Available online at www.sciencedirect.com

SCIENCE DIRECT®

European Journal of Operational Research 173 (2006) 600–616

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

www.elsevier.com/locate/ejor

Production, Manufacturing and Logistics

Lot-sizing two-echelon assembly systems with random yields and rigid demand

Abraham Grosfeld-Nir [a], Shoshana Anily [b,*], Tal Ben-Zvi [b]

[a] *Academic College of Tel-Aviv-Yaffo, 29 Melchet, Tel-Aviv 61560, Israel*
[b] *Faculty of Management, Tel Aviv University, Tel-Aviv 69978, Israel*

Received 1 September 2003; accepted 4 January 2005
Available online 25 March 2005

Abstract

We consider a two-echelon assembly system producing a single final product for which the demand is known. The first echelon consists of several parallel stages, whereas the second echelon consists of a single assembly stage. We assume that the yield at each stage is random and that demand needs to be satisfied in its entirety; thus, several production runs may be required. A production policy should specify, for each possible configuration of intermediate inventories, on which stage to produce next and the lot size to be processed. The objective is to minimize the expected total of setup and variable production costs.

We prove that the expected cost of any production policy can be calculated by solving a finite set of linear equations whose solution is unique. The result is general in that it applies to any yield distribution. We also develop efficient algorithms leading to heuristic solutions with high precision and, as an example, provide numerical results for binomial yields.

Keywords: Multiple lot sizing; Random yield; Rigid demand; Assembly system

1. Introduction

As manufacturing frequently requires the combined effort of several stages (machines), the efficient planning of production lots often becomes a crucial economic factor. The determination of production lots is particularly challenging when yields are random and demand needs to be satisfied in its entirety (i.e., is "rigid"). Then, several production runs may be necessary at some or all stages.

* Corresponding author. Tel.: +972 3 6408861; fax: +972 3 6409983.
E-mail addresses: agn@mta.ac.il (A. Grosfeld-Nir), anily@post.tau.ac.il (S. Anily), talbenz@post.tau.ac.il (T. Ben-Zvi).

0377-2217/$ - see front matter
doi:10.1016/j.ejor.2005.01.033

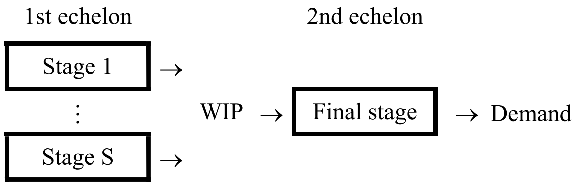

Fig. 1. A two-echelon assembly system.

We consider a two-echelon assembly system producing a single final product. The first echelon consists of at least two parallel stages and the second echelon consists of a single ("final") assembly stage (Fig. 1). We refer to output of the first echelon, waiting to be used by the assembly stage, as work-in-process (WIP).

We assume:

- Production at each stage is in lots involving fixed and variable processing costs.
- Production yield at each stage is random. Defective units, either WIP or final products, are worthless and must be scrapped.
- Demand for final goods is rigid.
- There is no value to excess WIP or conforming final products.

These assumptions fit environments where orders are for small quantities and products are custom-made. In particular, models leading to multiple production runs are realistic in environments where the variance of the yield rate is large. For more information see Grosfeld-Nir and Gerchak (2004).

For any outstanding demand level and any possible configuration of intermediate inventories (i.e., WIP), a production policy should specify the stage at which production takes place next and the lot size to be used at that stage. The objective is to minimize the expected total of setup and variable production costs.

Note that while our formulation below holds in general, i.e., for any yield distribution, our numerical study tests only the binomial distribution.

1.1. Literature review

The modeling of manufacturing systems with random yields attracted the attention of many researchers; see Yano and Lee (1995) for a literature review. Two variants of demand have been addressed in the literature: (i) "rigid demand", where an order must be satisfied in its entirety, possibly necessitating multiple production runs; see Grosfeld-Nir and Gerchak (2004) for a literature review of such models; and (ii) "non-rigid demand", where there is only one production run and a penalty for a shortage; see Yano and Lee (1995) and Section 9.4.8 of Zipkin (2000) for a description of such single-attempt scenarios.

The single stage with binomial yields and rigid demand has been analyzed since the mid-1950s, often under the label of "reject allowance" (Sepheri et al., 1986). For the binomial yield, Beja (1977) proved that the optimal lot size strictly increases in the demand and that the expected cost is quasi-convex in the lot size, which provides a convenient stopping rule in the quest for the optimal policy. Anily (1995) proved a similar result for the discrete uniform yield case. Grosfeld-Nir and Gerchak (1996) showed that the monotonicity of the lot size as a function of the demand does not hold for any yield distribution: in their study of some fundamental questions concerning the single stage, the authors provide examples where the optimal lot size may decrease in the demand and in the setup cost. Anily et al. (2002), analyzed the interrupted-geometric yield distribution, and proved that the optimal lot size never exceeds the demand and that an increase in the demand may induce a decrease in the optimal lot size.

Only few results are known concerning serial multi-stage systems with random yield and rigid demand. Grosfeld-Nir and Ronen (1993) and Grosfeld-Nir (1995) studied "single-bottleneck systems" (SBNS) which are described in Section 2.2. Wein (1992) assumed "perfect rework", i.e., if the number of usable units exiting a certain stage is insufficient, defective units can be perfectly reworked, thus necessitating at most two production runs at each stage. Grosfeld-Nir and Gerchak (2002) allow for unlimited reworks at each stage. Pentico (1994) analyzed a heuristic for the serial multi-stage system with binomial machines. He assumed that all usable units exiting a stage proceed to the next stage.

Grosfeld-Nir and Robinson (1995) studied a binomial two-stage system, providing an LP formulation for calculating the optimal policy. The LP formulation requires a large number of constraints; numerically they solved examples for demand equal one and two, and proposed a heuristic for a larger demand. Grosfeld-Nir (2005) studied "two-bottleneck systems" (2-BNS): multi-stage serial systems where setup costs for all but two stages are zero. The author proved the expected cost of any production policy can be evaluated by solving a finite set of linear equations. This result is important since it provides the means to compare the effectiveness of various policies and heuristics. In particular, one can search for the optimal policy by using a full enumeration procedure, i.e., by evaluation of all feasible policies within a subset of policies that contains an optimal solution. The author also developed a policy improvement algorithm for the two-stage system and reports that in all his testable experiments it reached the optimal solution.

In total, early studies prove that solving any system with more than one machine requires a considerable effort. As explained below, the two-stage problem is much harder and requires a specialized approach. An effective way to treat a serial system is yet unknown. In this context the two-stage assembly system is a natural avenue for research.

1.2. Summary of results

- We show that the expected cost associated with any given production policy, for the two-echelon assembly system can be evaluated by solving a finite set of linear equations, which has a unique solution.
- We develop modifications of algorithms, used for the two-stage system, to solve the more complex assembly system.
- We develop a new powerful "intermediate-demand-algorithm" (IDA) that is rapid and easy to implement.

2. Serial multi-stage systems

In this section we review several known results for serial multi-stage systems. Later we employ these results in the analysis of the two-echelon assembly system.

2.1. The single stage

We consider a production facility producing in lots, or runs, where the production of a lot N entails the cost $\alpha + \beta N$. We refer to α and β as setup and variable production costs, respectively. We denote $p(x, N)$, $0 \leqslant x \leqslant N$, the probability to obtain x conforming units from a lot of size N and define the following cost functions:

V_D is the optimal (minimal) expected cost required to fulfill an order of size $D \geqslant 1$.
$V_D(N)$ is the expected cost if a lot of size N is run whenever the demand is D and an optimal lot is processed whenever the remaining demand is less than D.

Therefore, $V_D = \min_N\{V_D(N)\} \equiv V_D(N_D)$; where $N_D \in \operatorname{argmin}\{V_D(N) : N \geqslant 1\}$. Without loss of generality we can assume that the salvage value is zero (see discussion in Anily et al. (2002)) and therefore we can define $V_D = 0$ for any $D \leqslant 0$. The following recursive formulation that was originally proposed by White (1965) holds for any yield distribution:

$$V_D(N) = \alpha + \beta N + p(0,N)V_D(N) + \sum_{x=1}^{D-1} p(x,N)V_{D-x}$$

or equivalently,

$$V_D(N) = \frac{\alpha + \beta N + \sum_{x=1}^{D-1} p(x,N)V_{D-x}}{1 - p(0,N)}. \tag{1}$$

Thus, the optimal lot size and the optimal expected cost can be calculated recursively in D, via a search over N. We refer to a stage as "binomial machine" if $p(x,N) = \binom{N}{x}\theta^x(1-\theta)^{N-x}$. We refer to a system with binomial machines only as a "binomial system".

2.2. A single-bottleneck system with binomial yields

Grosfeld-Nir and Ronen (1993) and Grosfeld-Nir (1995) refer to a stage with non-zero setup cost as a "bottleneck" (BN) and to a system with only one BN as a "single-bottleneck system" (SBNS). Similarly, a system with no BNs (all setups are zero) is referred to as a "zero-bottleneck system" (0-BNS). Typically, a SBNS consists of two 0-BNS and the BN. Fig. 2 demonstrates a binomial SBNS; θ_j is the success probability of machine j.

When a binomial 0-BNS faces a rigid demand D, it is optimal to process units one at a time until the demand is satisfied. The resulting expected cost is mD, where m, the minimal expected cost to satisfy a demand of one unit, is given by

$$m = \frac{\beta_1}{\theta_1 \ldots \theta_M} + \frac{\beta_2}{\theta_2 \ldots \theta_M} + \cdots + \frac{\beta_M}{\theta_M}. \tag{2}$$

When a binomial SBNS faces a rigid demand D, it is optimal to process units one at a time on the first 0-BNS until a certain batch size is ready to be processed on the BN. These units are then processed in one batch on the BN, and, finally, the usable units exiting the BN are processed, one at a time, on the second 0-BNS until the demand is satisfied or all units are exhausted. Therefore, the problem of optimally controlling a SBNS is completely characterized by the optimal lot to be processed by the BN.

We denote by m_1 and m_2 the expected cost of obtaining one good unit by the first and second 0-BNS, respectively (to be computed via (2)). Grosfeld-Nir (1995) proved that a binomial machine whose parameters are

$$\alpha = \alpha_k; \quad \beta = \beta_k + m_1; \quad \theta = \theta_k\theta_{k+1} \ldots \theta_M$$

is "equivalent" to the SBNS of Fig. 2. The equivalency is in that

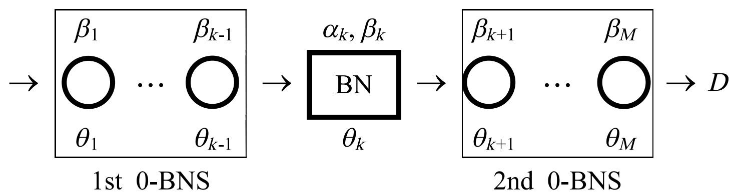

Fig. 2. A typical binomial SBNS.

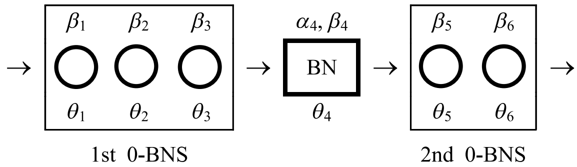

Fig. 3. A 6-stage binomial SBNS with a BN at the 4th stage.

(a) The optimal lot-size to enter the BN of a SBNS is the same as the optimal lot-size for the equivalent machine (for each D).
(b) The expected cost required to fulfill an order D by the SBNS equals the expected cost to fulfill the same order by the equivalent machine *plus* Dm_2.

For example, consider the 6-stage binomial SBNS of Fig. 3; the equivalent machine has parameters: $\alpha = \alpha_4$; $\beta = \beta_1/\theta_1\theta_2\theta_3 + \beta_2/\theta_2\theta_3 + \beta_3/\theta_3 + \beta_4$;$\theta = \theta_4\theta_5\theta_6$.

2.3. The two-stage system

In this subsection we consider a two-stage system consisting of two bottlenecks, M_1 and M_2 (Fig. 4). A production policy should specify, for each level of WIP, on which machine to produce next and the number of units to be processed.

The two-stage problem is hard to solve for the following reason (formal details can be found in Grosfeld-Nir (2005)): For each level of WIP it is necessary to decide whether to produce on M_1 or M_2. Production on M_1 may increase the WIP, thus the expected cost associated with such an action depends upon expected costs of higher levels of WIP, making it impossible to solve the problem recursively.

A "fixed policy" for the two-stage problem, as defined by Grosfeld-Nir (2005), is a specific production plan, i.e., for any level of WIP a fixed policy specifies on which machine to produce next and the lot-size to be processed. The author proves that the expected cost of such a policy can be calculated by solving a finite system of linear equations (below). Note that a fixed policy specifies the production plan for a specific D. In contrast, a "fixed production strategy" consists of a collection of fixed policies, for $D \geqslant 1$.

We wish to note that in his treatment of fixed policies Grosfeld-Nir (2005) does not address the issue of feasible and non-feasible production plans. For example a production policy instructing to always produce on $M1$ is non-feasible, as the demand will never be satisfied. On the other hand, any production policy using a control limit which instructs to produce on $M2$, whenever the WIP exceeds the control limit, is feasible. For obvious reasons we always require $p_i(0, N) < 1$, $i = 1, 2$.

Any feasible fixed policy can be defined by two disjoint sets Δ_D^1 and Δ_D^2 of WIP levels for which production takes place on M_1 or M_2, respectively. We denote by N_D (L) the lot size to be processed whenever the demand is D and the WIP is L. (The lot $N_D(L)$ is processed on M_1 if $L \in \Delta_D^1$ and on M_2 if $L \in \Delta_D^2$.) Note that Δ_D^1, Δ_D^2 and $N_D(L)$ completely define a fixed policy (see Example 1). Grosfeld-Nir (2005) implicitly assumed that the set Δ_D^1 is finite and proved the following:

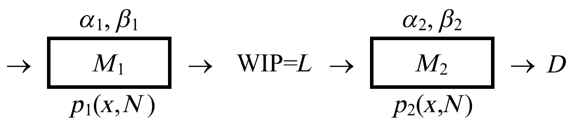

Fig. 4. A two-stage system.

Theorem 1. *Let $U_D(L)$ be the expected cost associated with a certain fixed policy whenever the demand is D and the WIP is $L \geqslant 0$, and suppose that $U_d(L)$, $L \geqslant 0$, are known for $0 < d < D$, and $U_d(L) = 0$ for $d \leqslant 0$ and any $L \geqslant 0$. Then,*

(a) *The expected costs, $U_D(L)$, $L \geqslant 0$, can be calculated via solving the finite set of linear equations defined in (3) and (4), below.*
(b) *The number of equations is $1 + \delta(D)$, where $\delta(D) \equiv \max\{L + N_D(L) : L \in \Delta_D^1\}$.*

Note that, starting with $L = 0$ (zero WIP), $\delta(D)$ is the maximum WIP level that can be reached. If production is on M_1 (i.e., $L \in \Delta_D^1$, and writing, for simplicity, N instead of $N_D(L)$), we have:

$$U_D(L) = \alpha_1 + \beta_1 N + \sum_{x=0}^{N} p_1(x, N) U_D(L + x). \tag{3}$$

If production is on M_2 (i.e., $L \in \Delta_D^2$), then:

$$U_D(L) = \alpha_2 + \beta_2 N + \sum_{x=0}^{N} p_2(x, N) U_{D-x}(L - N). \tag{4}$$

Note that these equations hold for any yield distribution.

Example 1. Consider the two-stage system of Fig. 4. The following is an example of a fixed policy for a given D, $D \geqslant 1$:

L	Action
0	Produce 4 units on M_1
1	Produce 1 unit on M_2
2	Produce 2 units on M_2
3	Produce 2 units on M_1
4	Produce 4 units on M_2
5	Produce 4 units on M_2

Note that a fixed policy is general in that it allows, for example, to produce on M_2 when $L = 2$ and on M_1 when $L = 3$. For this example, $\Delta_D^1 = \{0, 3\}$; $\Delta_D^2 = \{1, 2, 4, 5\}$; and $N_D(0) = 4$, $N_D(1) = 1$, etc'. Also, starting with $L = 0$, the maximum WIP level that can be reached is $\delta(D) = 5$. Thus, $U_D(0), \ldots, U_D$ (5) can be calculated by solving the following six linear equations:

$$U_D(0) = \alpha_1 + 4\beta_1 + \sum_{x=0}^{4} p_1(x, 4) U_D(x),$$

$$U_D(1) = \alpha_2 + \beta_2 + \sum_{x=0}^{1} p_2(x, 1) U_{D-x}(0),$$

$$U_D(2) = \alpha_2 + 2\beta_2 + \sum_{x=0}^{2} p_2(x, 2) U_{D-x}(0),$$

$$U_D(3) = \alpha_1 + 2\beta_1 + \sum_{x=0}^{2} p_1(x,2)U_D(3+x),$$

$$U_D(4) = \alpha_2 + 4\beta_2 + \sum_{x=0}^{4} p_2(x,4)U_{D-x}(0),$$

$$U_D(5) = \alpha_2 + 4\beta_2 + \sum_{x=0}^{4} p_2(x,4)U_{D-x}(1).$$

As a consequence of Theorem 1 we have

Corollary 1. *Given a fixed production strategy, i.e., a sequence of fixed policies for* $D = 1, 2, \ldots$ *The expected costs* $U_D(L)$, $L \geqslant 0$, *can be computed recursively, in* D, *by solving repeatedly the linear set of equations* (3) *and* (4) *starting with* $D = 1$, *then* $D = 2$ *and so on.*

While the linear equations (3) *and* (4), *above, were formulated in Grosfeld-Nir (2005), the author did not prove that these equations have a unique solution. We prove the uniqueness of the solution in Theorem 2 in Appendix A.*

Theorem 2. *For any given fixed policy, the set of* $\delta(D) + 1$ *linear equations defined by* (3) *and* (4) *has a unique solution.*

2.3.1. Algorithms for the two-stage system

We believe that any optimal policy for the two-stage system is such that the set Δ_D^1 is bounded. If this property holds true then we need to search the optimal policy among a finite number of distinct fixed policies. The most naive way for searching for the optimal policy over the set is by evaluating the expected cost of all fixed policies within the set. Clearly, this process is tedious as the number of fixed policies that need to be evaluated is large. If the respective expected cost function turns out to be quasi-convex in the lot size, then policy-improvement algorithms can be used.

Grosfeld-Nir (2005) considered the following plausible class of policies for the two-stage problem with binomial yields: For any demand level D, a control-limit C_D is associated, so that production takes place on M_2 if and only if WIP $\geqslant C_D$, i.e., $\Delta_D^1 = \{0, \ldots, C_D - 1\}$. The paper also proposes a Policy-Improvement Algorithm (PIA). With PIA both C_D and the production lots are decision variables. Clearly, PIA converges to the optimal policy if the expected cost functions are convex in the production lots for all levels of WIP. PIA was then combined with a *Fixed-Policy-Algorithm* (FPA) which is based on successive approximation and takes advantage of the special structure of (3) and (4). Grosfeld-Nir (2005) reported that in all the (binomial) two-stage problems he examined PIA reached the optimal policy.

In the sequel we used modifications of FPA and PIA to solve assembly systems. However, as even the simplest assembly system is significantly more complex than the two-stage system, computation time is prohibitive. This motivated us to develop a new method named the intermediate-demand-algorithm.

2.3.2. The intermediate-demand-algorithm (IDA)

We denote $N_D^{M_1}(N_D^{M_2})$ the optimal lot to be processed on $M_1(M_2)$, if it alone, as a single stage, faced a rigid demand D. These lots can be calculated by (1). The intermediate-demand-algorithm (IDA) is a heuristic that searches over a subset of control-limit policies. For demand D, IDA is defined in terms of a single decision variable K_D, playing the role of "intermediate demand": whenever production takes place on M_1 the lot size to be processed is determined as if it alone faced the rigid demand K_D. More precisely, let WIP $= L$, then IDA, with parameter K_D, is defined as follows:

Table 1
Numerical results for a binomial two-stage system with parameters: $\alpha_1 = 20$, $\beta_1 = 5$, $\theta_1 = 0.6$ and $\alpha_2 = 50$, $\beta_2 = 2$, $\theta_2 = 0.8$, as a function of D, when $L = 0$

D	$U_D(0)$		Gap (%)	$N_D^{M_1}(0)$		C_D		Time (seconds)	
	PIA	IDA		PIA	IDA	PIA	IDA	PIA	IDA
1	99.4	102.0	2.6	3	2	1	1	20	4
2	118.3	119.7	1.2	6	6	2	3	60	5
3	135.2	137.1	1.4	8	7	4	4	180	6
5	166.1	169.0	1.7	12	12	6	7	360	15
10	239.3	242.2	1.2	23	22	12	13	1260	60
15	311.8	313.0	0.4	34	32	16	19	2100	185
20	381.6	383.0	0.4	45	43	21	26	4320	450

- If $L \geqslant N_D^{M_2}$ a lot of size N_D^{M2} is processed on M_2.
- If $L < N_D^{M_2}$ and $L \geqslant K_D$ a lot of size L is processed on M_2.
- If $L < \min\{N_D^{M_2}, K_D\} \equiv C_D$ a lot of size $N_{K_D-L}^{M_1}$ is processed on M_1.

Therefore, IDA generates a control-limit policy, where production takes place on M_2 if and only if $L \geqslant C_D$. Also, IDA simplifies the computation as the lot sizes to be processed on M_1 and M_2 are obtained by solving single stage problems. IDA searches over all plausible values of K_D. Our numerical test shows that (i) PIA always produces a better solution than IDA; (ii) IDA is much faster and the gap between the solution produced by IDA and PIA is small.

Numerical results comparing PIA and IDA are exhibited in Table 1 ($N_D^{M_1}(0)$ is the lot size to be processed on M_1 whenever the demand is D and the WIP is zero; U_D (0) is the corresponding expected cost). Table 1 demonstrates that relative deviations from PIA decrease in D. As we are going to see, this remains true when solving assembly systems.

3. Two-echelon assembly systems

Referring to Fig. 5, we denote $\alpha_k + \beta_k N$ the cost of processing a lot of size N at stage k and by $p_k(x, N)$ the corresponding probability to obtain x conforming units out of this stage. We denote L_k the number of usable units that exited stage k, ready to enter the final stage. We assume in the sequel that one assembled unit requires one unit of each of the intermediate inventories from the first stage. General bill of material plans can be handled similarly.

3.1. Fixed policies for assembly systems

A fixed policy specifies for each possible configuration of given intermediate inventories, i.e., a vector $(L_1, L_2, \ldots, L_S)$, and an outstanding demand D, on which machine to produce next and the lot size to be processed. The objective is to minimize the expected total of setup and variable production costs required to satisfy the whole demand.

Similarly to the two-stage problem we define Δ_D^i to be the set of WIP levels for which production takes place on machine i, $i = 1, \ldots, S+1$, when the outstanding demand is D. Let $N_D(L_1, L_2, \ldots, L_S)$ denote the lot size to be processed whenever the demand is D and the WIP is $(L_1, L_2, \ldots, L_S)$, and $U_D(L_1, L_2, \ldots, L_S)$

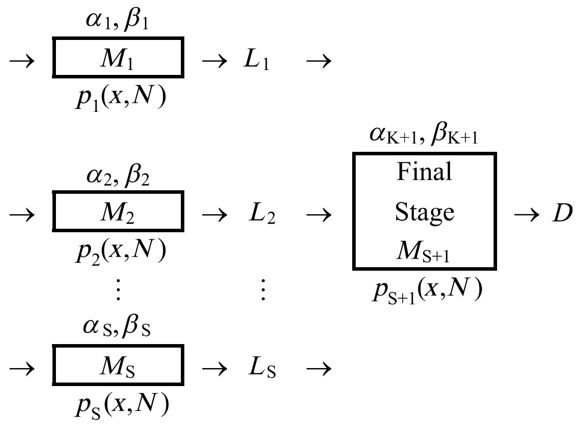

Fig. 5. A general two-echelon assembly system.

the corresponding expected cost. As for the two-stage serial systems, we consider only fixed policies for which the sets Δ_D^i for $i = 1, \ldots, S$ are bounded.

3.2. Calculating the expected cost of fixed policies

We refer to the most elementary assembly system, which consists of only two stages at the first echelon as *the basic assembly system* (Fig. 6).

Next we provide an example of a fixed policy, for $D = 1$ (Table 2), and write the corresponding set of linear equation.

Example 2. Note that $\Delta_D^1 = \{(0,0),(0,1)\}$, $\Delta_D^2 = \{(1,0),(2,0)\}$ and $\Delta_D^3 = \{(1,1),(2,1)\}$. Also, $N_D(0,0) = 2$, $N_D(0,1) = 2$, etc'. Thus, the expected costs, $U_1(L_1, L_2)$ can be calculated via solving the following set of six linear equations:

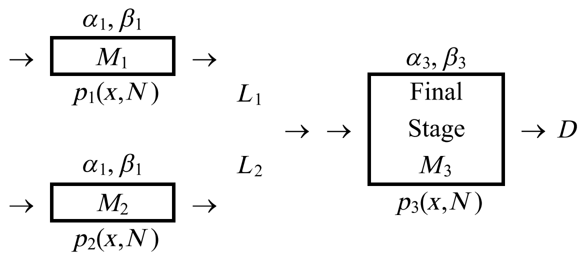

Fig. 6. The basic assembly system.

Table 2
A fixed policy for a basic assembly system with $D = 1$

	$L_2 = 0$	$L_2 = 1$
$L_1 = 0$	Produce 2 units on M_1	Produce 2 units on M_1
$L_1 = 1$	Produce 1 unit on M_2	Produce 1 unit on M_3
$L_1 = 2$	Produce 1 unit on M_2	Produce 1 unit on M_3

$$U_1(0,0) = \alpha_1 + 2\beta_1 + p_1(0,2)U_1(0,0) + p_1(1,2)U_1(1,0) + p_1(2,2)U_1(2,0),$$

$$U_1(0,1) = \alpha_1 + 2\beta_1 + p_1(0,2)U_1(0,1) + p_1(1,2)U_1(1,1) + p_1(2,2)U_1(2,1),$$

$$U_1(1,0) = \alpha_2 + \beta_2 + p_2(0,1)U_1(1,0) + p_2(1,1)U_1(1,1),$$

$$U_1(2,0) = \alpha_2 + \beta_2 + p_2(0,1)U_1(2,0) + p_2(1,1)U_1(2,1),$$

$$U_1(1,1) = \alpha_3 + \beta_3 + p_3(0,1)U_1(0,0),$$

$$U_1(2,1) = \alpha_3 + \beta_3 + p_3(0,1)U_1(1,0).$$

For any fixed policy for which the sets Δ_D^i for $i = 1, \ldots, S$ are bounded, the following S constants are defined: $\delta_i = \max\{L_i + N_D(L_1, \ldots, L_i, \ldots, L_S)$: $(L_1, \ldots, L_i, \ldots, L_S) \in \Delta_D^i\}$ for $i = 1, \ldots, S$. Note that δ_i is an upper bound on the maximum WIP of component i that can be realized by the given policy whenever production starts without any stock of WIP. Theorem 3 proves that the expected cost of any given fixed policy for the assembly system can be obtained by solving a finite set of linear equations.

Theorem 3. *Suppose we are given a fixed policy for an assembly system with demand D, such that Δ_D^i are bounded for $i = 1, \ldots, S$. In addition, suppose that the expected costs $U_d(L_1, \ldots, L_S)$, $(L_1, \ldots, L_S) \geqslant (0, \ldots, 0)$ are known for $d = 1, 2, \ldots, D-1$. Then the expected costs, $U_D(L_1, \ldots, L_S)$, $L_i \geqslant 0$ for $i = 1, \ldots, S$, can be calculated by solving a finite system of $\prod_{i=1}^{S}(1 + \delta_i)$ linear equations. Moreover, this system of equations is guaranteed to have a single solution.*

Proof. In order to prove the theorem we present the system of linear equations for solving $U_D(L_1, L_2, \ldots, L_S)$, $0 \leqslant L_i \leqslant \delta_i$, For simplicity, we write N instead of $N_D(L_1, \ldots, L_S)$.

If $(L_1, \ldots, L_S) \in \Delta_D^i$ for $i \in \{1, \ldots, S\}$ then the equations are of the following form:

$$U_D(L_1, L_2, \ldots, L_S) = \frac{\alpha_i + \beta_i N + \sum_{x=1}^{N} p_i(x, N) U_D(L_1, L_2, \ldots, L_{i-1}, L_i + x, L_{i+1}, \ldots, L_S)}{1 - p_i(0, N)}. \tag{5}$$

Otherwise, i.e., if $(L_1, L_2, \ldots, L_S) \in \Delta_D^{S+1}$, the equation is as follows:

$$U_D(L_1, L_2, \ldots, L_S) = \alpha_{S+1} + \beta_{S+1} N + \sum_{x=0}^{D-1} p_{S+1}(x, N) U_{D-x}(L_1 - N, L_2 - N, \ldots, L_S - N), \tag{6}$$

where $U_d(L_1, \ldots, L_S) = 0$, if $d \leqslant 0$. The number of equations is determined by the number of possible WIP levels. The uniqueness of the solution is proved similarly to the proof of Theorem 2 and is, therefore, omitted. □

In the next proposition we prove two basic properties of the expected cost of an optimal policy. Let $F_D(L_1, \ldots, L_S)$ denote the optimal (minimal) expected cost required to satisfy the demand D when the WIP $= (L_1, \ldots, L_S)$. Note that $F_D(L_1, \ldots, L_S)$ is different from $U_D(L_1, \ldots, L_S)$, as the latter corresponds to a fixed policy which may not be optimal.

Proposition 1

(a) *The optimal expected cost $F_D(L_1, \ldots, L_S)$ is non-decreasing in D for given $(L_1, \ldots, L_S)$.*
(b) *The optimal expected cost, $F_D(L_1, \ldots, L_S)$, is non-increasing in the WIP levels, $(L_1, \ldots, L_S)$.*

Proof

(a) Consider two levels of demand: $D1$ and $D2$, where $D2 > D1$. Then, when the demand is $D1$, the manufacturer has the option to follow, throughout, the same policy which is optimal for a demand level equal to $D2$ until the demand $D1$ is satisfied to its entirety, at a cost of maximum $F_{D2}(L_1, \ldots, L_S)$. Thus, $F_{D1}(L_1, \ldots, L_S) \leqslant F_{D2}(L_1, \ldots, L_S)$.

(b) Consider two levels of WIP: $(L_1, \ldots, L_S)$ and $(L'_1, \ldots, L'_S)$, such that $(L'_1, \ldots, L'_S) \geqslant (L_1, \ldots, L_S)$. Then, for the WIP $= (L'_1, \ldots, L'_S)$ the manufacturer has the option to follow, throughout, the same policy which is optimal if the WIP were $(L_1, \ldots, L_S)$, at a cost of $F_D(L_1, \ldots, L_S)$. Thus, $F_D(L'_1, \ldots, L'_S) \leqslant F_D(L_1, \ldots, L_S)$. □

4. Heuristics for the assembly problem

In this section we refer to computational aspects of heuristic policies for the assembly system. For this sake we generalize the concept of *control limit policies* that was proposed by Grosfeld-Nir (2005) for the two-stage system.

4.1. Control limit policies

The purpose of processing units on the first echelon, i.e., $M_1, \ldots, M_S$ is, to increase the WIP to levels that warrant production on M_{S+1}. We, therefore, conjecture that there exists an optimal policy of the following form: "Produce on M_{S+1} if and only if $L_i \geqslant C_D$ for $i = 1, \ldots, S$". Table 3 provides an example of a control limit policy for a basic assembly system.

Note that the control limit does not completely define a fixed policy. Additionally, it is required to specify the size of the lots to be processed at each stage. Also, in the case that production is possible on a number of machines, it is required to specify on which machine to produce first. This order of production may have an affect on the expected total cost if the lot size on one machine in the first echelon depends on the WIP of another component. However, to simplify matters we will restrict ourselves to control-limit policies where we produce on the lowest indexed machine in the first echelon, for which the WIP is below C_D. Thus, we adopt the following rule: If $\min\{L_1, \ldots, L_{i-1}\} \geqslant C_D$ and $L_i < C_D$ for a certain $i \leqslant S$, then produce on M_i. Otherwise, i.e., if $L_i \geqslant C_D$ for $i = 1, \ldots, S$, then produce on M_{S+1}. Specifically, for a basic assembly system the rule is as follows:

Table 3
The basic assembly system with $C_D = 2$

$L1$	$L2$					
	0	1	2	3	4	5
0	Produce on either M_1 or M_2		Produce on M_1			
1						
2	Produce on M_2		Produce on M_3			
3						
4						
5						

- If $L_1 < C_D$ produce on M_1.
- If $L_1 \geqslant C_D$ and $L_2 < C_D$ produce on M_2.
- If $L_1 \geqslant C_D$ and $L_2 \geqslant C_D$ produce on M_3.

For the remainder of this article we restrict ourselves to such control-limit policies.

4.2. Heuristics

In this section we propose modifications of PIA + FPA and IDA + FPA to solve assembly systems. We like to comment that we used FPA merely to solve the linear equations (5) and (6). Instead, MATLAB, for example, could be used.

It turns out that PIA can be used almost in the same way as with the two-stage system and it gives very good results. On the other hand, IDA must be modified to fit assembly systems. As we have mentioned, the problem with PIA is the time it consumes. For example, consider the binomial basic assembly system with the parameters

$$\alpha_1 = 20, \beta_1 = 5, \theta_1 = 0.7; \quad \alpha_2 = 50, \beta_2 = 2, \theta_2 = 0.9; \quad \alpha_3 = 30, \beta_3 = 10, \theta_3 = 0.8. \tag{7}$$

For $D = 4$ the running time on a Pentium was 275 minutes (see Table 4, below).

4.2.1. Modification of IDA for a general assembly system

Similar to the two-stage problem, IDA defines fixed policies in terms of a single decision variable K_D: For demand D, whenever production takes place on a machine of the first echelon, the lot to be processed is determined via solving a problem where a single stage aims to satisfy the demand K_D. More precisely, let $N_D^{M_i}$, for $i = 1, \ldots, S+1$, be the optimal lot size if machine M_i alone, as a single stage, faced a rigid demand D. Let also $L \equiv \min\{L_1, \ldots, L_S\}$. IDA with parameter K_D is defined as follows:

- If $L \geqslant N_D^{M_{S+1}}$ produce a lot of size $N_D^{M_{S+1}}$ on M_{S+1}.
- If $L < N_D^{M_{S+1}}$ and $L \geqslant K_D$ produce a lot of size L on M_{S+1}.

Otherwise, i.e., if $L < N_D^{M_{S+1}}$ and $L < K_D$.

- Identify machines of the first echelon for which $L_i < N_D^{M_{S+1}}$ and $L_i < K_D$; produce a lot of size $N_{K_D - L_i}^{M_i}$ on M_i.

Note that IDA defines a control limit $C_D = \min\{K_D, N_D^{M_{S+1}}\}$, so that production takes place on M_{S+1} if and only if $L \geqslant C_D$. Thus, with IDA, we calculate $U_D(0, 0, \ldots, 0)$ for $K_D = 1, 2, \ldots$, to stop when there is no more improvement. Selected numerical results for the system presented in (7), compared with PIA, are

Table 4
A comparison between PIA and IDA for the basic assembly problem (7)

D	$U_D(0,0)$		Gap (%)	C_D		Time (seconds)	
	PIA	IDA		PIA	IDA	PIA	IDA
1	144.5	145.5	0.7	1	1	120	6
2	177.1	180.0	1.6	2	3	900	20
3	206.4	209.3	1.4	3	4	4800	35
4	235.1	236.7	0.7	4	5	16,500	48

exhibited in Table 4. As can be seen, using PIA gives slightly better solutions at a much significant time increase. Indeed, the running time needed by PIA for assembly systems with even a few machines in the first echelon makes it prohibitive.

As mentioned, our computational study focuses on binomial machines. In order to test the quality of IDA we develop a simple lower bound on the optimal expected cost.

Proposition 2. *Let F_D be the minimal expected cost associated with the (binomial) assembly system of Fig. 5 starting with zero WIP, in the particular case where $\alpha_1 = \cdots = \alpha_S = 0$. Then*

(a) *F_D is equal to the optimal expected cost required for a binomial single stage with demand D and parameters: $\alpha = \alpha_{S+1}; \beta = \sum_{i=1}^{S}\beta_i/\theta_i + \beta_{S+1}; \theta = \theta_{S+1}$.*
(b) *$U_D^{\mathrm{LB}} \equiv F_D + \sum_{i=1}^{S}\alpha_i$ is a lower bound on the optimal expected cost for the assembly system.*

Proof. For (a) see Section 5; for (b) note that If $\alpha_i \neq 0$, the expected cost increases by at least α_i. □

An intuitive way to understand Proposition 2 is as follows: The optimal expected cost of the system of Fig. 5 is at least as large as if only one setup cost were incurred by each of the machines of the first echelon, and thereafter the remaining expected costs is minimized.

In Table 5, we demonstrate the effectiveness of IDA compared to the lower bound for the basic assembly system problem given in (7). As we can see, the gap between IDA and the lower bound tends to decrease as the demand increases.

Example 3. Consider an assembly system with three machines at the first echelon with the parameters presented in Table 6.

We solve Example 3 using IDA. A few numerical results are exhibited in Table 7.

Table 5
A comparison between the lower bound and IDA for the problem (7)

D	Lower bound	$U_D(0,0)$	Gap (%)	C_D	Time (seconds)
1	131.7	145.5	10.5	1	6
2	162.2	180.0	11.0	3	20
3	189.5	209.3	10.4	4	35
4	215.0	236.7	10.1	5	48
5	241.0	267.0	10.8	7	100
6	267.2	293.6	9.9	7	175
7	293.6	319.2	8.7	9	310
8	318.3	345.8	8.6	10	630
9	343.3	374.5	9.1	12	985
10	368.5	400.5	8.7	12	1440

Table 6
Data for an assembly system with three machines in the first echelon

	α_i	β_i	θ_i
M_1	50	1	0.8
M_2	40	2	0.9
M_3	30	3	0.8
M_4	20	4	0.9

Table 7
Numerical results for an assembly problem with three machines at the first echelon

Demand	Lower bound	$U_D(0,0,0)$	Gap (%)	C_D	Running time (minutes)
1	154.7	164.4	6.3	1	1.5
2	169.2	186.4	10.2	2	3
3	183.5	201.9	10.0	4	8
4	197.6	215.8	9.2	5	16.5
5	211.5	230.1	8.8	6	28

5. Extensions

In this section we consider a few variants of assembly systems where all the machines are binomial. We first show that assembly systems with additional 0-BN machines at either end (Fig. 7), can be analyzed using the methodology above. The situation then is similar to the SBNS of Section 2.2.

Referring to a binomial assembly system of this structure, we denote m_i, $i = 1, \ldots, S+1$, the expected cost of producing one good unit on the ith 0-BNS (to be computed via (2)). Note that as long as a policy dictates to produce on the first SBNS (consisting of the first 0-BNS followed by M_1), it is optimal to process units, one at a time, until a batch of conforming units of sufficient size is ready to enter M_1. Thus, the cost effect of the first 0-BNS is to add a cost m_1, to the variable cost of M_1. That is, the expected cost of processing a lot of size N on M_1 is $\alpha_1 + (m_1 + \beta_1)N$. The same holds true for the other SBNS of the first echelon. Finally, as a lot exits M_{S+1}, it is optimal to process these units, one at a time, until the demand is satisfied, or all units are exhausted. Thus, the effect of the $(S+1)$-st 0-BNS is to add the expected cost Dm_{S+1} to the expected cost of satisfying the demand. Also, let γ denote the probability that a unit processed by the $(S+1)$-st 0-BNS ends conforming. The next Theorem defines an equivalent two-echelon assembly system for any multi-echelon assembly system of the form depicted in Fig. 7.

Theorem 4. *The problem of optimally controlling the binomial assembly with SBNS at either end can be reduced to that of optimally controlling an equivalent assembly system with binomial machines* M_i^e *and parameters* $\alpha_i^e, \beta_i^e, \theta_i^e$ *for*

$$\alpha_i^e = \alpha_i, \quad \beta_i^e = \beta_i + m_i, \quad \theta_i^e = \theta_i \quad \text{for} \quad i = 1, \ldots, S, \quad \text{and} \quad \alpha_{S+1}^e = \alpha_{S+1}, \quad \beta_{S+1}^e = \beta_{S+1},$$
$$\theta_{S+1}^e = \gamma\theta_{S+1}.$$

The equivalency is in that

(a) *The optimal lot size to enter machines* M_i *for* $i = 1, \ldots, S$, *is the same as the optimal lot size to enter the equivalent machines* M_i^e *for* $i = 1, \ldots, S$.
(b) *The expected cost required to fulfill an order* D *by such an assembly system equals the expected cost to fulfill the same order by the equivalent assembly system plus* Dm_{S+1}.

We omit the proof as it is similar to Grosfeld-Nir (1995).

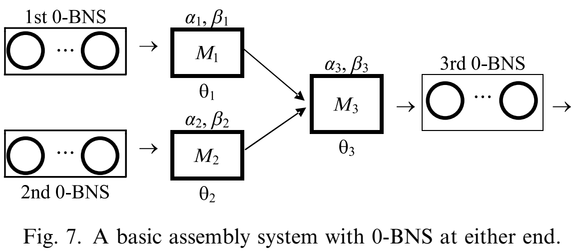

Fig. 7. A basic assembly system with 0-BNS at either end.

Corollary 2. *Consider the general assembly system of Fig. 5 with binomial machines, where M_i is replaced by a binomial 0-BNS, (i.e., the ith "branch" does not contain any BN machine) for some $i, i \leqslant S$. Then the optimal lot size and expected cost remain unchanged if this branch is omitted from the system and the cost m_i is added to the variable cost of the final stage. In particular, if all the machines of the first echelon are replaced by 0-BNSs, the problem of optimally controlling the assembly system becomes one of optimally controlling a single-stage binomial machine.*

6. Conclusions

Assembly production systems realistically mirror the level of complexity that manufacturers face in real life. These systems become extremely difficult to analyze when demand is rigid, because the optimal policy depends upon all possible levels of intermediate inventories. The fact that in spite of their importance the literature concerning such systems is very sparse is a reflection of the immense complexity of the problem. This study is among the first attempts to acquire a deeper understanding of such systems.

One of the main results of this paper is that the expected costs of fixed policies can be evaluated by solving a finite set of linear equations. The result is general in that it applies to any yield distribution. Also, we note that the fixed-policy-algorithm (FPA) provides a very useful tool to solve these equations.

In all our numerical experimentations IDA proved to be efficient: there was only a moderate deviation from the best policy obtained by the PIA. Also when compared to a lower bound on the optimal cost, IDA results in a reasonable expected cost. For assembly systems with binomial machines we show how more complex systems, i.e., systems in which each machine in the first echelon is preceded by a 0-BNS and the machine in the second echelon is followed by a 0-BNS, can be reduced to a two-echelon assembly system.

IDA has some very appealing properties that we wish to point out:

(a) IDA is fast, impressively accurate, and simple to implement as production lots are always determined by solving single-stage problems.
(b) With PIA, after each production run the manufacturer needs to observe the realized usable output after which he must choose the next machine to operate and determine the production lot. In contrast, with IDA, production lots are determined for each machine, independent of the other machines.
(c) With PIA, "simultaneous" manufacturing is impossible: the manufacturer must operate one machine at a time. In contrast, with IDA, the manufacturer can operate some or all the machines of the first echelon simultaneously.

Future research should characterize the structure of an optimal policy, and explore the dependency of the optimal lots and expected costs on the WIP levels and the outstanding demand level, for assembly systems with specific yield distributions. If there proves to be an optimal policy that is a control-limit policy, then it will be interesting to explore whether the control limits are monotone in the problem parameters and the demand.

Future research should also consider serial multi-stage production systems and more complex assembly systems where each "branch" consists of several BN machines.

Appendix A

We prove now Theorem 2, i.e., the set of $\delta(D) + 1$ linear equations defined by (3) and (4) has a single solution. These equations can be written in the form:

$$U_D(L) = b_D(L) + \sum_{\ell=0}^{\delta(D)} \theta^D_{L,\ell} U_D(\ell), \quad \text{for } L = 0, \ldots, \delta(D),$$

where $b_D(L)$ and $\theta^D_{L,\ell}$ for $\ell, L \in \{0, \ldots, \delta(D)\}$ are the constants:

$$b_D(L) = \begin{cases} \alpha_1 + \beta_1 N_D(L) & \text{if } L \in \Delta^1_D, \\ \alpha_2 + \beta_2 N_D(L) + \sum_{x=1}^{N_D(L)} p_2(x, N_D(L)) U_{D-x}(L - N_D(L)) & \text{if } L \in \Delta^2_D, \end{cases}$$

$$\theta^D_{L,\ell} = \begin{cases} p_1(\ell - L, N_D(L)) & \text{if } L \in \Delta^1_D \quad \text{and } 0 \leqslant \ell - L \leqslant N_D(L), \\ p_2(0, N_D(L)) & \text{if } L \in \Delta^2_D \quad \text{and } \ell = L - N_D(L), \\ 0 & \text{otherwise.} \end{cases}$$

Thus, (3) and (4) can be written in the vector/matrix form as $U_D = b_D + \Theta_D U_D$, where U_D and b_D are vectors of size $\delta(D) + 1$, and Θ_D is a square matrix of size $(\delta(D) + 1) \times (\delta(D) + 1)$.

Clearly, in order to prove the existence of a single solution we need to show that the matrix $(I - \Theta_D)^{-1}$ is not singular. Kemeny and Snell (1969) proved this holds true if $\lim_{n\to\infty}(\Theta_D)^n = 0$.

Note the properties of the matrix Θ_D: If $L \in \Delta^1$ then row L resembles a Markovian matrix (entries are probabilities whose sum is 1). If $L \in \Delta^2$ then the only non-zero entry in the row is $p_2(0, N_D(L))$ in the cell $(L, L - N_D(L))$. Since we consider only plausible policies that terminate in a finite number of setup, we assume that each $L \in \Delta^1$ can reach an $L' \in \Delta^2$ in a finite number of steps. That is, we assume that Δ^2 is reachable from every state $L \in \Delta^1$.

We augment the matrix Θ_D to become Markovian by adding a fictitious state R where the entry in cell (R, R) is 1; for any $L \in \Delta^1$, the entry in cell (L, R) is zero; and for any $L \in \Delta^2$, the entry in cell (L, R) is $1 - p_2(0, N_D(L)) > 0$.

The augmented matrix, named A_D, can be written as $A_D = \begin{bmatrix} \Theta_D & \vec{a} \\ 0 & 1 \end{bmatrix}$, where $\vec{a}$ is a vector of size $\delta(D) + 1$ whose entries are between zero and 1, and since $\Delta^2 \neq \emptyset$ at least one of its entries is strictly positive. Thus, $[A_D]^\infty = \begin{bmatrix} [\Theta_D]^\infty & - \\ - & 1 \end{bmatrix}$, where we do not specify the other entries of $[A_D]^\infty$. Note that in A_D all original states (of Θ_D) are transient. Thus, in the matrix $[A_D]^\infty$ all entries are zero except for the cells in column R which are all 1. In particular, $[\Theta_D]^\infty = 0$.

References

Anily, S., 1995. Single machine lot-sizing with discrete uniform yields and rigid demand: Robustness of optimal solution. IIE Transactions 27, 625–633.

Anily, S., Beja, A., Mendel, A., 2002. Optimal lot-sizes with geometric production yield and rigid demand. Operations Research 50, 424–432.

Beja, A., 1977. Optimal reject allowance with constant marginal production efficiency. Naval Research Logistic Quarterly 24, 21–33.

Grosfeld-Nir, A., 1995. Single bottleneck systems with proportional expected yields and rigid demand. European Journal of Operational Research 80, 297–307.

Grosfeld-Nir, A., 2005. A two-bottleneck system with binomial yields and rigid demand. European Journal of Operational Research 165, 231–250.

Grosfeld-Nir, A., Gerchak, Y., 1996. Production to order with random yields: Single-stage multiple lot-sizing. IIE Transactions 27, 669–676.

Grosfeld-Nir, A., Gerchak, Y., 2002. Multistage production to order with rework capability. Management Science 48, 652–664.

Grosfeld-Nir, A., Gerchak, Y., 2004. Multiple lot sizing, production to order (MLPO) with random yield: Review of recent advances. Annals of Operations Research. 126, 43–69.

Grosfeld-Nir, A., Robinson, L.W., 1995. Production to order on a tandem line with random yields and rigid demand. European Journal of Operational Research 80, 264–276.

Grosfeld-Nir, A., Ronen, B., 1993. A single bottleneck system with binomial yield and rigid demand. Management Science 39, 650–653.

Kemeny, J.G., Snell, J.L., 1969. Finite Markov Chains. D. Van Nostrand Company, Inc., pp. 22–23.

Pentico, D.W., 1994. Multistage production systems with random yield: Heuristics and optimality. International Journal of Production Research 32, 2455–2462.

Sepheri, M., Silver, E.A., New, C., 1986. A heuristic for multiple lot sizing for an order under variable yield. IIE Transactions 18, 63–69.

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

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

Yano, C.A., Lee, H.L., 1995. Lot sizing with random yields: A review. Operations Research 43, 311–334.

Zipkin, P.H., 2000. Foundations of Inventory Management. Irwin/McGraw-Hill.