IIE Transactions (1999) **31**, 1–9

Manufacturing control of a serial system with binomial yields, multiple production runs, and non-rigid demand: a decomposition approach

DAN BRAHA

Department of Industrial Engineering, Ben-Gurion University, P.O. Box 653, Beer-Sheva 84120, Israel
E-mail: brahad@bgumail.bgu.ac.il

Received March 1997 and accepted February 1998

In this paper, the problem of ordering a series of production runs to meet a single make-to-order demand, where the various stages in the production process have binomial yields, is addressed. The use of additional procurement of part-finished material or reworking defective material previously made to supplement yield at any stage is considered. The costs of this and manufacture, disposal of surplus or shortage are assumed to be proportional to the numbers involved. Set-up costs may or may not be incurred. The treatment follows from, and builds on, a previous single-run analytical result by developing an approximation in which the problem is decomposed into a series of single runs.

1. Introduction and review of previous work

In a recent paper, Barad and Braha [1] have considered a single production run modeling of discrete manufacturing systems with binomial yields and a make-to-order demand for non-defective units of a finished product. The problem was described as an N-stage serial system, representing a discrete and extended version of the formulation treated by Lee and Yano [2] and also Wein [3]. For a literature review of lot sizing with random yields please refer to Yano and Lee [4]. It was assumed that the demand was to be supplied at the end of a single production run, when a linear cost (for shortage or excess) was eventually incurred. Inspection stations, performing a perfectly reliable 100% inspection of the output, preceded each manufacturing stage. After ascertaining the available yield from the previous stage, a decision had to be made regarding the number of units to be processed at the next stage. The decision formulation allowed for three alternatives: (A_1) processing all the non-defective units available from the previous stage; (A_2) reducing the available input; and (A_3) increasing the available input by procuring additional units (through purchasing or rework). Alternatives A_1 and A_2 incurred additional respective costs per unit removed or procured, while unavailability to procure additional units at a given stage was modeled by a very large procuring price. It was shown that the optimal policy is determined at each stage i by two control limits (L_i lower limit and M_i upper limit) as follows: select A_1 when the available quantity is within the ith stage control limits; select A_2 when it is above the upper limit M_i; or select A_3 when it is below the lower limit L_i. The above optimal policy was derived based on establishing the 'discrete convexity' of the related expected cost functions.

Let us now consider the possibility that, if at the end of a production run the number of final, non-defective units falls short of the given demand, additional production runs may be carried out. Three important factors, namely time, cost and what might be called the 'customer's voice' are likely to influence the final decision.

Time appears here in terms of the delivery due date as compared to the total processing time of a production run. An additional production run may not be practical unless the length of the granted delivery period is at least three to four times that of the expected duration of a single production run.

In addition, to be economic to the producer, the decision to launch a new run has to consider the necessary set-up costs. These costs, though substantial, are typically lower than the initial preparatory costs [5]. They stem from reordering of parts and raw materials as well as from additional efforts devoted to equipment preparation. When viewed through a longer term economic perspective, a 'loud' customer's voice (reflected in pressure exerted by powerful customers that insist on having the entire order delivered) may play a highly significant role in the producer's decision on whether or not to launch a new production run. In other words, if it is imperative to supply the demand, short-term economics as reflected in a set-up cost may be over-ruled.

0740-817X © 1999 "IIE"

Purchasing or reworking activities at intermediate stages within a single production run, that are likely to be necessary in order to lower the probability of shortages, may pose some problems due to timeliness, in process storage space and other operational constraints. Hence, waiting for rework to be completed or for externally purchased units to be delivered, may not always be feasible. Several authors [5–8] have considered additional production runs to satisfy a *rigid* make-to-order demand (i.e., demand needs to be satisfied in full) for the *single* stage problem. In most of these papers it is assumed that the number of non-defective units in a batch obeys a binomial distribution. Bowman and Fetter [6] have considered set-up costs for each additional production run. Goode and Saltzman [9] permitted an infinite number of production runs (each incurring a set up) and solved the problem through a dynamic oriented algorithm. Hillier [10] used differences to find optimal solutions to problems whose expected costs could be represented in terms of convex functions. White [8] presented an approach where production runs were carried out until eventually, the make-to-order demand was satisfied. Grosfeld-Nir and Gerchak [11,12] have considered a stochastically proportional yield distribution version of the above approach and have shown that optimal solutions can be obtained through an inverse transform. Klein [7] used a cyclic Markov chain with an absorbing state to formulate the multi-production run problem. This state could be reached by either satisfying the demand or by completing the last allowable run. Another Markov based decision model was presented by Beja [13], who used an objective function that included both the unit manufacturing costs and also the set-up costs. In his formulation the decision was whether to produce another unit, or to inspect the already produced units and, if necessary, to initiate an additional production run incurring a set-up cost. Sepheri *et al.* [5] generalized the set-up modeling by considering both a limited and also an infinite number of allowable production runs. The problem was formulated in terms of dynamic programming.

An important class of problems in manufacturing management are *multistage* (serial) scenarios with random yields at each stage and a rigid demand. An early approach (and probably the first attempt) to the multistage lot sizing problem with rigid demand was addressed by Vachani [14]. These problems are only now being reexamined [3,15].

In this paper, the problem of ordering *multiple* production runs to meet a single made-to-order *non-rigid* demand, where the various stages in the production process have *binomial yields*, is addressed. While the single-run problem was recently solved by the author [1]. It is probably unrealistic to expect to find such exact analytical solutions for the more complex model of multiple production-runs. In addition, brute force numerical techniques are probably unsatisfactory because of the vast number of possible solutions. Consequently, a decomposition method is described which is based on the previous single-run analytical results. This increases the validity of the approach. The method is developed by representing the multiple production run problem as a series of single-run problems. A 'first version' of the problem, that does not specifically consider set-up costs (to launch additional production runs) is formulated. It is then shown that the first version can be straightforwardly modified to accommodate a set-up cost; the suggested procedure is altered accordingly and presented as a 'modified version'.

The paper is organized as follows: in the first section, the first version of the M production run problem (without set-up costs) is formulated, to be followed by the decomposition approach and an algorithmic solution procedure. Then, the modified version that considers set-up costs is described and a numerical example is also presented. Concluding remarks appear in the final section.

2. Model formulation (first version): no set-up costs for an additional production run

A make-to-order demand of non-defective finished units, whose manufacturing process is modeled as an N-stage serial system is given, and a limited number of production runs may be carried out to satisfy the demand. After performing perfectly reliable, 100% inspection output from the previous stage and discarding all the defective units, a decision has to be made regarding the number of units to be processed at the next stage.

2.1. *Basic assumptions and notation*

(1) $D =$ the given demand;
(2) $i =$ stage index chosen to represent the number of remaining stages till the end of the manufacturing process, $i = 1, 2, \ldots, N$;
(3) The non-defective units produced at any stage i, $i = 1, 2, \ldots, N$, are independently generated with an equal probability p_i, a stage dependent parameter, $0 \leq p_i \leq 1$;
(4) There is an unlimited available supply of raw material (to be used as input to the first manufacturing stage, N). Hence y_i – the available input (non-defective units) to stage i, is defined as follows:

for $i = 1, 2, \ldots, N-1$, y_i is equal to the yield (non-defective units) of stage $i+1$;
for $i = N \quad y_i = \infty$;

(5) $U_i =$ the decision variable at stage i, representing the input to the ith stage;
(6) $P_i(x \mid U_i) =$ the probability of a yield x arising from an input to the ith stage of U_i (on basic assumption 3 above, this probability represents the binomial probability law);

(7) w_i = the manufacturing costs per unit (including inspection) at stage i;
(8) h_i = the disposal cost per available non-defective unit at stage i;
(9) r_i = the procurement cost per non-defective semi-finished unit at stage i;
(10) h_0 = the average cost incurred by producing one finished unit above the demand (may be incurred at the end of any production run);
(11) π = the shortage penalty cost per unit of unsatisfied demand (may be incurred at the end of the last permitted production run);
(12) M = the maximum number of production runs that may be carried out to satisfy the demand, $M \geq 1$;
(13) t = production run index chosen to represent the number of remaining production runs allowed, $t = 0, 1, \ldots, M$;
(14) x_1^t = the number of non-defective units produced in production run t, $t \geq 1$;
(15) At the end of a production run $t+1$, $t = 0, 1, \ldots, M-1$, the unsatisfied demand is updated, to eventually set-up the target demand of the next production run t, if $t \geq 1$, or to estimate the shortage costs (before terminating the process), if $t = 0$. Hence, D^t = the unsatisfied demand at the start of production run t (for $t \geq 1$) or at process termination (for $t = 0$) is defined as follows:

$$D^t = \begin{cases} \max\{(D^{t+1} - x_1^{t+1}), 0\} & \text{for } t = 0, 1, \ldots, M-1; \\ D & \text{for } t = M. \end{cases} \tag{1}$$

2.2. *The decision sets*

Given y_i, $i = 1, 2, \ldots, N$ and D^t, $t = 1, 2, \ldots, M$, as respectively defined by basic assumptions (4) and (15), two decision sets can be established as follows:

Decision set 1: For any production run t, $t = 1, 2, \ldots, M$ and manufacturing stage i, $i = 1, 2, \ldots, N-1$, there are three decision alternatives as considered in Barad and Braha [1], namely;

- A_1: process all the non-defective units available from the previous stage;
- A_2: reduce the available input;
- A_3: increase the available input by purchasing (reworking) units.

Decision set 2: For manufacturing stage N, representing the beginning of a production run t, $t = 1, 2, \ldots, M$, on basic assumption (4) there is a unique alternative, namely;

- A_4: process a batch of optimal size, to accommodate the given updated unsatisfied demand. (Implicitly, if the unsatisfied demand is zero, manufacturing should be discontinued.) The procedure is pictured in Fig. 1.

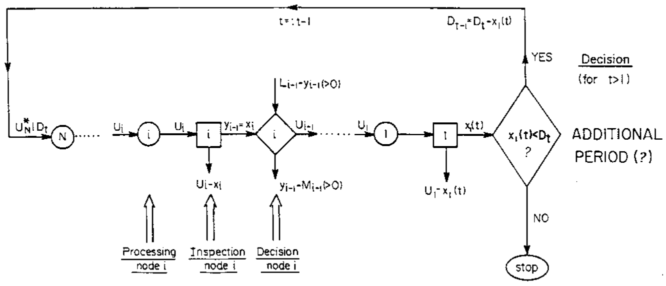

Fig. 1. The general topology of the multi-stage multi-production run system.

2.3. *The cost functions*

Similarly to the approach of Barad and Braha [1] let us now introduce two inter-related functions: $C_i^t(y_i, D^t)$ which is the minimal total expected cost given that the available input to stage i in production run t is y_i, the current unsatisfied demand is D^t, and an optimal policy is carried out in the subsequent stages $i-1, i-2, \ldots, 1$ in production run t, and within each of the remaining production runs, $t-1, t-2, \ldots, 1$.

$F_i^t(U_i, D^t)$ = the minimal total expected cost, given that the input to the ith stage is U_i, the current unsatisfied demand is D^t, and an optimal policy is carried out in the subsequent stages within production run t, as well as within each of the remaining production runs, $t-1, t-2, \ldots, 1$. For each t, $t = 1, 2, \ldots, M$, the connection between these two functions, as prescribed by our decision-modeling, is expressed by the following equations: $C_i^t(y_i, D^t)$ is defined for $i = 1, 2, \ldots, N-1$, as related to the three alternatives in decision set 1:

$$C_i^t(y_i, D^t) = \min\{F_i^t(U_i, D^t), \min_{j<y_i}[F_i^t(y_i - j, D^t) + jh_i], \min_{k>0}[F_i^t(y_i + k, D^t) + kr_i]\}. \tag{2}$$

$C_N^t(y_N, D^t)$ is defined for $i = N$, as related to the unique alternative in decision set 2:

$$C_N^t(y_N, D^t) = \min_{U_N} F_N^t(U_N, D^t) = F_N^t(U_N^*, D^t). \tag{3}$$

In Equations (2) and (3), $F_i^t(U_i, D^t)$ for $i > 1$ is given by:

$$F_i^t(U_i, D^t) = w_i U_i + \sum_{x=0}^{U_i} C_{i-1}^t(x, D^t) P_i(x \mid U_i).$$

$$F_1^t(U_1, D^t) = w_1 U_1 + \sum_{x_1^t < D^t} F_N^{t-1}(U_N^*, D^t - x_1^t) P_1(x_1^t \mid U_1) + h_0 \sum_{x_1^t \geq D^t} (x_1^t - D^t) P_1(x_1^t \mid U_1). \tag{4}$$

We define:

$$F_N^0(U_N^*, D^0) = \pi D^0$$
$$\text{(in Equation (1)} \quad D^0 = \max(D^1 - x_1^1, 0)). \tag{5}$$

Explanation of Equations (4) and (5) for $i = 1$: If $x_1^t < D^t$ and $t > 1$, an additional run $(t-1)$ will be carried out, with an optimal batch size U_N^*, for satisfying the updated demand D^{t-1}. If $t = 1$, there will be no more runs $(t - 1 = 0)$, a penalty π will be incurred per unit of unsatisfied demand by the end of stage 1 (the last stage according to the notation here). In this case, the single production run model considered in Barad and Braha [1] will be obtained.

Our objective here is to determine: (1) $C_N^M(y_N, D)$, the minimal total expected cost for supplying the demand D [which by Equation (3) is equivalent to $F_N^M(U_N^*, D)$]; and (2) the optimal policy at any stage i, $i = 1, 2, \ldots, N-1$ within each production run.

2.4. *The structure of the optimal solution for a single production run* $(M = 1)$

It was shown in Barad and Braha, [1] that by using a dynamic programming formulation, the structure of the optimal policy can be described as follows:

Theorem 1. *For the single production run case* (*with known demand*, D), *we may define* $F_i^1(U_i, D) \equiv F_i(U_i)$.

Let also $(\Delta F_i(U)$ defined as $F_i(U+1) - F_i(U)$. Then:

1. $F_i(U_i)$ is a discrete convex function (i.e., $\Delta F_i(U)$ is a monotone increasing) possessing a unique minimum value at $U_i = U_i^*$.
2. Every stage possesses critical numbers L_i, M_i satisfying the following properties:
   2.1 $L_i \le U_i^* \le M_i$;
   2.2 U_i^* is the lowest U_i for which $\Delta F_i(U_i) \ge 0$;
   2.3 L_i is the lowest U_i for which $\Delta F_i(L_i) \ge -r_i$;
   2.4 M_i is the lowest U_i for which $\Delta F_i(M_i) \ge h_i$;
3. The optimal input, U_i^o, is determined by the optimal policy:

$$U_i^o = \begin{cases} L_i & \text{for } y_i \le L_i, \\ y_i & \text{for } L_i \le y_i \le M_i, \\ M_i & \text{for } y_i \ge M_i. \end{cases}$$

3. Multiple production runs: the decomposition approach

Based on the single production run formulation presented in Barad and Braha [1] it is clear that if $F_i^t(U_i, D^t)$ were a discrete convex function for any t, $t = 1, 2, \ldots, M$, an optimal policy expressed in terms of control limits (two critical numbers), with a structure similar to the one formulated there, could have been determined. However, it has been shown by Braha [16] that for $t > 1$ non-convex functions may be obtained. Hence, an approximate scheme, based on a decomposition approach is proposed.

The basic reasoning behind our decomposition approach is introduced in Lemma 1, which is followed by our suggested decomposition approximation. Lemma 1 and the decomposition approximation are then used to formulate the proposed algorithmic procedure.

Lemma 1. *For every* $D^t > 0, t \ge 1, D^t = D_1^t + D_2^t$,

$$F_N^t(U_N^*, D^t) \le F_N^t(U_N^*, D_1^t) + F_N^t(U_N^*, D_2^t), \tag{6}$$

For $D^t = 0$ $F_N^t(U_N^*, D^t) = 0$.

Proof. The rationale of this statement is that the RHS of the above inequality is subjected to more constraints (each demand, D_1^t and D_2^t, has to be separately satisfied). Hence, under constant marginal probability for producing a non-defective unit, the probability of not fulfilling both constraints may not be lower than that of not fulfilling a single constraint (as in the LHS). The expected penalties for the respective occurrence of these events, as expressed above, follow the same rationale. ■

Applying Lemma 1 multiple times, as necessary to decompose the term $F_N^t(U_N^*, D^t)$ into a sum of D^t equal terms, $F_N^t(U_N^*, 1)$, yields:

$$\begin{aligned} F_N^t(U_N^*, D^t) &\le F_N^t(U_N^*, 1) + F_N^t(U_N^*, D^t - 1) \\ &\le \cdots \le F_N^t(U_N^*, 1) D^t. \end{aligned} \tag{7}$$

Let us now assume that for any t, $t = 1, 2, \ldots M$, $F_N^t(U_N^*, D^t)$ is approximately linear in D^t and suggest an equation to express this, namely:

$$\tilde{F}_N^t(U_N^*, D^t) = [\alpha^t F_N^t(U_N^*, 1)] D^t, \tag{8}$$

where $\tilde{F}_N^t(U_N^*, D^t)$ is the approximated function and α^t is a constant, which on inequality (7) satisfies $0 \le \alpha^t \le 1$, for $t \ge 1$.

By substituting Equation (8) into Equation (4) (for $t-1$), an approximation for $F_1^t(U_1, D^t), t \ge 1$, can be derived as follows:

$$\begin{aligned} \tilde{F}_1^t(U_1, D^t) = w_1 U_1 &+ \alpha^{t-1} F_N^{t-1}(U_N^*, 1) \\ &\times \sum_{x_1^t < D^t} (D^t - x_1^t) P_1(x_1^t \mid U_1) \\ &+ h_0 \sum_{x_1^t \ge D^t} (x_1^t - D^t) P_1(x_1^t \mid U_1). \end{aligned} \tag{9}$$

Let:

$$\pi^t = \alpha^{t-1} F_N^{t-1}(U_N^*, 1) \text{ for } t > 1, \text{ with } \pi^1 = \pi. \tag{10}$$

Then π^t may be interpreted as follows: it is proportional to the optimal total expected operating cost to be incurred given that an optimal policy is carried out in the next period, $t-1$, $t-1 > 0$, per given unsatisfied demand of one non-defective finished unit (i.e. $D^{t-1} = 1$.)

The decomposition approximation is obtained by substituting Equation (10) into Equation (8):

$$\tilde{F}_N^t(U_N^*, D^t) = [\alpha^t F_N^t(U_N^*, 1)]D^t = \pi^{t+1}D^t \text{ for } t \geq 1. \quad (11)$$

By Equation (5), $F_N^0(U_N^*, D^0) = \pi D^0$.

Substituting Equation (10) in Equation (9) for $t > 1$ yields:

$$\tilde{F}_1^t(U_1, D^t) = w_1 U_1 + \pi^t \sum_{x_1^t < D^t} (D^t - x_1^t) P_1(x_1^t \mid U_1) + h_0 \sum_{x_1^t \geq D^t} (x_1^t - D^t) P_1(x_1^t \mid U_1). \quad (12)$$

Thus, an approximate cost function, similar to $F_1^t(U_1, D^1)$ (as defined in Equation (4) for $t = 1$, and proved in Barad and Braha [1] to be a discrete convex function), is obtained in Equation (12) for $t > 1$. Its parameters, D^t and π^t, are respectively equivalent to D and π in Barad and Braha [1], while the other parameters, and among them h_0, remain unchanged. Hence, the arguments used there are still valid, indicating that, for $t > 1$, $\tilde{F}_1^t(U_1, D^t)$ (the derived approximate expression for $F_1^t(U_1, D^t)$) is 'discrete convex' and that a similar structure of the critical number as in Barad and Braha [1] i.e., dependent on D^t, can be established for each stage within each production run t. For $t = 1$ no approximation is involved, since $F_1^1(U_1, D^1)$ is directly obtained from Equation (4).

4. The Algorithmic procedure

Based on the above results, it is now possible to formulate an algorithm for the solution of the M multiple run, N-stage problem, comprising of two phases. Phase I (off-line) is intended to estimate the parameters α^t and π^{t+1} for the remaining periods t, $t = 1, 2, \ldots, M-1$, while phase II (on-line) consists of the operating procedure to determine the 'optimal' control limits.

Since π^t (as defined in Equation (10)) can only be estimated backward step-by-step, during phase I (off-line) of the algorithm, the system is operated backward, run by run. (The only difference amongst the runs will be their different penalty per unit shortage at their respective end.) Thus, the first run to be carried out for estimation purposes, will be the last permitted one, namely, $t = 1$ using the notation here, with π^1 given and where α^1 and π^2 are to be estimated. The last off-line run will be the actual second run, $t = M-1$, with π^{M-2} now given, and through which α^{M-1} and π^M are to be estimated. To ensure consistency, for each production run, the same target demands are defined. Since it seems reasonable that the linearity coefficient, in any production run, might be sensitive to the size of the demand (D^t), it is suggested that α^t be calculated for t, $t = 1, 2, \ldots, M-1$, over all the values of D^t in the interval $\{1, D\}$ and then estimated using the average value as follows:

$$\alpha_{\text{ave}}^t = (1/D) \sum_{D^t=1}^{D} F_N^t(U_N^*, D^t)/[D^t F_N^t(U_N^*, 1)]. \quad (13)$$

Accordingly, from Equation (10),

$$\pi_{\text{ave}}^{t+1} = \alpha_{\text{ave}}^t F_N^t(U_N^*, 1). \quad (14)$$

Phase II (on-line) of the algorithm will present a forward, run by run, operating procedure of the system. In the first on-line run t, namely $t = M$ using the notation here, the objective will be to supply $D^M = D$ non-defective units of the final product with a penalty cost per unit of unsatisfied demand π^M, (as estimated in phase I of the algorithm). Hence, the problem will be defined as a single-run, N-stage system, with parameters $(D, \pi^M, h_0, w, h, r, P)$, the critical numbers being calculated as in Theorem 1 [1]. In the subsequent runs to be carried out on-line, the demand, as well as the penalty per shortage, will be updated as follows.

At the end of any run t, the unsatisfied demand will be calculated using Equation (1), to yield the target demand D^{t-1} for the beginning of the next run, $t-1$. Provided D^{t-1} and $t-1 > 0$, an additional production run $t-1$, will be carried out. Otherwise, production will be discontinued. The penalty per unit shortage π^{t-1}, will be updated with t, i.e., will assume the respective value (for a given t) as estimated in the off-line phase. The respective 'control limits' (critical numbers) will be calculated for each stage within each production run (as detailed in Theorem 1).

4.1. *The detailed procedures*

Input $N, M, D, \pi, h_0, w, h, r, P$.

Phase I: ***Parameters' estimation procedure***
(Off-line backward estimation of parameters α^t and π^{t+1})
begin
Step 1. $\pi^1 := \pi$
Step 2. For $t = 1$ to $M-1$ **do**
Step 3. For $j = 1$ to D **do**
begin
Step 4. For each production run calculate the optimal operating cost, $F_N^1(U_N^*, j)$, of an N-stage system and single production run, with parameters $(j, \pi^t, h_0, \underline{w}, \underline{h}, \underline{r}, \underline{P})$.
Step 5. $\alpha^t(j) = F_N^t(U_N^*, j)/[jF_N^t(U_N^*, 1)]$
end
Step 6. $\alpha_{\text{ave}}^t = (1/D)\sum_{j=1}^{D} \alpha^t(j)$
Step 7. $\pi_{\text{ave}}^{t+1} = \alpha_{\text{ave}}^t F_N^t(U_N^*, 1)$
end

Phase II: *Optimal decision procedure*
(On-line forward operation of the system using the off-line estimated parameters)

Input π^t_{ave} from phase I, for $t = 2, 3, \ldots, M$ $(\pi^1 := \pi)$
begin
Step 1. For $t = M$ to 1 do
Step 2. While $D^t > 0$ and $t \geq 1$ do
begin
Step 3. For each production run with parameters $(D^t, \pi^t, h_0, \underline{w}, \underline{h}, \underline{r}, \underline{P})$, calculate the control limits L^t_i and M^t_i (as detailed in Theorem 1), for each stage $i = 1, 2, \ldots, N-1$.
Step 4. Start the tth production run with $U^{t,*}_N$ (the optimal batch U^*_N as detailed in Theorem 1).
Step 5. Update the unsatisfied demand by the end of production run t, D^{t-1}, using Equation (1).
end
Step 6. Terminate the manufacturing process.
end

5. Model formulation and algorithmic procedure (modified version): a set-up cost is incurred for each additional production run

As mentioned in the Introduction, the current decomposition is also compatible with a variation of the model described above. Instead of assuming that a fixed maximal number of production runs (M) has to be carried out (while the updated unsatisfied demand is positive), given that a set-up cost is involved, an economic decision is made at the end of each production run about whether or not it is worthwhile to start a new one.

Basic assumption (to be added to the decision sets on Section 2.2.)

A set-up cost, $A, A > 0$, is necessary in order to start an *additional* production run t, $1 \leq t \leq M-1$.
The above assumption will not affect decision set 1 (for manufacturing stage i, $i = 1, 2, ...N-1$), but only decision set 2, that will be amended as follows.

Decision set 2 (adjusted to incorporate the set-up cost A)
For manufacturing stage i, $i = N$, representing the beginning of a production run t, $t = 1, 2, ..., M-1$, there are two alternatives, namely:

A4: process a batch of optimal size (on basic assumption (5)) to accommodate the given updated unsatisfied demand;
A5: Discontinue the manufacturing process.

For $i = N$ and the first production run $t = M$, no set-up cost is involved. Thus only alternative A4 is considered.

For production runs $t = 1, 2, ...M-1$, we will now consider the two alternatives, i.e. carry out an additional run with set-up cost A, or terminate the process. Thus, Equation (3) is modified as follows:

$$C^t_N(y_N, D^t) = \min\{[F^t_N(U^*_N, D^t) + A], \pi D^t\}. \quad (15)$$

Applying the decomposition approximation (Equation (11)) to $F^t_N(U^*_N, D^t)$ yields:

$$F^t_N(U^*_N, D^t) \approx \pi^{t+1} D^t \text{ for } t \geq 1. \quad (16)$$

Hence, by substituting Equation (16) into Equation (15), the condition for an economic justification of an additional production run will be:

Produce an additional run if: $\pi^{t+1}D^t + A < \pi D^t$
This, in turn, can also be expressed as: $D^t > A/(\pi - \pi^{t+1})$ for $t = 1, \ldots, M-1$.

Phase I: ***Parameters' estimation procedure***
To accommodate this version, no changes are needed in phase I (off-line) of the algorithmic procedure described above. Phase II (on-line) of the algorithmic procedure, has to be slightly amended as detailed below.

Phase II: ***Optimal decision procedure*** (adjusted to incorporate the set-up cost A)
Input π^t_{ave} from phase I for $t = 2, 3, \ldots M$ $(\pi^1 := \pi)$
begin
Step 1. For $t = M$ to 1 do
Step 2. While $\{t = M\}$ **OR** $\{[D^t > A/(\pi - \pi^{t+1})]$ **AND** $[t \geq 1]\}$ do
begin
Step 3. For each production run with parameters $(D^t, \pi^t, h_0, \underline{w}, \underline{h}, \underline{r}, \underline{P})$, calculate the control limits L^t_i and M^t_i (as detailed in Theorem 1), for each stage $i = 1, 2, \ldots, N-1$.
Step 4. Start the tth production run with $U^{t,*}_N$ (the optimal batch size as detailed in Theorem 1)
Step 5. Update the unsatisfied demand by the end of production run t, D^{t-1}, using Equation (1)
end
Step 6. Terminate the manufacturing process.
end

6. A numerical example

Let $D = 40$, $N = 4$, $M = 3$, $h_0 = 20$ and the stage parameters listed in Table 1.
Let the procurement cost be $r = \{27, 19, 9, 1\}$ and assume the shortage penalty $\pi = 52$. Applying the off-line backward estimation procedure (phase I), yields, the values shown in Table 2.

When there are no set-up costs for starting an additional production run $(A = 0)$, forward operation of the four-stage system with $M = 3$ and $D = 40$ (according to Phase II of the algorithm), results in a total expected optimal cost of 1302.54. Assuming the same demand and the other conditions as detailed above, when a set-up cost is incurred, the expected costs of operating the system for varying numbers of runs M are listed in Table 3.

As logically anticipated, when $A = 0$ the total expected cost decreases with M. The above economic benefits resulting from an additional run may now be compared to a set-up cost (here $A = 30$ and $A = 50$) as incurred at the beginning of each additional run. It is seen that for $A = 0$ the 'optimal' number of production runs is the maximal number of runs allowed, i.e., $M^* = 3$, while for $A = 30$ and $A = 50$ it is $M^* = 2$ and $M^* = 1$, respectively.

Let us now examine the effect of procuring (versus not procuring) non-defectives at intermediate stages.

Calculations are performed for three shortage penalties, namely $\pi = 52, 100, 150$ and two sets of procurement costs (r). Set 1 obeys the conditions ensuring positive values of the lower control limits while the second set comprises high procurement costs (at all intermediate stages) resulting in zero valued lower control limits and thus prohibiting procurement. The results are presented in Tables 4 and 5.

Table 4 shows the shortage penalties π^t-(as estimated here) at the end of a given number of remaining production runs t, for $t = 1, 2, 3$, $\pi^1 = \pi$ for all combinations of the two sets of procurement costs and the three shortage penalty costs. When expressed in terms of a ratio of the final shortage penalty, it is seen (as expected) that these ratios are asymptotic with the number of remaining runs (t) for all cases. Larger reductions (smaller ratios) are exhibited by set 1 (procurement encouraged) as compared to set 2 (procurement prohibited). The same effect is shown by increasing π, the final shortage penalty.

In Table 5, the reduction in the total expected operating cost stemming from carrying out one additional production run is examined (the set up for the additional run is disregarded). The results are expressed as a percentage of the total expected operating cost of carrying out a single run. As might have been expected, it is seen that the reduction is higher for a higher shortage penalty. However, this reduction is slightly affected by active procurement at intermediate stages. This slight difference is decreasing with the shortage penalty level. At $\pi = 150$ very similar results were obtained for the two sets of procurement costs.

Table 1. Stage parameters

Stage i	w_i	h_i	P_i
1	2	2	0.8
2	2	2	0.8
3	6	2	0.8
4	6	2	0.8

Table 2. The off-line backward estimation procedure (phase 1)

t	$F_N^t(U_N^*, 1)$	α_{ave}^t	π_{ave}^t
1	40.65	0.92	52.00 (π)
2	35.54	0.97	37.40 ($= 0.92 \times 40.65$)
3			34.47 ($= 0.97 \times 35.54$)

Table 3. The expected costs of operating the system for varying numbers of runs

M (number of runs)	*Expected total 'optimal' cost*		
	$A = 0$	$A = 30$	$A = 50$
1	1364.12	1364.12	1364.12*
2	1320.24	1350.24*	1370.24
3	1302.54*	1362.54	

7. Concluding remarks

There are a number of ways for dealing with complex manufacturing systems in the literature. One approach is to consider relatively simple models such as a continuous-time fluid model of serial machines as in Perkins and Kumar [17]. The other approach is along the line of the so called hierarchical decomposition control approach; see details in Sethi and Zhang, [18].

In this paper we considered a decomposition approximation to approach the discrete, multi-production run, multi-stage systems with binomial yield. The decomposition approach enabled to accommodate two modeling variations of the system (with and without a set-up cost for starting an additional production run).

This paper contributes to resolving and understanding the problem of multistage scenarios with random yields at each stage and multiple production runs in several respects: (1) it provides the first attempt (to the author's knowledge) to model and solve a complex multistage system with more realistic and in-depth decision making criteria; (2) since the single-run problem can be analyzed

Table 4. The shortage penalties $\pi^t (t = 1, 2, 3)$ for the two sets of procuring costs and varying final shortage penalties

Procurement costs	*Shortage penalties* $\pi = 52$	$\pi = 100$	$\pi = 150$
Set 1: $r = (27, 19, 9, 1)$	$\pi^t = (52, 37.40, 34.47)$	$\pi^t = (100, 47.00, 36.40)$	$\pi^t = (150, 57.00, 38.40)$
(Strictly positive lower control limits)	$\pi^t/\pi = (1, 0.72, 0.66)$	$\pi^t/\pi = (1, 0.47, 0.36)$	$\pi^t/\pi = (1, 0.38, 0.26)$
Set 2: $r = (200, 200, 200, 200)$	$\pi^t = (52, 43.80, 38.96)$	$\pi^t = (100, 64.42, 51.13)$	$\pi^t = (150, 75.36, 55.63)$
(Zero valued lower control limits)	$\pi^t/\pi = (1, 0.84, 0.75)$	$\pi^t/\pi = (1, 0.64, 0.51)$	$\pi^t/\pi = (1, 0.50, 0.37)$

Table 5. The expected total operating costs with two runs, in percentage of the expected operating cost with a single run, for the two sets of procuring costs and varying final shortage penalties (the set-up is disregarded)

Procurement costs	*Shortage penalties* $\pi = 52$	$\pi = 100$	$\pi = 150$
Set 1: $r = (27, 19, 9, 1)$ (Strictly positive lower control limits)	96.78%	94.19%	93.16%
Set 2: $r = (200, 200, 200, 200)$ (Zero valued lower control limits)	97.59%	95.23%	93.37%

in more complex settings [1], the suggested decomposition approach, which uses the single-run problems as 'building blocks,' can also be incorporated to resolve the more complex multiple production run settings; (3) the decomposition approach is based on logical, sound reasoning and thus contributes to the advancement of yield management theory; and (4) the approach is also easy to implement and can thus be utilized to resolve practical industry problems.

The paper can be extended to include the following related issues:

- *Quality Control aspects*

(1) Inspection location stations and inspection reliability

In the current work we assumed that preceding each station, a perfectly reliable inspection station was located. These restrictions may be removed and the procedure may be modified to handle situations connected with locating the inspection stations, or coping with unreliable inspection. Optimal decisions should consider trade-offs between the cost of carrying out the inspection and the critical numbers (i.e., the 'control limits' imposed on the system).

(2) Quality improvement and yield modeling

The economic implications of improving the quality of the product at each stage, may be taken into consideration as another possible enhancement of the model. Such possibility may be especially fruitful in wafer fabrication processes, where some authors reported a U-shaped distribution of the yield [3] that may reasonably be a result of mixed batches of varying quality levels.

- *System complexity*

Treating the system as a serial one excluded some manufacturing structures, frequently encountered in industry, such as assembly lines and possibly more complex systems i.e., assembly coupled with serial lines. Hence, it might be desirable to enhance the decomposition approach as developed here, to accommodate such complex systems, by decomposing them into simpler components.

Acknowledgments

Thanks are due to Professor M. Barad of Tel-Aviv University for useful suggestions.

References

[1] Barad, M. and Braha D. (1996) Control limits for multi-stage manufacturing processes with binomial yield (single and multiple production runs). *Journal of the Operational Research Society*, **47**, 98–112.

[2] Lee, H.L. and Yano, C.A. (1988) Production control for multi-stage systems with variable yield losses. *Operations Research*, **36**, 269–278.

[3] Wein, A.S. (1992) Random yield, rework and scrap in a multi-stage batch manufacturing environment. *Operations Research*, **40**, 551–563.

[4] Yano, C.A. and Lee, H.L. (1995) Lot sizing with random yields: a review. *Operations Research*, **43**, 311–337.

[5] Sepehri, M., Silver, E.A. and New, C. (1986) A heuristic for multiple lot sizing for an order under variable yield. *IIE Transactions*, **18**, 63–69.

[6] Bowman, E.H. and Fetter, R.B. (1960) *Analysis for Production Management*, Irwin Inc., Homewood, Il, pp. 324–330.

[7] Klein, M. (1966) Markovian decision models for reject allowance problems. *Management Science*, **12**, 349–358.

[8] White, S. (1965) Dynamic programming and systems of uncertain duration. *Management Science*, **12**, 37–67.

[9] Goode, H.P. and Saltzman, S. (1961) Computing optimum shrinkage allowances of small order sizes. *Journal of Industrial Engineering*, **12**(1), 57–66.

[10] Hillier, F.S. (1963) Reject allowances for job lot orders. *The Journal of Industrial Engineering*, **14**, 311–316.

[11] Grosfeld Nir, A. and Gerchak, Y. (1990) Multiple lot-sizing with random common-cause yield and rigid demand. *Operations Research Letters*, **9**, 383–388.

[12] Grosfeld Nir, A. and Gerchak, Y. (1996) Production to order with random yields: single-stage multiple lot-sizing. *IIE Transactions*, **28**, 669–676.

[13] Beja, A. (1977) Optimal reject allowance with constant marginal production efficiency. *Naval Research Logistics Quarterly*, **24**, 21–33.

[14] Vachani, M. (1970) Determining optimum reject allowance for multistage job-lot manufacturing. *AIIE Transactions*, **2**, 70–77.

[15] Grosfeld Nir, A. and Gerchak, Y. (1992) Production to order with random yields: multiple lot sizing. Technical Report, Department of Management Sciences, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.

[16] Braha, D. (1995) Control limits for multi-stage processing systems with binomial yield. Technical Report, Department of Manufacturing Engineering, Boston University, Boston, MA02215.

[17] Perkins, J.R. and Kumar, P.R. (1998) Optimal control of pull manufacturing systems. *IEEE Transactions on Automatic Control*, (in press)

[18] Sethi, S.P. and Zhang, Q. (1994) *Hierarchical Decision Making in Stochastic Manufacturing Systems*, Birkhauser, Cambridge, MA.

Biography

Dan Braha received a Ph.D. in Industrial Engineering from Tel-Aviv University, Israel, in 1994. He served as a Post-Doctoral Fellow, Department of Manufacturing Engineering, Boston University, Boston MA, until September 1995. He is currently an Assistant Professor, Department of Industrial Engineering, Ben-Gurion University, Israel. His research within engineering design focuses on developing methods to help the designer move from the conceptual phase to the realization of the physical device. To achieve this objective, he has developed the formal design theory (FDT), which is a mathematical theory of design. He has developed methods to solve routine design with group technology, genetic algorithms, and simulated annealing, to define computerized architecture, structure, and databases for the conceptual design process, and to develop a framework of logic decomposition and case-based reasoning methodology for mechanical and manufacturing design. He also co-authored, with O. Maimon, a book on the foundations of engineering design. Additional interests include real-time scheduling of flexible manufacturing systems, inventory management, risk and decision analyses, and group technology. His papers have been published in several journals, including the *International Journal of Production Research, Journal of the Operational Research Society* and *IEEE Transactions on Systems, Man, and Cybernetics*.