THE STABILITY OF A CAPACITATED, MULTI-ECHELON PRODUCTION-INVENTORY SYSTEM UNDER A BASE-STOCK POLICY

PAUL GLASSERMAN
Columbia University, New York, New York

SRIDHAR TAYUR
Carnegie Mellon University, Pittsburgh, Pennsylvania

(Received August 1992; revisions received February, May 1993; accepted June 1993)

Most models of multilevel production and distribution systems assume unlimited production capacity at each site. When capacity limits are introduced, an ineffective policy may lead to increasingly large order backlogs: The *stability* of the system becomes an issue. In this paper, we examine the stability of a multi-echelon system in which each node has limited production capacity and operates under a *base-stock* policy. We show that if the mean demand per period is smaller than the capacity at every node, then inventories and backlogs are stable, having a unique stationary distribution to which they converge from all initial states. Under i.i.d. demands we show that the system is a Harris ergodic Markov chain and is thus wide-sense regenerative. Under a slightly stronger condition, inventories return to their target levels infinitely often, with probability one. We discuss cost implications of these results, and give extensions to systems with random lead times and periodic demands.

Most models of multi-echelon production and distribution systems assume unlimited production capacity and unlimited order size at each site. Under this assumption, various conditions on costs and model structure have been shown to imply the optimality of certain policies. Rather less is known about what happens when capacity limits are taken into account. For capacitated systems, a more fundamental question than optimality of a policy is stability: Does a given policy allow the system to meet demands, or does the system become increasingly backlogged?

In this paper, we analyze the stability of a multi-echelon system in which each node follows a *base-stock* policy, modified because of capacity constraints. Under a standard base-stock policy, the operation of each node is determined by a target level of safety stock. As demands deplete inventories, each node produces goods to restore inventories to their target levels. When production capacity is limited, it may take several periods of production to offset demand in a single period. Speaking loosely, the system is stable if, on average, it can produce finished goods at a greater rate than they are demanded.

We show that our system is indeed stable under the natural capacity condition, namely, that the mean demand per period be smaller than the per-period production capacity at every node. This condition is not itself surprising; the interest lies in determining just what it implies. We show that for general stationary demands, this condition suffices to ensure that the system has a unique stationary distribution to which it converges from all initial states. Under independent and identically distributed demands, we show that the state of the system constitutes a *Harris ergodic* Markov chain, and thus inherits the wide-sense regenerative structure of that class of processes. Under a slightly stronger condition, the system is regenerative in the classical sense and we identify explicit regeneration times. These properties have useful consequences for simulation, and it was the simulation-based optimization method of Glasserman and Tayur (1992a) that motivated this investigation. We also examine stability in the presence of lead times and demands influenced by a randomly fluctuating environment, including the case of periodic demands.

Our model is similar to those of Clark and Scarf (1960), Federgruen and Zipkin (1984), and Rosling

Subject classifications: Inventory/production, multistage: Stability under base-stock policy. Probability, regenerative processes: Harris recurrent inventory processes.
Area of review: Stochastic Processes and their Applications.

0030-364X/94/4204-0913 $01.25

(1989) in most respects, except for the capacity limits. A related continuous-time model is that of Svoronos and Zipkin (1991); other variants can be found in Graves, Rinnooy Kan and Zipkin (1992). The literature on capacitated systems is much more limited. For a *single-stage* capacitated system, Federgruen and Zipkin (1986a, b) show that a base-stock policy is optimal under rather general cost assumptions. Tayur (1992) provides a method for computing the optimal base-stock level.

Given the difficulty of finding optimal policies for general capacitated systems, it makes sense to restrict attention to a specific class of operating rules. Base-stock policies are attractive because they are simple and are known to be optimal in certain settings. The stability results given here are part of the justification for the gradient estimation method of Glasserman and Tayur (1992a), which can be used to find optimal basestock levels.

We know of no previous work on the stability of multi-echelon systems; but the stability of single-stage systems has been studied extensively, often in the setting of storage processes and dams. Prabhu (1965) includes some results of this type. General single-stage models are studied in Brockwell, Resnick and Tweedie (1982) in continuous time, and in Glynn (1989) in discrete time. Many additional references can be found in those papers.

Once our model is appropriately set up, existing general tools can be used to prove stability results. In this regard, a key step in our analysis is representing the state of the system through *echelon shortfalls*. These are differences between (cumulative) base-stock levels and (cumulative) inventories. The shortfalls satisfy a recursive equation that facilitates their analysis. In particular, through this recursion we are able to apply the method of Loynes (1962) to find a stationary distribution, as Baccelli, Massey and Towsley (1989) and Baccelli and Liu (1992) do for certain queueing systems. The recursion is also useful in establishing Harris ergodicity through a *coupling* argument, as in Thorisson (1983), Asmussen (1987), and Sigman (1988). While these techniques are reasonably familiar in queueing theory, they seem to be less well established in the stochastic production-inventory literature. One purpose of this paper is to show how they can be used in this setting as well.

The details of our model are presented in Section 1. Section 2 shows the existence of a stationary regime for the echelon shortfalls and convergence to stationarity from all initial conditions. In Section 3, after a brief review of Harris chains, we show that our system is Harris ergodic, then give conditions for explicit regeneration times. Section 4 discusses cost implications of our stability results. Section 5 covers systems with fixed lead times and two models of variable lead times, giving stability conditions in each case. In Section 6 we generalize the demand process, allowing demands to be influenced by a (possibly periodic) random environment.

1. THE MODEL

Our basic model is a serial system in which each stage has limited capacity and follows a base-stock policy for echelon inventory, i.e., for cumulative inventory downstream from that stage. Where applicable, we note extensions to an assembly system. In all cases, inventories are reviewed periodically (i.e., the system evolves in discrete time) and unfilled orders are backlogged. Demands are nonnegative but otherwise initially arbitrary; we introduce restrictions as they are needed. A discussion of lead times is postponed to Section 5.

1.1. The Base-Stock Policy

There are m stages, indexed by $i = 1, \ldots, m$. Stage 1 supplies external demands, stage $i + 1$ supplies stage i for $i = 1, \ldots, m - 1$, and stage m draws raw material from an unlimited source—an outside supplier. Within each period, events occur in the following order: First, production at stage $i + 1$ from the previous period advances to stage i, $i = 1, \ldots, m - 1$. Second, demands arrive at stage 1 and are filled or backlogged according to the available inventory. Lastly, the production level for the current period is set. This is the sequence of events in Clark and Scarf. Much of the subsequent literature assumes production levels are set before demands are revealed. The Clark-Scarf sequence simplifies our analysis.

To describe the operation of the system we use the following notation:

D_n = the demand in period n;

s^i = the base-stock level for echelon i;

c^i = the production capacity at stage i.

At stage 1,

I_n^1 = the inventory-backlog in period n,

and for $i = 2, \ldots, m$,

I_n^i = the installation inventory at stage i in period n.

Thus, $I_n^i \geqslant 0$, $i = 2, \ldots, m$ is the inventory available for production at stage $i - 1$, and I_n^1 is stock for external demands when it is positive and the size of the backlog when it is negative.

Under a (modified) basestock policy, stage i sets its production level in each period to try to restore the echelon inventory position

$$\left(\sum_{j=1}^{i} I_n^j\right) - D_n$$

to level s^i. Without capacity constraints, this would be achieved by setting production equal to the smaller of D_n and the available inventory. Since, however, production cannot exceed c^i, it may take multiple periods of production to offset demand in a single period, even if ample inventory is available for production.

To make this more explicit, we let

R_n^i = the production at stage i in period n.

Then the base-stock policy sets

$$R_n^i = \min\{s^i + D_n - (I_n^1 + \cdots + I_n^i), I_n^{i+1}, c^i\} \qquad i = 1, \ldots, m-1, \quad (1)$$

and

$$R_n^m = \min\{s^m + D_n - (I_n^1 + \cdots + I_n^m), c^m\}. \quad (2)$$

The first term inside the minimum in (1) is the difference between the target cumulative inventory s^i for stages 1 through i and the actual inventory $I_n^1 + \cdots + I_n^i - D_n$; stage i attempts to drive this difference to zero. The next two terms inside the minimum reflect the supply and capacity constraints, respectively. Since stage m draws raw material from an infinite source, the supply constraint is absent in (2). The evolution of the system is completely specified by (1), (2) and the following rules for the inventories:

$$I_{n+1}^1 = I_n^1 + R_n^1 - D_n;$$

$$I_{n+1}^i = I_n^i + R_n^i - R_n^{i-1} \quad i = 2, \ldots, m.$$

These reflect the downstream flow of material.

1.2. Echelon Shortfalls

Physical inventory levels are arguably the most natural descriptors of the state of the system. But, as is often the case in these types of systems, it turns out to be mathematically more convenient to work with echelon quantities. For $i = 1, \ldots, m$ define the period-n *shortfall* for echelon i by

$$Y_n^i = s^i - \sum_{j=1}^{i} I_n^j. \quad (3)$$

The shortfalls determine the inventories, because

$$I_n^1 = s^1 - Y_n^1; \quad (4)$$

$$I_n^i = (s^i - s^{i-1}) + (Y_n^{i-1} - Y_n^i) \quad i = 2, \ldots, m. \quad (5)$$

So, we may analyze the stability of $Y_n = (Y_n^1, \ldots, Y_n^m)$ and then interpret the results for inventories.

In general, the shortfall at stage i satisfies

$$Y_{n+1}^i = Y_n^i + D_n - R_n^i.$$

A base-stock policy attempts to reduce the shortfall to zero, while never driving it below zero. However, as in (1), production at stage i is constrained by the capacity c^i and, for $i < m$, by the available inventory. Since stage m draws raw material from an infinite source, we have

$$Y_{n+1}^m = Y_n^m + D_n - \min\{Y_n^m + D_n, c^m\}$$
$$= \max\{0, Y_n^m + D_n - c^m\}. \quad (6)$$

For $i = 1, \ldots, m-1$, the available inventory is limited to I_n^{i+1}, so we have

$$Y_{n+1}^i = Y_n^i + D_n - \min\{Y_n^i + D_n, c^i, I_n^{i+1}\}.$$

Using (5) and simplifying, we get

$$Y_{n+1}^i = \max\{0, Y_n^i + D_n - c^i, Y_n^{i+1} + D_n - (s^{i+1} - s^i)\}. \quad (7)$$

Equations 6 and 7 are the key to our analysis. The first of these is a Lindley equation, and this will be important in what follows. Compared with I_n^i and R_n^i, the shortfalls lend themselves more easily to an analysis of stability. Summarizing developments thus far, we have:

Lemma 1. *The echelon shortfalls satisfy* $Y_{n+1} = \phi(Y_n, D_n)$ *where* $\phi : \mathbf{R}_+^m \times \mathbf{R} \mapsto \mathbf{R}_+^m$ *is defined by* (6)–(7). *In particular,* ϕ *is increasing and continuous.*

Similar recursions hold in an assembly system, as we now explain. In an assembly system, each node i has a set $\pi(i)$ of predecessor nodes with indices greater than i. If i is a *root*, then $\pi(i)$ is empty and node i draws raw material from an infinite source. Otherwise, node i combines material from all nodes in $\pi(i)$ in equal quantities. Thus, period-n production at node i is limited by $\min\{I_n^j : j \in \pi(i)\}$. Proceeding as before, we obtain

$$Y_{n+1}^i = \max\{0, Y_n^i + D_n - c^i, \max_{j \in \pi(i)}\{Y_n^i + D_n - (s^j - s^i)\}\}, \quad (8)$$

where the maximum over an empty set is taken to be zero.

Remark. There is some similarity between the evolution of our serial system and that of queues in tandem. In both cases, material passes through a sequence of stages in series. However, the connection does not go beyond that. Notice, in particular, that in (ordinary) tandem queues the service mechanism at each stage does not depend on the status of other stages, whereas in our system the target production at each stage depends on the inventory at all downstream stages. Hence, there is no direct connection between (6)–(7) and recursions for quantities associated with tandem queues. At the same time, techniques used to analyze queueing systems serve, with modification, as the basis of our analysis.

2. THE STATIONARY REGIME

Suppose, now, that the demands form a stationary process. In this setting, through the method of Loynes, the conclusion of Lemma 1 is sufficiently strong to imply the existence of a stationary version of the echelon shortfalls. (In fact, it would suffice for ϕ to be increasing and continuous in its first argument for all values of its second argument.) Moreover, the natural stability condition

$$\mathsf{E}[D_0] < \min\{c^i : i = 1, \ldots, m\}, \tag{9}$$

ensures that there is just one finite stationary distribution. Here and throughout, $\Rightarrow$ denotes convergence in distribution.

Theorem 1. *Suppose that the demands* $\{D_n, n \geqslant 0\}$ *are stationary.*

i. *There exists a (possibly infinite) stationary process* $\{\tilde{Y}_n, n \geqslant 0\}$ *satisfying* $\tilde{Y}_{n+1} = \phi(\tilde{Y}_n, D_n)$ *for all* n, *such that if* $Y_0 = 0$ *almost surely (a.s.) then* $Y_n \Rightarrow \tilde{Y}_0$.
ii. *Suppose the demands are ergodic as well as stationary. If* (9) *holds, then* $\tilde{Y}_0$ *is almost surely finite; if, for some* i, $\mathsf{E}[D_0] > c^i$, *then* $\tilde{Y}_0^j = \infty$ *a.s., for all* $j = 1, \ldots, i$.
iii. *For stationary, ergodic demands satisfying* (9), $Y_n \Rightarrow \tilde{Y}_0$ *for all* Y_0.

Outline of Proof. A detailed proof of each of the assertions in the theorem is given in Glasserman and Tayur (1992b). Here, we outline how a proof may be constructed by appealing to related results in the literature.

Part i follows from Loynes via Lemma 1. In part ii, the assertion for Y^m follows from Loynes' analysis of the single-server queue. For Y^i, $i < m$, proceed by induction on i from m down to 1 to show, using (7), that (9) implies that the stationary distribution is finite; this step is similar to the analysis in Baccelli, Massey and Towsley of acyclic fork-join queues. For the converse, use the fact that $Y_{n+1}^i \geqslant Y_n^i + D_n - c^i$ to conclude that if $\mathsf{E}[D_0] > c^i$, then $\tilde{Y}_0^i$ must be infinite.

Part iii follows from part ii through a coupling argument. The process $\{Y_n, n \geqslant 0\}$ is said to *admit coupling* if for all pairs of initial states it is possible to construct two copies of the process started in those states in such a way that the two copies coincide after a finite (random) time. A process that admits coupling can have, at most, one stationary distribution. To show that Y admits coupling, note that Y^m does (again from Loynes) and argue from (7) that Y^i couples a finite, random time after $(Y^{i+1}, \ldots, Y^m)$ has coupled.

Remarks

i. The same argument works for the assembly system, proceeding by induction down the branches of the precedence tree.
ii. Theorem 1 could alternatively be proved by appealing to general results for (max, +) recursions in Baccelli and Liu (1992) or Glasserman and Yao (1992). The method of Baccelli and Liu associates a randomly weighted graph with recursions (like (6) and (7)) involving only max and +. Glasserman and Yao use a matrix formulation encompassing (max, +) and other types of recursions. In both approaches, the stability condition takes the form $\gamma < 0$, where the constant γ depends on the particular recursion and cannot be computed easily in general. It can be shown that for our model $\gamma = \mathsf{E}[D_0] - \min_i c^i$, so Theorem 1 is consistent with the general results.

3. REGENERATION

The previous section established conditions for the stability of the echelon shortfall process when demands are stationary and ergodic. We now examine the regenerative structure of $\{Y_n, n \geqslant 0\}$ when $\{D_n, n \geqslant 0\}$ is an i.i.d. sequence. (In Section 6 we relax the i.i.d. assumption.) Regenerative properties are valuable in establishing convergence of costs and also simulation estimators. Indeed, it was the simulation-based application in Glasserman and Tayur (1992a) that initially motivated this investigation.

We show that the stability condition of Section 2 suffices to ensure that $\{Y_n, n \geqslant 0\}$ possesses the regenerative structure of a *Harris ergodic* Markov chain. Under a stronger condition, we show that the vector of shortfalls returns to the origin infinitely often, with probability one.

3.1. Harris Recurrence

Many of the attractive properties of classical regenerative processes have been shown to hold for the somewhat weaker regenerative structure of Harris recurrent Markov chains. We briefly review key definitions and results of this framework to apply them to our model. More extensive coverage can be found in Nummelin (1984) and Asmussen (1987); the treatment in Sigman is particularly relevant to our application.

The general setting for Harris recurrence is a Markov chain $X = \{X_n, n \geq 0\}$ on a state-space $\mathbf{S}$ with Borel sets $\mathcal{B}$. Let P_x denote the law of X when $X_0 = x$. Then X is Harris recurrent if there exists a σ-finite measure ψ on $(\mathbf{S}, \mathcal{B})$, not identically zero, such that, for all $A \in \mathcal{B}$,

$$\psi(A) > 0 \Rightarrow P_x\left(\sum_{n=0}^{\infty} \mathbf{1}\{X_n \in A\} = \infty\right) = 1 \quad \text{for all } x \in \mathbf{S}. \tag{10}$$

Thus, every set of positive ψ-measure is visited infinitely often from all initial states. Every Harris recurrent Markov chain has an invariant measure π that is unique up to multiplication by a constant. The sets of positive π-measure are precisely those that are visited infinitely often from all initial states. If π is finite (hence a probability, without loss of generality), then X is called *positive* Harris recurrent. If, in addition, X is aperiodic, then it is Harris *ergodic*.

The connection with regeneration enters as follows. If X is Harris recurrent, then there exists a (discrete-time) renewal process $\{\tau_k, k \geq 1\}$ and an integer $r \geq 1$ such that

$$\{(X_{\tau_k+n}, n \geq 0), (\tau_{n+k+1} - \tau_{n+k}, n \geq 0)\}$$

has the same distribution for all $k \geq 1$ and is independent of

$$\{\tau_1, \ldots, \tau_k, (X_n, 0 \leq n \leq \tau_k - r)\}.$$

When $r > 1$, there may be dependence between consecutive cycles $\{X_n, \tau_{k-1} \leq n < \tau_k\}$, in contrast to the classical case of independent cycles (and this is indeed the case in our model). However, if X is positive Harris recurrent and if $f : \mathbf{S} \to \mathbf{R}$ is π-integrable, then the regenerative ratio formula

$$\mathsf{E}_\pi[f(X_0)] = \frac{\mathsf{E}\left[\sum_{n=\tau_{k-1}}^{\tau_k - 1} f(X_n)\right]}{\mathsf{E}[\tau_k - \tau_{k-1}]} \tag{11}$$

remains valid, as does the associated central limit theorem (under second-moment assumptions). Moreover, if X is Harris ergodic, then for all initial conditions the distribution of X_n converges to π in *total variation*; that is,

$$\sup_{A \in \mathcal{B}} |P_x(X_n \in A) - \pi(A)| \to 0$$

as $n \to \infty$ for all $x \in \mathbf{S}$. Indeed, this total variation convergence to a probability measure completely characterizes Harris ergodicity.

A powerful tool in the analysis of Harris ergodic Markov chains is a connection with coupling; see, for example, Thorisson (1983), Asmussen and Thorisson (1987), and Sigman (1988) for background. The main result is this: A Markov chain with an invariant probability measure admits coupling if and only if it is Harris ergodic. Since we already used a coupling argument for Y in Section 2, it is now easy to prove this:

Theorem 2. *Let demands $\{D_n, n \geq 0\}$ be i.i.d. with $\mathsf{E}[D_0] < \min_i c^i$. Then $\{Y_n, n \geq 0\}$ is a Harris ergodic Markov chain.*

Proof. Since $Y_{n+1} = \phi(Y_n, D_n)$, $n \geq 0$, Y is a Markov chain when D is i.i.d. We know from Theorem 1 that Y has an invariant (i.e., stationary) distribution and that Y admits coupling. Thus, Y is Harris ergodic.

As a result of Theorem 2, Y inherits the regenerative structure of Harris ergodic Markov chains and the attendant ratio formula and convergence results. The same holds for the inventory levels:

Corollary 1. *Under the conditions of Theorem 2, the inventory process $\{(I_n^1, \ldots, I_n^m), n \geq 0\}$ is a Harris ergodic Markov chain.*

Proof. Equations (3)–(5) put Y_n and $I_n = (I_n^1, \ldots, I_n^m)$ in one-to-one correspondence for all n. Consequently, $I = \{I_n, n \geq 0\}$ is Markov if Y is, and I is Harris ergodic if Y is.

3.2. Explicit Regeneration Times

While Harris recurrence ensures the existence of (wide-sense) regeneration times $\{\tau_k, k \geq 1\}$, it does not provide a means of identifying these times. Explicit regeneration times are not needed for convergence results, but they are useful in, for example, computing confidence intervals for simulation estimators. We now give a sufficient condition for $\{Y_n, n \geq 0\}$ to have readily identifiable regeneration times.

Theorem 3. *Let demands be i.i.d. with* $\mathsf{E}[D_0] < \min_i c^i$. *Define* $s^0 \equiv 0$ *and suppose that*

$$P(D_0 \leq s^i - s^{i-1}) > 0 \quad i = 1, \ldots, m. \tag{12}$$

Then Y returns to the origin infinitely often, with probability one.

Proof. If $\mathsf{E}[D_0] < c^i$, then $P(D_0 < c^1) > 0$. Consequently, under the conditions of the theorem there exists an ϵ with $\epsilon < \min_i c^i$ and $\epsilon \leq \min_i(s^i - s^{i-1})$ such that $\delta \triangleq P(D_0 \leq \epsilon) > 0$. Since Y has a finite stationary distribution, there exists a constant $b > 0$ such that the set $B_b \subseteq \mathbf{R}^m$ defined by

$$B_b = \{(y^1, \ldots, y^m) : 0 \leq y^i \leq b, i = 1, \ldots, m\}$$

is visited infinitely often by Y. We will show that there exists an integer $r \geq 0$ and a real p such that

$$P_x(Y_r = 0) \geq p > 0 \quad \text{for all } x \in B_b, \tag{13}$$

from which it follows that Y visits 0 infinitely often.

If $D_0 \leq \epsilon$, then either $Y_1^m = 0$ or $Y_1^m \leq Y_0^m - (c^m - \epsilon)$. Thus, every time a demand falls in $[0, \epsilon]$, the echelon-m shortfall is decreased by at least $c^m - \epsilon$, until it reaches zero. Starting in B_b, it takes at most $r_m = \lceil b/(c^m - \epsilon)\rceil$ consecutive such demands to drive that shortfall to zero. Thus, with $p_m = \delta^{r_m}$, we have $P_x(Y_{r_m}^m = 0) \geq p_m$ for all $x \in B_b$.

Suppose now that $Y_0^{i+1}, \ldots, Y_0^m = 0$ for some i and that $Y_0^i \leq b$. With probability at least δ^n, shortfalls $i + 1, \ldots, m$ remain at zero for the next n transitions. Moreover, for any n, if $Y_n^{i+1} = 0$ and $Y_n^i > 0$, then the inventory I_n^{i+1} available for use by stage i is strictly greater than $s^{i+1} - s^i$; see (5). Thus, if $D_n \leq \epsilon$, stage i cannot be constrained by inventory, and either $Y_{n+1}^i = 0$ or $Y_{n+1}^i \leq Y_n^i - (c^i - \epsilon)$. If we set $r_i = \lceil b/(c^i - \epsilon)\rceil$ then, with probability at least $p_i = \delta^{r_i}$, Y^i is driven to zero in r_i steps. We conclude that with probability at least $p = p_1 \cdots p_m$, $Y_{r_1+\cdots+r_m} = 0$ for any $Y_0 \in B_b$.

Corollary 2. *Under the conditions of Theorem* 3, the inventory process $\{(I_n^1, \ldots, I_n^m), n \geq 0\}$ *returns to* $(s^1, s^2 - s^1, \ldots, s^m - s^{m-1})$ *infinitely often, with probability one.*

The conclusion of Theorem 3 is not, in general, true without (12) or further distributional assumptions on demands. This is particularly clear when $s^{i+1} = s^i$ for some i; that is, stage $i + 1$ keeps no safety stock. In this case, the shortfall Y^i can never reach zero unless $D_0 = 0$ with positive probability.

4. COST IMPLICATIONS

The condition in Theorem 3 motivates an investigation into what ranges of parameters can be optimal when we impose costs. The stability results of the previous two sections also make it possible to give a partial characterization of infinite-horizon costs, and this may be useful in optimization.

To each echelon i we assign an inventory holding cost h^i, $i = 1, \ldots, m$. Backorders at stage 1 are penalized at rate p. There is no fixed cost for production in a period; if there were, a base-stock policy would be unattractive. Costs are incurred at the end of each period, so the cost in period n is

$$f(Y_n) \triangleq p(Y_n^1 - s^1)^+ + \sum_{i=1}^m h^i(s^i - Y_n^i)^+. \tag{14}$$

From the stability results of Sections 2 and 3, we obtain a partial characterization of the infinite-horizon average cost for any choice of parameters. Let $\tilde{Y}_0$ be as in Section 2, and define

$$F^i(y) = P(\tilde{Y}_0^i \leq y), \; y \in \mathbf{R},$$

for $i = 1, \ldots, m$. As a consequence of Theorem 1, we have

Corollary 3. *Under the conditions of Theorem* 1 iii,

$$n^{-1}\sum_{i=0}^{n-1} f(Y_i) \to \mathsf{E}[f(\tilde{Y}_0)]$$
$$= p\int_0^{s^1}(y - s^1)dF^1(y) + \sum_{i=1}^m h^i \int_0^{s^i}(s^i - y)dF^i(y),$$

with probability one for all Y_0. *The case* $\mathsf{E}[f(\tilde{Y}_0)] = \infty$ *is not excluded.*

This result is a direct consequence of the strong law of large numbers for ergodic stationary sequences and the fact that f is nonnegative. The form of $\mathsf{E}[f(\tilde{Y}_0)]$ is precisely what one would expect; our results guarantee that the limit holds, and may, therefore, be useful in finding optimal base-stock levels. In particular, this result can be used in the computation of optimal levels in the two cases where base-stock policies are known to be optimal: a multistage uncapacitated system and a single-stage capacitated system.

Superficially, the expression in Corollary 3 is the type required for the optimization algorithm of Van Houtum and Zijm (1990) for multistage uncapacitated serial systems. In the uncapacitated case, F^i can be expressed in terms of the demand distribution K and $s^j, j \geq i$, and these s^j appear only as location parameters. The shortfall distributions are *nested* because the system decomposes by stages. However, in the presence of capacities, each F^i depends on $s^{i+1}, \ldots, s^m$, and $c^i, \ldots, c^m$ in a more intricate way, and so the method of Van Houtum and Zijm is not applicable.

In the case of a single-stage capacitated system, Tayur provides an expression for F^1, the shortfall distribution. He shows that if the capacity is c, then F^1 solves the equation $F_c^1 * K = F^1$, where K is the demand distribution and where $F_c(x) = F(c + x)$ for any distribution F. Solving for s in the equation $F^1 * K(s) = p/(p + h)$ yields the optimal base-stock level.

The two results above imply that a multistage system that is uncapacitated but for stage m can be solved. The algorithm of Van Houtum and Zijm requires only one modification, namely, replacing the F^m for the uncapacitated system with the distribution found by Tayur. Unfortunately, other capacitated cases are not amenable to such straightforward analysis. However, some conclusions about optimal base-stock levels can still be drawn in special cases of multistage capacitated systems. Our next result shows that if capacities increase with the stage index, then it is never optimal to hold more safety stock between each pair of stages than the downstream stage can use in a single period:

Proposition 1. *Suppose that* $c^1 \leq c^2 \leq \cdots \geq c^m$. *If* $s^{i+1} - s^i \geq c^i$, $i = 1, \ldots, m - 1$, *then reducing each* $s^{i+1} - s^i$ *to* c^i, *leaving* s^1 *fixed, decreases costs.*

Proof. In light of Corollary 3, we may assume that the shortfall process starts at the origin. We claim that if $s^{i+1} - s^i \geq c^i$, $i = 1, \ldots, m - 1$, then for all n we have

$$Y_n^1 \geq Y_n^2 \geq \cdots \geq Y_n^m \tag{15}$$

and

$$Y_{n+1}^i = \max\{0, Y_n^i + D_n - c^i\} \quad i = 1, \ldots, m. \tag{16}$$

By hypothesis, (15) holds at $n = 0$. From (7) we see that (15) implies (16) whenever $c^i \geq s^{i+1} - s^i$, and, in turn, (16) implies that (15) holds at $n + 1$ if the c^i's are increasing.

Under the assumption that $Y_0 = 0$, (16) shows that the evolution of Y is independent of the base-stock levels, so long as they satisfy $s^{i+1} - s^i \geq c^i$. However, the echelon inventories $s^i - Y^i$, $i = 1, \ldots, m$ increase with the base-stock levels. Thus, if we reduce s^{i+1} to $s^1 + c^1 + \cdots + c^i$, $i = 1, \ldots, m - 1$, we lower holding costs without increasing backorders, since Y^1 is unchanged.

Suppose now that the capacity levels are subject to control, possibly within a range of values. For example, it might be possible to physically re-allocate capacity from one stage to another, or else a stage may modify its policy, choosing a maximum production level less than its capacity. This has the same effect as changing some c^i. The following result gives a necessary condition for a set of optimal capacity levels.

Proposition 2. *An optimal* $(c^1, \ldots, c^m)$ *satisfies* $c^{i+1} \leq c^i$, $i = 1, \ldots, m$. *More precisely, given any set of* $(c^1, \ldots, c^m)$, *if* $c^{i+1} \geq c^i$ *then replacing* c^{i+1} *with* c^i *does not increase costs.*

Proof. Let Y be the shortfall process under the original capacities and let $\bar{Y}$ be the shortfall process when c^{i+1} is reduced to c^i. Initialize the two processes with $\bar{Y}_0^j = Y_0^j$, $j \neq i + 1$ and

$$\bar{Y}_0^{i+1} = \max\{Y_0^{i+1}, Y_0^i + (s^{i+1} - s^i) - c^i\}. \tag{17}$$

For all n, we claim that $\bar{Y}_n^j = Y_n^j$, $j \neq i + 1$, and that (17) holds with zero replaced by n. Stages $i + 2, \ldots, m$ are unaffected by the change in c^{i+1}, so the claim certainly holds for those stages. Assuming the claim is valid at some fixed n, we have

$$\begin{aligned}\bar{Y}_{n+1}^i &= \max\{0, Y_n^i + D_n - c^i, \bar{Y}_n^{i+1} + D_n \\ &\qquad - (s^{i+1} - s^i)\} \\ &= \max\{0, Y_n^i + D_n - c^i, Y_n^{i+1} + D_n \\ &\qquad - (s^{i+1} - s^i)\} \\ &= Y_{n+1}^i;\end{aligned}$$

the first equality uses (7) with c^{i+1} replaced by c^i and the second equality substitutes (17) evaluated at n into the first equality. It follows that the claim holds at $n + 1$ for $Y^1, \ldots, Y^i$. A similar argument shows that (17) is preserved at each transition.

It follows from the claim just proved that reducing c^{i+1} to c^i does not decrease any shortfalls; hence, it does not increase any echelon inventory levels. Moreover, since Y^1 is unchanged, backorder penalties are not increased; so, total costs are not increased.

5. LEAD TIMES

We now examine variants of our basic model in which it may take several periods for production at stage i to become available inventory at stage $i - 1$. We show that for fixed lead times, our results continue to hold essentially without modification. When each order draws a random lead time and moves in parallel with other orders, it suffices to add that the mean lead time be finite. When shipments between stages are FIFO (in a sense to be made precise), a stronger condition is needed for stability.

5.1. Fixed Lead Times

Suppose that production at stage i in period n becomes available at stage $i - 1$ in period $n + \ell^i + 1$, $i = 2, \ldots, m$. At stage 1, ℓ^1 is the lead time from final production to external availability. Our earlier model used $\ell^i = 0, i = 1, \ldots, m$. We now let the lead times be any fixed, nonnegative integers.

Once we introduce lead times, installation inventories no longer give a complete description of the physical state of the system: We must record, as well, inventories in transit. As in Section 1, let R_n^i denote production at stage i in period n. The physical state is now

$$(I_n^i, R_{n-1}^i, \ldots, R_{n-\ell^i}^i)_{i=1}^m. \tag{18}$$

The variables $R_{n-\ell^i-1+j}^i$, $j = 1, \ldots, \ell^i$, indicate how much new inventory becomes available at stage $i - 1$ in the next ℓ^i periods.

Consider the system illustrated on the left side in Figure 1. There is a lead time of $\ell^i = 2$ between stages i and $i - 1$, illustrated by the line segment between the production facility (the circle) and the storage facility (the square) for stage i. Think of this line segment as being divided into $\ell^i = 2$ positions. A quantity R_{n-2}^i is placed in the first position during period $n - 2$; this is the stage-i production in that period. In period $n - 1$, the quantity R_{n-2}^i advances one position and a quantity R_{n-1}^i is placed in the first position. In period n, the quantity R_{n-2}^i arrives at the storage facility and increments I_n^i; the quantity R_{n-1}^i advances to the second lead time position; and R_n^i is placed in the first lead time position. Thus, stage-i production advances by one lead time position each period. The same holds for arbitrary ℓ^i.

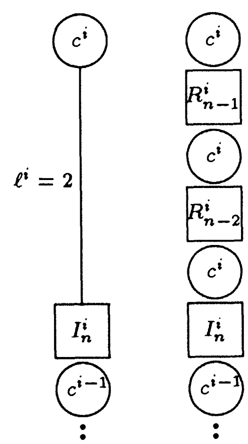

Figure 1. Replacing a lead time with dummy nodes. Stock advances by one dummy node each period, thus mimicking the effect of a lead time.

We reduce the operation of this system to one with additional stages but no lead times. This reduction rests on the following lemma, for a system without lead times.

Lemma 2. *Suppose that for some $i > 1$, we have $c^i = c^{i-1}$ and $s^i = s^{i-1}$. If $I_1^i \leq c^i$, then $R_n^{i-1} = I_n^i$ for all $n \geq 1$ and, consequently, $R_n^{i-1} = R_{n-1}^i$ for all $n \geq 2$. In other words, in each period stage $i - 1$ produces exactly as much as stage i produced in the previous period.*

Proof. Since stages i and $i - 1$ have the same base-stock level, the shortfall for echelon $i - 1$ equals the shortfall for echelon i plus the inventory between the two stages; i.e., $Y_n^{i-1} = Y_n^i + I_n^i$ for all n. In particular, in period 1 the echelon-($i - 1$) shortfall is at least I_1^i; so, $R_1^{i-1} = I_1^i$, under our hypothesis that $I_1^i \leq c^i = c^{i-1}$. Now suppose that $R_k^{i-1} = I_k^i$ for all $k = 1, \ldots, n - 1$. Then the only inventory between stages i and $i - 1$ at the start of period n is the previous period's production at stage i; that is, $I_n^i = R_{n-1}^i$, which is no greater than $c^{i-1} = c^i$. As noted, $Y_n^{i-1} \geq I_n^i$, so in period n, stage $i - 1$ produces as much as the supply of inventory allows; i.e., $R_n^{i-1} = I_n^i$. The first assertion is thus proved by induction. The second assertion follows: If stage $i - 1$ depletes its inventory in each period, then $R_n^{i-1} = R_{n-1}^i$ for all n.

With this result, we can mimic the operation of a system with fixed lead times using additional stages and no lead times. Introduce ℓ^i dummy stages between (genuine) stages i and $i - 1$, each having capacity c^i and base-stock level s^i. This augmented system operates as an ordinary serial system without lead times. (The right side of Figure 1 illustrates the insertion of dummy nodes.) From Lemma 2 we see that the effect of these dummy nodes is to advance production at stage i by one node each period. So, period-n production at stage i becomes available at stage $i - 1$ in period $n - \ell^i + 1$. This reproduces the effect of the lead time ℓ^i. The assumption in Lemma 2 that the initial inventory I_1^i does not exceed c^i is not a restriction: In the lead time model, the quantity one position downstream from stage i is just the previous period's production at stage i and so cannot exceed c^i.

As before, our analysis simplifies if we work with shortfalls rather than inventories. For $i = 1, \ldots, m$ and $j = 1, \ldots, \ell^i$, denote by Y^{ij} the echelon shortfall corresponding to the jth dummy node upstream from stage $i-1$ and let $Y^{i,\ell^i+1} = Y^i$ denote the echelon shortfall corresponding to stage i. Paralleling (3) we have

$$Y_n^{ij} = s^i - \sum_{k=1}^{i-1}\left(I_n^k + \sum_{r=0}^{\ell^k-1} R_{n-\ell^k-r}^k\right) - \left(I_n^i + \sum_{r=1}^{j-1} R_{n+1-\ell^i-r}^i\right) \qquad j = 1, \ldots, \ell^i+1,$$

and replacing (5) we have

$$R_{n-\ell^i-1+j}^i = Y_n^{i,j} - Y_n^{i,j+1} \quad j = 1, \ldots, \ell^i;$$
$$I_n^i = Y_n^{i-1} - Y_n^{i,1} + (s^i - s^{i-1}),$$

with the obvious modifications for $i = 1$. Since the system with dummy nodes has no lead times, the augmented set of shortfalls $\{(Y_n^{ij}, j = 1, \ldots, \ell^i+1, i = 1, \ldots, m), n \geq 0\}$ satisfies a recursion of exactly the type studied in previous sections. The next result thus follows.

Theorem 4. *If the lead times ℓ^i, $i = 1, \ldots, m$, are fixed, nonnegative integers, then Theorems 1 and 2 hold for the augmented shortfalls.*

As an immediate consequence, we have:

Corollary 4. *If the demands $\{D_n, n \geq 0\}$ are i.i.d. with $\mathsf{E}[D_0] < \min_i c^i$, then the Markov chain $\{(I_n^i, R_{n-1}^i, \ldots, R_{n-\ell^i}^i)_{i=1}^m, n \geq 0\}$ is Harris ergodic.*

If any of the lead times is strictly positive, then there are at least two stages in the augmented system with the same base-stock level: If, say, $\ell^i \geq 1$, then the ℓ^i dummy nodes between (genuine) stages i and $i-1$ all use level s^i. Consequently, condition (12) in Theorem 3 cannot be satisfied unless the demand distribution has mass at zero. In this setting, (12) is actually necessary for installation inventories to return infinitely often to their full-inventory level $(s^1, s^2 - s^1, \ldots, s^m - s^{m-1})$. For if $P(D_0 = 0) = 0$ and $\ell^i \geq 1$, then for all n, either $Y_n^i > 0$ or else there is inventory in transit through stage i.

5.2. Parallel Lead Times

We now consider a system with random lead times. Production in period n at stage i becomes available stock at stage $i-1$ after L_n^i periods; the sequence of vectors $\{(L_n^1, \ldots, L_n^m), n \geq 0\}$ is stationary. We refer to this mechanism as parallel lead times because different shipments do not interfere with each other. Overtaking is possible.

If the lead times were bounded, then with minor modification this model could be fit into the framework of subsection 5.1. Without this restriction, we need to introduce infinitely many dummy nodes between each pair of stages. The dummy capacities and base-stock levels are as in subsection 5.1. Also, as before, Y^{ij} denotes the echelon shortfall for the jth dummy node upstream from stage $i-1$, but now j has no upper bound. The stage-i shortfall is still Y^i. The state of the shortfalls in period n is $(Y_n^i; Y_n^{ij}, j = 1, 2, \ldots)_{i=1}^m$. Mimic the operation of the original system as follows. If, in the original system, a quantity of production R^i at stage i draws lead time L^i, then this quantity moves (at the end of the period) directly to the finished inventory at the (L^i+1)-st dummy node upstream from node $i-1$. The shortfalls Y^{ij}, $j = L^i+1, L^i+2, \ldots$, all drop by R^i, and the shortfalls Y^{ij}, $j = 1, \ldots, L^i$ remain unchanged. Subsequently, the quantity R^i advances by one dummy stage in each period, and so becomes available for use by genuine stage $i-1$ exactly L^i periods after it is completed at stage i.

Let us call an array $\{y^i; y^{ij}, j = 1, 2, \ldots; i = 1, \ldots, m\}$ of shortfalls *finite* if all entries are finite and if, in addition, all but finitely many increments $y^{ij} - y^{i,j+1}$ are zero. The second condition means that there are only finitely many dummy stages with inventory. We now have:

Theorem 5. *Suppose $\{(D_n, L_n), n \geq 0\}$ is stationary and ergodic. Suppose that $\mathsf{E}[D_0] < \min_i c^i$ and $\mathsf{E}[L_0^i] < \infty$, $i = 1, \ldots, m$. Then the augmented shortfall process $\{(Y_n^i; Y_n^{ij}, j = 1, 2, \ldots, i = 1, \ldots, m), n \geq 0\}$ has a unique finite stationary distribution to which it converges from all finite initial states.*

Proof. If Y_n denotes the array of period-n shortfalls, then Y_{n+1} is completely determined by Y_n, D_n, and L_n. Moreover, the mapping from (Y_n, D_n, L_n) to Y_{n+1} is component-wise increasing and continuous in Y_n for all values of D_n and L_n. It follows as in the system without lead times that there is a stationary process $\{\tilde{Y}_n, n \geq 0\}$, satisfying the same recursion, such that if $Y_0 = 0$, then $Y_n \Rightarrow \tilde{Y}_0$. We now argue that $\tilde{Y}_0$ is finite, almost surely.

The evolution of Y^m is still governed by the Lindley equation, so, under our stability condition, $\tilde{Y}_0^m$ is finite, almost surely. We claim that for all $j = 1, 2, \ldots, \tilde{Y}_0^{mj}$ is finite, almost surely. To prove this, we construct an auxilliary $G/G/\infty$ queue, modeling the movement of inventory from stage m to stage $m-1$. The queue evolves in discrete time. There is an arrival at time n precisely if there is production at stage m in period n; i.e., if $Y_n^m + D_n > 0$. The service time of the customer arriving at time n is L_n^m. The number of customers in this auxilliary queue is the number of shipments in progress from stage m to stage $m-1$, and multiplying the number in system by c^m gives an upper bound on the total inventory in transit between these stages. We know that Y^m couples with its stationary version in finite time. Hence, the arrival process to our infinite-server queue couples with a stationary version in finite time. For a $G/G/\infty$ system with stationary arrivals and service times and integrable service times, we know from Theorem 2.3.1 of Franken et al. (1982) that the queue has a unique finite stationary distribution (and couples with its stationary version in finite time).

Returning to the shortfalls, the event $\{\tilde{Y}_0^{mj} = \infty\}$ has probability zero or one if demands are ergodic. Since $\tilde{Y}_0^m$ is finite, the only way to have $\tilde{Y}_0^{mj}$ infinite is to have infinite inventory in dummy stages $j+1, j+2, \ldots$. However, this inventory is bounded by the auxilliary $G/G/\infty$ queue and cannot be infinite; so, $\tilde{Y}_0^{mj}$ is finite almost surely, $j = 1, 2, \ldots$.

We now argue that only finitely many $\tilde{Y}_0^{mj} - \tilde{Y}_0^{m,j+1}$, $j = 1, 2, \ldots$, are nonzero. Each such increment is at least as great as production in stage m in some previous period. When the stage-m shortfall is stationary, so is the sequence of production levels at stage m. But then for the total dummy-node inventory

$$\sum_{j=1}^{\infty} (\tilde{Y}_0^{m,j} - \tilde{Y}_0^{m,j+1})$$

to be finite, only finitely many terms can be nonzero. This further implies that for any finite Y_0, $(Y_n^{mj}, j = 1, 2, \ldots)$ couples with its stationary version in finite time: These shortfalls couple with those of an initially zero system in N periods, where N is the largest j for which $Y_0^{mj} < Y_0^{m,j-1}$; i.e., N is the largest index among the dummy nodes with inventory. This mirrors the fact that the $G/G/\infty$ queue couples with an initially-empty system once all customers present at time zero have departed.

Now consider stage $m-1$. Its evolution is governed by the dummy stage that immediately precedes it, with no intervening lead time. Moreover, we have shown that $\tilde{Y}_0^{m1}$ is finite. Hence, as in Theorem 1 we conclude that $\tilde{Y}_0^{m-1}$ must also be finite, almost surely.

We can now repeat the argument for stages $i = m-1, \ldots, 1$, but with one modification. At stage i, there is an arrival to the auxilliary $G/G/\infty$ queue in period n if the shortfall is strictly positive *and* inventory is available; i.e., $Y_n^i + D_n > 0$ and $Y_n^{i+1,1} < Y_n^i$. But this arrival process also couples with a stationary version (because $Y^{i+1,1}$ and Y^i do), so the argument still applies. We conclude, by induction, that all components of $\tilde{Y}_0$ are finite almost surely, and (by coupling) that this is the only finite stationary distribution and that Y_n converges to it from all finite Y_0.

When $\{(D_n, L_n), n \geq 0\}$ are i.i.d., $\{Y_n, n \geq 0\}$ is a Markov chain. Coupling, together with the existence of a stationary distribution, proves Harris ergodicity. The state space for $\{Y_n, n \geq 0\}$, $\mathbf{R}^{m\times\infty}$ is more complicated than those we considered earlier, but with the topology of component-wise convergence $\mathbf{R}^{m\times\infty}$ is metrizable as a complete, separable metric space (Billingsley 1968, p. 218) and this suffices for general results on Harris chains. As in our previous models, regeneration of the shortfalls implies regeneration of the inventory levels.

5.3. FIFO Lead Times

In our final model of lead times, each shipment from stage i must wait until all previous shipments from i have been transported before initiating its transition. This models a system in which a single vehicle moves stock between each pair of stages; the vehicle completes a roundtrip for each period in which the upstream stage has production. For period-n production at stage i, the roundtrip travel time is L_n^i. For this system, much stronger and less easily verified conditions are needed for stability. We use the notation of subsection 5.2.

Theorem 6. *Under the conditions of Theorem 5, there exists a stationary version $\{\tilde{Y}_n, n \geq 0\}$ of the shortfalls for which $Y_n \Rightarrow Y_0$ if $Y_0 = 0$. If, in addition,*

$$P(\tilde{Y}_0^i + D_0 > 0,\ \tilde{Y}_0^{i+1,1} + s^{i+1} - s^i < \tilde{Y}_0^i) < 1/\mathsf{E}[L_0^i],\quad i = 1, \ldots, m-1; \tag{19}$$

$$P(\tilde{Y}_0^m + D_0 > 0) < 1/\mathsf{E}[L_0^m], \tag{20}$$

then $\tilde{Y}_0$ is finite and $Y_n \Rightarrow \tilde{Y}_0$ for all finite initial conditions.

Proof. The argument of Theorem 5 applies with minor modification. The existence of $\{\tilde{Y}_n, n \geq 0\}$ and convergence to it from $Y_0 = 0$ is just as before. To establish finiteness, we now model the movement of stock between stages as a $G/G/1$ queue. The service times are the L_n^i's; there is an arrival to the queue between stages i and $i-1$ at time n if there is production at stage i in period n. The probabilities appearing in (19) and (20) are precisely the arrival rates to the auxilliary queues, so the inequalities there are just the familiar conditions for stability of these queues. When the auxilliary queues are stable, their stationary versions have finite queue lengths and couple in finite time with versions starting in any other state. Finiteness of the queue implies finiteness of the shortfalls, as in the proof of Theorem 5.

A shortcoming of Theorem 6 is that the probabilities in (19)–(20) are generally unknown. Ordinarily, we would expect them to be close to 1, and in any case 1 is a simple upper bound. This suggests that $\mathsf{E}[L_0^i]$ must typically be less than 1 for stability, implying that the lead times are often zero; i.e., less than one period.

The key step in Theorems 5 and 6 is bounding pipeline inventories through an auxilliary stationary system. Other models of lead times can be analyzed similarly. For a general discussion of stochastic lead times, see Zipkin (1986) and Svoronos and Zipkin (1991).

6. RANDOM ENVIRONMENTS AND PERIODIC DEMANDS

We now return to the basic model of Section 1 to consider systems with more general demand patterns and, correspondingly, more general production rules. Our new assumption is that demands are influenced by an environment that is itself subject to random fluctuations. Base-stock levels may be adjusted to changes in the environment.

We model the environment as a Markov chain with a general state space. This is no real restriction; rather, it means that the state of the environment is sufficiently rich to include all relevant information about the past. We first require the environment to be Harris ergodic, then allow it to be periodic, thus capturing, e.g., seasonal demand patterns.

Models of this type are not new to inventory theory. Iglehart and Karlin (1962) find optimal policies when the demand distribution is governed by a finite-state Markov chain. More recently, Song and Zipkin (1993) consider a countable-state Markov environment and show that an environment-dependent basestock policy is optimal for their cost structure. Song and Zipkin also discuss modeling applications and review related work.

6.1. Ergodic Environment

Throughout this section $\Theta = \{\Theta_n, n \geq 0\}$ is a Harris ergodic Markov chain representing the state of the world. Demands vary with Θ, so we let the base-stock levels vary too. Denote by $S_n = (S_n^1, \ldots, S_n^m)$ the vector of base-stock levels in period n. Our key assumption is that $(D_n, S_n) = g(\Theta_n)$ for some function g. In the terminology of Sigman, demands and base-stock levels are *governed* by the environment.

To define echelon shortfalls, we need to assume that the (now random) base-stock levels have upper bounds. Suppose, then, that there are constants $\bar{s}^i$, $i = 1, \ldots, m$, for which $S_n^i \leq \bar{s}^i$, almost surely, for all n and i. Define *virtual* shortfalls with respect to these upper bounds:

$$Y_n^i = \bar{s}^i - \sum_{j=1}^{i} I_n^j \quad i = 1, \ldots, m.$$

Production decisions at stage i are based on the *actual* shortfall $Y_n^i - \bar{s}^i + S_n^i$; this is the difference between the echelon-i inventory and the current base-stock level S_n^i. Production is set to try to reduce the actual shortfall to zero. A drop in the base-stock level from one period to the next can make the actual shortfall negative, whereas the virtual shortfalls can never be less than zero. Arguing just as in (6) and (7), we obtain

$$Y_{n+1}^m = \max\{\bar{s}^m - Y_n^m, Y_n^m + D_n - c^m\} \tag{21}$$

$$Y_{n+1}^i = \max\{\bar{s}^i - S_n^i, Y_n^i + D_n - c^i, Y_n^{i+1} + D_n - (\bar{s}^{i+1} - \bar{s}^i)\}. \tag{22}$$

We now give conditions for stability. Let $\{\tilde{\Theta}_n, -\infty < n < \infty\}$ be a stationary version of Θ and let $\tilde{\mathsf{E}}$ denote expectation with respect to this stationary version.

Theorem 7. *Suppose that the environment $\{\Theta_n, n \geq 0\}$ is a Harris ergodic Markov chain and that demands and base-stock levels are governed by Θ. Suppose the base-stock levels are bounded above. If $\tilde{\mathsf{E}}[S_0] < \infty$ and $\tilde{\mathsf{E}}[D_0] < \min_i c^i$, then $\{(Y_n, \Theta_n), n \geq 0\}$ is a Harris ergodic Markov chain.*

Proof. That $\{(Y_n, \Theta_n), n \geq 0\}$ is Markov follows from (21)–(22) and the fact that $(D_n, S_n) = g(\Theta_n)$, just as in Lemma 3.1 of Sigman. The result follows once we show that this Markov chain has a stationary distribution and admits coupling.

To construct a stationary distribution, drive the system with $\{\tilde{\Theta}_n, n \geq 0\}$, and thus stationary demands and base-stock levels. Equations 21 and 22 show that Y_{n+1} is increasing and continuous in Y_n for all values of the other arguments in these recursions. This shows that the distribution of Y_n converges to a stationary distribution $\tilde{Y}_0$ if $Y_0 = 0$; see Theorem 1. Moreover, $\tilde{Y}_0$ can be constructed on the same probability space as $\tilde{\Theta}$ to make $(\tilde{Y}_0, \tilde{\Theta}_0)$ stationary for $\{(Y_n, \Theta_n), n \geq 0\}$. The proof that $\tilde{Y}_0$ is finite proceeds much as in Theorem 1.

To show coupling, observe that because Θ is Harris ergodic there exists a finite random time N at which Θ couples with its stationary version. Subsequently, any two copies of Y driven by the same Θ are driven by the same stationary version. It suffices to show that any such copy of Y couples with one started at zero. At some finite $N_m > N$, the maximum in (21) is attained by the first term; otherwise Y^m would decrease to $-\infty$. Subsequently, Y^m agrees with a copy started at zero. Now proceed by induction on i from m down to 1. Some time after $(Y^{i+1}, \ldots, Y^m)$ has coupled with a copy started at zero, the maximum in (22) must be attained by either the first or the third term, and at that time Y^i couples.

Remark. A referee points out that the extension from Theorem 2 to Theorem 7 can be argued based on a general coupling result of Borovkov and Foss (1992).

6.2. Periodic Demands

Perhaps the greatest limitation of the usual assumption of demand stationarity is that it rules out seasonal or, more generally, periodic effects. We now introduce periodicity in demands through periodicity in the environment.

Theorem 8. *Let the conditions of Theorem 7 be in effect, except that now Θ is positive Harris recurrent with period $d \geq 1$. Then $\{(Y_n, \Theta_n), n \geq 0\}$ is positive Harris recurrent.*

Proof. As in subsection VI.3 of Asmussen or subsection 2.4 of Nummelin, the state space of Θ can be partitioned into d sets $E_1, \ldots, E_d$ such that $\Theta^i \equiv \{\Theta_{nd+i}, n \geq 0\}$ has state-space E_i, $i = 1, \ldots, d$. By Proposition 3.14 of Nummelin, each Θ^i is Harris recurrent on E_i, $i = 1, \ldots, d$, and is, in fact, positive Harris recurrent because Θ is positive Harris recurrent. By construction, each Θ^i is aperiodic and thus Harris ergodic. Now, just as in Theorem 7, $\{(Y_n, \Theta_n), n \geq 0\}$ has a stationary distribution $(\tilde{Y}_0, \tilde{\Theta}_0)$; the proof of this step did not use Harris ergodicity, just the existence of a stationary version of Θ. Let Z^i be the process $\{(Y_{nd+i}, \Theta_{nd+i}), n \geq 0\}$ on $\mathbf{R}_+^m \times E_i$, $i = 1, \ldots, d$. Then the distribution π^i defined by

$$\pi^i(\cdot) = P((\tilde{Y}_0, \tilde{\Theta}_0) \in \cdot | \tilde{\Theta}_0 \in E_i)$$

is stationary for Z^i. Moreover, $\{(Y_n, \Theta_n), n \geq 0\}$ admits coupling from any two initial states in $\mathbf{R}_+^m \times E_i$, because Θ admits coupling from any two states in E_i (by Harris ergodicity of Θ^i) and Y couples once Θ couples, as in Theorem 7. It follows that each Z^i, $i = 1, \ldots, d$ admits coupling and is thus Harris ergodic. Taking $\psi = \pi^1 + \cdots + \pi^d$ satisfies condition 10 and shows that $\{(Y_n, \Theta_n), n \geq 0\}$ is Harris recurrent. Since each π^i is finite, so is the stationary distribution of (Y, Θ), which is therefore positive Harris recurrent.

ACKNOWLEDGMENT

The first author is supported by the National Science Foundation under grant MSS-9216490.

REFERENCES

ASMUSSEN, S. 1987. *Applied Probability and Queues*. John Wiley, New York.

ASMUSSEN, S., AND H. THORISSON. 1987. A Markov Chain Approach to Periodic Queues. *J. Appl. Prob.* **24,** 215–225.

BACCELLI, F., AND Z. LIU. 1992. On a Class of Stochastic Recursive Sequences Arising in Queueing Theory. *Anns. Prob.* **20,** 350–374.

BACCELLI, F., W. A. MASSEY AND D. TOWSLEY. 1989. Acyclic Fork-Join Queueing Networks. *J. ACM* **36,** 615–642.

BOROVKOV, A., AND S. FOSS. 1992. Siberian Adv. in Math. **1,** 16–81.

BROCKWELL, P. J., S. I. RESNICK AND R. L. TWEEDIE. 1982. Storage Processes With General Release Rule and Additive Inputs. *Adv. Appl. Prob.* **14,** 392–433.

BILLINGSLEY, P. 1968. *Convergence of Probability Measures*. John Wiley, New York.

CLARK, A. J., AND H. SCARF. 1960. Optimal Policies for a Multi-Echelon Inventory Problem. *Mgmt. Sci.* **6,** 475–490.

FEDERGRUEN, A., AND P. ZIPKIN. 1984. Computational Issues in an Infinite-Horizon Multiechelon Inventory Model. *Opns. Res.* **32,** 818–836.

FEDERGRUEN, A., AND P. ZIPKIN. 1986a. An Inventory Model With Limited Production Capacity and Uncertain Demands, I: The Average Cost Criterion. *Math. Opns. Res.* **11,** 193–207.

FEDERGRUEN, A., AND P. ZIPKIN. 1986b. An Inventory Model With Limited Production Capacity and Uncertain Demands, II: The Discounted Cost Criterion. *Math. Opns. Res.* **11,** 208–215.

FRANKEN, P., D. KONIG, U. ARNDT AND V. SCHMIDT. 1982. *Queues and Point Processes*. John Wiley, Chichester, England.

GLASSERMAN, P., AND S. TAYUR. 1992a. Sensitivity Analysis for Base-Stock Levels in Multi-Echelon Production-Inventory Systems. *Mgmt. Sci.* (to appear).

GLASSERMAN, P., AND S. TAYUR. 1992b. The Stability of a Capacitated, Multi-Echelon Production-Inventory System Under a Base-Stock Policy, Working Paper 92-36, GSIA, Carnegie Mellon University, Pittsburgh, Penn.

GLASSERMAN, P., AND D. D. YAO. 1992. The Stochastic Vector Equation $Y_{n+1} = A_n \otimes Y_n \oplus B_n$ With Stationary Coefficients. *Adv. Appl. Prob.* (to appear).

GLYNN, J. E. 1989. A Discrete Time Storage Process With a General Release Rule. *J. Appl. Prob.* **26,** 566–583.

GRAVES, S., A. RINNOOY KAN AND P. ZIPKIN (EDS.). 1992. *Logistics of Production and Inventory*, Vol. 4. North-Holland, Amsterdam.

IGLEHART, D. L., AND S. KARLIN. 1962. Optimal Policy for Dynamic Inventory Process With Nonstationary Stochastic Demands. In *Studies in Applied Probability and Management Science*, K. Arrow, S. Karlin and H. Scarf (eds.). Stanford University Press, Stanford, Calif.

LOYNES, R. M. 1962. The Stability of a Queue With Non-Independent Interarrival and Service Times. *Proc. Cambridge Philos. Soc.* **58,** 497–520.

NUMMELIN, E. 1984. *General Irreducible Markov Chains and Non-Negative Operators*. Cambridge University Press, London.

PRABHU, U. 1965. *Queues and Inventories*. John Wiley, New York.

ROSLING, K. 1989. Optimal Inventory Policies for Assembly Systems Under Random Demands. *Opns. Res.* **37,** 565–579.

SIGMAN, K. 1988. Queues as Harris Recurrent Markov Chains. *QUESTA* **3,** 179–198.

SONG, J. S., AND P. ZIPKIN. 1993. Inventory Control in a Fluctuating Demand Environment. *Opns. Res.* **41,** 351–370.

SVORONOS, A., AND P. ZIPKIN. 1991. Evaluation of One-for-One Replenishment Policies for Multiechelon Inventory Systems. *Mgmt. Sci.* **37,** 68–83.

TAYUR, S. 1992. Computing the Optimal Policy for Capacitated Inventory Models. *Comm. Statist. Stoch. Models* **9,** 585–598.

THORISSON, H. 1983. The Coupling of Regenerative Processes. *Adv. Appl. Prob.* **15,** 531–561.

VAN HOUTUM, G. J., AND W. H. M. ZIJM. 1990. Computational Procedures for Stochastic Multi-Echelon Production Systems. Centre for Quantitative Methods, Nederlandse Philips Bedrijven, B.V.

ZIPKIN, P. 1986. Stochastic Leadtimes in Continuous Time Inventory Models. *Naval Res. Logist. Quart.* **33,** 763–774.