Analysis of a Decentralized Production-Inventory System

René Caldentey • Lawrence M. Wein
Stern School of Business, New York University, New York, New York 10012
Graduate School of Business, Stanford University, Stanford, California 94306
rcaldent@stern.nyu.edu • lwein@stanford.edu

We model an isolated portion of a competitive supply chain as a $M/M/1$ make-to-stock queue. The retailer carries finished goods inventory to service a Poisson demand process, and specifies a policy for replenishing his inventory from an upstream supplier. The supplier chooses the service rate, i.e., the capacity of his manufacturing facility, which behaves as a single-server queue with exponential service times. Demand is backlogged and both agents share the backorder cost. In addition, a linear inventory holding cost is charged to the retailer, and a linear cost for building production capacity is incurred by the supplier. The inventory level, demand rate, and cost parameters are common knowledge to both agents. Under the continuous-state approximation where the $M/M/1$ queue has an exponential rather than geometric steady-state distribution, we characterize the optimal centralized and Nash solutions, and show that a contract with linear transfer payments replicates a cost-sharing agreement and coordinates the system. We also compare the total system costs, the agents' decision variables, and the customer service levels of the centralized versus Nash versus Stackelberg solutions.
(*Make-to-Stock Queue; Game Theory*)

1. Introduction

Within many supply chains, a devoted upstream agent, referred to here as the *supplier*, produces goods for a downstream agent, called the *retailer*, in a make-to-stock manner. Broadly speaking, the performance (e.g., service levels, cost to produce and hold items) of this isolated portion of the supply chain is dictated by three factors: (i) retailer demand, which is largely exogenous but can in some cases be manipulated via pricing and advertising, (ii) the effectiveness of the supplier's production process and the subsequent transportation of goods, and (iii) the inventory replenishment policy, by which retailer demand is mapped into orders placed with the supplier. If the supplier and retailer are under different ownership or are independent entities within the same firm, then their competing objectives can lead to severe coordination problems: The supplier typically wants the retailer to hold as much inventory as possible, while the retailer prefers to hold very little inventory and desires rapid response from the supplier. These tensions may deteriorate overall system performance.

The recent explosion in the academic supply chain management literature is aimed at this type of multi-agent problem. Almost without exception, the papers that incorporate stochastic demand employ variants of one of two prototypical operations management models: the newsvendor model, or the Clark-Scarf (1960) multiechelon inventory model. One-period and two-period versions of newsvendor supply chain models have been studied intensively to address the three factors above; see Agrawal et al. (1999), Cachon (1999), and Lariviere (1999), for recent reviews. Although many valuable insights have been generated by this work, these models are primarily useful for style goods and products with very short

1523-4614/03/0501/0001$05.00
1526-5498 electronic ISSN

life cycles. More complex (multiperiod, and possibly multiechelon and positive lead time) supply chain models have been used to analyze the case where a product experiences ongoing production and demand. Of the three factors in the last paragraph, these multiperiod supply chain models successfully capture the replenishment policy and have addressed some aspects of retailer demand, for example, information lead times in the Clark-Scarf model (Chen 1999), pricing in multiechelon models with deterministic demand and ordering costs (Chen et al. 2001), and forecast updates (Anupindi and Bassok (1999) in a multiperiod newsvendor model, and Tsay and Lovejoy (1999) in a multistage model). However, the Clark-Scarf model simplifies the supplier's production process by assuming that lead times are independent of the ordering process, or equivalently, that the production process is an infinite-server queue.

In this paper, we use an alternative prototypical model, an $M/M/1$ make-to-stock queue, to analyze a supply chain. Here, the supplier is modeled as a single-server queue, rather than an infinite-server queue, and the retailer's optimal inventory replenishment strategy is a base-stock policy. Because the production system is explicitly incorporated, these make-to-stock queues are also referred to as *production-inventory systems*. The $M/M/1$ make-to-stock queue was introduced by Morse (1958), but lay mysteriously dormant for the next three decades, perhaps because the multiechelon version of it lacked the attractive decomposition property of the Clark-Scarf model and traditional (i.e., make-to-order) queueing networks, except under some restrictive inventory policies (Rubio and Wein 1996). Make-to-stock queueing systems have experienced a revival in the 1990s, including multiproduct queues with (e.g., Federgruen and Katalan 1996, Markowitz et al. 2000) and without (e.g., Zheng and Zipkin 1990, Wein 1992) setups, and single-product, multistage systems in continuous time (e.g., Buzacott et al. 1992, Lee and Zipkin 1992) and discrete time (e.g., Glasserman and Tayur 1995 and Gavirneni et al. 1996, building on earlier work by Federgruen and Zipkin 1986). Although these papers either undertake a performance analysis or consider a centralized decision maker (Gaverneni et al. (1996) analyze their system under various informational structures, but not in a game-theoretic setting), the make-to-stock queue is amenable to a competitive analysis because it explicitly captures the trade-off between the supplier's capacity choice and the retailer's choice of base-stock level. However, the model treats the third key factor in a naive way, by assuming that retailer demand is an exogenous Poisson process. Moreover, we assume that the system state, the demand rate and the cost parameters are known by each agent. While this assumption is admittedly crude, we believe it is an appropriate starting point for exploring competitive make-to-stock queues. In the only other contemporaneous multiagent production-inventory study that we are aware of, Plambeck and Zenios (1999) take a significant step forward by analyzing a dynamic system with information asymmetry. Subsequent work includes Cachon (1999b), where both agents choose base-stock levels in a single-stage lost sales model, and Duenyas and Tsai (2001), where two profit-maximizing agents manage each stage of a tandem system that incurs demand for intermediate and end products.

In an attempt to isolate—and hence understand—the impact of incorporating capacity into a supply chain model, we intentionally mimic Cachon and Zipkin (1999). Their two-stage Clark-Scarf model is quite similar to our $M/M/1$ make-to-stock queue: Both models have two players, assume linear backorder and holding costs for retailer inventory (where the backorder costs are shared by both agents), employ steady-state analysis, and ignore fixed ordering costs. The key distinction between the two models is that the production stage is an infinite-server queue and the supplier controls his (local or echelon) inventory level in Cachon and Zipkin (1999), whereas in our paper, the production stage is modeled as a single-server queue and the supplier controls the capacity level, which in turn dictates a steady-state lead-time distribution. While Cachon and Zipkin's supplier incurs a linear inventory holding cost, our supplier is subjected to a linear capacity cost. A minor difference is that our queueing model is in continuous time, while Cachon and Zipkin's inventory model is in discrete time. In fact, to make our results more transparent and to maintain a closer match of the two models, we use a continuous-state approximation, essentially replacing the geometric steady-state distribution of the $M/M/1$ queue by an exponential distribution with the same mean.

After defining the model in §2, we derive the centralized solution in §3, where a single decision maker optimizes system performance, and the unique Nash equilibrium in §4, where the supplier and retailer minimize their own cost. The two solutions are compared in §5. We show that the Nash solution is always inefficient and that the magnitude of the inefficiency tends to be minimized when backorder costs are evenly split between the supplier and the retailer. We also analyze the Nash equilibrium in terms of service performance that is measured by the fraction of customers that are backlogged. The Nash solution provides a poorer service than the centralized one and the difference is minimized when the retailer absorbs most of the backorder costs. In this respect, we conclude that customers prefer that the penalty for shortages be absorbed primarily by the agent in direct contact with them. In §6, we describe a family of contracts based on transfer payments between the two players that coordinate the system; for example, it allows the decentralized system to achieve the same cost as the centralized system. The payments are selected in such a way that each player transfers a fixed fraction of its own cost to the other player. The resulting cost structure mimics a cost-sharing agreement and thus coordinates the system. In §7, we analyze the Stackelberg games, where one agent has all the bargaining power. In terms of efficiency, the retailer's Stackelberg game dominates the supplier's Stackelberg game. However, from a customer service perspective, the supplier's Stackelberg game is preferable. Concluding remarks, including a comparison of our results to those of Cachon and Zipkin (1999), are presented in §8.

2. The Model

Our idealized supply chain consists of a supplier providing a single product to a retailer. Retailer demand is modeled as a homogeneous Poisson process with rate λ. The retailer carries inventory to service this demand, and unsatisfied demand is backordered. Because we assume that there are no fixed ordering costs, the retailer's optimal replenishment policy (given the supplier's irreversible capacity decision, described below) is characterized by a $(s-1, s)$ base-stock policy. That is, the inventory initially contains s units, and the retailer places an order for one unit with the supplier at each epoch of the Poisson demand process.

The supplier's production facility is modeled as a single-server queue with service times that are exponentially distributed with rate μ. The supplier is responsible for choosing the parameter μ, which will also be referred to as the *capacity*. The server is only busy when retailer orders are present in the queue. The supplier's facility behaves as a $M/M/1$ queue because the demand process is Poisson and a base-stock policy is used. This setting, in which the supplier does not store finished goods, often holds when both agents are within the same company, or the supplier's sole customer is the retailer.

In our model, the selling price that the retailer charges the customer, the wholesale price that the retailer pays to the supplier, and the supplier's variable production cost, are assumed to be given and constant. These conditions hold, for example, if we assume that the retailer and supplier operate in competitive markets. Because demand is exogenous, unsatisfied demand is backordered, and we use a long-run average cost criterion, it follows that the agents' revenues are independent of their actions, as long as their utilities are nonnegative. Hence, profit maximization and cost minimization lead to the same solution, and we employ a cost-minimization framework.

Each backordered unit generates a cost b per unit of time for the production-inventory system. As in Cachon and Zipkin (1999), this backorder cost is split between the two agents, with a fraction $\alpha \in [0, 1]$ incurred by the retailer. The parameter α, which we refer to as the *retailer's backorder share*, is exogenously specified in our model. Because much of the academic literature assumes $\alpha = 1$, this assumption requires some discussion. The value of α depends on a variety of factors, such as the structures of the market and distribution channel, and the customers' expectations (e.g., Stern et al. 1996). At one extreme, if the supplier has a monopolistic (or well established) position in the market and uses an *intensive* distribution strategy with many competing retailers, then customers facing an inventory shortage at one retailer may buy the product from a different retailer. In this case, α will be near 1 and the retailer will bear the brunt of the backorder costs. At the other extreme, suppose the supplier is in a competitive market and either the retailer has a monopoly at the distribution level or the supplier uses an exclusive distribution strategy with a single retailer. In this situation, poor customer service at the retailer level will mostly harm the supplier; for example, α will be near 0. Of course, various intermediate situations generate less extreme values of α.

In addition, the retailer incurs a cost for holding units in inventory. The retailer's holding cost includes both physical and financial components, and is, in general, not the same in the centralized and decentralized systems. More specifically, the financial holding cost is proportional to the retailer's purchase cost, which is the supplier's production cost in the centralized system and the sum of the production cost and the supplier's profit margin in the decentralized system. Therefore, we expect the holding cost to be larger in the decentralized case, which is a source of inefficiency with respect to the centralized chain (e.g., Gallego and Boyaci 2002). For simplicity, however, we do not make this distinction here. That is, we assume that there is single holding cost h per unit of inventory per unit of time paid by the retailer independent of the supply chain configuration. This assumption is reasonable if the supplier's margin is small, which may hold if the supplier and retailer belong to the same company or are in industries in which physical holding costs dominate financial holding costs and/or there is a competitive market of low-margin suppliers.

On the other hand, the supplier pays the fixed cost of building production capacity. The capacity cost parameter c is per unit of product, so that $c\mu$ represents the amortized cost per unit of time that the supplier incurs for having the capacity μ; this fixed cost rate is independent of the demand level.

To make our results more transparent, we normalize the expected variable cost per unit time by dividing it by the holding cost rate h. Toward this end, we normalize the cost parameters as follows:

$$\tilde{h} = \frac{h}{h} = 1, \qquad \tilde{b} = \frac{b}{h}, \qquad \tilde{c} = \frac{\lambda c}{h}. \tag{1}$$

To ease the notation, we hereafter omit the tildes from these cost parameters.

Let N be the steady-state number of orders at the supplier's manufacturing facility. If we assume for now that $\mu > \lambda$ (this point is revisited later), then N is geometrically distributed with mean ν^{-1}, where

$$\nu = \frac{\mu - \lambda}{\lambda}. \tag{2}$$

This parameter, which represents the normalized excess capacity, is the supplier's decision variable in our analysis, and we often refer to it simply as capacity. To simplify our analysis, we assume that N is a continuous random variable, and replace the geometric distribution by an exponential distribution with parameter ν. This continuous-state approximation can be justified by a heavy traffic approximation (e.g., §10 of Harrison 1988), and generates mean queue lengths that coincide with $M/M/1$ results for all server utilization levels. The heavy traffic approximation allows the incorporation of general interarrival time and service time distributions; in the nonexponential case, ν in Equation (2) would be divided by one-half of the sum of the squared coefficients of variation of the interarrival and service time distributions, $(c_a^2 + c_s^2)/2$, but we do not pursue this avenue here. Although this continuous-state approximation leads to slightly different quantitative results (the approximation tends to underestimate the optimal discrete base-stock level), it has no effect on the qualitative system behavior, which is the object of our study.

The steady-state expected normalized variable cost per unit time for the risk-neutral retailer (C_R) and supplier (C_S) in terms of the two decision variables are given by

$$\begin{aligned} C_R(s, \nu) &= E[(s-N)^+] + \alpha\, b\, E[(N-s)^+] \\ &= s - \frac{1-e^{-\nu s}}{\nu} + \alpha b \frac{e^{-\nu s}}{\nu}, \end{aligned} \tag{3}$$

and

$$\begin{aligned} C_S(s, \nu) &= (1-\alpha)\, b\, E[(N-s)^+] + c\nu \\ &= (1-\alpha) b \frac{e^{-\nu s}}{\nu} + c\nu. \end{aligned} \tag{4}$$

3. The Centralized Solution

As a reference point for the efficiency of the two-agent system, we start by finding the optimal solution to the centralized version of the problem, where there is a single decision maker that simultaneously optimizes the base-stock level s and the normalized excess capacity ν. The steady-state expected normalized cost per unit time C for this decision maker is

$$C(s,\nu) = C_R(s,\nu) + C_S(s,\nu) = c\nu + s - \frac{1-(b+1)e^{-\nu s}}{\nu}. \tag{5}$$

The centralized solution is given in Proposition 1; see the Appendix for the proof.

Proposition 1. *The optimal centralized solution is the unique solution to the first-order conditions*

$$\frac{\partial C(s,\nu)}{\partial s} = 0 \Longleftrightarrow \nu s = \ln(1+b), \tag{6}$$

$$\frac{\partial C(s,\nu)}{\partial \nu} = 0 \Longleftrightarrow -(b+1)(\nu s+1)\frac{e^{-\nu s}}{\nu^2} + \frac{1}{\nu^2} + c = 0, \tag{7}$$

and is given by

$$\nu^* = \sqrt{\frac{\ln(1+b)}{c}} \quad \text{and} \quad s^* = \sqrt{c\ln(1+b)}. \tag{8}$$

The resulting cost is

$$C(s^*,\nu^*) = 2\sqrt{c\ln(1+b)}. \tag{9}$$

By relation (6), the ratio of the base-stock level, s, to the supplier's mean queue length, ν^{-1}, satisfies $\nu s = \ln(1+b)$ at optimality. (The corresponding first-order conditions for the discrete inventory problem is $\ln(\nu+1)s = \ln(1+b)$, and so our continuous approximation can be viewed as using the Taylor series approximation $\ln(\nu+1) \approx \nu$.) Although this ratio is independent of the capacity cost c, the optimal solution depends on c via $s = \nu c$ according to Equation (8).

As expected, the optimal capacity level decreases with the capacity cost and increases with the backorder-to-holding cost ratio b. Similarly, because capacity and safety stock provide alternative means to avoid backorders, the optimal base-stock level increases with the capacity cost and with the normalized backorder cost b. Finally, as expected, α plays no role in this single-agent optimization, because transfer payments between the retailer and the supplier do not affect the centralized cost.

4. The Nash Solution

Under the Nash equilibrium concept, the retailer chooses s to minimize $C_R(s,\nu)$, assuming that the supplier chooses ν to minimize $C_S(s,\nu)$; likewise, the supplier simultaneously chooses ν to minimize $C_S(s,\nu)$, assuming the retailer chooses s to minimize $C_R(s,\nu)$. Because each agent's strategy is a best response to the other's, neither agent is motivated to depart from this equilibrium.

In anticipation of subsequent analysis, we express the Nash equilibrium in terms of the retailer's backorder share α. Let us also define the auxiliary function

$$f_\alpha(b) = \sqrt{\frac{(1-\alpha)b(\ln(1+\alpha b)+1)}{(1+\alpha b)\ln(1+b)}}, \tag{10}$$

which plays a prominent role in our analysis.

Proposition 2. *The unique Nash equilibrium is*

$$\nu^*_\alpha = f_\alpha(b)\nu^*, \tag{11}$$

$$s^*_\alpha = \left(\frac{\ln(1+\alpha b)}{\ln(1+b)f_\alpha(b)}\right)s^*. \tag{12}$$

The resulting costs, $C^*_R(\alpha)$ *and* $C^*_S(\alpha)$*, are*

$$C^*_R(\alpha) = C_R(s^*_\alpha,\nu^*_\alpha) = s^*_\alpha, \tag{13}$$

$$C^*_S(\alpha) = C_S(s^*_\alpha,\nu^*_\alpha) = \left(\frac{\ln(1+\alpha b)+2}{\ln(1+\alpha b)+1}\right)c\nu^*_\alpha. \tag{14}$$

Proof. Let $s^*(\nu)$ be the retailer's reaction curve, for example, the optimal base-stock level given a capacity ν installed by the supplier. Because Equation (3) is concave in s, $s^*(\nu)$ is characterized by the first-order condition

$$\nu s^*(\nu) = \ln(1+\alpha b). \tag{15}$$

Using a similar argument, we find that the supplier's reaction curve $\nu^*(s)$ satisfies

$$e^{-\nu^*(s)s}\left(\frac{\nu^*(s)s+1}{(\nu^*(s)s)^2}\right) = \frac{c}{(1-\alpha)bs^2}. \tag{16}$$

The unique solution to Equations (15)–(16) is Equations (11)–(12), and substituting this solution into Equations (3)–(4) yields Equations (13)–(14). □

Because $f_\alpha(b)$ is decreasing in α and $\ln(1+\alpha b)$ is increasing in α for $b>0$, it follows that as α increases, the retailer becomes more concerned with backorders and increases his base-stock level, while the supplier cares less about backorders and builds less excess capacity.

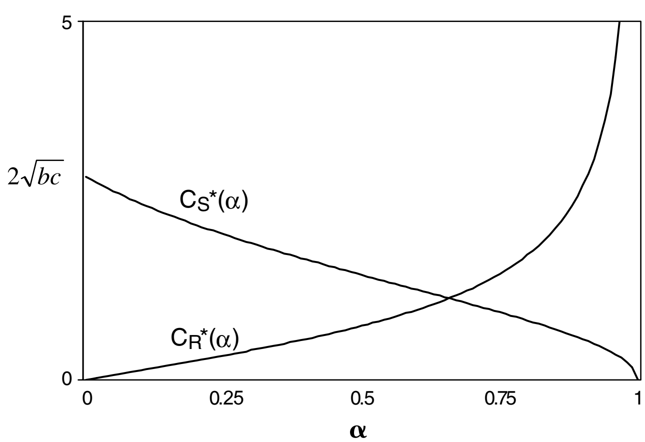
The supplier's variable cost $C_S^*(\alpha)$ is a monotonically decreasing function of α that satisfies $C_S^*(0)=2\sqrt{bc}$ and $C_S^*(1)=0$, and $C_R^*(\alpha)$ is a monotonically increasing function of α that satisfies

$$C_R^*(0)=0, \quad \lim_{\alpha\to 1} C_R^*(\alpha)\to +\infty, \tag{17}$$

as shown in Figure 1. (Many of the limits taken in this paper, for example, $\alpha\to 1$, are implicitly taken to be one-sided.)

To understand the unbounded retailer losses in Equation (17), note that for the extreme case $\alpha=1$, the supplier does not face any backorder cost and consequently has no incentive to build excess capacity, for instance, $\nu_1^*=0$. This corresponds to the null recurrent case of a queueing system with an arrival rate equal to its service rate. Consequently, the retailer is unable to maintain finite inventory (backorder plus holding) costs, and the supply chain is not operational in the Nash equilibrium.

Figure 1 **The Retailer's (C_R) and Supplier's (C_S) Costs in the Nash Equilibrium as a Function of the Retailer's Backorder Share α**



5. Comparison of Solutions

In this section, we compare the centralized solution and the Nash equilibrium with respect to the total system cost, the agents' decisions, and the customer service level.

The Nash Equilibrium Is Inefficient. The centralized solution is not achievable as a Nash equilibrium. By Equations (6) and (15), the first-order conditions are $\nu s=\ln(1+b)$ in the centralized solution and $\nu s=\ln(1+\alpha b)$ in the Nash solution. Hence, the two solutions are not equal when $\alpha<1$, and the Nash equilibrium in the $\alpha=1$ case is an unstable system, as discussed earlier. The source of this inefficiency is the inability of the agents to fully replicate the centralized cost structure.

The magnitude of the inefficiency of a Nash equilibrium is typically quantified by comparing the costs under the centralized and Nash solutions. We follow Cachon and Zipkin (1999) and compute the *competition penalty*, which is defined as the percentage increase in variable cost of the Nash equilibrium over the centralized solution. By Equations (5) and (8), the variable cost for the centralized solution is

$$\begin{aligned} C(s^*,\nu^*) &= \left(s^*-\frac{1-e^{-\nu^*s^*}}{\nu^*}\right)+b\frac{e^{-\nu^*s^*}}{\nu^*}+c\nu^* \\ &= 2\sqrt{c\ln(1+b)}, \end{aligned}$$

and the variable cost C_α^* associated with the Nash equilibrium is, by Equations (3)–(4) and Proposition 2,

$$\begin{aligned} C_\alpha^* &= C_S(s_\alpha^*,\nu_\alpha^*)+C_R(s_\alpha^*,\nu_\alpha^*) \\ &= \left[f_\alpha(b)\left(\frac{\ln(1+\alpha b)+2}{\ln(1+\alpha b)+1}\right)+\frac{\ln(1+\alpha b)}{\ln(1+b)f_\alpha(b)}\right]s^*. \end{aligned}$$

Hence, the competition penalty is

$$\frac{C_\alpha^*-C(s^*,\nu^*)}{C(s^*,\nu^*)}\times 100\%,$$

where

$$\begin{aligned} P_\alpha(b) &\triangleq \frac{C_\alpha^*-C(s^*,\nu^*)}{C(s^*,\nu^*)} \\ &= \frac{1}{2}\left[f_\alpha(b)\left(\frac{\ln(1+\alpha b)+2}{\ln(1+\alpha b)+1}\right)\right. \\ &\qquad \left.+\frac{\ln(1+\alpha b)}{\ln(1+b)f_\alpha(b)}\right]-1. \end{aligned} \tag{18}$$

Surprisingly, the competition penalty in Equation (18) is independent of the supplier's cost of capacity. This occurs because the centralized variable cost and the Nash variable cost are both proportional to $\sqrt{c}$ at optimality, which is a consequence of the particular functional form arising from the make-to-stock formulation. This penalty is a function of α and b, and we can simplify Equation (18) for the limiting values of these two parameters. The function $f_\alpha(b)$ is decreasing in α and $f_1(b) = 0$. Hence, the competition penalty goes to ∞ as $\alpha \to 1$. This inefficiency occurs because, as the retailer bears more of the backorder cost, the supplier builds less excess capacity, and in the limit the lack of excess capacity causes instability of the queueing system. At the other extreme, $f_\alpha(b) \to \sqrt{b/(\ln(1+b))}$ as $\alpha \to 0$, and the competition penalty in this case is given by

$$\sqrt{\frac{b}{\ln(1+b)}} - 1 \quad \text{for} \quad b > 0. \tag{19}$$

This function is increasing in b, approaches zero as $b \to 0$ and grows to ∞ as $b \to \infty$. Hence, when the supplier incurs most of the backorder cost, the retailer holds very little inventory and the competition penalty depends primarily on the backorder cost; if this cost is low then the supplier has little incentive to build excess capacity, which leads to a small competition penalty because the centralized planner holds neither safety stock nor excess capacity in this case. In contrast, if the backorder cost is very high, the supplier cannot overcome the retailer's lack of safety stock, and his backorders get out of control, leading to high inefficiency.

Turning to the backorder cost asymptotics, $f_\alpha(b) \to \sqrt{(1-\alpha)/\alpha}$ as $b \to \infty$, and the competition penalty approaches

$$\frac{1}{2\sqrt{\alpha(1-\alpha)}} - 1. \tag{20}$$

This quantity vanishes at $\alpha = 0.5$, is symmetric about $\alpha = 0.5$, and approaches ∞ as $\alpha \to 0$ and $\alpha \to 1$. Thus, when backorders are very expensive, this cost component dominates both agents' objective functions when they care equally about backorders ($\alpha = 0.5$), and their cost functions—and hence decisions—coincide with the centralized solution. However, when there is a severe imbalance in the backorder allocation (α is near 0 or 1), one of the agents does not build enough of his buffer resource, and the other agent cannot prevent many costly backorders, which is highly inefficient from the viewpoint of the entire supply chain. Finally, for the case $b \to 0$, the competition penalty is given by

$$\frac{2-\alpha}{2\sqrt{1-\alpha}} - 1, \tag{21}$$

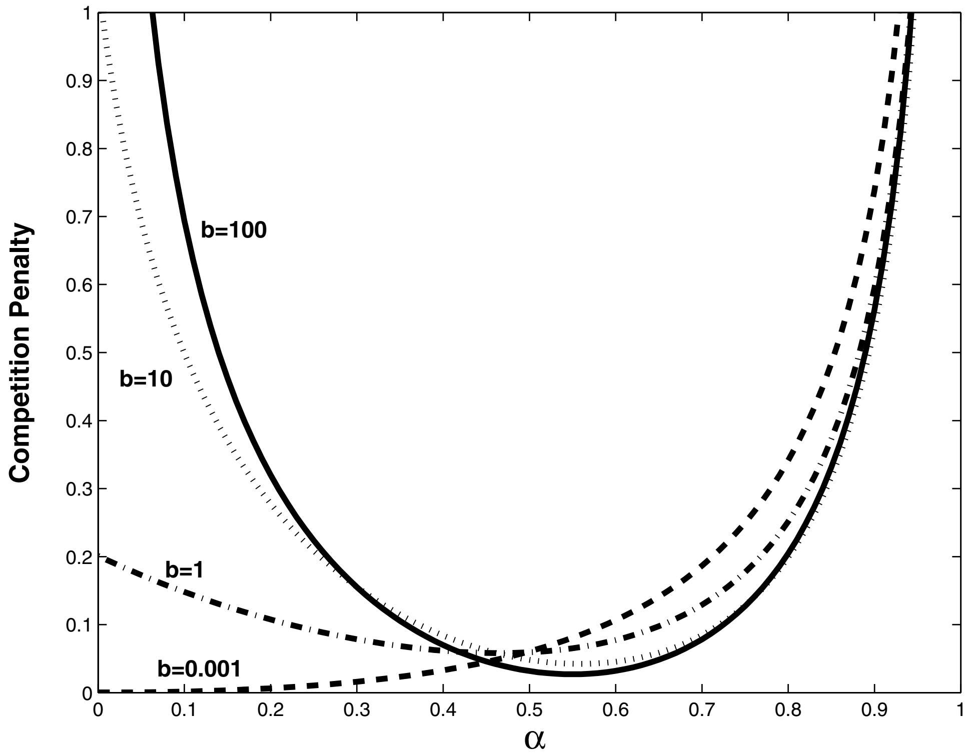
which is an increasing function of α. Consistent with the previous analysis, this penalty function vanishes as $\alpha \to 0$ and approaches ∞ as $\alpha \to 1$.

In summary, there are two regimes, ($\alpha = 0.5, b \to \infty$) and ($\alpha \to 0, b \to 0$), where the Nash equilibrium is asymptotically efficient, and two regimes, $\alpha \to 1$ and ($\alpha \to 0, b \to \infty$), where the inefficiency of the Nash solution is arbitrarily large. However, because Equation (18) does not consider the agents' participation constraints, some of the large inefficiencies in the latter regimes may not be attainable by the supply chain (see Caldentey and Wein 1999).

To complement these asymptotic results, we plot in Figure 2 the competition penalty in Equation (18) for various values of b as a function of α. A new insight emerges from Figure 2. The competition penalty is minimized by α near 0.5 (i.e., the backorder cost is split evenly) when the backorder-to-holding cost ratio $b \geq 1$.

Comparison of Decision Variables. Figure 3 depicts the optimal Nash production capacity ν_α^* and the optimal Nash base-stock level s_α^* as a function of α, and allows us to compare these functions to the centralized solutions, ν^* and s^*. Excess capacity and the base-stock level are alternative ways for the supplier and retailer, respectively, to buffer against demand uncertainty, and Figure 3 shows that the inefficiency of the Nash solution does not necessarily imply that these agents have less buffer resources in the Nash solution than in the centralized solution. For both decision variables, there exist thresholds on the value of α, denoted by α_s and α_ν in Figure 3, that divide the regions where the agents have more or less buffer resources than the optimal centralized solution. However, as shown in the next proposition, at least one

Figure 2 **The Competition Penalty in Equation (18) in the Absence of Participation Constraints for Different Values of the Backorder-to-Holding Cost Ratio b as a Function of α**



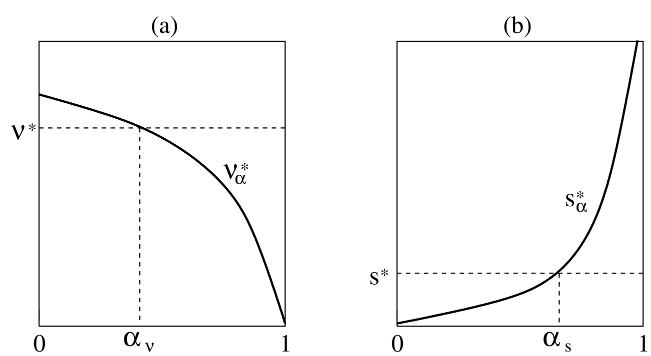
agent in the Nash equilibrium possesses less of his buffer resource than the central planner.

Proposition 3. *For α_s and α_ν defined in Figure* 3, *we have $\alpha_s > \alpha_\nu$.*

Proof. By Figure 3, if $\alpha_s \le \alpha_\nu$, then there exists $\hat{\alpha} \in [\alpha_s, \alpha_\nu]$ such that $\nu^*_{\hat{\alpha}} s^*_{\hat{\alpha}} \ge \nu^* s^*$. However, this inequality together with Equations (6) and (15) implies that $f_{\hat{\alpha}}(b) = \ln(1+b)$, for example, $\hat{\alpha} = 1$. But for $\alpha = 1$ the supply chain is unstable and does not operate. Hence, $\nu^*_\alpha s^*_\alpha < \nu^* s^*$ for $\alpha \in [0, 1)$, and consequently $\alpha_s > \alpha_\nu$. □

The parameters α_s and α_ν can be used in the following way. Let α^* be the value of the retailer's backorder share that minimizes the competition penalty in Equation (18). Then computational experiments show that for $b \ge 1$, the approximation $\alpha^* \approx (\alpha_s + \alpha_\nu)/2$ has an absolute error bounded by 0.02. Thus, in the case where the backorder cost is well quantified (e.g., a price discount for waiting, or the financial cost of delayed payments), knowledge of α_s and α_ν can be used to negotiate a contract that shifts the retailer's backorder share towards α^*. For instance, if $\alpha < \alpha^*$ then the contract should include a fee of $(\alpha^* - \alpha)b$ that the retailer pays to the supplier per unit of backorder per unit time, which would change the retailer's backorder share to α^*. Similarly, if $\alpha > \alpha^*$ then the supplier

Figure 3 **The Optimal Nash Production Capacity (ν^*_α) and the Optimal Nash Base-Stock Level (s^*_α) as a Function of the Retailer's Backorder Share α—The Centralized Solutions Are ν^* and s^*.**



should pay the retailer a fee of $(\alpha - \alpha^*)b$ per unit of backorder per unit time. This contract, while suboptimal, is conceptually simpler than the linear transfer payment in §6.

We cannot solve for α_s and α_ν in closed form, except when b takes on a limiting value. By Equation (12), α_s satisfies

$$\frac{\ln(1+\alpha b)}{\ln(1+b) f_\alpha(b)} = 1. \tag{22}$$

As $b \to 0$, we have $f_\alpha(b) \to \sqrt{1-\alpha}$ and $\ln(1+\alpha b)/\ln(1+b) \to \alpha$. Therefore, as $b \to 0$, α_s satisfies $(\alpha/\sqrt{1-\alpha}) = 1$, or $\alpha_s = ((\sqrt{5}-1)/2) \approx 0.618$, which is the inverse of the *golden-section number* that arises in a variety of disciplines (e.g., Vajda 1989). As $b \to \infty$, we have $f_\alpha(b) \to \sqrt{(1-\alpha)/\alpha}$, and $\ln(1+\alpha b)/\ln(1+b)$. In this case, α_s satisfies $\sqrt{(\alpha/(1-\alpha))} = 1$, or $\alpha_s = 0.5$. Numerical computations reveal that α_s is unimodal in b, achieving a maximum of 0.627 at $b = 1.48$, and is rather insensitive to moderate values of the backorder-to-holding cost ratio b (e.g., $\alpha_s \geq 0.61$ for $b \in [1, 10]$). In other words, for moderate values of the normalized backorder cost, the Nash retailer holds more inventory than optimal when his share of the backorder cost is more than about 61%.

By Equation (11), α_ν solves $f_\alpha(b) = 1$; for example, when the normalized backorder cost is small, the Nash supplier holds a less-than-optimal level of capacity. As $b \to 0$, this condition becomes $\sqrt{1-\alpha} = 1$, which gives $\alpha_\nu = 0$. As $b \to \infty$, the condition becomes $\sqrt{((1-\alpha)/\alpha)} = 1$, which is solved by $\alpha_\nu = 0.5$. Note that $\alpha_s = \alpha_\nu = 0.5$ as $b \to \infty$ is consistent with our previous claim that the Nash equilibrium is asymptotically efficient in the regime $(\alpha = 0.5, b \to \infty)$. A numerical study reveals that α_ν is more sensitive than α_s to the value of b. As the backorder-to-holding cost ratio b varies from 1 to 10, α_ν ranges from 0.28 to 0.49.

Customer Service Level. The exponential distribution of the queue length implies that the steady-state probability that a customer is forced to wait because of retailer shortages is equal to $Pr(N \geq s) = e^{-\nu s}$; consequently, we refer to $(1 - e^{-\nu s}) \times 100\%$ as the *service level*. By Equations (6) and (15), the stockout probability $e^{-\nu s}$ equals $(1+b)^{-1}$ in the centralized solution and $(1+\alpha b)^{-1}$ in the Nash solution. Hence, customers receive better service in the centralized solution than in the Nash equilibrium. This is because (see Figure 3) the product of the two buffer resources (normalized excess capacity and base-stock level) is always smaller in the Nash equilibrium than in the centralized solution, and the customers suffer from this less-than-optimal level of collective buffer resource; this degradation in customer service in decentralized systems also occurs in the traditional bilateral monopoly model (e.g., Tirole 1997), where double marginalization leads to a higher price charged to the customer and less goods sold. Finally, even though the system is not stable for $\alpha = 1$, customers generally desire a larger value of α; for example, they prefer that the penalty for shortages be absorbed primarily by the agent in direct contact with them.

6. Contracts

We showed in §5 that the Nash equilibrium is always inefficient when the supply chain operates. In this section, we analyze a coordinating mechanism based on static transfer payments between the agents. Specifically, we assume that the supplier (or the retailer) agrees to transfer a fixed amount τ per unit time to the retailer (supplier) in order to improve his operations and eventually the operation of the whole chain. In its most general form, this transfer τ depends on both s and ν. As in our earlier analysis, this information is assumed to be common knowledge. Cachon and Zipkin (1999) also use a linear transfer payment based on inventory levels to coordinate their supply chain, and readers are referred to §1.5 of Cachon (1999a) for a survey of alternative types of contracts in the multiechelon inventory setting. We do not impose an explicit constraint that forces either agent to build a predefined level of its buffer resource. Using Cachon and Lariviere's (2001) terminology, we assume a *voluntary compliance* regime, where both the retailer and the supplier choose their buffer resource levels to maximize their own profits.

The transfer payment $\tau(s, \nu)$ modifies the cost functions in Equations (3) and (4) for the retailer and supplier, respectively, to

$$\tilde{C}_R(s, \nu) \triangleq C_R(s, \nu) - \tau(s, \nu) \quad \text{and}$$
$$\tilde{C}_S(s, \nu) \triangleq C_S(s, \nu) + \tau(s, \nu). \tag{23}$$

The choice of $\tau(s,\nu)$ that coordinates the system is not unique. One possibility is to define the transfer in such a way that the modified cost functions replicate a *cost-sharing* situation. That is, we can set $\tau(s,\nu)$ such that $\tilde{C}_S(s,\nu)=\gamma\, C(s,\nu)$ and $\tilde{C}_R(s,\nu)=(1-\gamma)C(s,\nu)$, where $\gamma\in[0,1]$ is a splitting factor and $C(s,\nu)$ is the centralized cost function defined in Equation (5). This cost sharing forces the transfer payment to satisfy

$$\tau(s,\nu)=\gamma\, C_R(s,\nu)-(1-\gamma)C_S(s,\nu). \tag{24}$$

In this situation, both agents have an objective that is a scaled version of the centralized one. Therefore, supplier and retailer have aligned objectives and the centralized solution (s^*,ν^*) is the unique Nash equilibrium. We notice that the transfer payment has two components. First, $\gamma\, C_R(s,\nu)$ is a payment made by the supplier to the retailer proportional to the retailer's cost. Similarly, $-(1-\gamma)C_S(s,\nu)$ corresponds to a payment made by the retailer to the supplier proportional to the supplier's cost. In other words, each agent transfers a fraction of its own cost to the other player.

Although we appear to have a degree of freedom in splitting the costs via γ, both agents must be better off under the Nash equilibrium with the transfer payments than under the Nash equilibrium without the transfer payments, for example,

$$\begin{aligned} C_R(s^*_\alpha,\nu^*_\alpha) &\geq (1-\gamma)C(s^*,\nu^*),\\ C_S(s^*_\alpha,\nu^*_\alpha) &\geq \gamma C(s^*,\nu^*). \end{aligned} \tag{25}$$

This condition can be rewritten as $\gamma\in[\underline{\gamma}_\alpha(b),\bar{\gamma}_\alpha(b)]$, where

$$\begin{aligned} \underline{\gamma}_\alpha(b) &= 1-\frac{\ln(1+\alpha b)}{2\ln(1+b)f_\alpha(b)} \quad \text{and}\\ \bar{\gamma}_\alpha(b) &= \frac{f_\alpha(b)}{2}\left(\frac{\ln(1+\alpha b)+2}{\ln(1+\alpha b)+1}\right). \end{aligned} \tag{26}$$

In addition, γ has to be in the $[0,1]$ range. If not, one of the players would have a negative cost function. In such a situation, the optimal strategy for this player is to increase as much as possible the absolute value of its cost. For example, if $\gamma<0$, then the supplier has a negative cost function and his/her best strategy is to pick ν as large as possible independently of the value of s.

To summarize, the centralized solution can be achieved using the transfer payment (Equation 24) if and only if $\gamma\in[\underline{\gamma}_\alpha(b),\bar{\gamma}_\alpha(b)]\cap[0,1]$. It is not hard to show that $\underline{\gamma}_\alpha(b)\leq\bar{\gamma}_\alpha(b)$, $\underline{\gamma}_\alpha(b)<1$, and $\bar{\gamma}_\alpha(b)\geq 0$; hence, coordination is always possible using this type of contract. This conclusion stems from the fact that the Nash equilibrium is inefficient and the cost reduction obtained using the centralized solution can be split, making both agents better off. After some algebra we get

$$\underline{\gamma}_\alpha(b)=\frac{C_S(s^*_\alpha,\nu^*_\alpha)}{C(s^*,\nu^*)}-P_\alpha(b) \quad \text{and} \quad \bar{\gamma}_\alpha(b)=\frac{C_S(s^*_\alpha,\nu^*_\alpha)}{C(s^*,\nu^*)},$$

where $P_\alpha(b)$ is the competition penalty defined in Equation (18). Notice that the range of *admissible* γ (those that effectively coordinate the system), and hence the ease of negotiation, increases with the magnitude of the competition penalty.

Two extreme cases deserve particular attention: $\underline{\gamma}_\alpha(b)<0$ and $\bar{\gamma}_\alpha(b)>1$. In these situations, there exists a coordinating contract having one of the agents fully subsidizing the other agent's operation, for example, $\gamma=0$ or $\gamma=1$. When $\underline{\gamma}_\alpha(b)<0$ it is profitable for the retailer to absorb the supplier's variable cost (the reverse is true when $\bar{\gamma}_\alpha(b)>1$)[1]. This type of behavior, where large manufacturers pay for their supplier's capital equipment, has been observed in the automobile industry (e.g., Dyer and Ouchi 1993, Dyer et al. 1998), although for reasons related to maintaining long-term supplier relationships and bearing the risk of low future demand.

As b goes to infinity the lower and upper bounds converge to

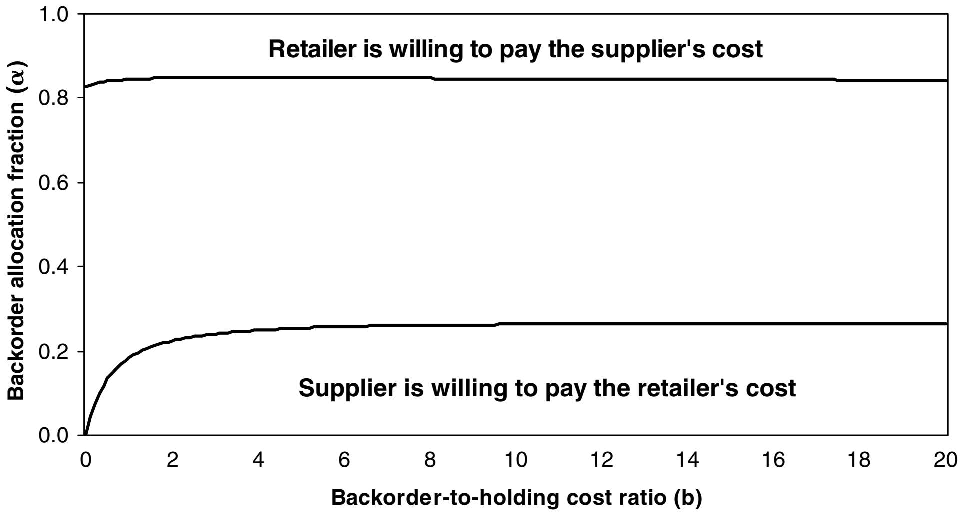
$$\lim_{b\to\infty}\underline{\gamma}_\alpha(b)=1-\frac{1}{2}\sqrt{\frac{\alpha}{1-\alpha}} \quad \text{and} \quad \lim_{b\to\infty}\bar{\gamma}_\alpha(b)=\frac{1}{2}\sqrt{\frac{1-\alpha}{\alpha}}.$$

Hence as $b\to\infty$ the conditions $\underline{\gamma}_\alpha(b)=0$ and $\bar{\gamma}_\alpha(b)=1$ hold for $\alpha=0.8$ and $\alpha=0.2$, respectively. At the other extreme, as b goes to zero the bounds are

$$\lim_{b\to 0}\underline{\gamma}_\alpha(b)=1-\frac{\alpha}{2\sqrt{1-\alpha}} \quad \text{and} \quad \lim_{b\to 0}\bar{\gamma}_\alpha(b)=\sqrt{1-\alpha}.$$

[1] If an agent is fully subsidized then any strategy is optimal and so it is not necessarily true that he/she will pick the optimal centralized solution. To avoid this technicality, it suffices that the player that absorbs all the cost offers a small fee to the subsidized player if he/she picks the optimal centralized solution.

Figure 4 **Solution to $\underline{\gamma}_\alpha(b) = 0$ and $\bar{\gamma}_\alpha(b) = 1$**



In this case, the roots for $\underline{\gamma}_\alpha(0) = 0$ and $\bar{\gamma}_\alpha(0) = 1$ are $\alpha = 2(\sqrt{2} - 1) \approx 0.828$ and $\alpha = 0$, respectively. Figure 4 plots the solution to $\underline{\gamma}_\alpha(b) = 0$ (upper curve) and $\bar{\gamma}_\alpha(b) = 1$ (lower curve). Roughly speaking, we can show that for $\alpha \geq 0.8$ the retailer is willing to pay the supplier's cost in order to achieve coordination and the opposite holds for $\alpha \leq 0.2$.

As we pointed out, the transfer payment in (24) replicates a cost-sharing situation in which both agents agree upfront to split the total cost of the supply chain. From a practical perspective, however, a direct cost-sharing negotiation based on total cost probably dominates the transfer-payment approach because it does not require any special accounting of cost (in terms of capacity and backordering costs) and it is more familiar, transparent, robust (with respect to model misspecifications), and easier to implement.

7. The Stackelberg Games

We conclude our study of this two-stage supply chain by considering the case where one agent dominates.

Supplier's Stackelberg Game. When the supplier is the Stackelberg leader, he chooses ν to optimize $C_S(s, \nu)$ in Equation (4), given the retailer's best response, $s^*(\nu)$ in Equation (15). This straightforward computation leads to the following proposition.

Proposition 4. *In the absence of participation constraints, the equilibrium in the supplier's Stackelberg game is*

$$\bar{s}_\alpha = s^*_\alpha \sqrt{1 + \ln(1 + \alpha b)}, \qquad \bar{\nu}_\alpha = \frac{\nu^*_\alpha}{\sqrt{1 + \ln(1 + \alpha b)}}. \tag{27}$$

The agents' costs are

$$C_S(\bar{s}_\alpha, \bar{\nu}_\alpha) = 2\sqrt{\frac{(1-\alpha)bc}{1+\alpha b}}, \qquad C_R(\bar{s}_\alpha, \bar{\nu}_\alpha) = \bar{s}_\alpha. \tag{28}$$

Equation (27) implies that $\bar{\nu}_\alpha \bar{s}_\alpha = \nu^*_\alpha s^*_\alpha$, and hence the customer service level is the same under the Stackelberg and Nash equilibria. Because the first-order conditions of the centralized problem dictate the service level, it also follows that the Stackelberg equilibrium is inefficient relative to the centralized solution. Not surprisingly, the supplier builds less capacity and the retailer holds more safety stock in Equation (27) than in the Nash equilibrium. The discrepancy between the Stackelberg and Nash solutions increases as α and b increase.

Now we compare the variable cost of each agent and the entire system under the Nash and Stackelberg equilibria. By Equation (14), the supplier's cost in the Nash equilibrium can be written as

$$C_S(s^*_\alpha, \nu^*_\alpha) = 2\sqrt{\frac{(1-\alpha)bc}{1+\alpha b}} \left(\frac{\ln(1+\alpha b) + 2}{2\sqrt{\ln(1+\alpha b) + 1}} \right).$$

The function $((x+2)/(2\sqrt{x+1}))$ is strictly increasing in $[0,\infty)$, and is equal to 1 when $x=0$. Thus, it is always the case that $C_S(\bar{s}_\alpha, \bar{\nu}_\alpha) \le C_S(s^*_\alpha, \nu^*_\alpha)$; this is to be expected, because the supplier incorporates the retailer's best response when selecting his level of capacity. However, $C_S(\bar{s}_\alpha, \bar{\nu}_\alpha) = C_S(s^*_\alpha, \nu^*_\alpha)$ when $\alpha=0$, $\alpha=1$ or $b=0$, and so the supplier does not benefit from being the leader in these extreme cases. When $\alpha=1$ or $b=0$, the supplier does not face any backorder costs and builds no excess capacity ($\nu=0$). On the other hand, when $\alpha=0$ the retailer—incurring no backorder costs—holds no safety stock ($s=0$). Because these choices ($\nu=0$ and $s=0$) are independent of the bargaining power of the supplier in these cases, the Stackelberg and Nash equilibria provide the same cost to the supplier.

By Equations (13) and (27), the difference in the retailer's cost between the Stackelberg equilibrium and the Nash equilibrium is

$$C_R(\bar{s}_\alpha, \bar{\nu}_\alpha) - C_R(s^*_\alpha, \nu^*_\alpha) = s^*_\alpha\left(\sqrt{1+\ln(1+\alpha b)}-1\right). \tag{29}$$

As expected, the retailer is worse off in the supplier's Stackelberg equilibrium than in the Nash equilibrium. By Equations (8), (10), (12), and (29), the increase in the retailer's cost from being the follower vanishes as $\alpha \to 0$ and $b \to 0$ (similar to the reasons given in the previous paragraph), and grows with α and b. As α gets larger, the supplier cares less about backorders and builds less capacity, leaving the retailer in a vulnerable situation.

A comparison of the total system cost shows that $C_R(s^*_\alpha, \nu^*_\alpha) + C_S(s^*_\alpha, \nu^*_\alpha) \le C_R(\bar{s}_\alpha, \bar{\nu}_\alpha) + C_S(\bar{s}_\alpha, \bar{\nu}_\alpha)$ if and only if

$$\frac{(1+\alpha b)\ln(1+\alpha b)}{(1-\alpha)b} - \sqrt{1+\ln(1+\alpha b)}+1 \ge 0. \tag{30}$$

Condition (30) holds for large values of α, but is not true in general. Because the left side of condition (30) equals zero when $\alpha=0$, is increasing in $\alpha \ge \alpha_0$ if it is increasing in α at α_0, and has a derivative with respect to α equal to $1-(b/2)$ when $\alpha=0$, we conclude that for a backorder-to-holding cost ratio $b \le 2$ the Nash solution achieves a lower system cost than the Stackelberg equilibrium for any value of α. If $b > 2$, the Nash solution is more efficient if and only if $\alpha > \bar{\alpha}$, where $\bar{\alpha}$ is the unique positive value of α that solves condition (30) with equality. Hence, overall system performance suffers when the retailer incurs most of the expensive backorder costs, and—as the follower—has less power than in the Nash equilibrium to control these costs.

Retailer's Stackelberg Game. The Stackelberg problem is less tractable when the retailer is the leader. The following proposition (see the Appendix for a proof) characterizes the solution.

Proposition 5. *Let $(\hat{\nu}_\alpha, \hat{s}_\alpha)$ be the equilibrium when the retailer is the Stackelberg leader in the absence of participation constraints. Define $\hat{\beta} \ge 0$ to be the unique nonnegative solution of*

$$\beta^2 + (\beta+2)\left(1-e^{-(\beta-\nu^*_\alpha s^*_\alpha)}\right) = 0. \tag{31}$$

Then the Stackelberg solution is

$$\hat{\nu}_\alpha = \sqrt{\frac{(1-\alpha)b(\hat{\beta}+1)e^{-\hat{\beta}}}{c}}, \qquad \hat{s}_\alpha = \frac{\hat{\beta}}{\hat{\nu}_\alpha}. \tag{32}$$

Although we do not have a closed-form solution to the retailer's Stackelberg game, the next proposition (see the Appendix for a proof) provides a comparison between this equilibrium and the Nash equilibrium.

Proposition 6. *The following five inequalities hold*:

$$\hat{\nu}_\alpha \hat{s}_\alpha \le \nu^*_\alpha s^*_\alpha, \tag{33}$$

$$\hat{\nu}_\alpha \ge \nu^*_\alpha, \tag{34}$$

$$\hat{s}_\alpha \le s^*_\alpha, \tag{35}$$

$$C_R(\hat{s}_\alpha, \hat{\nu}_\alpha) \le C_R(s^*_\alpha, \nu^*_\alpha), \tag{36}$$

$$C_S(\hat{s}_\alpha, \hat{\nu}_\alpha) \ge C_S(s^*_\alpha, \nu^*_\alpha). \tag{37}$$

While inequalities (34)–(37) mirror our results for the supplier Stackelberg game, inequality (33) states that the customer service level, $(1-e^{-\nu s}) \times 100\%$, is lower in the retailer's Stackelberg equilibrium than in the supplier's Stackelberg equilibrium (and the Nash equilibrium). Inequality (33) also implies that the retailer Stackelberg equilibrium is inefficient relative to the centralized solution. Analytical approximations (using $e^{-x} \approx 1-x$ in Equation 31) and numerical computations reveal that when the service level is close to 0 or 100%, both Stackelberg games have asymptotically the same service level. The maximum difference is approximately 9.5%, and is achieved when the service level is 76.0% for the supplier's Stackelberg game and 66.5% for the retailer's Stackelberg game. In a more practical example, if the supplier's Stackelberg service level is 90.0% then the retailer's Stackelberg service level is approximately 82%. Hence, the deterioration in customer service is not trivial, and if there is a leader the customers prefer that it is the supplier. This is because the base-stock level is more effective than the capacity level at controlling the customer service level, and the retailer chooses a small base-stock level when he is the leader.

Figure 5 shows the competition penalty P for the Nash and Stackelberg games. The left graph plots P as a function of the backorder cost b for $\alpha = 0.5$. Consistent with our discussion in §5, for $\alpha = 0.5$ the competition penalty decreases with b for the Nash game. In contrast, for both Stackelberg games the competition penalty increases with the backorder rate. The other plot in Figure 5 displays the competition penalty as a function of α for $b = 2$. The inefficiency for the three games is U-shaped, and is minimized near $\alpha = 0.5$. Moreover, all three games have a similar level of inefficiency as $\alpha \to 0$ and $\alpha \to 1$. Finally, Figure 5 also suggests that the retailer's Stackelberg game achieves a lower total system cost than the supplier's Stackelberg game. Evidently, the supplier, by owning the cost-driving resource, wreaks more havoc on his opponent as the leader than does the retailer. He holds less capacity, leading the retailer to incur considerable holding costs. The retailer holds less inventory (which is the customer-service-driving resource) when he leads, and the supplier's increase in capacity is not enough to compensate for the retailer's lower base-stock level; consequently, customer service suffers.

8. Concluding Remarks

The distinguishing feature of our simple supply chain model is that congestion at the supplier's manufacturing facility is explicitly captured via a single-server queue. Each agent has a resource at his disposal (the

Figure 5 **Competition Penalty for the Nash Game (*N*), the Supplier's Stackelberg Game (*S*), and the Retailer's Stackelberg Game (*R*)**

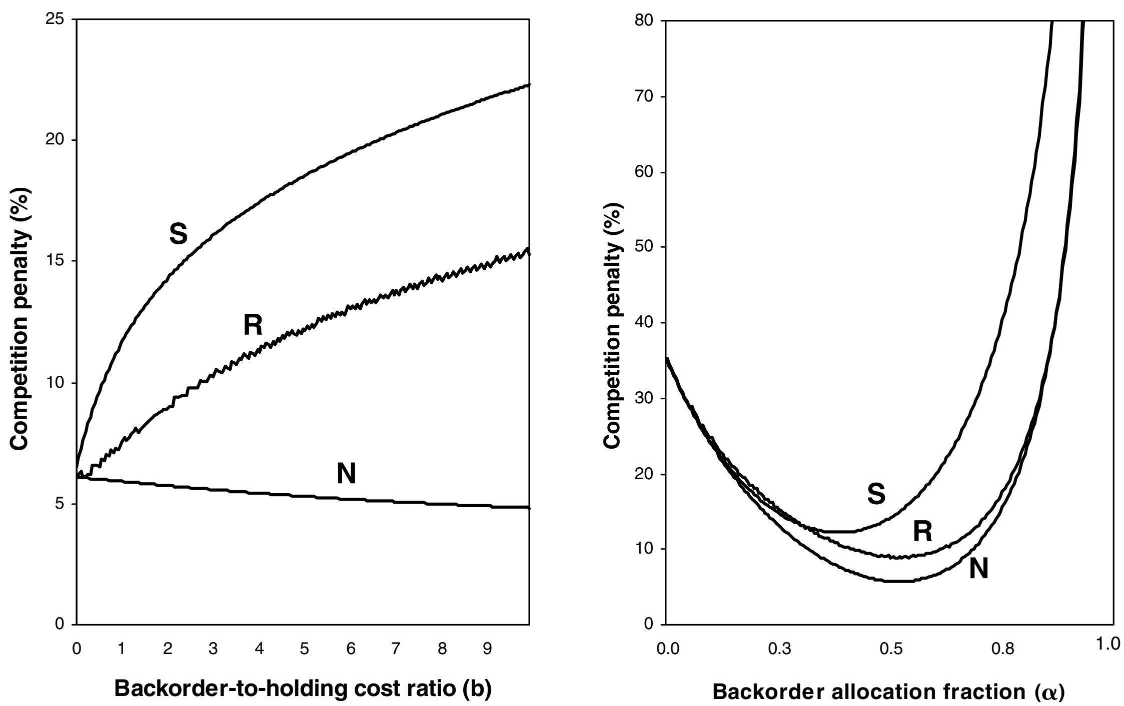


Table 1 A Summary of the Asymptotic Results

	$b \to 0$	$b \to \infty$	$0 < b < \infty$
Competition Penalty (Nash Inefficiency)	0 if $\alpha \to 0$ ∞ if $\alpha \to 1$	∞ if $\alpha \to 0$ 0 if $\alpha = 0.5$ ∞ if $\alpha \to 1$	U-shaped as α ranges from 0 to 1
Customer Service Level $(1 - \exp(-\nu s))$	Centr. = Nash	Centr. = Nash	Centr. $\gg$ Nash if $\alpha \to 0$ Centr. = Nash if $\alpha \to 1$
$s^*_\alpha \geq s^*$	$\alpha \geq 0.618$	$\alpha \geq 0.5$	$\alpha \geq \alpha_s =$ root of (22)
$\nu^*_\alpha \geq \nu^*$	$\alpha = 0$	$\alpha \leq 0.5$	$\alpha \leq \alpha_\nu =$ root of $f_\alpha(b) = 1$
Contract:			
Supplier Pays Retailer's Cost If	$\alpha = 0$	$\alpha \leq 0.2$	$\alpha \leq \underline{\gamma}(b)$ see condition (30)
Retailer Pays Supplier's Cost If	$\alpha \geq 0.828$	$\alpha \geq 0.8$	$\alpha \geq \bar{\gamma}(b)$ see condition (30)

supplier chooses the capacity level and the retailer chooses the base-stock level) that buffers against backorders. When the inventory backorder cost is incurred entirely by the retailer (i.e., the retailer's backorder share $\alpha = 1$), the supplier has no incentive to build any excess capacity, which leads to system instability. When the supplier incurs some backorder cost ($\alpha \in [0, 1)$), there is a unique Nash equilibrium that is always inefficient: The agents' selfish behavior degrades overall system performance. However, the Nash equilibrium is asymptotically efficient (a summary of all our asymptotic results is provided in Table 1) in two cases: (i) The backorder cost goes to zero and the supplier incurs all of the backorder cost, and (ii) the backorder cost goes to infinity and is split evenly between the two agents. The Nash equilibrium has an arbitrarily high inefficiency in two cases: (i) The backorder cost goes to infinity and the supplier incurs all of the backorder cost, and (ii) the retailer incurs all of the backorder cost. For most practical cases (i.e., backorder-to-holding cost ratio is greater than 2), the inefficiency of the Nash solution is smallest when the backorder costs are split relatively evenly between the two agents. This is likely to occur, for example, if the propensity of customers to switch brands (suppliers) or switch retailers is similar. Relative to the centralized solution, the agents in the Nash equilibrium have more buffer resources when they care sufficiently about backorders: The supplier builds more capacity than optimal when $\alpha < \alpha_\nu$ (and $\alpha_\nu \geq 0.28$ if backorders are more expensive than holding inventory) and the retailer has a larger than optimal base-stock level when $\alpha > 0.63$ (and, in some cases, an even smaller threshold). However, at least one of the agents in the Nash equilibrium holds a lower-than-optimal level of his buffer resource. Finally, customers receive better service in the centralized solution than in the Nash equilibrium, and customer service improves in the Nash setting when the retailer incurs most—but not all—of the backorder cost.

A linear transfer payment that induces cost sharing between the players coordinates the system. In the absence of participation constraints, coordination is always possible. A coordinating contract requires the parties to negotiate on the value of a single parameter (γ) that defines the split of the total system cost. The range of admissible values for γ increases with the magnitude of the inefficiency of the Nash equilibrium. Hence, negotiations should be easier when the backorder costs are not split too evenly. However, if an agent absorbs 80% or more of the system backorder cost then this agent is willing to sign a contract that fully subsidizes the other agent's operation. Consequently, our model predicts that vertical integration would preclude the widespread prevalence of decentralized supply chains with extreme values of α. Although we did not incorporate participation constraints in the model, a straightforward analysis (see Caldentey and Wein (1999)) in a profit-maximization framework shows that there are values of α (in the proximity of 0 and 1) for which the contract will lead to the operation of an otherwise inoperative supply chain, for example, the extra system profit generated by the contract is sufficient to entice the nonparticipating agent into playing.

Finally, when one of the agents is the Stackelberg leader, he builds less of his buffer resource and faces a lower cost than in the Nash equilibrium, and the other agent builds more of his buffer resource and pays a higher cost. The total system cost is less in the retailer Stackelberg game than in the supplier Stackelberg game. Because scarce capacity causes a nonlinear increase in delay, the supplier's leadership power causes more harm to the supply chain than the retailer's leadership power. Customer service is the same in the Nash equilibrium as when the supplier is the Stackelberg leader, but customers fare worse when the retailer is the leader. This latter effect is because the retailer holds less inventory when he is the leader, and the base-stock level has a more direct impact on customer service than the supplier's capacity. Taken together, these results provide operations managers with a comprehensive understanding of the competitive interactions in this system, and offer guidelines for when (i.e., for which sets of problem parameters) and how to negotiate contracts to induce participation and increase profits.

Recall that our model is quite similar to the two-stage inventory model of Cachon and Zipkin (1999): The main difference is the single-server vs. infinite-server model (i.e., queueing vs. inventory model) for the manufacturing process. Indeed, these two models can be viewed as the simplest prototypes for single-product, steady-state supply chains with no fixed ordering costs. In the extreme case when the supplier does not care about backorders ($\alpha = 1$), he builds no excess capacity in our queueing model, whereas he holds no inventory in Cachon and Zipkin's inventory model. The effect of the former is an unstable system, while the effect of the latter is to turn the supply chain into a stable—albeit ineffective—make-to-order system. In Cachon and Zipkin's echelon inventory game, the Nash solution is indeed highly inefficient when $\alpha = 1$, but in the local inventory game the median inefficiency in their computational study is only 1%. When $\alpha = 1$ in the local inventory game, the supplier's base-stock level offers him little control over the system's cost, whereas the capacity level in our model impacts the entire system in a more profound way. On the other hand, both models predict that the inefficiency is small when the backorder costs are shared equally. Another qualitative difference between the results in these two papers is that Cachon and Zipkin's agents typically hold less inventory (for the echelon inventory game) in the Nash equilibrium than in the centralized solution, whereas our agents build/hold a higher-than-optimal level of their buffer resource when their share of the backorder cost is large, as suggested in the management literature by Buzzell and Ortmeyer (1995) and others.

In our view, the make-to-stock queue is an attractive operations management model to embed into a game-theoretic framework. The model complements the newsvendor model by considering products with long life cycles, and is about as complex as—but considerably more tractable than—a two-stage Clark-Scarf model. It also allows us to capture the nonlinear effect of capacity on the supplier's lead times. Of course, none of these models attempt to mimic the complexities of an actual supply chain. Nevertheless, to the extent that queueing effects are present in manufacturers' production facilities, the make-to-stock queue is a parsimonious and tractable model for deriving new insights into multiagent models for supply chain management.

Acknowledgments
The authors are grateful to Gérard Cachon for valuable discussions and to the Senior Editor and two referees for their extraordinary efforts, which significantly improved the paper's content and presentation.

Appendix

Proof of Proposition 1. The function $C(s, \nu)$ defined in Equation (5) is continuously differentiable and bounded below by 0 in $X = \{(s, \nu) \mid s \geq 0, \nu > 0\}$. Thus, a global minimum is either a local interior minimum that satisfies the first-order conditions or an element of the boundary of X; alternatively, there could be no global minimum if the function decreases as $s \to \infty$ or $\nu \to \infty$.

However, we have checked that $\lim_{s\to\infty} C(s, \nu) \to \infty$ for $\nu > 0$, and $\lim_{\nu\to\infty} C(s, \nu) \to \infty$ for $s \geq 0$, which implies that a global minimum exists. From the first-order conditions (6) and (7), the only interior point that is a candidate for the global minimum is (s^*, ν^*). In addition, the Hessian of $C(s, \nu)$ at (s^*, ν^*) is given by

$$H(s^*, \nu^*) = \begin{pmatrix} \nu^* & s^* \\ s^* & \dfrac{c(\ln(1+b)+2)}{\nu^*} \end{pmatrix}.$$

Because $\ln(1+b) > 0$ for $b > 0$, the Hessian is positive definite and (s^*, ν^*) is the unique local minimum in the interior of X. The resulting cost is $C(s^*, \nu^*) = 2\sqrt{c\ln(1+b)}$. Finally, $\lim_{\nu\to 0} C(s, \nu) \to \infty$ for $s \geq 0$, and

$$C(0, \nu) = \frac{b}{\nu} - c\nu \geq 2\sqrt{cb} > C(s^*, \nu^*) \quad \text{for} \quad \nu > 0, b > 0.$$

Thus, (s^*, ν^*) is the unique global minimum for $\Pi(s, \nu)$. □

Proof of Proposition 5. To derive the Stackelberg equilibrium, we find it convenient to define

$$\beta = \nu s, \tag{38}$$

and rewrite the supplier's reaction curve (Equation 16) as

$$e^{-\beta}\left(\frac{\beta+1}{\beta^2}\right) = \frac{c}{(1-\alpha)bs^2}. \tag{39}$$

The one-to-one correspondence between the base-stock level s and the service level parameter β (recall that the service level is $e^{-\beta} \times 100\%$) allows the retailer in this Stackelberg game to choose β rather than s. By Equations (3) and (38), the retailer's cost is

$$C_R(\beta, \nu) = \left(\frac{\beta - 1 + (1+\alpha b)e^{-\beta}}{\nu}\right). \tag{40}$$

Solving Equation (39) for s and using Equation (38) gives

$$\nu(\beta) = \sqrt{\frac{(1-\alpha)b(\beta+1)e^{-\beta}}{c}}. \tag{41}$$

Substituting Equation (41) into Equation (40) yields the retailer's cost as the following convex function of $\beta \geq 0$:

$$C_R(\beta) = \sqrt{\frac{c}{(1-\alpha)b}}\left(\frac{\beta - 1 + (1+\alpha b)e^{-\beta}}{\sqrt{(\beta+1)e^{-\beta}}}\right). \tag{42}$$

Therefore, the first-order condition

$$\frac{\beta^2 + (\beta+2)(1-(1+\alpha b)e^{-\beta})}{2(\beta+1)^{\frac{3}{2}}e^{\frac{-\beta}{2}}} = 0 \tag{43}$$

is sufficient for optimality. Because $1 + \alpha b = e^{\nu_\alpha^* s_\alpha^*}$ and the denominator of condition (43) is always positive, condition (43) is equivalent to Equation (31). Hence, by Equations (31), (38), and (41), the Stackelberg equilibrium is given by Equations (31) and (32). □

Proof of Proposition 6. To prove Equation (33), note that the left side of Equation (31) is positive if $\beta > \nu_\alpha^* s_\alpha^*$. Thus, the root $\hat\beta$ of Equation (31) must satisfy $\hat\beta \leq \nu_\alpha^* s_\alpha^*$, for example, $\hat\nu_\alpha \hat s_\alpha \leq \nu_\alpha^* s_\alpha^*$.

To show that $\hat\nu_\alpha \geq \nu_\alpha^*$ and $\hat s_\alpha \leq s_\alpha^*$, we first observe that $\nu(s) = \arg\min_{\nu>0}\{C_R(s, \nu)\}$ and $(\partial^2 C_R(s, \nu))/(\partial s \partial \nu) = (1-\alpha)bse^{-\nu s} \leq 0$. Thus (e.g., Chapter 2 of Topkis 1998), $C_R(s, \nu)$ satisfies the decreasing difference property,

$$\frac{d\nu(s)}{ds} \leq 0. \tag{44}$$

In addition, the function $e^{-\beta}(\beta+1)/\beta^2$ is decreasing in $\beta > 0$. Hence, from Equation (39) and inequality (33), we conclude that $\hat s_\alpha \leq s_\alpha^*$. Finally, Equation (44) and $\hat s_\alpha \leq s_\alpha^*$ implies that $\hat\nu_\alpha \geq \nu_\alpha^*$.

The retailer's cost in Equation (42) is an increasing function of β for $\beta \geq \hat\beta$. Hence, inequality (36) follows from inequality (33). To prove Equation (37), for example, $C_S(\beta^*, \nu_\alpha^*) \leq C_S(\hat\beta, \hat\nu_\alpha)$, we first use Equation (4) to rewrite the supplier's cost as

$$C_S(\beta, \nu) = c\nu + \left(\frac{(1-\alpha)be^{-\beta}}{\nu}\right).$$

The function $C_S(\beta, \nu)$ is decreasing in β for $\nu > 0$, and so inequality (33) implies that $C_S(\beta^*, \hat\nu_\alpha) \leq C_S(\hat\beta, \hat\nu_\alpha)$. Hence the proof of Equation (37) will follow if we can show that $C_S(\beta^*, \nu_\alpha^*) \leq C_S(\beta^*, \hat\nu_\alpha)$. For any fixed nonnegative β, the function $C_S(\beta, \nu)$ is convex in ν, achieves its only minimum at

$$\nu(\beta) = \sqrt{\frac{(1-\alpha)be^{-\beta}}{c}},$$

and is increasing for $\nu \in [\nu(\beta), \infty)$. In particular, we have

$$\nu(\beta^*) = \sqrt{\frac{(1-\alpha)b}{c(1+\alpha b)}} \leq \sqrt{\frac{(1-\alpha)b}{c(1+\alpha b)}}\sqrt{1+\ln(1+\alpha b)} = \nu_\alpha^*.$$

This inequality and Equation (34) imply that $C_S(\beta^*, \nu_\alpha^*) \leq C_S(\beta^*, \hat\nu_\alpha)$, which completes the proof of Equation (37). □

References

Agrawal, N., S. Nahmias, A. Tsay. 1999. Modeling supply chain contracts: A review. S. Tayur, M. Magazine, R. Ganeshan, eds. *Quantitative Models for Supply Chain Management*. Kluwer Academic Publishers, Boston, MA.

Anupindi, R., Y. Bassok. 1999. Supply contracts with quantity commitments and stochastic demand. S. Tayur, M. Magazine, R. Ganeshan, eds. *Quantitative Models for Supply Chain Management*. Kluwer Academic Publishers, Boston, MA.

Buzacott, J. A., S. Price, J. G. Shanthikumar. 1992. Service levels in multistage MRP and base stock controlled production systems. G. Fandel, T. Gulledge, A. Jones, eds. *New Directions for Operations Research in Manufacturing*. Springer, Berlin, Germany.

Buzzell, R., G. Ortmeyer. 1995. Channel partnerships streamline distribution. *Sloan Management Rev.* **36** 85–96.

Cachon, G. 1999a. Competitive supply chain inventory management. S. Tayur, M. Magazine, R. Ganeshan, eds. *Quantitative Models for Supply Chain Management*. Kluwer Academic Publishers, Boston, MA.

——. 1999b. Competitive and cooperative inventory management in a two-echelon supply chain with lost sales. Fuqua School of Business, Duke University, Durham, NC.

——, M. Lariviere. 2001. Contracting to assure supply: How to share demand forecasts in a supply chain. *Management Sci.* **47** 629–646.

——, P. Zipkin. 1999. Competitive and cooperative inventory policies in a two-stage supply chain. *Management Sci.* **45** 936–953.

Caldentey, R., L. M. Wein. 1999. Analysis of a decentralized production-inventory system. Working paper, Sloan School of Management, MIT, Cambridge, MA.

Chen, F. 1999. Decentralized supply chains subject to information delays. *Management Sci.* **45** 1076–1090.

——, A. Federgruen, Y. Zheng. 2001. Coordination mechanisms for a distribution system with one supplier and multiple retailers. *Management Sci.* **47** 693–708.

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

Duenyas, I., C.-Y. Tsai. 2001. Centralized and decentralized control of a two-stage tandem manufacturing system with demand for intermediate and end products. Working paper, Business School, University of Michigan, Ann Arbor, MI.

Dyer, J., W. Ouchi. 1993. Japanese-style partnerships: Giving companies a competitive edge. *Sloan Management Rev.* **34** 51–63.

——, D. S. Cho, W. Chu. 1998. Strategic supplier segmentation: The next "best practice" in supply chain management. *California Management Rev.* **40** 57–77.

Federgruen, A., Z. Katalan. 1996. The stochastic economic lot scheduling problem: Cyclical base stock policies with idle times. *Management Sci.* **42** 783–796.

——, P. Zipkin. 1986. An inventory model with limited production capacity and uncertain demands, I: The average-cost criterion. *Math. Oper. Res.* **11** 193–207.

Gavirneni, S., R. Kapuscinski, S. Tayur. 1996. Value of information in capacitated supply chains. *Management Sci.* **45** 16–24.

Glasserman P., S. Tayur. 1995. Sensitivity analysis for base stock levels in multi-echelon production-inventory systems. *Management Sci.* **41** 263–281.

Harrison, J. M. 1988. Brownian models of queueing networks with heterogeneous customer populations. W. Fleming, P. L. Lions, eds. *Stochastic Differential Systems, Stochastic Control Theory and Applications*, IMA vol. 10. Springer-Verlag, New York.

Lariviere, M. 1999. Supply chain contracting and coordination with stochastic demand. S. Tayur, M. Magazine, R. Ganeshan, eds. *Quantitative Models for Supply Chain Management*. Kluwer Academic Publishers, Boston, MA.

Lee, Y. J., P. Zipkin. 1992. Tandem queues with planned inventories. *Oper. Res.* **40** 936–947.

Markowitz, D. M., M. I. Reiman, L. M. Wein. 2000. The stochastic economic lot scheduling problem: Heavy traffic analysis of dynamic cyclic policies. *Oper. Res.* **48** 136–154.

Morse, P. 1958. *Queues, Inventories and Maintenance*. Wiley, New York.

Plambeck, E. L., S. A. Zenios. 1999. Incentive efficient control of a make-to-stock production system. Working paper, Graduate School of Business, Stanford University, Stanford, CA.

Rubio, R., L. M. Wein. 1996. Setting base stock levels using product-form queueing networks. *Management Sci.* **42** 259–268.

Stern, L. W., I. El-Ansary, A. T. Coughlan. 1996. *Marketing Channels.* Prentice Hall, Inc., Englewood Cliffs, NJ.

Tirole, J. 1997. *The Theory of Industrial Organization*. MIT Press, Cambridge, MA.

Topkis, D. M. 1998. *Supermodularity and Complementarity*. Princeton University Press, Princeton, NJ.

Tsay, A. A., W. S. Lovejoy. 1999. Quantity flexibility contracts and supply chain performance. *Manufacturing and Service Oper. Management* **1** 89–111.

Vajda, S. 1989. *Fibonacci and Lucas Numbers, and the Golden Section: Theory and Applications*. Ellis Horwood Limited, England.

Wein, L. M. 1992. Dynamic scheduling of a multiclass make-to-stock queue. *Oper. Res.* **40** 724–735.

Zheng, Y., P. Zipkin. 1990. A queueing model to analyze the value of centralized inventory information. *Oper. Res.* **38** 296–307.

The consulting Senior Editor for this manuscript was William Lovejoy. This manuscript was received on November 30, 1999, and was with the authors 710 days for 5 revisions. The average review cycle time was 71 days.