Analysis of Decentralized Production-Inventory System

René Caldentey
Lawrence M. Wein

November 22, 1999
Working Paper Number 4099

Contact Address:
Prof. Lawrence M.Wein
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA 02142-1343

Lwein@mit.edu
http://web.mit.edu/lwein/www/

Analysis of a Decentralized Production-Inventory System

René Caldentey*   Lawrence M. Wein [†]

November 22, 1999

Abstract

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., 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 that 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 based on backorder, inventory and capacity levels coordinates the system in the absence of participation constraints. We also derive explicit formulas to assess the inefficiency of the Nash equilibrium, compare the agents' decision variables and the customer service level of the centralized versus Nash solutions, and identify conditions under which a coordinating contract is desirable for both agents.

---

*Department of Industrial Engineering, University of Chile, República 701, Santiago, and Sloan School of Management, MIT, Cambridge, MA 02139, (617) 253-8094 Fax: (617) 258-7579, rcaldent@mit.edu

[†]Sloan School of Management, MIT, Cambridge, MA 02139, (617) 253-6697 Fax: (617) 258-7579, lwein@mit.edu

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 exogeneous 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 to build as little capacity as possible and receive excellent demand forecasts and/or a steady stream of orders, 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) multi-echelon 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 life cycles. More complex (multi-period, and possibly multi-echelon 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 multi-period supply chain models successfully capture the replenishment policy and have addressed some aspects of retailer demand, e.g., information lead times in the Clark-Scarf model (Chen 1999), pricing

in multi-echelon models with deterministic demand and ordering costs (Chen *et al.* 1999), and forecast updates (Anupindi and Bassok 1999 in a multi-period newsvendor model and Tsay and Lovejoy 1999 in a multi-stage model). However, the Clark-Scarf model, and indeed all of traditional inventory theory, takes a crude approach towards 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 mysteriouly dormant for the next three decades, perhaps because the multi-echelon 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 multi-product queues with (e.g., Federgruen and Katalan 1996, Markowitz *et al.* 1999) and without (e.g., Zheng and Zipkin 1990, Wein 1992) setups, and single-product, multi-stage 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.* 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 multi-agent production-inventory study that we are aware of, Plambeck and Zenios (1999), contemporaneously to us, analyze a more complex dynamic system with information asymmetry.

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 analyses, 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, 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. Another deviation in the formulations is that Cachon and Zipkin's agents minimize cost, while our agents maximize profit; this allows us to explicitly incorporate participation (i.e., nonnegative profits) constraints. 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 Nash equilibrium in §4, where the supplier and retailer maximize their own profit. The two solutions are compared in §5. In §6, we describe the contract that *coordinates* the system; i.e., allows the decentralized system to achieve the same profit as the centralized system. In §7, we analyze the Stackelberg games, where one agent has all the bargaining power. Concluding remarks are presented in §8.

4

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. The retailer uses a $(s-1, s)$ base stock policy to replenish his inventory. 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. Because we assume that there are no fixed ordering costs, the retailer's optimal replenishment policy is indeed characterized by the base stock level s.

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.

The selling price r that the retailer charges to the end customers and the wholesale price w that the retailer pays to the supplier are fixed. These conditions implicitly assume that the retailer and supplier operate in competitive markets. Each backordered unit generates a cost b per unit of time for the production-inventory system. As in Cachon and Zipkin, 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 *backorder allocation fraction*, is exogenously specified in our model. In addition, the retailer incurs a holding cost h per unit of inventory per unit of time. The supplier is responsible for building production capacity and the capacity cost c is per unit of product, so that $c\mu$ is cost per unit time. We assume $r > w > c$, so that positive profits are not unattainable. To make our results more transparent, we normalize the expected profit per unit time by dividing it by the holding cost rate h. Towards this end, we normalize the cost parameters as follows:

$$\tilde{h} = \frac{h}{h} = 1, \quad \tilde{b} = \frac{b}{h}, \quad \tilde{c} = \frac{\lambda c}{h}, \quad \tilde{r} = \frac{\lambda r}{h}, \quad \tilde{w} = \frac{\lambda w}{h}. \tag{1}$$

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 leads to slightly different quantitative results (the approximation tends to underestimate the optimal discrete base stock level). However, it has no effect on the qualitative system behavior, which is the object of our study.

Because the revenues for each agent are fixed, profit maximization and cost minimization lead to the same solution. We employ profit maximization to explicitly incorporate the agents' *participation constraints*, which take the form of nonnegative expected profits. However, we introduce some variable cost notation (C_R and C_S) in equations (3)-(4) for future reference when discussing the inefficiency of the Nash solution (§5) and contracts that coordinate the system (§6). In these equations, the quantities $r - w$ and $w - c$ are independent of the supply chain decisions (c is the normalized capacity cost if no excess capacity is built) and represent fixed profits for the respective agents. The steady-state expected normalized profit per unit time for the risk-neutral retailer (Π_R) and supplier (Π_S) in terms of the two decision variables are given by

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

and

$$\Pi_S(s, \nu) = w - c - C_S(s, \nu) \tag{4}$$

6

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

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 profit per unit time Π (defined in terms of the total variable cost $C = C_R + C_S$) for this decision maker is

$$
\begin{aligned}
\Pi(s, \nu) &= \Pi_R(s, \nu) + \Pi_S(s, \nu) && (5) \\
&= r - c - C(s, \nu) \\
&= r - c(1 + \nu) - s + \frac{1 - (b + 1)e^{-\nu s}}{\nu}.
\end{aligned}
$$

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

Proposition 1 *If $r - c \geq 2\sqrt{c \ln(1 + b)}$, then the optimal centralized solution is the unique solution to the first-order conditions*

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

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

and is given by

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

The resulting profit is

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

If $r - c < 2\sqrt{c \ln(1 + b)}$, then the system generates negative profits and the optimal centralized solution is to not operate the supply chain.

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, which corresponds to a Pareto frontier for the selection of s and ν. (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 point on this Pareto frontier depends on c via $s = \nu c$ according to (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, neither w nor α play any role in this single-agent optimization, because transfer payments between the retailer and the supplier do no affect the centralized profit.

4   The Nash Solution

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

Our results are most easily presented by deriving the Nash equilibrium in the absence of participation constraints, which is done in the next proposition, and then incorporating the participation constraints, $\Pi_R \geq 0$ and $\Pi_S \geq 0$. In anticipation of subsequent analysis, we express the Nash equilibrium in terms of the backorder allocation fraction α. Let us also define the auxilliary 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 *In the absence of participation constraints, 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 profits, $\Pi_R^*(\alpha)$ and $\Pi_S^*(\alpha)$, are

$$\Pi_R^*(\alpha) = \Pi_R(s_\alpha^*, \nu_\alpha^*) = r - w - s_\alpha^*, \tag{13}$$

$$\Pi_S^*(\alpha) = \Pi_S(s_\alpha^*, \nu_\alpha^*) = w - c - \left(\frac{\ln(1+\alpha b)+2}{\ln(1+\alpha b)+1}\right)c\nu_\alpha^*. \tag{14}$$

Proof: Let $s^*(\nu)$ be the retailer's reaction curve, i.e., the optimal base stock level given a capacity ν installed by the supplier. Because (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 (15)-(16) is (11)-(12), and substituting this solution into (3)-(4) yields (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.

As mentioned previously, we assume that the two agents do not participate in the game unless their expected normalized profits in (13)-(14) are nonnegative. Hence, if either of these profits are negative, the Nash equilibrium (in the presence of participation constraints) is an inoperative supply chain. The remainder of this section is devoted to an analysis of these profits as a function of α. The supplier's profit $\Pi_S^*(\alpha)$ is an increasing function of α that satisfies

$$\Pi_S^*(0) = w - c - 2\sqrt{bc}, \qquad \Pi_S^*(1) = w - c,$$

and $\Pi_R^*(\alpha)$ is a decreasing function of α that satisfies

$$\Pi_R^*(0) = r - w, \qquad \lim_{\alpha \to 1} \Pi_R^*(\alpha) \to -\infty, \tag{17}$$

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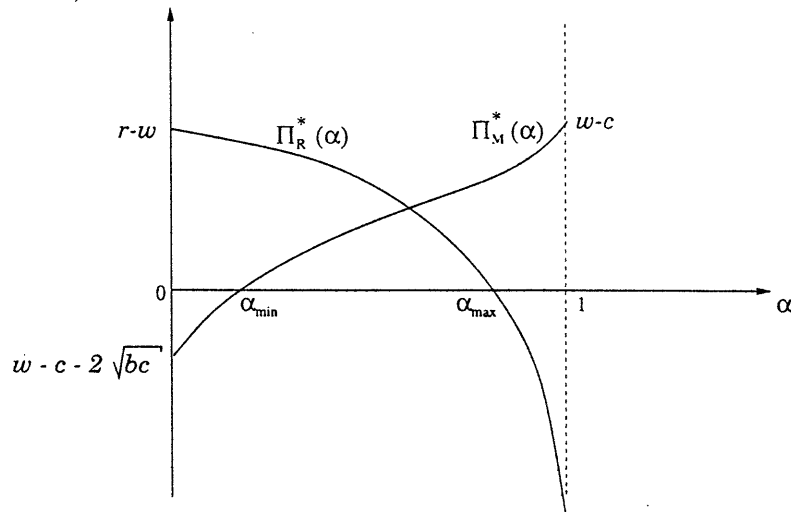

Figure 1: The retailer's (Π_R) and supplier's (Π_S) profits in the Nash equilibrium as a function of the backorder allocation fraction α.

To understand the unbounded retailer losses in (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, i.e, $\nu_1^* = 0$. This corresponds to the null recurrent case of a queueing system with an arrival rate equal to its service rate, and $s_1^* = \infty$. There is no base stock level that allows the retailer to maintain finite inventory (backorder plus holding) costs. Hence, this production-inventory system is unstable when $\alpha = 1$ and the Nash equilibrium is that the retailer does not participate, and the supply chain does not operate.

More generally, there exist $\alpha_{\min}$ and $\alpha_{\max}$ such that $\Pi_R^*(\alpha) \geq 0$ and $\Pi_S^*(\alpha) \geq 0$ if and only if $\alpha \in [\alpha_{\min}, \alpha_{\max}]$. That is, the Nash equilibrium is an inoperative supply chain when $\alpha < \alpha_{\min}$ or $\alpha > \alpha_{\max}$. The threshold $\alpha_{\max} \in (0, 1)$, and solves

$$r - w = \left(\frac{\ln(1 + \alpha b)}{\ln(1 + b) f_\alpha(b)} \right) s^*. \tag{18}$$

10

If $w - c \geq 2\sqrt{bc}$, then $\alpha_{\min} = 0$. Otherwise, $\alpha_{\min}$ solves

$$w - c = \left(\frac{\ln(1 + \alpha b) + 2}{\ln(1 + \alpha b) + 1} \right) f_\alpha(b) s^*. \tag{19}$$

We have been unable to explicitly solve for $\alpha_{\max}$ and $\alpha_{\min}$ in (18)-(19). However, to increase our understanding of these two equations, we investigate the solution in two extreme cases: When backorders are much less costly than holding inventory ($b << 1$), we have

$$\alpha_{\max} \approx \frac{\sqrt{(r - w)^4 + 4(r - w)^2 bc} - (r - w)^2}{2bc}, \qquad \alpha_{\min} \approx 1 - \frac{(w - c)^2}{4bc}. \tag{20}$$

When backorder costs are very large,

$$\alpha_{\max} \approx \frac{(r - w)^2}{\ln(1 + b)c + (r - w)^2}, \qquad \alpha_{\min} \approx \frac{\ln(1 + b)c}{\ln(1 + b)c + (w - c)^2}. \tag{21}$$

Even under the assumption $r > w > c$, it is possible that $\alpha_{\min} > \alpha_{\max}$ in (20)-(21). In this situation, even though each agent is willing to participate for some values of α, it is not possible for the retailer and supplier to simultaneously earn nonnegative profits.

5 Comparison of Solutions

In this section, we compare the centralized solution and the Nash equilibrium with respect to the total system profit, the agents' decisions, and the consumers of the product.

The Nash equilibrium is inefficient. As in §4, it is convenient to first quantify the inefficiency of the Nash equilibrium in the absence of participation constraints, and then to incorporate them later. In the absence of participation constraints, 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 magnitude of the inefficiency of a Nash equilibrium is typically quantified by comparing the profits under the centralized and Nash solutions. Because the profits $r - w$ and

$w - c$ are fixed in (3)-(4), it is more natural to restrict ourselves to the variable costs. This assumption also allows us to follow Cachon and Zipkin and compute the *competition penalty*, which is defined as the percentage increase in variable cost of the Nash equilibrium over the centralized solution. By (5) and (8), the variable cost for the centralized solution is

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

and the variable cost C_α^* associated with the Nash equilibrium in the absence of participation constraints is, by (3)-(4) and Proposition 2,

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

Hence, the competition penalty in the absence of participation constraints is $\frac{C_\alpha^* - C(s^*, \nu^*)}{C(s^*, \nu^*)} \times 100\%$, where

$$\frac{C_\alpha^* - C(s^*, \nu^*)}{C(s^*, \nu^*)} = \frac{1}{2} \left[f_\alpha(b) \left(\frac{\ln(1 + \alpha b) + 2}{\ln(1 + \alpha b) + 1} \right) + \frac{\ln(1 + \alpha b)}{\ln(1 + b) f_\alpha(b)} \right] - 1. \tag{22}$$

Surprisingly, the competition penalty in (22) is independent of the supplier's cost of excess capacity. However, this penalty is a function of α and b, and we can simplify equation (22) 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 $\alpha \to 1$. At the other extreme, $f_\alpha(b) \to \sqrt{\frac{b}{\ln(1+b)}}$ as $\alpha \to 0$, and the competition penalty in this case is given by

$$\sqrt{\frac{b}{\ln(1 + b)}} - 1 \quad \text{for} \quad b > 0. \tag{23}$$

This function is increasing and concave in b, approaches zero as $b \to 0$ and grows to ∞ as $b \to \infty$.

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

$$\frac{1}{2\sqrt{\alpha(1 - \alpha)}} - 1. \tag{24}$$

This quantity vanishes at $\alpha = 0.5$, is symmetric about $\alpha = 0.5$, is convex for $\alpha \in (0, 1)$, and approaches ∞ as $\alpha \to 0$ and $\alpha \to 1$. Finally, for the case $b \to 0$, the competition penalty is

given by

$$\frac{2-\alpha}{2\sqrt{1-\alpha}} - 1, \tag{25}$$

which is an increasing convex function of α. Consistent with the previous analyses, 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 (22) does not consider the agents' participation constraints, some of the large inefficiencies in the latter regimes are not attainable by the supply chain.

To complement these asymptotic results, we compute in Table 1 the competition penalty in (22) for various values of α and b. Our asymptotic results agree with the numbers around the four edges of this table. Two new insights emerge from Table 1. First, the competition penalty is minimized by α near 0.5 when $b \geq 1$. Second, the competition penalty appears to be an increasing function of b for $\alpha < 0.5$, and a U-shaped function of b for $\alpha > 0.5$.

Comparison of decision variables. Figure 2 plots 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 2 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 2, 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 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 2, we have $\alpha_s > \alpha_\nu$.*

	α						
b	0	0.1	0.3	0.5	0.7	0.9	1
10^{-3}	0.0%	0.2%	1.6%	6.1%	18.7%	73.9%	$\infty\%$
1	20.1%	14.8%	7.8%	5.9%	12.9%	59.5%	$\infty\%$
2	34.9%	23.8%	10.7%	5.8%	11.1%	56.1%	$\infty\%$
3	47.1%	30.1%	12.3%	5.6%	10.2%	54.8%	$\infty\%$
4	57.6%	34.9%	13.3%	5.4%	9.6%	54.1%	$\infty\%$
5	67.0%	38.7%	14.0%	5.3%	9.2%	53.8%	$\infty\%$
6	75.6%	41.7%	14.5%	5.2%	9.0%	53.6%	$\infty\%$
7	83.5%	44.3%	14.9%	5.1%	8.8%	53.5%	$\infty\%$
8	90.8%	46.4%	15.2%	5.0%	8.7%	53.5%	$\infty\%$
9	97.7%	48.3%	15.4%	4.9%	8.5%	53.5%	$\infty\%$
10	104.2%	49.9%	15.6%	4.8%	8.4%	53.5%	$\infty\%$
10^{200}	$4.6 \times 10^{100}\%$	66.7%	9.2%	0.0%	9.1%	66.5%	$\infty\%$

Table 1: The competition penalty in (22) in the absence of participation constraints for different values of the backorder-to-holding cost ratio b and the backorder allocation fraction α.

Proof: By Figure 2, if $\alpha_s \leq \alpha_\nu$ then there exists $\hat{\alpha} \in [\alpha_s, \alpha_\nu]$ such that $\nu_{\hat{\alpha}}^* s_{\hat{\alpha}}^* \geq \nu^* s^*$. However, this inequality together with (6) and (15) implies that $f_{\hat{\alpha}}(b) = \ln(1+b)$, i.e., $\hat{\alpha} = 1$. 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 $\alpha_s > \alpha_\nu$. ∎

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

$$\frac{\ln(1 + \alpha b)}{\ln(1 + b) f_\alpha(b)} = 1. \qquad (26)$$

As $b \to 0$, we have $f_\alpha(b) \to \sqrt{1 - \alpha}$ and $\frac{\ln(1+\alpha b)}{\ln(1+b)} \to \alpha$. Therefore, as $b \to 0$, α_s satisfies $\frac{\alpha}{\sqrt{1-\alpha}} = 1$, or $\alpha_s = \frac{\sqrt{5}-1}{2} \approx 0.618$, which is the inverse of the *golden-section number* that
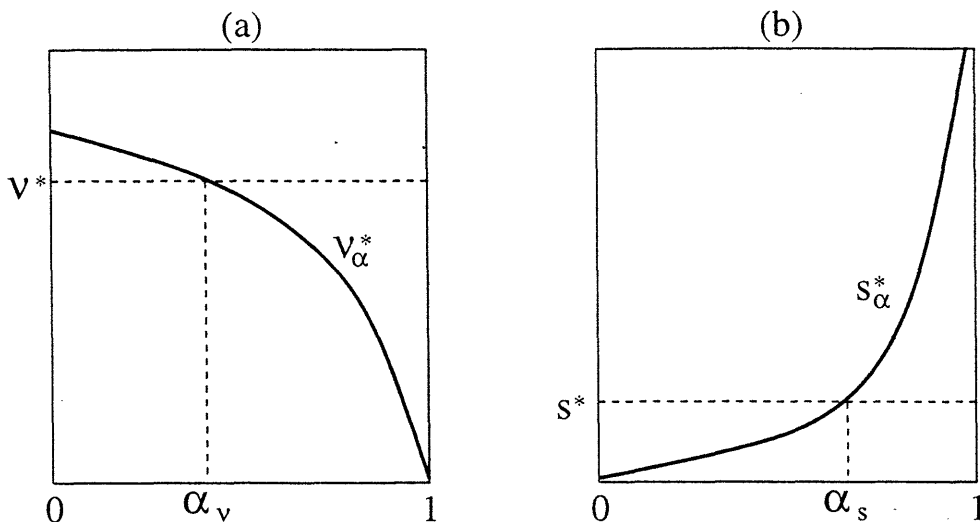
14

Figure 2: The optimal Nash production capacity (ν_α^*) and the optimal Nash base stock level (s_α^*) as a function of the backorder allocation fraction α. The centralized solutions are ν^* and s^*.

arises in a variety of disciplines (e.g., Vajda 1989). As $b \to \infty$, we have $f_\alpha(b) \to \sqrt{\frac{1-\alpha}{\alpha}}$, and $\frac{\ln(1+\alpha b)}{\ln(1+b)}$. In this case, α_s satisfies $\sqrt{\frac{\alpha}{1-\alpha}} = 1$, 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 b (e.g., $\alpha_s \geq 0.61$ for $b \in [1, 10]$).

By (11), α_ν solves $f_\alpha(b) = 1$. As $b \to 0$, this condition becomes $\sqrt{1 - \alpha} = 1$, which gives $\alpha_\nu = 0$. As $b \to \infty$, the condition becomes $\sqrt{\frac{1-\alpha}{\alpha}} = 1$, 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 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 $Pr(N \geq s) = e^{-\nu s}$; hence, we refer to $e^{-\nu s} \times 100\%$ as the *service level*. By equations (6) and (15), this quantity 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. Even though the system is not stable for $\alpha = 1$, customers generally

15

desire a larger value of α; i.e., 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 construct a coordinating contract that specifies linear tranfer payments based on retailer inventory and backorder levels, the capacity level and the cost parameters. As in our earlier analysis, this information is assumed to be common knowledge. Cachon and Zipkin also use a linear transfer payment based on inventory levels to coordinate their supply chain, and readers are referred to §1.5 of Cachon for a survey of alternative types of contracts in the multi-echelon 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 (1997) terminology, we assume a *voluntary compliance* regime, where both the retailer and the supplier choose their buffer resource levels to maximize their own profits.

Although we have used profit maximization thus far, because the revenues are fixed our presentation of the contract analysis is simpler – and perhaps more natural – in the setting of variable cost minimization. Consequently, we first present the contract in the absence of participation constraints, and at the end of this section we incorporate the revenues via the agents' participation constraints. Because the model contains three cost components, the most general linear transfer payment (without loss of generality, the payment is from the supplier to the retailer) contains cost coefficients for the holding cost, backorder cost and capacity cost, which are denoted by γ_h, γ_b and γ_c, respectively. The steady-state expected normalized (recall from (1) that $h = 1$) transfer payment per unit time is given by

$$T(s, \nu) = \gamma_h \left(E[(s - N)^+] \right) + \gamma_b \left(bE[(N - s)^+] \right) + \gamma_c c\nu. \tag{27}$$

This transfer payment modifies the profit functions in (3)-(4) for the retailer and supplier,

respectively, to

$$\tilde{C}_R(s,\nu) = C_R(s,\nu) - T(s,\nu) = (1-\gamma_h)\left(s - \frac{1-e^{-\nu s}}{\nu}\right) + b(\alpha - \gamma_b)\frac{e^{-\nu s}}{\nu} - \gamma_c c\nu, \quad (28)$$

$$\tilde{C}_S(s,\nu) = C_S(s,\nu) + T(s,\nu) = \gamma_h\left(s - \frac{1-e^{-\nu s}}{\nu}\right) + b(1-\alpha+\gamma_b)\frac{e^{-\nu s}}{\nu} + (1+\gamma_c)c\nu. \quad (29)$$

The following proposition provides a general result for the coordination of static games with additive utility structures, and may be applicable to other supply chain problems.

Proposition 4 *Consider a static game with n players, where player i has action space X_i for $i = 1,\ldots,n$. For any action $x \in X_1 \times X_2 \times \cdots \times X_n$, the utility of player i is given by $u_i(x)$. The utility function for the centralized planner problem is $\sum_{i=1}^{n} u_i(x)$, and let $x^* \in argmax_x \sum_{i=1}^{n} u_i(x)$ be an optimal centralized solution. For $i \neq j$, let $T_{ij}(x)$ be a linear transfer payment that player i pays to player j if the action is x. If $T_{ij}(x) = \gamma_j u_i(x)$, where $\sum_{j=1}^{n} \gamma_j = 1$, then x^* is a Nash equilibrium of the modified static game in which player i's utility function is given by*

$$\tilde{u}_i(x) = u_i(x) - \sum_{j \neq i} T_{ij}(x) + \sum_{j \neq i} T_{ji}(x).$$

Proof: By our assumptions on $T_{ij}(x)$, the modified utility function for player i can be written as

$$\tilde{u}_i(x) = \left(1 - \sum_{j \neq i} \gamma_j\right) u_i(x) + \gamma_i \sum_{j \neq i} u_j(x) \quad (30)$$

$$= \gamma_i \sum_{j=1}^{N} u_j(x). \quad (31)$$

Hence, $\sum_{i=1}^{n} \tilde{u}_i(x) = \sum_{i=1}^{n} u_i(x)$, and x^* is a Nash equilibrium of the modified game. ∎

Applying Proposition 4 to equations (28)-(29) shows that coordination in the absence of participation constraints can be achieved if

$$-\gamma_c = \alpha - \gamma_b = 1 - \gamma_h. \quad (32)$$

A comparison of equations (3)-(4) and (28)-(29) implies that the modified cost functions are given by

$$\tilde{C}_R(s,\nu) = (1-\gamma_h)C(s,\nu), \qquad \tilde{C}_S(s,\nu) = \gamma_h C(s,\nu). \quad (33)$$

17

This is a consequence of the more general result in (31), which shows that any split of the total profit is possible by selecting appropiate values of the $\{\gamma_i\}$ parameters. Note that γ_h need not be in the interval $[0, 1]$.

Although we appear to have a degree of freedom in splitting the profits via γ_h, two conditions must be met by γ_h to guarantee that both agents will enter into the contract. First, both agents must be better off under the Nash equilibrium with the transfer payments than under the Nash equilibrium without the transfer payments, i.e.,

$$C_R(s_\alpha^*, \nu_\alpha^*) \geq (1 - \gamma_h)C(s, \nu),$$
$$C_S(s_\alpha^*, \nu_\alpha^*) \geq \gamma_h C(s, \nu).$$

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

$$\underline{\gamma}_\alpha(b) = 1 - \frac{\ln(1 + \alpha b)}{2\ln(1 + b)f_\alpha(b)},$$
$$\bar{\gamma}_\alpha(b) = \frac{f_\alpha(b)}{2}\left(\frac{\ln(1 + \alpha b) + 2}{\ln(1 + \alpha b) + 1}\right).$$

The second condition on γ_h requires that both agents achieve a nonnegative profit. The resulting inequalities can be calculated using equations (3)-(4), (9) and (28)-(29). Combining these two conditions, we can characterize the range of coordinating contracts that are attractive to both agents as

$$\max\left\{1 - \frac{r - w}{2\sqrt{c\ln(1 + b)}}, \underline{\gamma}_\alpha(b)\right\} \leq \gamma_h \leq \min\left\{\frac{w - c}{2\sqrt{c\ln(1 + b)}}, \bar{\gamma}_\alpha(b)\right\}. \qquad (34)$$

If condition (34) is satisfied, we say that the *system can be coordinated* (by the contract), and the remainder of this section is devoted to an analysis of this condition. First we note that it is always possible to coordinate the system if both agents are willing to participate in the Nash equilibrium in §4. This conclusion stems from the fact that if both players are willing to participate in the Nash equilibrium, then the additional profits from the centralized solution can be split so that each agent is still willing to participate and is at least as well off as in the Nash solution. Hence, the nontrivial cases to analyze are $\alpha \in [0, \alpha_{\min})$ and

18

$\alpha \in (\alpha_{\max}, 1]$; recall that these two thresholds characterize the participation constraints in the Nash equilibrium and are defined in (18)-(19).

We can analyze (34) when α approaches one of its extreme values. As $\alpha \to 0$, we have

$$\underline{\gamma}_\alpha(b) \to 1 \quad \text{and} \quad \bar{\gamma}_\alpha(b) \to \sqrt{\frac{b}{\ln(1+b)}}. \tag{35}$$

Thus, because $\alpha \to 0$ implies that $\alpha \leq \frac{r-w}{\sqrt{bc}}$, equation (34) reduces to

$$1 \leq \gamma_h \leq \min\left\{\frac{w-c}{2\sqrt{c\ln(1+b)}}, \sqrt{\frac{b}{\ln(1+b)}}\right\}. \tag{36}$$

If $w - c \geq 2\sqrt{bc}$, then the second term in the brackets in (36) achieves the minimum, and the interval in (36) is nonempty because $b \geq \ln(1+b)$ for $b > 0$. Because $f_\alpha(b) \to \sqrt{\frac{b}{\ln(1+b)}}$ as $\alpha \to 0$, the inequality $w - c \geq 2\sqrt{bc}$ can be recognized as the supplier's participation constraint in (19) in the $\alpha \to 0$ case. If, on the other hand, $w - c < 2\sqrt{bc}$, then (36) reduces to

$$1 \leq \gamma_h \leq \frac{w-c}{2\sqrt{c\ln(1+b)}}. \tag{37}$$

By (9), the interval in (37) is nonempty if and only if $\Pi(s^*, \nu^*) \geq r - w$.

To summarize, when the supplier absorbs almost all of the backorder cost (i.e., $\alpha \to 0$), a coordinating contract is always possible if the supplier is willing to participate in the absence of the contract (i.e., $w - c \geq 2\sqrt{bc}$). If the centralized system profit $\Pi(s^*, \nu^*)$ is less than the retailer's fixed profit, $r - w$, then the supplier will not enter into the contract. Most interesting is the intermediate case, $w - c \in (2\sqrt{c\ln(1+b)}, 2\sqrt{bc})$, where the system profit is bigger than the retailer's fixed profit, but the supplier is unwilling to participate in the absence of a contract. Here, the excess system profit enables the contract to entice the supplier to participate in the supply chain.

Turning to the $\alpha \to 1$ case, we have $f_\alpha(b) \to 0$ and therefore $\underline{\gamma}_\alpha(b) \to -\infty$ and $\bar{\gamma}_\alpha(b) \to 0$. Thus, condition (34) reduces to

$$1 - \frac{r-w}{2\sqrt{c\ln(1+b)}} \leq \gamma_h \leq 0, \tag{38}$$

19

Equation (9) shows that the interval in (38) is nonempty if and only if $\Pi(s^*, \nu^*) \geq w - c$. Hence, when the retailer incurs almost all of the backorder cost, the contract is attractive to both parties if and only if the centralized system profit exceeds the supplier's fixed profit, $w - c$. In this case, the contract coefficient $\gamma_h \leq 0$, and (33) implies that $\tilde{C}_S(s, \nu) \leq 0$; i.e., the retailer subsidizes the supplier's entire operation.

The analysis is more difficult in the general case, $\alpha \in (0, \alpha_{\min}) \cup (\alpha_{\max}, 1)$. Here, we consider the extreme values of b. As $b \to \infty$, $\underline{\gamma}_\alpha(b) \to 1 - \frac{1}{2}\sqrt{\frac{\alpha}{1-\alpha}}$ and $\bar{\gamma}_\alpha(b) \to \frac{1}{2}\sqrt{\frac{1-\alpha}{\alpha}}$. Even though $\underline{\gamma}_\alpha(b) < \bar{\gamma}_\alpha(b)$ for $\alpha \in [0, 1]$ in this case (this can be derived with the change of variable $x = \sqrt{\frac{\alpha}{\alpha-1}}$), neither of these quantities are binding in (34), and coordination becomes impossible because the lower bound in (34) is at least 1 and the upper bound in (34) goes to 0. In contrast, as $b \to 0$, $\underline{\gamma}_\alpha(b)$ and $\bar{\gamma}_\alpha(b)$ are binding in (34), and $\underline{\gamma}_\alpha(b) \to 1 - \frac{\alpha}{\sqrt{1-\alpha}}$ and $\bar{\gamma}_\alpha(b) \to \sqrt{1-\alpha}$. Because $1 - \frac{\alpha}{\sqrt{1-\alpha}} \leq \sqrt{1-\alpha}$ for $\alpha \in [0, 1]$, coordination is always possible as $b \to 0$.

In conclusion, system coordination is most difficult when α takes on an extreme value. We can show (the proof is omitted, but the result follows from our analyses of the $\alpha \to 0$ and $\alpha \to 1$ cases) that a sufficient (but not necessary) condition for coordination is that the optimal centralized profit exceeds the fixed profits of both agents, $\Pi(s^*, \nu^*) \geq \max\{r - w, w - c\}$. Moreover, if we consider w as endogenous, then the likelihood of coordination is maximized by minimizing the right side of this inequality. This is achieved by $w = (r + c)/2$, which splits the fixed profits evenly. By (9), coordination is always possible in this case if $r - c \geq 4\sqrt{c \ln(1 + b)}$.

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 $\Pi_S(s, \nu)$ in (4), given the retailer's best response, $s^*(\nu)$ in (15). This straight-

forward computation leads to the following proposition.

Proposition 5 *In the absence of participation constraints, the equilibrium in the supplier's Stackelberg game is*

$$\bar{s}_\alpha = s_\alpha^* \sqrt{1 + \ln(1 + \alpha b)}, \qquad \bar{\nu}_\alpha = \frac{\nu_\alpha^*}{\sqrt{1 + \ln(1 + \alpha b)}}. \tag{39}$$

The agents' profits are

$$\Pi_S(\bar{s}_\alpha, \bar{\nu}_\alpha) = w - c - 2\sqrt{\frac{(1 - \alpha)bc}{1 + \alpha b}}, \qquad \Pi_R(\bar{s}_\alpha, \bar{\nu}_\alpha) = r - w - \bar{s}_\alpha. \tag{40}$$

Equation (39) 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 (39) than in the Nash equilibrium. The discrepancy between the Stackelberg and Nash solutions increases as α and b increase.

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

$$\Pi_S(s_\alpha^*, \nu_\alpha^*) = w - c - 2\sqrt{\frac{(1 - \alpha)bc}{1 + \alpha b}} \left(\frac{\ln(1 + \alpha b) + 2}{2\sqrt{\ln(1 + \alpha b) + 1}} \right).$$

The function $\frac{x+2}{2\sqrt{x+1}}$ is strictly increasing in $[0, \infty)$, and is equal to 1 when $x = 0$. Thus, it is always the case that $\Pi_S(\bar{s}_\alpha, \bar{\nu}_\alpha) \geq \Pi_S(s_\alpha^*, \nu_\alpha^*)$; this is to be expected, because the supplier incorporates the retailer's best response when selecting his level of capacity. However, $\Pi_S(\bar{s}_\alpha, \bar{\nu}_\alpha) = \Pi_S(s_\alpha^*, \nu_\alpha^*)$ when $\alpha = 0$, $\alpha = 1$ or $b = 0$, and so the supplier does not benefit from being the leader in these extreme cases.

By (13) and (39), the difference in the retailer's profit between the Nash equilibrium and the Stackelberg equilibrium is

$$\Pi_R(s_\alpha^*, \nu_\alpha^*) - \Pi_R(\bar{s}_\alpha, \bar{\nu}_\alpha) = s_\alpha^* \left(\sqrt{1 + \ln(1 + \alpha b)} - 1 \right). \tag{41}$$

21

As expected, the retailer is worse off in the supplier's Stackelberg equilibrium than in the Nash equilibrium. By (8), (10), (12) and (41), the reduction in the retailer's profit from being the follower vanishes as $\alpha \to 0$ and $b \to 0$, and increases with α and b.

A comparison of the total system profit shows that $\Pi_R(s_\alpha^*, \nu_\alpha^*) + \Pi_S(s_\alpha^*, \nu_\alpha^*) \geq \Pi_R(\bar{s}_\alpha, \bar{\nu}_\alpha) + \Pi_S(\bar{s}_\alpha, \bar{\nu}_\alpha)$ if and only if

$$\frac{(1 + \alpha b) \ln(1 + \alpha b)}{(1 - \alpha) b} - \sqrt{1 + \ln(1 + \alpha b)} + 1 \geq 0. \tag{42}$$

Condition (42) holds for large values of α, but is not true in general. Because the left side of (42) equals zero when $\alpha = 0$, is increasing in $\alpha \geq \alpha_0$ if it is increasing in α at α_0, and has a derivative with respect to α equal to $1 - \frac{b}{2}$ when $\alpha = 0$, we conclude that for $b \leq 2$ the Nash solution achieves a higher system profit 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 (42) with equality.

Now we turn to the participation constraints. We start with the follower (i.e., the retailer) because he performs the inner maximization in this game. By (8), (10), (12) and (39), we can express the optimal Stackelberg base stock level as

$$\bar{s}_\alpha = \ln(1 + \alpha b) \sqrt{\frac{(1 + \alpha b) c}{(1 - \alpha) b}}.$$

Hence, it follows from (40) that the retailer's participation constraint, $\Pi_R(\bar{s}_\alpha, \bar{\nu}_\alpha) \geq 0$, is equivalent to

$$(r - w) \sqrt{\frac{b}{c}} \geq \ln(1 + \alpha b) \sqrt{\frac{1 + \alpha b}{1 - \alpha}}. \tag{43}$$

The right side of (43) is increasing in $\alpha \geq 0$, and so there exists a threshold, call it $\bar{\alpha}_{\max}$, such that the participation constraint is satisfied if and only if $\alpha \leq \bar{\alpha}_{\max}$ (i.e., $\bar{\alpha}_{\max}$ solves (43) with equality). For future reference, let us also define the threshold

$$\bar{\alpha}_{\min} = \frac{4bc - (w - c)^2}{4bc + b(w - c)^2}. \tag{44}$$

The supplier's profit $\Pi_S(\bar{s}_\alpha, \bar{\nu}_\alpha)$ in (40) is nonnegative if and only if $\alpha \geq \bar{\alpha}_{\min}$.

There are three cases to examine: $\alpha > \bar{\alpha}_{\max}$, $\alpha \in [\bar{\alpha}_{\min}, \bar{\alpha}_{\max}]$ and $\alpha < \bar{\alpha}_{\min}$. If $\alpha > \bar{\alpha}_{\max}$ then the retailer's participation constraint is violated. To avoid an inoperative supply chain, which would give zero profits to both agents, the supplier – as the leader – has the luxury of selecting a capacity level that raises the retailer's expected profit to zero. That is, by (15) the supplier chooses $\bar{\nu}_\alpha = \frac{\ln(1+\alpha b)}{r-w}$, and the retailer subsequently selects $\bar{s}_\alpha = r - w$. If the supplier's resulting profit (see (4)) is nonnegative, i.e.,

$$w - c - \frac{(r-w)(1-\alpha)b}{(1+\alpha b)\ln(1+\alpha b)} - \frac{\ln(1+\alpha b)c}{r-w} \geq 0, \tag{45}$$

then this is the supplier's Stackelberg equilibrium for the $\alpha > \bar{\alpha}_{\max}$. If (45) is violated, then the equilibrium is an inoperative supply chain.

In the second case, $\alpha \in [\bar{\alpha}_{\min}, \bar{\alpha}_{\max}]$, both participation constraints are satisfied and the Stackelberg equilibrium is given by (39). In the last case, $\alpha < \bar{\alpha}_{\min}$, the supplier is unable to earn a nonnegative expected profit, and decides not to participate. Finally, as in §4, it is possible that $\bar{\alpha}_{\min} > \bar{\alpha}_{\max}$; in this case, there is no value of α that simultaneously provides nonnegative profits for both agents. For brevity's sake, we do not pursue asymptotics for $\bar{\alpha}_{\min}$ and $\bar{\alpha}_{\max}$.

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

Proposition 6 *Let $(\hat{\nu}_\alpha, \hat{s}_\alpha)$ be the equilibirum when the retailer is the Stackelberg leader in the absence of participation constraints. Define $\hat{\beta} \geq 0$ to be the unique nonnegative solution of*

$$\beta^2 + (\beta + 2)\left(1 - e^{-(\beta - \nu_\alpha^* s_\alpha^*)}\right) = 0. \tag{46}$$

Then the Stackelberg solution is

$$\hat{\nu}_\alpha = \sqrt{\frac{(1-\alpha)b(\hat{\beta}+1)e^{-\hat{\beta}}}{c}}, \qquad \hat{s}_\alpha = \frac{\hat{\beta}}{\hat{\nu}_\alpha}. \tag{47}$$

23

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 7 *The following five inequalities hold:*

$$\hat{\nu}_\alpha \hat{s}_\alpha \leq \nu_\alpha^* s_\alpha^*, \tag{48}$$

$$\hat{\nu}_\alpha \geq \nu_\alpha^*, \tag{49}$$

$$\hat{s}_\alpha \leq s_\alpha^*, \tag{50}$$

$$\Pi_R(\hat{s}_\alpha, \hat{\nu}_\alpha) \geq \Pi_R(s_\alpha^*, \nu_\alpha^*), \tag{51}$$

$$\Pi_S(\hat{s}_\alpha, \hat{\nu}_\alpha) \leq \Pi_S(s_\alpha^*, \nu_\alpha^*). \tag{52}$$

While inequalities (49)-(52) mirror our results for the supplier Stackelberg game, inequality (48) states that the customer service level, $e^{-\nu s} \times 100\%$, is lower in the retailer's Stackelberg equilibrium than in the supplier's Stackelberg equilibrium (and the Nash equilibrium). Inequality (48) also implies that the retailer Stackelberg equilibrium is inefficient relative to the centralized solution. Analytical approximations (using $e^{-x} \approx 1 - x$ in (46)) 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.

Finally, numerical experiments under a wide range of values for α and b suggest that the retailer's Stackelberg game achieves a higher total profit than the supplier's Stackelberg game. However, we have been unable to provide a proof.

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 supplier chooses the capacity level and the retailer chooses the base stock level) that buffers against expensive backorders of the retailer's inventory. When the inventory backorder cost is incurred entirely by the retailer (i.e., the backorder allocation fraction $\alpha = 1$), 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 in the absence of participation constraints. The Nash equilibrium is always inefficient: The agents' selfish behavior degrades overall system performance. The Nash equilibrium is asymptotically efficient in two cases: (i) The backorder cost goes to zero and the retailer is not charged for backorders, and (ii) the backorder cost goes to infinity and is split evenly between the two agents. In the absence of participation constraints, 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. 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.

We assume that the agents only participate if their expected profits are nonnegative. In the Nash equilibrium, the retailer refuses to participate when α gets sufficiently large (the inventory costs become arbitrarily large when $\alpha = 1$), and the supplier may also (depending

on the cost and revenue parameters) refuse to participate when α gets too small. A simple linear transfer payment, which is based on actual inventory and backorder levels, the capacity level and the cost parameters, coordinates the system in the absence of participation constraints. We derive bounds on α for when the coordinating contract is attractive to both parties, in that each agent achieves a nonnegative profit that is no smaller than his Nash equilibrium profit. There are values of α for which the contract will lead to the operation of an otherwise inoperative supply chain; i.e., the extra system profit generated by the contract is sufficient to entice the nonparticipating agent into playing. Overall, we find that a contract is more likely to be entered into by both agents when the system is reasonably profitable (i.e., the optimal profit of the centralized system is large) and relatively well-balanced (i.e., α is near 0.5 and the wholesale price w is intermediate between the retailer's selling price r and the supplier's manufacturing cost c). Finally, when one of the agents is the Stackelberg leader, he builds less of his buffer resource and receives a higher profit than in the Nash equilibrium, and the other agent builds more of his buffer resource and receives a smaller profit. 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.

Recall that our model is quite similar to the two-stage inventory model of Cachon and Zipkin: The two main differences are the single-server vs. infinite-server model for the manufacturing process, and our inclusion of revenue, and hence participation constraints. A key difference in the results of the two models occurs in the $\alpha = 1$ case. When the supplier does not care about retailer backorders, 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 affects 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 our models is that Cachon and Zipkin's agents hold less inventory 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. Again, this may be due to the larger nonlinear effect that capacity has in a make-to-stock queue relative to the upstream inventory in a two-stage inventory model.

In our view, the make-to-stock queue is an attractive operations management model to embed into a game-theoretic framework. The model is in most ways richer than the newsvendor model 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 and the impact of the retailer's order process 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 multi-agent models for supply chain management.

Appendix

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

However, we have checked that $\lim_{s \to \infty} \Pi(s, \nu) \to -\infty$ for $\nu > 0$, and $\lim_{\nu \to \infty} \Pi(s, \nu) \to -\infty$ for $s \geq 0$, which implies that a global maximum exists. From the first-order conditions (6) and (7), the only interior point that is a candidate for the global maximum is

(s^*, ν^*). In addition, the Hessian of $\Pi(s, \nu)$ at (s^*, ν^*) is given by

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

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

$$\Pi(0, \nu) = r - c - \frac{b}{\nu} + c\nu \leq r - c - 2\sqrt{cb} < \Pi(s^*, \nu^*) \quad \text{for} \quad \nu > 0, b > 0.$$

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

Proof of Proposition 6. To derive the Stackelberg equilibrium, we find it convenient to define

$$\beta = \nu s, \tag{53}$$

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

$$e^{-\beta} \left(\frac{\beta + 1}{\beta^2} \right) = \frac{c}{(1 - \alpha)bs^2}. \tag{54}$$

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 (3) and (53), the retailer's profit is

$$\Pi_R(\beta, \nu) = r - w - \left(\frac{\beta - 1 + (1 + \alpha b)e^{-\beta}}{\nu} \right). \tag{55}$$

Solving (54) for s and using (53) gives

$$\nu(\beta) = \sqrt{\frac{(1 - \alpha)b(\beta + 1)e^{-\beta}}{c}}. \tag{56}$$

Substituting (56) into (55) yields the retailer's profit as the following concave function of $\beta \geq 0$:

$$\Pi_R(\beta) = r - w - \sqrt{\frac{c}{(1 - \alpha)b}} \left(\frac{\beta - 1 + (1 + \alpha b)e^{-\beta}}{\sqrt{(\beta + 1)e^{-\beta}}} \right). \tag{57}$$

28

Therefore, the first-order condition

$$\frac{\beta^2 + (\beta + 2)(1 - (1 + \alpha b)e^{-\beta})}{2(\beta + 1)^{\frac{3}{2}}e^{\frac{-\beta}{2}}} = 0 \tag{58}$$

is sufficient for optimality. Because $1 + \alpha b = e^{\nu_\alpha^* s_\alpha^*}$ and the denominator of (58) is always positive, condition (58) is equivalent to (46). Hence, by (46), (53) and (56), the Stackelberg equilibrium is given by (46)-(47). ∎

Proof of Proposition 7. To prove (48), note that the left side of (46) is positive if $\beta > \nu_\alpha^* s_\alpha^*$. Thus, the root $\hat{\beta}$ of (46) must satisfy $\hat{\beta} \leq \nu_\alpha^* s_\alpha^*$, i.e., $\hat{\nu}_\alpha \hat{s}_\alpha \leq \nu_\alpha^* s_\alpha^*$.

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

$$\frac{d\nu(s)}{ds} \leq 0. \tag{59}$$

In addition, the function $e^{-\beta}(\beta + 1)/\beta^2$ is decreasing in $\beta > 0$. Hence, from (54) and inequality (48), we conclude that $\hat{s}_\alpha \leq s_\alpha^*$. Finally, (59) and $\hat{s}_\alpha \leq s_\alpha^*$ implies that $\hat{\nu}_\alpha \geq \nu_\alpha^*$.

The retailer's profit in (57) is a decreasing function of β for $\beta \geq \hat{\beta}$. Hence, inequality (51) follows from (48). To prove (52), i.e., $\Pi_S(\beta^*, \nu_\alpha^*) \geq \Pi_S(\hat{\beta}, \hat{\nu}_\alpha)$, we first use (4) to rewrite the supplier's profit as

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

The function $\Pi_S(\beta, \nu)$ is increasing in β for $\nu > 0$, and so inequality (48) implies that $\Pi_S(\beta^*, \hat{\nu}_\alpha) \geq \Pi_S(\hat{\beta}, \hat{\nu}_\alpha)$. Hence the proof of (52) will follow if we can show that $\Pi_S(\beta^*, \nu_\alpha^*) \geq \Pi_S(\beta^*, \hat{\nu}_\alpha)$. For any fixed nonnegative β, the function $\Pi_S(\beta, \nu)$ is concave in ν, achieves its only maximum at $\nu(\beta) = \sqrt{\frac{(1-\alpha)be^{-\beta}}{c}}$, and is decreasing 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 (49) imply that $\Pi_S(\beta^*, \nu_\alpha^*) \geq \Pi_S(\beta^*, \hat{\nu}_\alpha)$, which completes the proof of (52). ∎

References

1. Agrawal, N., S. Nahmias, A. Tsay. 1999. Modeling supply chain contracts: A review. Chapter 10 in Tayur S., M. Magazine, R. Ganeshan (Eds.), *Quantitative models for supply chain management.* Kluwer Academic Publishers.

2. Anupindi, R., Y. Bassok. 1999. Supply contracts with quantity commitments and stochastic demand. Chapter 7 in Tayur S., M. Magazine, R. Ganeshan (Eds.), *Quantitative models for supply chain management.* Kluwer Academic Publishers.

3. Buzacott, J. A., S. Price, J. G. Shanthikumar. 1992. Service levels in multistage MRP and base stock controlled production systems. In Fandel, G., T. Gulledge, A. Jones (Eds.), *New directions for operations research in manufacturing.* Springer, Berlin, 445-463.

4. Cachon, G. 1999. Competitive supply chain inventory management. Chapter 5 in Tayur S., M. Magazine, R. Ganeshan (Eds.), *Quantitative models for supply chain management.* Kluwer Academic Publishers.

5. Cachon, G., M. Lariviere. 1997. Contracting to assure supply or what did the supplier know and when did he know it? Working paper, Fuqua School of Business, Duke University.

6. Cachon, G., P. Zipkin. 1999. Competitive and cooperative inventory policies in a two-stage supply chain. *Management Science* **45**, 936-953.

7. Chen, F. 1999. Decentralized supply chains subject to information delays. To appear in *Management Science.*

8. Chen, F., A. Federgruen, Y. Zheng. 1999. Coordination mechanisms for a distribution system with one supplier and multiple retailers. Graduate School of Business, Columbia University, New York, NY.

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

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

11. Federgruen, A., P. Zipkin. 1986. An inventory model with limited production capacity and uncertain demands I: The average-cost criterion. *Math. Operations Research* **11**, 193-207.

12. Gavirneni, S., R. Kapuscinski, S. Tayur. 1996. Value of information in capacitated supply chains. Working paper, GSIA, Carnegie Mellon University, Pittsburgh, PA.

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

14. Harrison, J. M. 1988. Brownian models of queueing networks with heterogeneous customer populations. In *Stochastic differential systems, stochastic control theory and applications*, W. Fleming, P. L. Lions (eds.). IMA Volume **10**, Springer-Verlag, New York, 147-186.

15. Lariviere, M. 1999. Supply chain contracting and coordination with stochastic demand. Chapter 8 in Tayur S., M. Magazine, R. Ganeshan (Eds.), *Quantitative models for supply chain management*. Kluwer Academic Publishers.

16. Lee, Y. J., P. Zipkin. 1992. Tandem queues with planned inventories. *Operations Research* **40**, 936-947.

17. Markowitz, D. M., M. I. Reiman, L. M. Wein. 1999. The stochastic economic lot scheduling problem: Heavy traffic analysis of dynamic cyclic policies. To appear in *Operations Research*.

18. Morse, P. 1958. *Queues, inventories and maintenance.* Wiley, New York.

19. 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.

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

21. Topkis, D. M. 1998. *Supermodularity and complementarity.* Princeton University Press, Princeton, New Jersey.

22. Tsay, A. A., W. S. Lovejoy. 1999. Quantity flexibility contracts and supply chain performance. To appear in *Manufacturing & Service Operations Management.*

23. Vajda, S. 1989. *Fibonacci & Lucas numbers, and the golden section: Theory and applications.* Ellis Horwood Limited, England.

24. Wein, L. M. 1992. Dynamic scheduling of a multiclass make-to-stock queue. *Operations Research* **40**, 724-735.

25. Zheng, Y., P. Zipkin. 1990. A queueing model to analyze the value of centralized inventory information. *Operations Research* **38**, 296-307.