different entries in the table $\underline{T}$, implementing whichever pairwise interchange results in the *best* improvement, if any. The procedure terminates with a solution T^{PI} at the first iteration where no improvement can be obtained from a pairwise interchange.

Table II
Set 1, $N = 3$

ρ		r_1	ϵ_1	ϵ_2	ϵ_3	ϵ_4	Δ_1	Δ_2	Δ_3
0.15	mean	0.6864	0.0517	0.0213	0.0132	0.1068	0.0300	0.0080	0.0259
	stdev	0.2588	0.0522	0.0229	0.0176	0.1146	0.0522	0.0114	0.0553
	max	1.0892	0.2438	0.0749	0.0722	0.4737	0.2390	0.0492	0.2862
0.25	mean	0.7396	0.0521	0.0226	0.0099	0.1060	0.0287	0.0125	0.0220
	stdev	0.2586	0.0668	0.0472	0.0290	0.1138	0.0407	0.0362	0.0502
	max	1.1076	0.2772	0.1865	0.1569	0.4036	0.1398	0.1813	0.2045
0.35	mean	0.7327	0.0570	0.0158	0.0073	0.0773	0.0407	0.0084	0.0386
	stdev	0.2954	0.0547	0.0259	0.0167	0.0741	0.0488	0.0222	0.0692
	max	1.2416	0.2337	0.0960	0.0796	0.2550	0.2282	0.0937	0.2532
0.45	mean	0.8094	0.0425	0.0180	0.0047	0.0778	0.0240	0.0132	0.0112
	stdev	0.2759	0.0409	0.0307	0.0102	0.0598	0.0267	0.0270	0.0201
	max	1.3611	0.1470	0.1157	0.0398	0.1947	0.0959	0.0869	0.0924
0.55	mean	0.7845	0.0561	0.0066	0.0012	0.0442	0.0491	0.0055	0.0241
	stdev	0.2834	0.0479	0.0129	0.0037	0.0525	0.0428	0.0128	0.0357
	max	1.1352	0.1452	0.0459	0.0154	0.1640	0.1366	0.0459	0.1205
0.65	mean	0.8148	0.0516	0.0172	0.0059	0.0426	0.0338	0.0111	0.0322
	stdev	0.3158	0.0524	0.0290	0.0122	0.0634	0.0431	0.0217	0.0436
	max	1.4558	0.1864	0.1380	0.0440	0.3116	0.1549	0.0921	0.1388
0.75	mean	0.7283	0.0776	0.0173	0.0061	0.0348	0.0588	0.0111	0.0622
	stdev	0.2966	0.0885	0.0397	0.0175	0.0516	0.0676	0.0342	0.1017
	max	1.1591	0.2802	0.1726	0.0699	0.2187	0.2120	0.1726	0.4474
0.85	mean	0.6778	0.0795	0.0295	0.0127	0.0564	0.0481	0.0161	0.0503
	stdev	0.2140	0.0911	0.0557	0.0300	0.0620	0.0589	0.0305	0.0718
	max	0.9857	0.3156	0.1983	0.1026	0.1927	0.2078	0.1010	0.2625
0.90	mean	0.6443	0.0629	0.0079	0.0047	0.0292	0.0544	0.0031	0.0628
	stdev	0.2459	0.0766	0.0253	0.0202	0.0379	0.0671	0.0101	0.0932
	max	1.0028	0.2799	0.1283	0.1060	0.1283	0.2333	0.0453	0.3829
Set 1	mean	0.7353	0.0590	0.0174	0.0073	0.0639	0.0408	0.0099	0.0333
	stdev	0.1056	0.0129	0.0080	0.0059	0.0784	0.0211	0.0021	0.0048
	max	1.4558	0.3156	0.1983	0.1569	0.4737	0.2390	0.1813	0.4474

Let $C(T)$ denote the long-run average cost associated with a solution T, and Z^S, Z^N the approximated cost obtained in problem (P′) under the Scarf- and Normal- approximation, respectively. Let T^C denote the pure cyclical schedule and C^{OPT}—the cost of the best production table found, i.e., the table found by the extensive search over the 10^3, 5^5, or 5^7 neighbors of the solution T^E. In each set, we have recorded for each problem instance the following performance measures:

r_1 = the ratio between the minimum approximated cost and C^{OPT}
$= \min\{Z^S, Z^N\}/C^{\text{OPT}}$,

ϵ_1 = the relative optimality gap between the initial solution and C^{OPT}
$= \min\{C(T^S), C(T^N)\}/C^{\text{OPT}} - 1$,

ϵ_2 = the relative optimality gap between the solution obtained after the limited search and C^{OPT}
$= C(T^L)/C^{\text{OPT}} - 1$,

ϵ_3 = the relative optimality gap between the solution obtained after the extensive search and C^{OPT}
$= C(T^E)/C^{\text{OPT}} - 1$,

ϵ_4 = the relative optimality gap between the pure cyclical schedule and C^{OPT}
$= C(T^C)/C^{\text{OPT}} - 1$,

Δ_1 = the relative improvement obtained after the limited search
$= \min\{C(T^S), C(T^N)\}/C(T^E) - 1$,

Δ_2 = the relative improvement obtained after the extensive search
$= C(T^L)/C(T^E) - 1$,

Δ_3 = the relative improvement obtained after the pairwise interchange procedure
$= C(T^L)/C(T^{PI}) - 1$.

Tables II, II, and IV report the averages, standard deviations, and maxima of these performance measures for all categories of problem instances with the same total demand rate, ρ, in sets 1, 2, and 3, respectively, as well as their summary statistics for the entire set. Figures 1 and 2 exhibit the histograms of the four ϵ-optimality gap measures.

Focusing first on the results for systems with three items ($N = 3$, set 1) we observe that the initial solution obtained from the three-phase procedure is on *average* only slightly better than the cyclical base-stock policy with table T^C. However the performance of this initial solution is already significantly more *reliable* than that of the cyclical policies. The standard deviation of the optimality gap vis-à-vis C^{OPT} is more than six times smaller for the initial periodic table as compared to T^C. Our second major observation is that the relatively inexpensive limited local search eliminates most of the optimality gap vis-à-vis C^{OPT}, reducing it to an

Table III
Set 2, $N = 5$

ρ		r_1	ϵ_1	ϵ_2	ϵ_3	ϵ_4	Δ_1	Δ_2	Δ_3
0.25	mean	1.1209	0.1016	0.0152	0.0064	0.1101	0.0850	0.0087	0.0421
	stdev	0.1295	0.0636	0.0167	0.0139	0.0891	0.0574	0.0045	0.0549
	max	1.2378	0.1956	0.0566	0.0429	0.2793	0.1707	0.0142	0.1449
0.45	mean	1.1484	0.0864	0.0123	0.0065	0.0831	0.0732	0.0057	0.0685
	stdev	0.1117	0.0666	0.0103	0.0103	0.0648	0.0641	0.0044	0.1104
	max	1.2982	0.1905	0.0308	0.0303	0.2464	0.1684	0.0124	0.3367
0.65	mean	1.1584	0.0598	0.0097	0.0042	0.1237	0.0494	0.0055	0.0250
	stdev	0.1582	0.0418	0.0085	0.0077	0.0836	0.0352	0.0046	0.0275
	max	1.4761	0.1242	0.0264	0.0228	0.2661	0.0953	0.0131	0.0666
0.85	mean	1.1039	0.0946	0.0185	0.0113	0.1292	0.0747	0.0071	0.0507
	stdev	0.1896	0.0501	0.0236	0.0230	0.1026	0.0430	0.0052	0.0545
	max	1.2909	0.1412	0.0673	0.0639	0.2518	0.1267	0.0133	0.1341
Set 2	mean	1.1329	0.0856	0.0139	0.0071	0.1115	0.0706	0.0067	0.0466
	stdev	0.1453	0.0563	0.0156	0.0145	0.0843	0.0507	0.0047	0.0677
	max	1.4761	0.1956	0.0673	0.0639	0.2793	0.1707	0.0142	0.3367

average of 1.74%. (See the summary statistics for the measures ϵ_2 and Δ_2.) The same relative comparisons apply to set 2 for systems with $N = 5$ items and the more limited set 3 for the case where $N = 7$. Note however that in set 2 the cyclical base-stock policy is on average 11.15% more costly than C^{OPT} while the initial table found by the three-phase procedure is on average 8.56% more costly.

The histograms in Figures 1 and 2 show that the average optimality gap between T^L or T^E and C^{OPT}, i.e., the average value of ϵ_2 and ϵ_3, is in fact biased upward due to the tail of the distributions: for example, in set 1, the limited search eliminates the *complete* optimality gap vis-à-vis C^{OPT} in more than 50% of the problem instances.

The above results gauge the impact of determining appropriate production frequencies. Additional improvements can be achieved by altering the *sequence* in which the items are produced in the production table. The pairwise interchange procedure achieves in set 1 ($N = 3$), on average, a 3.33% reduction of the cost achieved under the table T^L and an average cost reduction of 4.66% for set 2 ($N = 5$).

We note, in addition that the above discussed relative performance of the different heuristics for Phases I and II, i.e., the cost the initial solutions (T^S, T^N), the cyclical table T^C, the solutions obtained after the limited and the extensive search T^L and T^E, as well as C^{OPT}, are relatively insensitive to the total utilization rate ρ; no specific patterns are apparent as ρ is varied from 0.15 to 0.9 in set 1 and from 0.25 to 0.85 in set 2.

On the other hand the potential for cost improvements due to alterations in the *sequence* in which the items are produced, tends to increase as ρ increases reflecting the fact that in systems with a high utilization rate ρ, the lengths and variabilities of the intervisit times $\{I_{i,j}\}$ for a given item i have a larger impact on system-wide performance.

We now turn to the eighteen problem instances considered in the "setup time" problem study of Markowitz et al. (1995). Fourteen of the 18 instances have $N = 2$ items. For these 14 instances, the authors report that cyclic base stock policies exhibit an average optimality gap of 11.1%, as opposed to an average optimality gap of 6.5% for their proposed *dynamic* cyclic policies. (The cyclical base stock policies are referred to as "generalized base stock policies" and represent a slight variant of the cyclical base stock policies considered here; after setting up for a given item, one starts production only if its inventory level is at least some y units below the base stock level and idles otherwise. The same minimum batch size y is used for all items. Partially based on the results mentioned below, we conjecture that little, if anything, is gained by the adoption of these minimum batch sizes as compared to static idle times.) The incremental 4.6% optimality gap is somewhat

Table IV
Set 3, $N = 7$

ρ, ω	r_1	ϵ_1	ϵ_2	ϵ_3	ϵ_4	Δ_1	Δ_2
0.25, 1	0.6438	0.0388	0.0243	0.0000	0.0749	0.0141	0.0243
0.25, 2	0.7367	0.0098	0.0014	0.0000	0.0108	0.0083	0.0014
0.25, 3	0.7533	0.0026	0.0026	0.0012	0.0066	0.0000	0.0013
0.45, 1	0.5533	0.0258	0.0048	0.0004	0.0552	0.0209	0.0043
0.45, 2	0.5887	0.0260	0.0040	0.0025	0.0501	0.0219	0.0015
0.45, 3	0.8379	0.0157	0.0157	0.0063	0.0260	0.0000	0.0093
mean	0.6857	0.0198	0.0088	0.0018	0.0373	0.0109	0.0071
stdev	0.1087	0.0130	0.0092	0.0024	0.0271	0.0098	0.0090
max	0.8379	0.0389	0.0243	0.0063	0.0750	0.0219	0.0243

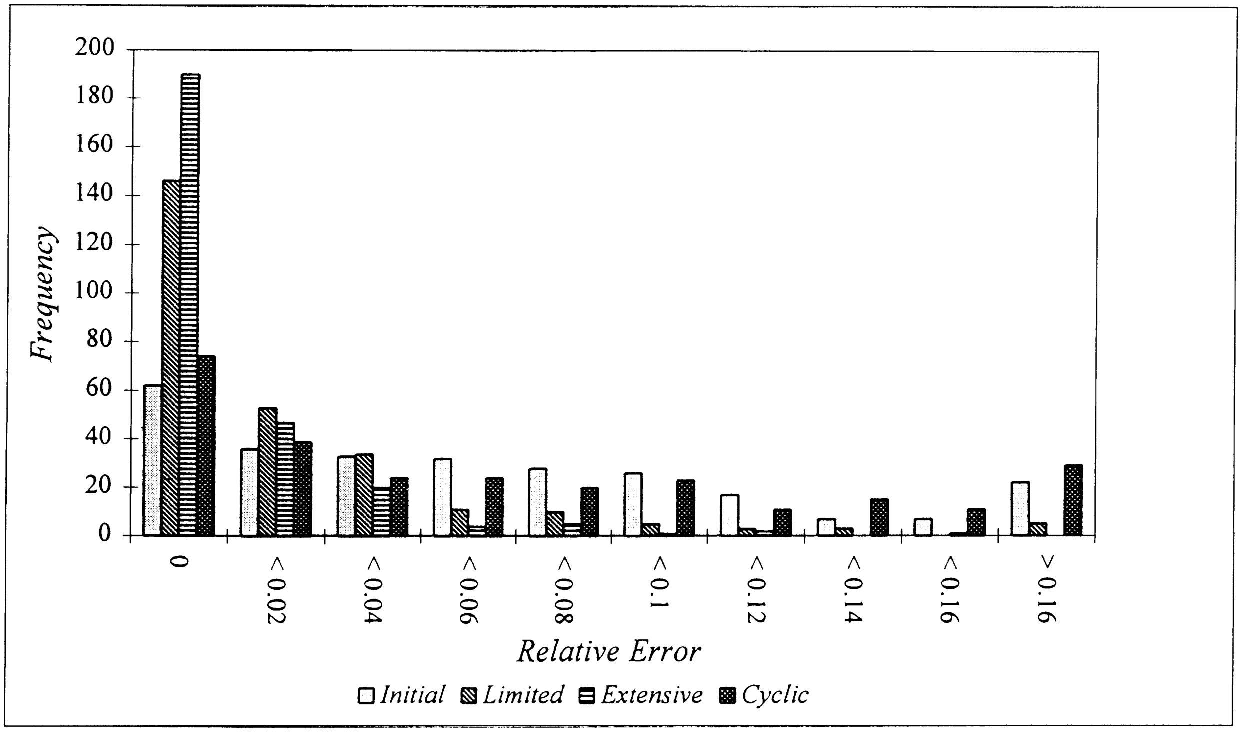

Figure 1. Distribution of relative error set 1, $N = 3$.

inflated by the fact that dynamic cyclic policies allow for *preemption* while base stock policies do not. Note that with $N = 2$ items, only the cyclical sequence is relevant, i.e., no improvements can be expected from a noncyclical item sequence.

The remaining four instances have $N = 5$ items. Two of these are symmetric, i.e., the items have identical characteristics. Here the authors report that the generalized base-stock policies (slightly) outperform the proposed dynamic cyclic policies. Below in Table V, we show that the

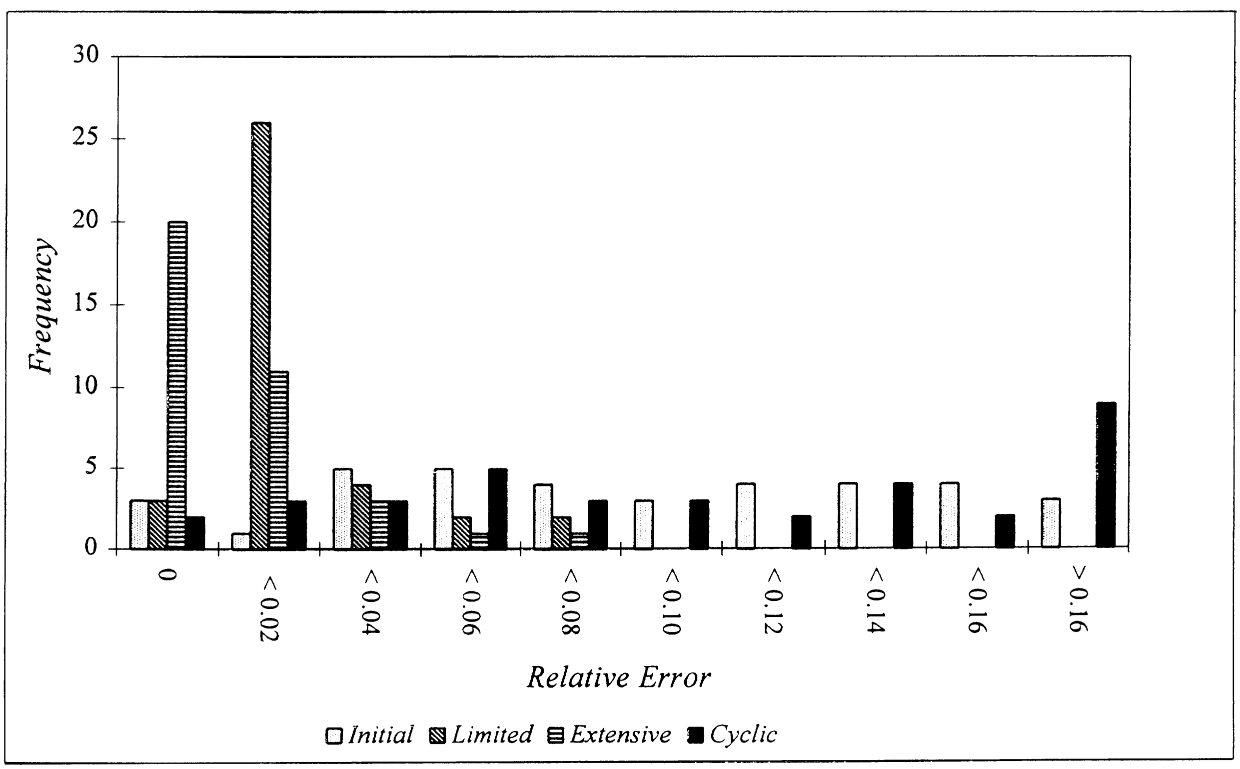

Figure 2. Distribution of relative error set 2, $N = 5$.

Table V
Comparison for the Five-item Cases with Markowitz et al. (1995) (MRW)

p_i/h_i-ratio	Cost Structure	Cost of Dynamic Policy (MRW)	Cost of Gen. Base-stock Policy (MRW)	Cost of Standard Cyclical Base-stock Policy	Cost of Periodic Base-stock Policy
5	Symmetric	215.4 (±4.9)	214.1 (±2.6)	214.3	NA
10	Symmetric	264.7 (±10.4)	260.2 (±4.7)	260.6	NA
5	Asymmetric	610.8 (±8.9)	661.0 (±9.1)	642.7	633.8
10	Asymmetric	737.4 (±18.7)	791.7 (±16.1)	781.9	769.1

same applies to the standard cyclic base stock policies. Clearly, for symmetric instances, the cyclical sequence is optimal among all periodic sequences.

This leaves us with two asymmetric instances with $N = 5$ items. Even in these two instances, all items share identical demand rates, setup times and production time distributions, and zero setup costs i.e., they differ only in their holding and backlogging cost rates: $h_i = i$, $i = 1, \ldots, 5$ and $p_i = 5h_i$, $[10h_i]$ for the first [second] instance. For these two instances the cyclic base stock policy is 5.2% and 6.0% more expensive than the proposed dynamic cyclic policy. (Note that the standard cyclical base stock policy is again somewhat cheaper than the generalized cyclic base stock policy.) On the other hand, the nonperiodic sequence (5, 1, 2, 5, 3, 4) with $M_1 = M_2 = M_3 = M_4 = 1$ and $M_5 = 2$ is only 3.8% and 4.3% more expensive. The frequency vector $M^* = (1, 1, 1, 1, 2)$ is optimal among all 5^5 vectors $\{1, 2, 3, 4, 5\}^5$. This vector and the corresponding sequence (5, 1, 2, 5, 3, 4) is indeed found by our three-phase heuristic.

REFERENCES

→ ANUPINDI, R., AND S. TAYUR. 1998. Managing Stochastic Multi-product Systems: Models, Measures and Analysis. *Op. Res.* **46,** S98–S111.

BERTSIMAS, D., AND J. NIÑO-MORA. 1993. Conservation Laws, Extended Polymatroids and Multi-armed Bandit Problems: A Unified Polyhedral Approach. *Proceedings of the Third Conference on Integer Programming and Combinatorial Optimization*, 355–385.

BERTSIMAS, D., I. PASCHALIDIS, AND J. TSITSIKLIS. 1994. Optimization of Multiclass Queueing Networks: Polyhedral and Nonlinear Characterizations of Achievable Performance. to appear in *Annals of Applied Probability*.

BERTSIMAS, D., AND H. XU. 1993. Optimization of Polling Systems and Dynamic Vehicle Routing Problems on Networks. Working Paper, Sloan School of Management and Operations Research Center, MIT, Cambridge, MA.

BHATTACHARYA, P. P., L. GEORGIADIS, AND P. TSOUCAS. 1991. Extended Polymatroids: Properties and Optimization. *Proceedings of International Conference on Integer Programming and Combinatorial Optimization*. Carnegie Mellon University, Mathematical Programming Society, 298–315.

BOURLAND, K. E., AND C. A. YANO. 1991. 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.

→ BOURLAND, K. E., AND C. A. YANO. 1994. The Strategic Use of Capacity Slack in the Economic Lot Scheduling Problem with Random Demand, *Mgmt. Sci.* **40,** 1690–1704.

BOXMA, O. J., H. LEVY, AND J. A. WESTSTRATE. 1990. Optimization of Polling Systems. In *Performance 90*, P. J. B. King, I. Mitrani, and R. J. Pooley (eds.). Elsevier Science Publishers B.V., North Holland.

BOXMA, O. J., H. LEVY, AND J. A. WESTSTRATE. 1991. Efficient Visit Frequencies for Polling Tables: Minimization of Waiting Cost. *Queueing Systems,* **9,** 133–162.

BOXMA, O. J., H. LEVY, AND J. A. WESTSTRATE. 1993. Efficient Visit Orders for Polling Systems. *Performance Evaluation*, **18,** 103–123.

BROWNE, S., AND U. YECHIALI. 1989a. Dynamic Routing in Polling Systems. *Teletraffic Science for New Cost-Effective Systems, Networks and Services, ITC-12*. M. Bonatti (ed.), Elsevier Science Publishers, North Holland, Amsterdam, 1455–1466.

BROWNE, S., AND U. YECHIALI. 1989b. Dynamic Priority Rules for Cyclic-type Queues. *Advances in Applied Probability*. **21,** 2, 432–450.

CHOUDHURY, G. L., AND W. WHITT. 1994. Computing Transient and Steady-state Distributions in Polling Models by Numerical Transform Inversion. Working Paper, AT&T Bell Laboratories, Holmdel, NJ.

→ COFFMAN, E., AND I. MITRANI. 1980. A Characterization of Waiting Time Performance Realizable by Single Server Queues. *Opns. Res.* **28,** 810–821.

DOBSON, G. 1987. The Economic Lot Scheduling Problem: Achieving Feasibility Using Time-varying Lot Sizes. *Opns. Res.* **35,** 5, 764–771.

DUENYAS, I., AND M. P. VAN OYEN. 1995. Stochastic Scheduling of Parallel Queues with Setup Costs. *Queueing Systems*, **19,** 421–444.

→ DUENYAS, I., AND M. P. VAN OYEN. 1996. Heuristic Scheduling of Parallel Heterogeneous Queues with Setups. *Mgmt. Sci.* **42,** 814–829.

FEDERGRUEN, A., AND H. GROENEVELT. 1988a. Characterization and Optimization of Achievable Performance in General Queueing Systems. *Opns. Res.* **36,** 736–741.

→ FEDERGRUEN, A., AND H. GROENEVELT. 1988b. M/G/c Queueing Systems with Multiple Customer Classes: Characterization and Control Achievable Performance under Nonpreemptive Priority Rules. *Mgmt. Sci.* **34,** 1121–1138.

FEDERGRUEN, A., AND Z. KATALAN. 1994. Approximating Queue Size and Waiting Time Distributions in General Polling Systems. *Queueing Systems*, **18,** 353–386.

FEDERGRUEN, A., AND Z. KATALAN. 1995. Make-to-Stock or Make to Order: That is the Question; Novel Answers to an Ancient Debate. Working Paper, Graduate School of Business, Columbia University, New York, NY.

FEDERGRUEN, A., AND Z. KATALAN. 1996a. The Stochastic Economic Lot Scheduling Problem: Cyclical Base-stock Policies with Idle Times. *Mgmt. Sci.* **42,** 783–796.

FEDERGRUEN, A., AND Z. KATALAN. 1996b. The Impact of Setup Times on the Performance of Multi-class Service and Production Systems. *Opns. Res.* **44,** 989–1001.

FOX, B. 1958. Discrete Optimization via Marginal Analysis. *Mgmt. Sci.* **4,** 136–153.

FREDERICKSON, G. N., AND D. B. JOHNSON. 1982. The Complexity of Selection and Ranking in X + Y and Matrices with Sorted Columns. *J. Comput. Sys. Sci*. 197–208.

FUHRMANN, S. W., AND R. B. COOPER. 1985. Stochastic Decompositions in the M/G/1 Queue with Generalized Vacations. *Opns. Res.* **33,** 5, 1117–1129.

GALIL, Z., AND N. MEGIDDO. 1979. A Fast Selection Algorithm and the Problem of Optimum Distribution of Efforts. *J. ACM*, **26,** 58–64.

GALLEGO, G., AND I. MOON. 1993. The Distribution Free Newsboy Problem: Review and Extensions. *J. Opl. Res. Soc.* **44,** 825–834.

GALLEGO, G. 1998. New Bounds and Heuristics for (Q, r) Policies. *Mgmt. Sci.* **44,** 219–233.

GROENEVELT, H. 1986. Resource Allocation Problems with Decreasing Marginal Returns to Scale, Ph.D. Dissertation, Columbia University, New York, NY.

GROSS, O. 1956. A Class of Discrete-type Minimization Problems. Working Paper, RM-1644-PR, The Rand Corporation, Santa Monica, CA.

HA, A. 1992. Optimal Scheduling and Stock Rationing Policies in Make-to-stock Production Systems. Ph.D. Dissertation, Stanford University, Stanford, CA.

KATEHAKIS, M. N., AND A. F. VEINOTT. 1987. The Multiarmed Bandit Problem: Decomposition and Computation. *Math. O. R.* **12,** 262–268.

KLEINROCK, L., AND H. LEVY. 1988. The Analysis of Random Polling Systems. *Opns. Res.* **36,** 716–732.

KLINCEWICZ, J. G., AND W. WHITT. 1984. On Approximation for Queues, II: Shape Constraints. *AT&T Bell Laboratories Technical J.* **63,** 139–161.

KNUTH, D. E. 1973. *The Art of Computer Programming, Vol. 3.* Addison-Wesley, Reading, MA.

KONHEIM, A. G., H. LEVY, AND M. M. SRINIVASAN. Descendant Set: An Efficient Approach for the Analysis of Polling Systems. *IEEE Trans. Communications*, **42,** 2/3/4, 1245–1253.

LEACHMAN, R. C., AND A. GASCON. 1988. A Heuristic Scheduling Policy for Multi-item, Single-machine Production Systems with Time-varying Stochastic Demands. *Mgmt. Sci.* **34,** 3, 377–390.

LEACHMAN, R. C., Z. K. XIONG, A. GASCON, AND K. PARK. 1991. Note: An Improvement to the Dynamic Cycle Lengths Heuristic for Scheduling the Multi-item, Single Machine. *Mgmt. Sci.* **37,** 9, 1201–1205.

MARKOWITZ, D. M., M. I. REIMAN, AND L. M. WEIN. 1995. The Stochastic Economic Lot Scheduling Problem: Heavy Traffic Analysis of Dynamic Cyclic Policies. Working Paper, Operations Research Center, MIT, Cambridge, MA.

QUI, J., AND R. LOULOU. 1995. Multiproduct Production Inventory Control under Random Demands. *IEEE Trans. Automatic Control*, **40,** 350–356.

SARKAR, D., AND W. I. ZANGWILL. 1991. Variance Effects in Cyclic Production Systems. *Mgmt. Sci.* **37,** 444–453.

SCARF, H. 1958. A Min-Max Solution of an Inventory Problem. In *Studies in the Mathematical Theory of Inventory and Production*. Stanford University Press.

SHANTHIKUMAR, G., AND D. D. YAO. 1992. Multiclass Queueing Systems: Polymatroidal Structure and Optimal Scheduling Control. *Opns. Res.* **40,** Supplement 2, S293–299.

TAKAGI, H. 1986. *Analysis of Polling Systems*. MIT Press, Cambridge, MA.

TAYUR, S. 1994. A JIT Implementation at a Laminate Plant: Some New Insights. Working Paper, GSIA, Carnegie Mellon University, Pittsburgh, PA.

TIJMS, H. C. 1986. *Stochastic Modeling and Analysis: A Computational Approach*. John Wiley & Sons, Great Britain.

ZANGWILL, W. I. 1992. The Limits of Japanese Production Theory. *Interfaces*, **22,** 14–25.

ZANGWILL, W. I. 1994. Comments on "The Limits of Japanese Production Theory." *Interfaces*, **24,** 77–94.

ZIPKIN, P. 1982. Exact and Approximate Cost Functions for Product Aggregates. *Mgmt. Sci.* **28,** 1002–1012.

ZIPKIN, P. 1986. Models for Design and Control of Stochastic, Multi-item Batch Production Systems. *Opns. Res.* **34,** 1, 91–104.

ZIPKIN, P. 1991. Computing Optimal Lot Sizes in the Economic Lot Scheduling Problem. *Opns. Res.* **39,** 1, 56–63.