Manufacturing & Service Operations Management

Serial Production/Distribution Systems Under Service Constraints

Tamer Boyaci, Guillermo Gallego,

Serial Production/Distribution Systems Under Service Constraints

Tamer Boyaci • Guillermo Gallego
Faculty of Management, McGill University, 1001 Sherbrooke Street, Montreal, Quebec H3A 1G5, Canada
Dept. of IEOR, Columbia University, 500 West 120th St., New York, New York 10027

We analyze the problem of minimizing average inventory costs subject to fill-rate type of service-level constraints in serial and assembly production/distribution systems. We propose optimal and heuristic procedures to solve this problem. Our model and solution procedures can be used to manage the fill rate or fill rate within a ''time window'' service measures. We also relate our service-constrained model to the traditional model with backorder costs and show that it is possible to prespecify backorder cost rates to achieve desired service levels. We explore the inventory cost impact of such a practice, and we find that the cost penalty can be very high.

(*Inventory/Production; Multistage; Serial; Fill Rate; Base-Stock Policy; Solution and Heuristics*)

1. Introduction

We consider the following single-item supply system: There are *J* stocking points (stages) arranged in series. The first stage receives supplies from an external source. All other stages are internally linked; each stage supplies the next one. The customer demand arrives at the final stage, and the unmet demand at any stage is backordered. Inventory in the supply system is reviewed continuously and controlled with a base-stock policy. We concentrate on a basic system where the demand is a Poisson process, each stage generates a constant leadtime, there are no fixed costs of ordering/production, the horizon is infinite, and all data and cost parameters are stationary.

For the above serial supply system, we analyze the problem of minimizing average inventory costs subject to fill-rate type of service-level constraints. This problem is a difficult one because one cannot make use of convexity and separability results. Indeed, even the form of the optimal policy is unknown. Our analysis focuses on base-stock policies. We develop bounds on the total system stock and on base-stock levels of each stage, and we incorporate them in an algorithm to find the optimal base-stock inventory policy. We also develop heuristics for this problem that are computationally more efficient for large systems. Our model and solution procedures can be used to manage the fill rate (i.e., fraction of demand immediately met from stock), fill rate within a ''time window,'' and similar constraints such as the limit probability of nonnegative inventory (PONI). The transformation in Rosling (1989) extends the applicability of our model to assembly systems as well.

There is a close connection between our service-constrained model and the traditional backorder-cost (BC) models in the inventory literature. Traditional BC models charge, in addition to the linear holding costs at every stage, a linear backorder penalty cost at the final stage for backordered customer demand, and they minimize the sum of holding and backorder costs. The use of backorder penalties in BC models is often justified by the service levels they induce. For example, for the single-stage version of the problem considered here, it is well known (from the newsvendor solution) that setting the backorder penalty rate as the product of the holding cost and

1523-4614/01/0301/0043$05.00
1526-5498 electronic ISSN

$\beta/(1 - \beta)$ guarantees a $100\beta\%$ service level. Furthermore, in this case, the two modeling approaches are equivalent in that they result in the same inventory policy and holding cost. We show that essentially the same service connection exists for multistage systems, which suggests that it is possible to specify backorder penalty rates in BC models to achieve desired service levels. However, for multistage systems, the two approaches are not equivalent. We address the cost effectiveness of this approach and find that the inventory cost penalty can be very significant.

2. Brief Literature Review

We refer the reader to Federgruen (1993) and the survey article by Van Houtum et al. (1996) for an extensive review of the literature on the BC model for serial systems, and we refer to Diks et al. (1996) and Axsäter (1993) for a survey of a related literature that concentrates on policy evaluation.

The majority of the research dealing with fill-rate type service measures are performance estimation or sensitivity analysis models. There is a limited literature on service-constrained optimization models for multistage systems. Van Houtum and Zijm (1991) consider a periodic review serial system and briefly mention minimization of inventory costs subject to a constraint on (effectively) expected backorders. Lee and Billington (1993) and Ettl et al. (1995) present 2 inventory-queue models that consider service-constrained optimization for large scale, multistage, and multiitem systems operating under continuous review base-stock policies. The focus of both models is on characterizing leadtimes and building normal approximations to leadtime demands, fill rates, and costs. Using these approximations, they minimize inventory costs subject to a fill-rate constraint. Chapter 5 of Wang (1998) also discusses service-constrained optimization models for assemble-to-order systems, building on the approximations for high fill rates in Glasserman and Wang (1996). Unlike earlier approximation models, we present optimal and heuristic solutions that are based on exact cost and service definitions for the single-item, multistage serial system introduced in §1.

3. Preliminaries

In this section, we introduce notation, describe the system dynamics, and present the formulation of our model. For $j = 1, \ldots, J$, we define

$$L_j = \text{leadtime generated at Stage } j,$$

$$s_j = \text{local base-stock level at Stage } j.$$

Define the following steady-state random variables,

$$D_j = \text{the leadtime demand at Stage } j$$
$$(D_j = \text{Poisson}(\lambda L_j)),$$

$$I_j = \text{inventory on hand at Stage } j,$$

$$IT_j = \text{inventory in transit to Stage } j.$$

Let h_j, denote the holding cost rate at Stage j. We assume that $h_1 \leq \cdots \leq h_J$. Then, the expected average holding cost for a given base-stock vector $s = (s_1, \ldots, s_J)$ can be written as

$$C(s) = \sum_{j=1}^{J} h_j E[I_j].$$

Notice that we are charging the holding cost only to those units on hand at Stage j, not to the units in transit to Stage j. However, the steady-state expected number of units in transit to Stage j is $E[IT_j] = E[D_j]$ regardless of the choice of the base-stock policy, and, hence, these costs can be ignored in determining optimal base-stock levels. It is useful to define the following steady-state random variables recursively,

$$\tilde{D}_1(s) = D_1$$

$$\tilde{D}_j(s) = D_j + [\tilde{D}_{j-1}(s) - s_{j-1}]^+ \quad \text{for } j = 2, \ldots, J,$$

where $\tilde{D}_j(s)$ represents the "modified" steady-state distribution of leadtime demand of Stage j, which is equal to the actual leadtime demand plus the backorders at Stage $j - 1$.

With this notation, we can express I_j as

$$I_j = [s_j - \tilde{D}_j(s)]^+ \quad \text{for } j = 1, \ldots, J.$$

Hence, the average inventory cost of the system can be written as

$$C(s) = \sum_{j=1}^{J} h_j E[s_j - \tilde{D}_j(s)]^+. \tag{1}$$

For a given base-stock vector s, the fill rate (FR) is denoted by $\beta(s)$ and is equal to the limit probability of having *positive* on-hand inventory at the last stage:

$$\beta(s) = P(\tilde{D}_j(s) < s_J). \qquad (2)$$

When arrivals see time averages (e.g., when demand is Poisson), $\beta(s)$ is also equal to the long-run fraction of demands that see positive inventory levels and is, therefore, equal to the fill rate. The FR model minimizes $C(s)$ subject to a given minimum desired fill rate β.

We analyze the FR model first for a two-stage serial system, and then the general J-stage system. In line with the Poisson demand assumption, our analysis will be based on discrete demand distributions. The definitions and results extend (even more sharply) to the continuous demand case. The continuous counterpart of our discussion is omitted here for the sake of brevity, but it can be found in Boyaci (1998).

4. Two-Stage Serial System

For the two-stage serial system, we can state the FR model as follows:

$$\min_{s_1, s_2} C(s_1, s_2) = h_1 E[s_1 - D_1]^+ + h_2 E[s_2 - (D_2 + (D_1 - s_1)^+)]^+ \qquad (3)$$

subject to

$$\beta(s_1, s_2) = P(D_2 + (D_1 - s_1)^+ < s_2) \geq \beta.$$

Suppose that s_1 is fixed. Then, the cost function (3) is increasing in s_2; hence, the optimal value of s_2, say $s_2(s_1)$, is given as the minimum base-stock level that satisfies the FR constraint: $s_2(s_1) = \min\{s_2: \beta(s_1, s_2) \geq \beta\}$. Let $C(s_1) = C(s_1, s_2(s_1))$. Unfortunately, $C(s_1)$ is not necessarily convex in s_1; hence, the optimal value of s_1 must be found through a search. The following lemma is useful in developing bounds to restrict this search.

LEMMA 1. $s_2(s_1) - 1 \leq s_2(s_1 + 1) \leq s_2(s_1)$.

PROOF. From the definition of $s_2(s_1)$, we have

$$\beta(s_1, s_2(s_1) - 1) < \beta \leq \beta(s_1, s_2(s_1)).$$

Because

$$\beta(s_1 + 1, s_2(s_1)) \geq \beta(s_1, s_2(s_1)) \geq \beta,$$

we have $s_2(s_1 + 1) \leq s_2(s_1)$. We show $s_2(s_1) - 1 \leq s_2(s_1 + 1)$ by contradiction. Suppose that $s_2(s_1) - 1 > s_2(s_1 + 1)$. Then we must have $\beta(s_1, s_2(s_1 + 1) + 1) < \beta$. However,

$$\beta(s_1, s_2(s_1 + 1) + 1) \geq \beta(s_1 + 1, s_2(s_1 + 1)) \geq \beta,$$

is a contradiction. □

Applying Lemma 1, we obtain the upper bound $s_2^u = s_2(0) = \min\{s_2: P(D_1 + D_2 < s_2) \geq \beta\}$ and the lower bound $s_2^l = s_2(\infty) = \min\{s_2: P(D_2 < s_2) \geq \beta\}$ on s_2, which are both based on feasibility constraints. Now, $s_1^u = \min\{s_1: s_2(s_1) = \underline{s}_2\}$ is an upperbound on the optimal value of s_1 because for any $s_1 > s_1^u$, $s_2(s_1) = s_2^l$, and $C(s_1, s_2)$ is increasing in s_1 for fixed s_2.

It is possible to improve on these bounds by including cost considerations. Consider the following equations:

$$C(s_1 + 1, s_2) = C(s_1, s_2) + h_1 P(D_1 \leq s_1) + h_2 \sum_{k=1}^{s_2} P(D_1 = s_1 + k) \times P(D_2 \leq s_2 - k), \qquad (4)$$

$$C(s_1 + 1, s_2 - 1) = C(s_1, s_2) + P(D_1 \leq s_1)[h_1 - h_2 P(D_2 \leq s_2 - 1)]. \qquad (5)$$

Equation (4) reveals that a move from (s_1, s_2) to $(s_1 + 1, s_2)$ can only increase cost. On the other hand, Equation (5) reveals that a move to a FR-feasible point $(s_1 + 1, s_2 - 1)$ reduces cost whenever $P(D_2 < s_2) \geq h_1/h_2$. Let

$$s_2^c = \min\left\{s_2: P(D_2 \leq s_2) \geq \frac{h_1}{h_2}\right\}.$$

It can be easily verified that Equation (5) together with Lemma 1 implies that all policies with $s_2 < s_2^c$ can be ignored. Therefore, defining

$$\underline{s}_2 = \max\{s_2^l, s_2^c\},$$

we obtain an improved lower bound on s_2 and, hence, an improved upper bound on s_1:

$$\bar{s}_1 = \min\{s_1: s_2(s_1) = \underline{s}_2\}.$$

Computing the optimal policy $s^* = (s_1^*, s_2(s_1^*))$ now requires a simple line search on $s_1 \in [0, \bar{s}_1]$. Furthermore, not all values of s_1 need to be considered in this search. From (4), the only candidates for s_1^* are those s_1 values where the function $s_2(s_1)$ jumps down: $s_2(s_1) = s_2(s_1 - 1) - 1$. An efficient algorithm to perform this search is relatively easy to develop and is, therefore, omitted. We will use one such algorithm later on in a heuristic for the general J-stage problem.

5. *J*-Stage Serial System

In this section we extend our analysis to $J (> 2)$ stages. We develop bounds on the policy variables and incorporate them in an algorithm to compute the optimal base-stock policy. We also develop cost-effective and computationally efficient heuristics.

5.1. Optimal Policies for Serial Systems

The obvious way of finding the optimal base-stock policy for serial systems is through enumeration. It is possible to improve on enumeration by incorporating upper and lower bounds on the policy variables and total system stock. We start by developing upper and lower bounds on the total system stock, $s_T = s_1 + \cdots + s_J$.

Notice that in a serial system the highest customer service protection is provided by the last stage. Therefore, for a given total stock s_T, the highest FR is achieved by placing all the stock in the last stage. Finding the minimum stock level that maintains feasibility yields the lower bound $\underline{s}_T$. Clearly,

$$\underline{s}_T = \min\{s: P(D_1 + \cdots + D_J < s) \geq \beta\}. \tag{6}$$

An upper bound on s_T can be developed similarly. Let $\underline{s}_J$ denote the minimum stock level that the last stage should hold under any feasible base-stock policy:

$$\underline{s}_J = \min\{s: P(D_J < s) \geq \beta\}.$$

Let

$$\bar{s}_1 = \min\{s: P([D_1 - s_1]^+ + D_2 + \cdots + D_J \leq \underline{s}_J) < \beta\}.$$

An upper bound $\underline{\bar{s}}_T$ on the total system stock can now be defined as

$$\bar{s}_T = \bar{s}_1 + \underline{s}_J, \tag{7}$$

because for any base-stock policy with $s_T > \bar{s}_T$ the average inventory cost can be decreased by decreasing at least 1 base-stock level while still being feasible.

Surprisingly, the difference between the maximum desired and minimum required total safety stock ($\bar{s}_T - \underline{s}_T$) turns out to be quite small.[1] This finding indicates that in achieving a given FR it is not a matter of *how much* to stock, but rather a matter of *where* to stock. This is consistent with the observation made in Gallego and Zipkin (1999) for the BC model.

The task of finding optimal base-stock policy among those that satisfy $\underline{s}_T \leq s_T \leq \bar{s}_T$ requires evaluation of

$$\binom{s_T + J - 1}{s_T}$$

base-stock policies. However, some of these policies would not be feasible because of the FR restraint. Hence, there is a need to more cleverly allocate s_T across stages. One way of achieving this is to incorporate upper bounds on the local base-stock levels.

Let s_T be fixed. Consider a partial allocation $s = (s_1, \ldots, s_{k-1}, 0, \ldots, 0)$ with $s_P = s_1 + \cdots + s_{k-1} \leq s_T$. For this partial allocation FR is maximized by placing the remaining stock $s_T - s_P$ in Stage J. Then an upper bound on s_k can be found by finding the maximum value of s_k that maintains feasibility:

$$\begin{aligned}\bar{s}_k = \max\{s_k: \beta(s_1, \ldots, s_{k-1}, s_k, 0, \ldots, 0, s_t - s_p - s_k)\\ \geq \beta\}.\end{aligned} \tag{8}$$

These bounds on local base-stock levels can be used dynamically in a procedure to find the optimal base-stock level for a given value of total stock s_T. The recursive procedure Proc(s_T, s, k) below starts with a partial allocation $s = (s_1, s_2, \ldots, s_{k-1}, 0 \ldots, 0)$ and computes the optimal base-stock levels for Stages k,

[1] For the problems we tested, this gap ranged between 0 and 4; the average was around 1.5. In more than 60% cases the gap was at most 1.

. . . , J for a given s_T, evaluating only the feasible policies.

Proc(s_T, s, k) = Procedure To Find Optimal Base Stock Policy

```
1. Set $s_k = 0$, calculate $\bar{s}_k$ based on partial vector
   $(s_1, \ldots, s_{k-1}, 0, \ldots, 0)$.
2. DO WHILE $s_k \le \bar{s}_k$
     IF $k < J - 1$,
       Call Proc $(s_T, s, k + 1)$.
     ELSE
       Set $s_J = s_T - s_{J-1} - \cdots - s_1$,
       Cost $= C(s_1, \ldots, s_J)$,
       IF Cost $< C(s_T)$, THEN
         $C(s_T) =$ Cost,
         $s(s_T) = (s_1, \ldots, s_J)$.
       ENDIF
     ENDIF
     SET $s_k = s_k + 1$
     ENDO
```

The optimal policy for fixed s_T can be found by making a single call to procedure Proc(s_T, s, 1), after initializing s, $s(s_T)$, and $C(s_T)$. At the termination of the procedure, the optimal cost is given by $C(s_T)$ and the optimal base-stock policy is given by $s(s_T)$. The optimal base-stock policy for FR model can now be found by searching over s_T, and by making repeated use of the above procedure.

5.2. Heuristics for Serial Systems

One plausible heuristic is to restrict the number of stages that hold inventory. This heuristic is motivated by the observation in Gallego and Zipkin (1999) that such managerially convenient restrictions have little cost impact. Our lower and upper bounds on the total system stock have provided further support for this observation. Specifically, we can restrict ourselves to the case where only two stages hold inventory, the last stage and some other Stage j, $j < J$. The heuristic chooses the best location $j < J$. We call this the two-stage heuristic (TSH). TSH requires solving $J - 1$ two-stage problems, which can be done efficiently using an algorithm as discussed in §4.

The second heuristic is based on the notion of majorization (Marshal and Olkin 1979). Given total system stock $s_T \in [\underline{s}_T, \bar{s}_T]$, the majorization heuristic (MH) finds the least majorized policy among all feasible policies. This policy is found efficiently through a *greedy* procedure that initially places all the Stock s_T at Stage J and then moves maximum possible stock to Stage $J - 1$ while retaining feasibility, and it repeats the procedure for $J - 1$, $J - 2$, etc. Notice that this procedure simply tries to push inventory upstream (where holding costs are less expensive) and, hence, does not depend on the holding cost data. The MH finds the least majorized feasible policy for each value of $s_T \in [\underline{s}_T, \bar{s}_T]$ and chooses the policy that minimizes the average cost. It is worth noting that MH is guaranteed to find the optimal policy for a two-stage system, see Boyaci (1998).

The third heuristic modifies the sequential push mechanism of MH. Instead of moving maximum possible stock from Stage j directly to $j - 1$, the heuristic compares the potential cost savings of moving maximum stock to all possible upstream Stages 1, . . . , $j - 1$, chooses the best location (say j^*), and repeats the procedure for Stages 1, . . . , j^*. In this sense, the heuristic can be viewed to combine MH with TSH. We call this heuristic the mixed heuristic (MXH). MXH does not dominate MH costwise, and requires no more than $J - 1$ times the computational effort of MH.

6. Connection with the BC Model

There is a strong connection between the traditional BC model and a service-constrained model closely related to our FR model. Letting b (\$/*unit*/*time*) denote the backorder cost rate, the BC model can be stated as

$$\text{(BC)} \qquad \min_{s} \tilde{C}(s) = \sum_{j=1}^{J} h_j E[s_j - \tilde{D}_j(s)]^+ + bE[s_J - \tilde{D}_J(s)]^-, \qquad (9)$$

where $E[s_J - \tilde{D}_J(s)]^-$ is the expected number of customer backorders. An efficient ($O(J)$) algorithm for solving the BC model is originally due to Clark and Scarf (1960), which was later refined and extended by Axsäter and Rosling (1993) and Chen and Zheng (1994), among others.

Recall that fill rate is defined as the limit probability of positive inventory, $P(\tilde{D}_J(s) < s_J)$. Consider now a slight variation of the fill rate, namely PONI, defined as the limit probability of *nonnegative* inventory, $P(\tilde{D}_J(s) \le s_J)$. Clearly, PONI and fill-rate service measures differ only when demands are *discrete.* The proposition below establishes that BC models have indeed implied PONI service levels.

Proposition 1. *BC model has an implied PONI of at least $b/(b + h_J)$.*

Proof. Suppose that $(s_1^*, s_2^*, \ldots, s_{J-1}^*)$ are the optimal local base-stock levels for BC model for Stages $j = 1, \ldots, J - 1$. Then in the BC model, s_J^* is chosen to minimize

$$h_J E[s_J - \tilde{D}_J(s)]^+ + bE[s_J - \tilde{D}_J(s)]^-.$$

This a newsvendor problem, and the optimal solution is given as

$$s_J^* = \left\{\min_{s_J}: P(\tilde{D}_J(s) \le s_J) \ge b/(b + h_J)\right\}. \quad \square$$

Proposition 1 encourages inventory managers to use the BC model as a heuristic for the FR model. There are two problems with this approach. First, strictly speaking, BC models guarantee PONI service levels only. Second, and perhaps more important, this approach can result in significantly higher inventory holding costs. Our numerical study sheds light on the disadvantages of this approach and indicates that practitioners can benefit from our optimal algorithm, and in many cases from our simple heuristics, for better managing service levels and associated inventory costs.

7. Numerical Illustrations

In this section we illustrate the results of our optimal algorithm and the performance of the heuristics. We also compare them with the BC heuristic (BCH) with the backorder penalty rate chosen as $b = h_J\beta/(1 - \beta))$ to induce β service levels. To compare the solutions on an equal footing, we solved the PONI version of our model and heuristics. This amounts to replacing ">" with "≥" in the fill-rate expression, and does not impact the nature of our model and procedures or the performance of our heuristics.

Table 1 Comparison with the Optimal Solution

	MH	MXH	TSH	BCH
Average Penalty (%)	1.15	2.88	5.42	6.19
Maximum Penalty (%)	3.76	6.31	9.20	15.19
Minimum Penalty (%)	0.00	0.09	1.68	1.62

The numerical study is based on the set of problems in Gallego and Zipkin (1999). We fix the time scale so that total leadtime $L = 1$ and the monetary unit make the last stage's holding cost rate 1. The stages are placed symmetrically such that $L_j = 1/J$, for $j = 1, \ldots, J$. We test three demand rates, $\lambda = 16$, 32, 64 (corresponding to low, medium, high level of demand); and we test two target PONI levels $\beta =$ 90%, 97.5% (corresponding to $b = 9$, 39). As in Gallego and Zipkin (1999), we test four holding cost structures across the stages: linear, affine, kink, jump.

We have solved 24 problem instances with $J = 4$, optimally. Table 1 provides a summary of the performance of the heuristics reporting each heuristic's average, maximum, and minimum cost penalty over the optimal policy. The results indicate that MH performs best, with a 1.15% average cost penalty over the optimal cost. In 25% of the cases the MH solution coincided with the optimal solution. The BCH performed consistently worse than MH and MXH—the second best heuristic, with an average 6.19% cost penalty over the optimal solution.

We also compared the heuristics on an expanded set of problems including $J = 8$, 16. The results indicate similar relative performance, with the exception of TSH, which performed (as one would expect) 0.31% worse than BCH. MH and MXH performed 2.23% and 1.79% better than BCH, respectively. We remark that MXH tends to perform better when MH does not, and a heuristic that chooses the best of MH and MXH would increase the savings over BCH to 3.07%.

8. Fill Rate Within a Time Window

In practice managers are also interested in the proportion of orders that are filled (fill rate) within a time window (FRW). Our FR model and heuristics can be easily adapted to manage this service measure. Let w be delivery time window, and define $\hat{D}_j(w)$ as the distribution of leadtime demands when the total leadtime in the system is reduced by w, starting from the last stage.

Proposition 2. *The FRW of a given policy and time window (w) is equal to the fill rate of the same policy under the transformed leadtime demands $\hat{D}_j(w)$.*

We emphasize that the transformation should be done *only* to compute FRW; the average cost is incurred according to the original leadtimes generated by the system and, hence, should be based on them. We refer the reader to Boyaci (1998) for a proof of Proposition 2, and we note that this result is equivalently established for two-stage systems in Hariharan and Zipkin (1995).

9. Conclusion

In this paper we have studied the problem of achieving desired customer service levels in serial production/distribution systems. We formulate and analyze a model that can be used to manage the fill rate and fill rate within a time window service measures. We developed optimal and heuristic procedures to solve this problem. We also formalized the connection between our model and the traditional BC model. This connection establishes the possibility of prespecifying penalty rates in BC models to achieve desired fill-rate/PONI levels. However, our numerical examples suggest that this approach can be costly; i.e., the optimal algorithm, and in many cases the heuristics, provide the same desired service levels at lower inventory costs.

Our model and results also provide insights into how much to stock and where to stock under fill-rate type service constraints. The upper and lower bounds on the total system stock are quite tight, which suggests that the total amount of stock in the system can be predicted with high precision. Consequently, achieving desired service levels is more a matter of stock positioning, i.e., *where* to stock. Our optimal algorithm and heuristics can be viewed as alternative ways of allocating/positioning the total system stock. The allocation scheme that greedily pushes stock to upstream stages is very close to optimal; in fact, it is frequently the optimal allocation. Interestingly, restricting the number of stocking locations is not necessarily very costly (especially for low/moderate number of stages), provided that the locations are chosen carefully.

In our FR-constrained model, we have optimized over the class of base-stock policies. It is worth noting that the form of the optimal policy for this model remains unknown. From a theoretical point of view, it would be of interest to determine the form of the optimal policy.

References

Axsäter, S. 1993. Continuous review policies for multilevel inventory systems with stochastic demand. Graves et al.

———, K. Rosling. 1993. Installation vs. echelon stock policies for multilevel inventory control. *Management Sci.* **39** 1274–1280.

Boyaci, T. 1998. Supply chain coordination and service level management. Ph.D. thesis, Dept. IEOR, Columbia University, New York.

Chen, F., Y. Zheng. 1994. Lower bounds for multiechelon stochastic inventory systems. *Management Sci.* **40** 1426–1443.

Clark, A., H. Scarf. 1960. Optimal policies for a multiechelon inventory problem. *Management Sci.* **6** 475–490.

Diks, E. B., A. G. de Kok, A. G. Lagodimos. 1996. Multiechelon systems: A service measure perspective. *Eur. J. Oper. Res.* **95** 241–263.

Ettl, M., G. E. Feigin, G. Y. Lin, D. D. Yao. 1995. A supply-chain model with base-stock control and service level requirements. IBM research report. IBM T.J. Watson Research Center, Yorktown Heights, New York.

Federgruen, A. 1993. Centralized planning models for multiechelon inventory systems under uncertainty. Graves et al.

Gallego, G., P. Zipkin. 1999. Stock positioning and performance estimation in serial production-transportation systems. *Manufacturing and Service Oper. Management J.* **1** 77–88.

Glasserman, P., Y. Wang. 1996. Leadtime-inventory trade-offs in assemble-to-order systems. Working Paper, Graduate School of Business, Columbia University, New York.

Graves, S., A. Rinnooy Kan, P. Zipkin, eds. 1993. Logistics of production and inventory. *Handbooks in Operations Research and Management Sci.* Vol. 4. Elsevier, North-Holland, Amsterdam.

Hariharan, R., P. Zipkin. 1995. Customer-order information, leadtimes, and inventories. *Management Sci.* **41** 1599–1607.

Lee, H., C. Billington. 1993. Material management in decentralized supply chains. *Oper. Res.* **41** 835–847.

Marshall, A. W., I. Olkin. 1979. *Inequalities: Theory of Majorization and Its Applications.* Academic Press, New York.

Rosling, K. 1989. Optimal inventory policies for assembly systems under random demands. *Oper. Res.* **37** 565–579.

Van Houtum, G. J., K. Inderfurth, W. H. M. Zijm. 1996. Materials coordination in stochastic multiechelon production systems. *Euro. J. Oper. Res.* **95** 1–23.

———, W. H. M. Zijm. 1991. Computational procedures for stochastic multiechelon production systems. *Internat. J. Production Econom.* **23** 223–237.

Wang, Y. 1998. Service levels in production-inventory networks: Bottlenecks, trade-offs, and optimization. Ph.D. thesis, Graduate School of Business, Columbia University, New York.

The consulting Senior Editor for this manuscript was Paul Zipkin. This manuscript was received on June 17, 1999, and was with the authors 233 days for 5 revisions. The average review cycle time was 41.2 days.