Selling to Strategic Consumers When Product Value is Uncertain: The Value of Matching Supply and Demand

Robert Swinney
Graduate School of Business, Stanford University
518 Memorial Way, Stanford, CA 94305-5015, swinney@stanford.edu

May, 2008

Abstract

Quick response inventory practices—which combine reduced production leadtimes, sophisticated information systems, and continuous demand forecasting improvement—are often discussed as a potential remedy to the negative aspects of supply and demand mismatches. We address the value of these practices when selling to a forward-looking consumer population with uncertain, heterogeneous values for a product. Consumers have the option of purchasing the product early, before its value has been learned, or delaying the purchase decision until a time at which valuation uncertainty has been resolved, a trade-off frequently made by consumers shopping for new or innovative products. While individual consumer valuations are uncertain \textit{ex ante}, the market size is uncertain to the firm. The firm may either commit to a single production run at a low unit cost prior to learning demand, or commit to a quick response strategy which allows additional production (at a higher unit cost) after learning additional demand information. We find that it is possible for a quick response strategy to decrease the profit of the firm, though whether this occurs depends on various characteristics of the market; specifically, we identify conditions under which quick response always increases profits (when prices are increasing, when dissatisfied consumers can return the product) and conditions under which quick response may decrease profits (when prices are constant or when consumer returns are not allowed). Finally, we demonstrate that our model is also applicable to a manufacturer selling to a population of strategic retailers, and we discuss factors that influence the manufacturer’s incentives to adopt a short leadtime manufacturing strategy.

1 Introduction

In October of 2007, Susan faced a dilemma. Her three year-old son, Ryan, had recently developed a strong interest in the Little Einstein line of toys produced by Fisher-Price. Susan knew that one particular toy—the Pat-Pat Rocket—was rumored to be a “hot toy” for toddlers during the 2007 holiday shopping season. By chance, Susan one day came across a store with several Pat-Pat Rockets in stock. She knew that if she purchased the toy immediately, there was no guarantee that Ryan would still enjoy Little Einstein products in three months time; his taste in toys seemed to change weekly. If she did not buy the toy now, however, she believed her chances of finding it in
the future may be slim, particularly if the toy turned out to be a hot holiday gift. Susan ultimately chose to purchase the Pat-Pat Rocket; the risk of not obtaining the toy was too great to delay her decision until closer to the holidays.

Parents increasingly participate in this unfortunate holiday ritual (Slatalla 2002), trading off the risk of buying early and facing uncertain value for the product (i.e., possibly buying a toy that turns out to be a “dud” or that their child does not want) with the risk of buying late and facing uncertain availability for the product (i.e., experiencing a stock-out). Recent history provides numerous examples of hot holiday toys for which demand outstrips expectations, from Barbie to Elmo to the Nintendo Wii video game system. Stock-outs during the holidays are assumed to be particularly costly to firms, as consumers shopping for gifts are more likely to switch to a competitor’s product rather than wait for an inventory replenishment that occurs after the holidays have concluded. Indeed, the perception is that potential losses due to inventory shortages during the holidays can be enormous—Richtel (2007), for example, reports estimates that Nintendo experienced lost sales in excess of $1 billion due to unsatisfied demand on the Wii video game system during the 2007 holiday season.

Long production and shipping leadtimes are often cited as key causes for holiday gift shortages, particularly on those products manufactured in Asia and exported to the US or Europe. Due to these long leadtimes, demand forecasts must be made far in advance of the selling season, when uncertainty concerning final demand is high. Thus, if leadtimes could be reduced—via, for example, localized production, increased capacity, improved information systems, and expedited shipping methods—allowing for a rapid response to updated demand information closer to (or during) the selling season, supply and demand could be more closely matched, reducing or eliminating costly shortages. Such techniques (often known as reactive capacity or quick response systems) can be expensive due to expedited production or transportation costs, but are known to provide significant value to firms by better matching supply with uncertain demand (Fisher and Raman 1996). The consensus is that quick response systems are beneficial to a firm: indeed, in the absence of fixed costs related to the implementation of such systems, the opportunity to procure additional inventory after learning updated demand information is an option which always possesses positive value.

In this paper, we consider whether (and under what conditions) quick response inventory practices do in fact benefit a firm. Motivated by our example of holiday gifts, we consider a product
characterized by initially uncertain consumer value. This may be the case if, for instance, the product is a new or innovative item (e.g., a complex or innovative product such as a Nintendo Wii, an Apple iPhone, or an automobile), a media item (such as books, movies, music, or video games), or the consumer’s requirements for the item are uncertain (e.g., snow skis for a potential weekend trip in two months when weather is unknown, or a gift for a child whose preferences frequently change). Over time, consumers learn more information about the product and gain a better sense of its value; for example, via channels such as professional product reviews from web sites and magazines, the reviews of fellow consumers (e.g., from online retailers such as Amazon.com), the experiences of friends and family who may have purchased the same product, or via the resolution of intrinsic uncertainty in product value (e.g., the weather affecting the value of a pair of skis is known the day of the ski trip). Individual consumers thus makes a decision on when and whether to purchase the product: the later the customer waits to buy, the more information she will have about product value and the greater the risk of a stock-out.

When consumers experience this type of time dependent learning, greater availability resulting from an improved matching of supply and demand encourages consumers to delay purchasing the product: by reducing the likelihood of a stock-out, the firm decreases the riskiness of waiting to learn more information about product value. Thus, there is a clear interaction of consumer learning (of product value) and firm learning (of product demand). We explore this interaction by addressing how the responsiveness of the firm’s supply chain—its ability to respond to improved demand information—affects consumer purchasing behavior. We analyze models with a single firm selling to a rational, forward-looking consumer population. Consumers choose to either purchase early—prior to learning their value for a product—or purchase late, after learning their value. The firm chooses to either commit to a single production run in advance of learning product popularity or to adopt the ability to rapidly produce inventory after stochastic demand is revealed.

Using this stylized framework, we demonstrate that the basic intuition that quick response provides an option with purely positive value may be incorrect; that is, even without fixed costs it is possible for a firm to be worse off if it has an additional procurement opportunity after receiving updated demand information. This occurs when consumers—cognizant of the results of the firm’s operating policies, in particular the inventory availability—modify their own purchasing behavior to account for the implementation of a quick response system. In other words, while quick response
does better match supply and demand, demand itself can be negatively affected once consumers
become aware of the increased availability resulting from quick response and optimize their own
behavior accordingly. The net effect may decrease firm profits, though we demonstrate that
whether this occurs (and to what degree it occurs) depends heavily on several characteristics of the
selling environment; specifically, when prices increase over time or when dissatisfied consumers can
return the product for a full refund, quick response always increases firm profit, whereas if prices
are constant or decline over time or if consumers cannot return the product for a full refund, quick
response may decrease firm profit. Finally, we demonstrate that our model of a firm selling to
multiple consumers with uncertain value is analogous to a manufacturer selling to multiple retailers
facing uncertain demand, and we derive additional insights concerning this interpretation of the
model.

The remainder of the paper is organized as follows. §2 provides a review of the literature. §3
introduces the model, and §§4–5 analyze the consumer decision and the equilibrium to the game,
respectively. Three extensions are then analyzed: consumer returns in §6, pricing in §7, and a
model of a manufacturer selling to multiple retailers in §8. §9 concludes the paper with a discussion
of the results.

2 Related Literature

There are two areas of the literature that are of particular relation to this analysis: the first
concerns forward-looking or “strategic” consumer behavior. Explicitly modeling the intertemporal
purchasing decision of rational consumers has received increased attention in recent years; see, for
example, Aviv and Pazgal (2007), Liu and van Ryzin (2005), Su and Zhang (2005), and Jerath et al.
(2007). The most relevant papers to our own from this stream of research are those addressing
uncertain consumer valuations. DeGraba (1995) explains why a firm may intentionally understock
to induce consumers to purchase when valuations are uncertain and learned over time. Unlike
our model, there is no demand uncertainty to the firm. Xie and Shugan (2001) demonstrate
that selling to consumers prior to the determination of value and consumption (e.g., with advance
ticket sales) can substantially increase firm profits. Alexandrov and Lariviere (2006) consider the
problem of a restaurant choosing whether to offer reservations (guaranteed seats) to customers who
may or may not value dining on a given night, demonstrating when reservations increase the profit of the firm. Dana (1998) and Akan et al. (2007) discuss optimal pricing to screen heterogeneous consumers whose values are revealed over time. In these papers, in contrast to our model, inventory (or capacity) is either infinite, exogenously set, or fixed throughout the selling season, and hence issues of inventory replenishment after receiving updated demand information are not considered. An exception is Debo and van Ryzin (2007), who consider a periodic review inventory problem. However, in their model the base-stock level is exogenously given, as is the decision of how often to replenish, whereas in our model, inventory levels and the decision of whether to obtain a midseason replenishment are endogenously determined.

Uncertain consumer valuations are also a hallmark of the herding literature—see, e.g., Bikhchandani et al. (1992). Typically in this literature, the actions of the firm are fixed, while consumers observe the sequence of sales and use Bayesian updating to determine whether to purchase a product with uncertain value. An exception is Debo (2007), which incorporates both the consumer learning characteristics of the herding literature and the firm’s decision to price optimally with fixed inventories. In contrast to the herding literature, our model does not allow consumers to observe the sequence of sales of the product and infer value from this information; rather, valuations are exogenously revealed over time, e.g., via external channels such as expert product reviews. We abstract from the consumer learning dynamics of the herding literature to focus instead on the inventory game between the firm and consumers, as well as the decision to adopt quick response inventory practices.

In §6, we address consumer returns policies which allow customers who buy prior to learning their value to return the product should their realized valuation turn out to be low. Such policies have received attention in the literature: Davis et al. (1995) and Moorthy and Srinivasan (1995) analyze the value of money-back guarantees when selling to consumers with uncertain value; Gallego and Şahin (2006) discuss multiperiod pricing of a single-consumption good with fixed capacity and unknown value to consumers, and demonstrate that selling call options on capacity can increase firm revenue (call options being analogous to costly product returns); Su (2007) provides an analysis of how consumer product returns affect inventory decisions when valuations are learned after the purchase of an item (e.g., experience goods); and Coughlan et al. (2007) address the role of returns policies in competitive retail settings. These papers do not consider the impact of consumer returns
policies on a firm's incentives to adopt a rapid procurement strategy, however.

The second broad stream of research related to our own is the quick response literature—see, for example, Fisher and Raman (1996), Eppen and Iyer (1997), Iyer and Bergen (1997), and Fisher et al. (2001). In the absence of strategic consumer behavior, these papers demonstrate the value engendered by the ability of a firm to react quickly to updated demand information. In a related paper (Cachon and Swinney 2007), we address the value of quick response systems in a fashion retail setting with forward-looking consumers. The model in Cachon and Swinney (2007) is characterized by markdowns and known consumer valuations, and consumers may strategically delay purchasing in order to “get a good deal” when the item goes on sale. The present model, by contrast, is characterized by constant (or increasing, as discussed in §7) prices and unknown consumer valuations—as a result, consumers delay purchasing to obtain better information about product value. Consequently, while Cachon and Swinney (2007) is applicable to setting in which value is easily judged and markdowns are likely (e.g., fashion), the present analysis is applicable to more complex products in which value is difficult to judge or inherently stochastic.

3 Model

A single firm sells a single product at an exogenous price\(^1\) \(p\) to a consumer population of uncertain size, \(N\). Uncertainty in market size is the result of an exogenous stochastic process; that is, \(N\) is a random variable with positive support, distribution function \(F(\cdot)\) and density \(f(\cdot)\). The product is sold over two periods. At the start of the first period, neither the firm nor consumers know the value of \(N\). At some point in the first period (e.g., after observing early sales), the firm exogenously and perfectly learns \(N\).\(^2\)

In addition to market size uncertainty, consumers face uncertainty about their own private valuations for the product. Nature moves first (prior to the start of the game) and decides the “type” of each consumer: a fraction \(\theta\) of the population has positive value \(v > p\) for the item, while a fraction \(1 - \theta\) has zero value. If a consumer possesses value \(v\) for the product, we refer to her as

---

\(^1\)In §7, we relax the assumption of exogenous prices and consider endogenous pricing.

\(^2\)In reality, forecast updating and refinement may be the result of an endogenous process, e.g., monitoring early sales and imputing total demand, or performing market research. To avoid issues outside the scope of this analysis—e.g., demand estimation based on stochastic arrivals—we assume that the revelation of \(N\) is exogenous and perfect.
a “high type” consumer, whereas if she possesses zero value for the product, we refer to her as a “low type” consumer.

In the first period, consumers do not know their private valuation for the product (their type). In the second period, consumers exogenously learn their value for the product (e.g., via product reviews from professionals and other consumers, experiences with demonstration units in-store, etc.). While consumers do not know their individual valuations in period one, they do know the underlying probability structure that determines their type (i.e., they know \( \theta \)); given this structure, from the point of view of an individual consumer (absent any additional information), the consumer is high type with probability \( \theta \) and low type with probability \( 1 - \theta \).

In the first period, each consumer receives a noisy private signal that is an indication of her type. We define \( \alpha \) to be the quality of the signal, i.e., the probability that the signal is correct. For example, a high type consumer receives a signal of high product value with probability \( \alpha \), and a low type consumer receives a signal of low product value with the same probability. We allow consumers to be heterogeneous in the quality of their private signals by letting \( \alpha \) be distributed among the population (independently of consumer type) according to the continuous distribution \( G(\cdot) \) with support on the interval \( (1/2, 1) \). Such heterogeneity in the quality of the signal may represent, for example, domain expertise of the population in the product category (e.g., some consumers are highly technical and capable of accurately judging the quality of a new, high tech product, while some less sophisticated consumers receive more noisy signals that leave them less sure of product value). Consumers are aware of their individual values of \( \alpha \), and the distribution \( G(\cdot) \) and density \( g(\cdot) \) of consumer signal strengths is known to the firm.

After receiving their private signals, consumers arrive at the firm throughout the first period. Each consumer updates her beliefs of product value (via Bayes’ rule) and calculates the expected utility of purchasing early (before knowing product value) and the expected utility of purchasing late (after learning product value), based on her private signal and individual signal strength. In order to evaluate the expected surplus of delaying a purchase until the late period, consumers must also form expectations on the probability that a unit will be obtained in the late period (i.e., the second period availability), which we denote \( \hat{\phi} \).

Consumers are risk-neutral expected utility maximizers that discount future consumption at rate \( \delta \in [0, 1] \), and hence consumers choose to purchase in the period that maximizes their total
The firm produces initial inventory.

If firm has QR capabilities, market size \((N)\) is learned. A second production run occurs and arrives immediately.

Valuations revealed to consumers.

Consumers arrive, receive a private signal of product value, and choose to either purchase immediately or wait until period 2.

Consumers who voluntarily waited for period 2 arrive and purchase if they have positive value and the product is available.

---

Figure 1. Sequence of events.

---

expected discounted surplus (expected product value minus purchase price). All consumers who arrive at the store in the second period know their value and purchase if and only if they have positive surplus and the product is in-stock, and any consumer who does not obtain a unit receives zero surplus.

We consider two potential operating regimes for the firm: the single procurement regime (SP), and the quick response regime (QR). In the single procurement regime, all production occurs in advance of the selling season, and the firm chooses an inventory level \(q\) before learning market size \((N)\). There is a linear unit production or ordering cost \(c_1\). In the quick response regime, the firm is allowed to procure some inventory \(q\) prior to the realization of \(N\), but may also produce or procure additional inventory—at an increased cost—after monitoring initial sales and updating demand forecasts in the first period. Inventory procured using quick response is assumed to arrive immediately and prior to any stock-out in period one. In this regime, the unit procurement cost prior to the realization of \(N\) is \(c_1\), and the unit procurement cost after \(N\) has been realized is \(c_2\), where \(c_1 \leq c_2\). The firm is assumed to have infinite reactive capacity, though capacity constraints may easily be added without qualitatively affecting any results. The firm chooses an initial inventory level \(q\) at the start of the game that maximizes total expected profit over both periods, and chooses an inventory replenishment (in the quick response regime) which maximizes profit-to-go. Excess inventory is costlessly carried from period one to period two, and we assume that excess inventory remaining at the end of period two has zero value. The sequence of events is summarized in Figure 1.
4 The Consumer Decision: Wait or Buy

In this section we analyze the consumer decision: whether to wait or buy. We begin by discussing the nature of consumer expectations of second period availability ($\hat{\phi}$). We assume that consumers form rational expectations (see, e.g., Muth 1961, Su and Zhang 2005, and Cachon and Swinney 2007) concerning the availability of the product (i.e., consumers possess beliefs about the chance of obtaining the product period two that are consistent with the equilibrium availability in the second period). Rational expectations may be formed by repeated interaction with a firm over time; for instance, consumers have come to expect that video game manufacturer Nintendo is incapable of rapid inventory replenishment to meet demand (Richtel 2007) and hence future availability is low. On the other hand, consumers have come to expect that General Motors will satisfy demand on hit products and hence future availability is high, a belief that GM is now actively trying to change (Stoll 2007). We further assume that the allocation of inventory in the late period is random (all consumers have an equal chance of procuring a unit). Because expectations are rational and consumers have an equal chance of obtaining a unit in the late period, all consumers must have identical expectations of $\hat{\phi}$.

In analyzing the consumer decision, the relevant unit of analysis is a consumer who arrives in period one, finds a unit in-stock, and considers purchasing the product immediately (which ensures that a unit will be obtained, but not that value will be high) or delaying the purchase decision until period two (which ensures that the consumer will only purchase if she has high value for the product, but does not ensure that she will successfully obtain a unit). The expected surplus of a period one purchase is $s_v p$, where $s_v p$ is the posterior probability that the consumer has high value for the product, conditional on a signal $s \in \{l, h\}$ (i.e., low or high value) and signal strength $\alpha$. For a consumer receiving a high value signal, this posterior probability is

$$
\gamma_h(\alpha) = \frac{\Pr(\text{High Type and High Signal})}{\Pr(\text{High Signal})} = \frac{\alpha\theta}{\alpha\theta + (1-\alpha)(1-\theta)}.
$$

(1)

Note that $\gamma_h(\alpha)$ is increasing in $\alpha$. Similarly, if the consumer receives a signal indicating that the

---

3 If any consumer finds the firm out-of-stock, the game essentially over; due to our assumption that the firm’s QR order arrives prior to any potential stock-out, if a consumer finds the firm out-of-stock, all subsequent consumers will as well, regardless of the operating regime.
product is low value, the posterior probability is

\[ \gamma_t(\alpha) = \frac{(1-\alpha)\theta}{(1-\alpha)\theta + \alpha(1-\theta)}. \]  

(2)

Note that \( \gamma_t(\alpha) \) is decreasing in \( \alpha \). If \( \gamma_t(\alpha) v - p > 0 \) for some \( \alpha \), consumers receiving a low signal may receive positive surplus from an early period purchase, whereas if \( \gamma_t(\alpha) v - p < 0 \), they always receive negative surplus. In the following analysis, we assume that the latter case holds for all \( \alpha \), though this assumption may be relaxed without qualitatively changing the results.

Due to this assumption, all consumers receiving a low signal have negative expected surplus from purchasing in the early period. It follows that all low signal consumers will delay purchasing until the second period, and only those consumers who receive a high signal will consider a purchase in period one. For these consumers, the expected surplus from waiting until the late period is

\[ \delta (v - p) \frac{\alpha \hat{\phi}}{\alpha \theta + (1-\alpha)(1-\theta)}. \]  

(3)

Because \( \hat{\phi} \leq 1 \), it is true that (3) is increasing in \( \alpha \) at a slower rate than first period surplus. Furthermore, if \( \alpha = 1 \), then early period surplus is strictly greater than late period surplus, while if \( \alpha = 1/2 \), the opposite relationship holds (from our assumption that \( \gamma_l(1/2) v - p < 0 \)). Thus, given any expectation of \( \hat{\phi} \), there exists a unique \( \alpha \) such that early period surplus exactly equals late period surplus. As a result, consumers who have high signal quality (accurately judge product value) will purchase in the early period, while consumers who have low signal quality (poorly judge product value) will delay until the late period. In other words, in any equilibrium there exists some critical signal quality, below which consumers wait for the late period and above which consumers purchase in the early period.\(^4\) The firm, which forms beliefs on consumer behavior in order to estimate demand in each period, thus possesses a belief \( \hat{\alpha} \) on the critical signal strength. We assume that these expectations are consistent with the equilibrium consumer actions, i.e., the expectations are rational.

\(^4\) We assume that consumers who are indifferent purchase in the early period.
5 Equilibrium and the Value of Quick Response

From the analysis in the preceding section, we conclude that we seek an equilibrium to the game that consists of consumer purchasing behavior and an inventory decision on the part of the firm. Such an equilibrium will be characterized by values of $q$ (the firm’s inventory level) and $\alpha$ (the critical signal strength of the consumer population). Let the superscript * denote a generic equilibrium parameter (replacing * with $SP$ or $QR$ when referring specifically to the single procurement or quick response case). We then formally define the equilibrium as follows:

**Definition 1** A equilibrium $(q^*, \alpha^*)$ with rational expectations to the game between the firm and the consumer population satisfies:

1. The firm chooses an initial inventory level $q^*$ to maximize total expected profit, conditional on beliefs about consumer behavior, $\hat{\alpha}$;

2. The consumer population determines the critical signal strength $\alpha^*$, conditional on beliefs about second period availability $\hat{\phi}$;

3. Beliefs are rational, $\hat{\alpha} = \alpha^*$ and $\hat{\phi} = \phi(q^*, \alpha^*)$, where $\phi(q, \alpha)$ is the second period fill rate given initial inventory $q$ and critical signal strength $\alpha$.

The critical signal strength is determined by calculating the surplus from an immediate purchase by a consumer who arrives at the store and finds a unit in-stock, and equating that surplus with the expected surplus of delaying the purchase until the late period, yielding

$$\alpha^* = \frac{(1 - \theta)p}{(1 - \theta)p + \theta(v - p)(1 - \hat{\delta}\hat{\phi})}.$$ (4)

Because the actions of all consumers may be summarized by a single variable ($\alpha^*$, the critical signal strength), there are essentially two actions in the game: the firm chooses an inventory level (which depends upon how many consumers purchase early and how many purchase late), and consumers determine the critical signal quality, which depends upon the expected second period availability (and hence the inventory level of the firm). The rational expectations hypothesis thus implies the game is one of simultaneous moves with two players.
We must prove that the equilibrium to the game exists (and that such an equilibrium is unique) in order to discuss its properties; the following lemma accomplishes this for the SP regime.

**Lemma 1** When the firm operates in the single procurement regime, an equilibrium \((q^{sp}, \alpha^{sp})\) exists and is unique. The equilibrium total demand to the firm is

\[
D = N \left( \theta + (1 - \theta) \int_{\alpha^{sp}}^{1} (1 - x) g(x) \, dx \right) .
\]  

(5)

**Proof.** In order to determine the equilibrium to the game, we must first derive the firm’s best reply to a belief \(\hat{\alpha}\) concerning customer behavior. All consumers who receive a high value signal with a signal strength greater than \(\hat{\alpha}\) purchase in the first period, thus first period demand is composed of two types of consumers: those with high value (probability \(\theta\)) and correct signals (probability \(\alpha\)), and those with low value (probability \(1 - \theta\)) and incorrect signals (probability \(1 - \alpha\)). Let

\[
\xi_1 = \theta \int_{\hat{\alpha}}^{1} xg(x) \, dx + (1 - \theta) \int_{\hat{\alpha}}^{1} (1 - x) g(x) \, dx .
\]

The total demand in the early period is thus \(N\xi_1\). All consumers with signal strengths less than \(\hat{\alpha}\) delay purchasing until the late period, at which time only those consumers with high value will purchase the product. Demand in this period will thus consist of all consumers who have high value (probability \(\theta\)) and received a low value signal in period one (probability \(1 - \alpha\)), and consumers who have high value (probability \(\theta\)), received correct signals in period one (probability \(\alpha\)), and chose to delay their purchase (signal strength between 1/2 and \(\hat{\alpha}\)). Let

\[
\xi_2 = \theta \int_{1/2}^{1} (1 - x) g(x) \, dx + \theta \int_{1/2}^{\hat{\alpha}} xg(x) \, dx ,
\]

such that the total late period demand is \(N\xi_2\). The total demand is thus \(D = N(\xi_1 + \xi_2)\), where

\[
\xi_1 + \xi_2 = \theta + (1 - \theta) \int_{\hat{\alpha}}^{1} (1 - x) g(x) \, dx .
\]

The firm’s expected profit is \(\pi(q) = \mathbb{E}[p \min(q, D) - c_1q]\), which is a concave function of \(q\) yielding an optimal inventory level satisfying \(\Pr(D < q) = (p - c_1) / p\). Substituting for \(D\), we see that the
best reply function is

\[ q(\alpha) = \left( \theta + (1 - \theta) \int_{\alpha}^{1} (1 - x) g(x) \, dx \right) F^{-1} \left( \frac{p - c_1}{p} \right). \]

We may now derive the equilibrium to the game by imposing the rational expectations hypothesis, which implies \( \alpha = \alpha^{sp} \) and \( \phi = \phi(q^{sp}, \alpha^{sp}) \). The firm best reply is decreasing in \( \alpha^{sp} \) (the more consumers that purchase early, the higher the demand and thus the inventory level of the firm). The actual fill rate \( \phi(q, \alpha) \) is increasing in \( \alpha \) (intuitively, if more consumers wait until the late period, total demand decreases and hence the fill rate given a fixed inventory level increases–see the technical appendix for a detailed derivation). This implies that a unique consumer best reply exists, determined by the solution to (4) with \( \phi = \hat{\phi}(q^{sp}, \alpha^{sp}) \), and further that best reply is increasing in the inventory level of the firm (since higher inventory implies a higher fill rate). Because one best reply is increasing and one is decreasing, they must intersect at a unique point, hence an equilibrium to the game exists and is unique.

From (5), the equilibrium demand of the firm is decreasing in \( \alpha^{sp} \). It is apparent, then, that the firm prefers more consumers to purchase early as this increases total demand; the firm enjoys the benefits of valuation uncertainty due to the rationing risk created by limited inventories in the late period. This result is sometimes referred to as the advance selling phenomenon–see Xie and Shugan (2001)–in which a firm exploits consumer valuation uncertainty by inducing some consumers to purchase the product before learning their value that will ultimately be dissatisfied (have low valuation).

We next move to the game in which the firm operates in the QR regime. Recall that the firm behaves in a subgame perfect manner; that is, when determining the number of units to produce using quick response, the firm chooses an inventory level that maximizes total profit. As a result, if the firm has quick response capabilities, rational consumers must believe that the fill rate at that firm is equal to 1; after learning the true value of demand, the firm cannot credibly commit to satisfying anything less than the total demand it receives.\(^5\) Consequently, quick response increases

\(^5\)It is worthwhile to consider what would happen if the firm did not maximize revenue when placing the quick response order–for instance, the firm might plan \textit{ex ante} to fulfill some fraction of the total demand while leaving some residual rationing risk in order to “train” repeat customers to expect limited availability. If the firm is capable of such credible commitment, it is still true that quick response increases overall availability and increases the consumer incentive to wait until the late period, just not to the extent that it would if the firm behaves in a subgame perfect manner. Unless the firm commits to not use quick response at all (in which case, the firm essentially operates without...
the expected surplus of late period consumers and strengthens the incentive for consumers to wait. All else being equal, this will shift demand from the early period to the late period, which will in turn decrease the amount of advance selling that occurs.

The story does not end with the effect of quick response on consumer behavior, however; QR also offers value by better matching supply and demand under uncertainty. Thus, it remains to be seen how QR affects the profit of the firm in equilibrium. Before we answer this question, we must first demonstrate that an equilibrium exists and is unique when the firm operates in the QR regime. The following lemma does this, in addition to comparing the equilibrium outcomes (critical signal strength and inventory level) to the single procurement regime.

**Lemma 2** When the firm operates in the quick response regime, a subgame perfect equilibrium \((q^{qr}, \alpha^{qr})\) exists and is unique. In equilibrium, more consumers wait for the late period \((\alpha^{sp} \leq \alpha^{qr})\) and the firm sets a lower inventory level \((q^{qr} \leq q^{sp})\) than in the single procurement regime.

**Proof.** Because the firm operates in the QR regime, the only rational belief of the consumer population is that \(\hat{\phi} = 1\); because the quick response procurement is subgame perfect, the firm will satisfy all second period demand. Hence, the consumer best reply is independent of any firm actions, and is dictated by the solution to (4) with \(\hat{\phi} = 1\), which implies

\[
\alpha^{qr} = \frac{(1 - \theta) p}{(1 - \theta) p + \theta (v - p) (1 - \delta)}.
\]

This is clearly a unique consumer best reply, and it is immediately apparent that \(\alpha^{sp} \leq \alpha^{qr}\) for any equilibrium fill rate in the single procurement regime. The firm’s profit function is \(\pi (q) = \mathbb{E} [pD - c_1 q - c_2 (D - q)^+]\), where \(D = N (\xi_1 + \xi_2)\) and \(\xi_1\) and \(\xi_2\) as are in the proof of Lemma 1.

It follows that the firm best reply exists and is unique, given by

\[
q (\alpha^{qr}) = \left( \theta + (1 - \theta) \int_{\alpha^{qr}}^{1} (1 - x) g (x) dx \right) F^{-1} \left( \frac{c_2 - c_1}{c_2} \right),
\]

hence the equilibrium existence and uniqueness results follow. This furthermore implies

\[
(1 - \theta) \int_{\alpha^{sp}}^{1} (1 - x) g (x) dx \geq (1 - \theta) \int_{\alpha^{qr}}^{1} (1 - x) g (x) dx,
\]

quick response), its adoption will generally increase availability and the incentive to wait until the late period.
and hence it follows that total equilibrium demand to the firm is greater in the SP regime than in the QR regime, yielding \( q^{qr} \leq q^{sp} \). ■

Note that the equilibrium with quick response is subgame perfect due to our assumptions on the behavior of the firm when making the second inventory procurement. Having demonstrated that equilibria exist and are unique in both regimes, we may now address the value of quick response: the incremental increase in profit due to the adoption of a quick response system.

**Theorem 1** Let \( \Delta = \pi^{qr} - \pi^{sp} \) be the incremental equilibrium value of quick response. \( \Delta \) is strictly decreasing in the cost of quick response \( (c_2) \), and if \( c_2 = p \), \( \Delta \leq 0 \).

**Proof.** Define \( \pi^{qr} (q) = \mathbb{E} [pD - c_1q - c_2(D - q)] \), where \( D \) is the total demand at the firm (a function of \( \alpha^{qr} \)). Let \( \pi^{qr} \) be the equilibrium profit of the firm with quick response, and let \( \pi^{sp} \) be the equilibrium profit without QR. Differentiating \( \pi^{qr} \) with respect to \( c_2 \), we have, from the Envelope Theorem,

\[
\frac{d\pi^{qr}}{dc_2} = \frac{\partial \pi^{qr} (q)}{\partial c_2} \bigg|_{q=q^{qr}} + \frac{\partial \pi^{qr} (q)}{\partial q} \frac{dq^{qr}}{dc_2} = \frac{\partial \pi^{qr} (q)}{\partial c_2} \bigg|_{q=q^{qr}}.
\]

Note that since \( \alpha^{qr} \) contains no dependence on \( c_2 \), there is no derivative term with respect to \( \alpha^{qr} \). This implies

\[
\frac{d\pi^{qr}}{dc_2} = - \Pr (D > q^{qr}) = -1 + \frac{c_2 - c_1}{c_2} < 0.
\]

Thus, the equilibrium profit of the firm is decreasing in \( c_2 \). Note that, in the limit as \( c_2 \to p \), the margin on each unit sold that is procured via QR goes to zero. Hence, the firm’s profit effectively becomes the same as if it did not have QR capabilities, with one caveat: in equilibrium, more consumers will wait for the late period than if the firm did not have QR. Thus, \( \lim_{c_2\to p} \pi^{qr} = \pi^{sp}|_{\alpha=\alpha^{qr}} \leq \pi^{sp}|_{\alpha=\alpha^{sp}} \), i.e., for large \( c_2 \) QR yields lower expected profits than the SP regime. ■

The first part of Theorem 1 is natural—the value of a quick response system is decreasing in the marginal cost of a midseason replenishment. A more surprising result is provided by the second part of the theorem, which demonstrates that the value of a quick response option \( (\Delta) \) can be negative if \( c_2 = p \). Combining both parts of the theorem implies that quick response may reduce the profit of the firm even if the marginal procurement cost is strictly less than the selling price. This result stands in contrast to the existing literature on quick response: with non-strategic
Figure 2. The incremental value of quick response ($\Delta$) as a function of the cost of an expedited procurement ($c_2$) when $c_1 = 2$ and $p = 10$, separated into component factors. Matching supply and demand provides positive value while shifting demand provides negative value.

consumers (e.g., Fisher and Raman 1996) or with strategic consumers in the absence of learning (Cachon and Swinney 2007), quick response always provides non-negative value if the margin on a unit procured using quick response is weakly positive (i.e., if $c_2 \leq p$). Theorem 1 shows that this need not be the case when consumers learn about their valuations over time: it is possible for quick response to yield a positive margin on each unit sold while simultaneously yielding lower expected profit to the firm than the single procurement regime.

The key to this result is that quick response has two effects: shifting demand and matching supply with demand. These two effects pull the equilibrium profit of the firm in opposite directions. Shifting demand to the late period reduces profits by decreasing the amount of advance selling. Matching supply with demand increases profits by eliminating lost sales—all demand is captured, albeit at a higher unit procurement cost—and reducing the chance of overstock. Hence, the firm only values quick response so long as the cost of shifting demand is exceeded by the gain from better matching supply with demand; see Figure 2.

This is not to say that quick response is always harmful to the firm when consumers learn about value over time. As Theorem 1 and Figure 2 demonstrate, quick response can increase the profitability of the firm if $c_2$ is small enough. Nevertheless, a result of Theorem 1 is that it may be in the best interests of the firm to forgo quick response tactics and the option to procure additional inventory, and further to ensure that consumers are aware of this operating regime. This relates to the rationing risk results in the literature on strategic consumer purchasing—see, e.g., Su and Zhang.
(2005) and Debo and van Ryzin (2007). In contrast to the mere reduction of inventory described in this literature, Theorem 1 implies that the firm may be better off with an entirely different operating policy (Single Procurement vs. Quick Response) when consumers behave strategically—the inability to react to updated demand information in a timely and responsive way can benefit the firm by generating a credible mismatch between supply and demand and inducing more consumers to purchase prior to learning their value.

The fact that it may be optimal for the firm to operate without quick response lends justification to publicized “limited edition” runs of certain products: such tactics induce consumers who are otherwise “on the fence” to purchase the product prior to learning if they truly value it, lest no inventory remain once valuations are revealed. For example, Disney is famous for releasing its classic films on video for very limited periods of time, after which the films are “placed in the Disney vault,” not to be released again for a period of several years. However, as we shall see in the following sections, this result is sensitive to at least two key assumptions regarding the nature of the product.

6 Consumer Returns

The preceding analysis assumed that any consumer who purchased an item early had no recourse if their value for that item turned out to be low—that is, the possibility that a consumer could return a product if she is dissatisfied was excluded. In some industries, this assumption is appropriate. For example, with most types of media (e.g., movies, music, video games, or computer software) returns are forbidden once an item has been opened (often due to fears of piracy), and Amazon.com does not allow returns on large televisions due to the logistical challenges of return shipping.

In some cases, however, product returns are a common and important component of firm strategy. Satisfaction guarantees abound in many settings (clothing, electronics, etc.), with firms encouraging customers to try new products “risk free” while promoting generous return policies. At both Amazon.com and the electronics retailer Best Buy, for example, returns are allowed for full refunds on most items within a 30 day period; during the holidays this return window is extended up to a maximum of 90 days. Such policies increase the consumer incentive to purchase early by reducing the consequences of buying a product which is not valued.
In this section, we consider the effect of returns policies on our model of consumer and firm learning. Such policies have been addressed—see the discussion in §2—though unlike some previous papers, we do not address the issue of designing the optimal return policy, but rather we assume that the firm offers full refunds to any dissatisfied customer (possibly for marketing or competitive reasons) due to the ubiquity of this type of return policy in retailing. (Implications of partial refunds are discussed at the end of this section.)

Returns occur immediately following the early period, prior to the arrival of late period demand. In addition to assuming that returns are for full refunds, we assume that returned products are resalable—that is, the firm may repackaged and resell in the late period any returns that occur from early period demand. For generality, we assume that consumers who make a return incur a hassle cost $h \geq 0$, and that returns are costly to the firm, incurring a restocking fee of $r \geq 0$ on each returned item. We assume that $p - h \geq 0$, i.e., a dissatisfied consumer benefits from a return. This implies that if $\gamma_h (\alpha) v - p \geq 0$, then

$$
\gamma_h (\alpha) v - p + (1 - \gamma_h (\alpha)) (p - h) \geq \gamma_h (\alpha) v - p \geq 0,
$$

i.e., with returns, high signal consumers have greater incentive to purchase in the early period. We assume also that returns are enough of a hassle ($h$ is large enough) that low signal consumers still do not purchase in the early period, i.e., that $\gamma_l (1/2) v - p + (1 - \gamma_l (1/2)) (p - h) < 0$.

We are interested in how the addition of the described return policy changes the results of §5, specifically the results provided in Theorem 1. As we might expect from (6), by increasing expected surplus in the early period, returns encourage more consumers to purchase early. While this would seem to benefit the firm, the increase in advance purchasing comes at a price: consumers who purchase in the early period and are dissatisfied are costly to the firm, due to the fact that each returned unit costs the firm the price of the refund, $p$, and the restocking fee, $r$. Thus, the value of quick response practices—which as we have already mentioned shift demand from the early period to the late period by lessening the availability risk associated with delaying a purchase—will differ from that derived in the model without returns. The following theorem formalizes this argument.

**Theorem 2** Let $\Delta_r = \pi^{0q}_r - \pi^{sp}_r$ be the incremental equilibrium value of quick response with consumer returns. $\Delta_r$ is strictly decreasing in the cost of quick response ($c_2$), and if $c_2 = p$, $\Delta_r \geq 0$. 
Proof. The proofs of equilibrium existence and uniqueness are similar to Lemmas 1 and 2, and are hence omitted. First, we note that with consumer returns, any consumers who purchase in the early period and are dissatisfied with the product will return the item. Because we assume that these products are returned at the start of the late period and are resalable, the total demand to the firm is simply $\theta N$. Thus, the expected profit (without quick response) is

$$\pi^p_r (q) = \mathbb{E} \left[ p \theta N - \frac{\theta N}{p} N \int_0^1 (1 - x) g (x) \, dx \right],$$

where $\alpha_{sp}$ refers to the equilibrium critical consumer signal strength with returns, determined by equating expected first and second period surplus, yielding

$$\alpha_{sp}^{qr} = \frac{h (1 - \theta)}{h (1 - \theta) + \theta (v - p) (1 - \delta \bar{\phi})}.$$

Immediately we see that (conditional on identical second period fill rates), $\alpha_{sp}^{qr} \leq \alpha_{sp}$ if $h \leq p$, i.e., holding inventory availability constant, returns encourage more consumers to buy early. As in the case without returns, quick response induces $\bar{\phi} = 1$, hence

$$\alpha_{qr}^{qr} = \frac{h (1 - \theta)}{h (1 - \theta) + \theta (v - p) (1 - \delta)}$$

and $\alpha_{sp}^{qr} \leq \alpha_{qr}^{qr}$ for any equilibrium belief concerning the fill rate in the SP regime. Thus, the expected profit with quick response is

$$\pi^{qr}_r (q) = \mathbb{E} \left[ p \theta N - c_2 (\theta N - q)^+ - c_1 q - r (1 - \theta) N \int_{\alpha_{sp}^{qr}}^1 (1 - x) g (x) \, dx \right].$$

Clearly $\pi^{qr}_r$ is strictly decreasing in $c_2$ (and hence $\Delta_r$ is as well). In addition, for any $q$,

$$\pi^{qr}_r (q) - \pi^{sp}_r (q) = \mathbb{E} \left[ (p - c_2) (\theta N - q)^+ + r (1 - \theta) N \left( \int_{\alpha_{sp}^{qr}}^1 (1 - x) g (x) \, dx - \int_{\alpha_{qr}^{qr}}^1 (1 - x) g (x) \, dx \right) \right].$$

Because $\alpha_{sp}^{qr} \leq \alpha_{qr}^{qr}$, this number is weakly positive if $p - c_2 \geq 0$, hence for any fixed quantity, quick response provides non-negative value (and thus the same holds at the optimal quantities), proving
the result. ■

When the firm allows consumer returns, quick response always increases profits ($\Delta_r \geq 0$) if the margin on a unit procured using QR is positive ($c_2 \leq p$). Recall that in the absence of consumer returns, quick response may decrease profits ($\Delta \leq 0$) even if $c_2 \leq p$; see Theorem 1. The two consequences of quick response—shifting demand from the early period to the late period, and matching supply and demand—move the firm’s profit in opposite directions when returns are not allowed. Shifting demand in particular hurts the firm because it means that consumers who would have purchased with an expectation of positive surplus in the early period may instead not purchase in the late period after learning their value.

With consumer returns, however, shifting demand increases firm profit, because consumers who purchase early and are not satisfied with the product are costly to the firm (due to the restocking fee), and the firm would rather these consumers delay purchasing until they learn their valuations. Indeed, the above result extends to the case when some or all of the returned goods are not resalable—in that case, returns are even more costly to the firm due to the lost opportunity of reselling a returned product. This result is due to the tendency of consumers to hoard inventory: given that returns are possible, a consumer would rather purchase an item early and run the risk of having to return the product, as opposed to delaying the purchase and risking a stock-out. Quick response reduces the amount of consumer hoarding by increasing overall availability, which in turn decreases the number of costly returns. See Figure 3 for a graphical depiction of this effect on firm profit.

This result differs from the existing literature. DeGraba (1995) shows, for instance, that limiting availability increases the profit of the firm when consumers learn their value over time. Theorem 2 implies that if consumer returns are allowed, exactly the opposite is true: the firm prefers the highest level of availability possible (that created by quick response, which provides one hundred percent availability) in order to minimize the consumer tendency to hoard inventory. The reason for this difference is that we have assumed that returns are costly to the firm. Advance selling by limiting availability (DeGraba 1995) or reducing initial prices (Xie and Shugan 2001) provides value precisely because dissatisfied consumers cannot return the product for a full refund. Indeed, Xie and Shugan (2001) discuss how returns can benefit the firm with advance selling—provided that refunds are not for the full selling price, and indeed that refunds are small enough that the firm
profits from every returned unit. In general, if returns are not full refunds, whenever returns are costly (i.e., whenever the firm restocking costs plus the difference between the return amount and the purchase price is positive), Theorem 2 holds, whereas whenever returns are profitable (the restocking cost plus the difference between the return amount and purchase price is negative), the result mirrors that of Theorem 1.

While partial refunds may be reasonable in a booking context (such as airline tickets) in which service fees for changing reservations are customary, in retailing the vast majority of returns are for full (or nearly full) refunds due to competitive pressure, and are subsequently costly to firms—see Stock et al. (2006) for a discussion of how firms actively attempt to minimize returns, and Moorthy and Srinivasan (1995) for a discussion of costly returns. In our model, the interaction of two effects—consumer learning and costly product returns—implies that the firm would rather sell to high type consumers alone in the late period than sell to both types of consumers in the early period and suffer a large number of returns. Quick response provides a tool to induce precisely such behavior, while simultaneously providing value by better matching supply and demand; as a result, when consumer returns are an issue, quick response yields significant value to the firm.

Figure 3. The incremental value of quick response as a function of the cost of an expedited procurement ($c_2$) when $c_1 = 2$ and $p = 10$, separated into component factors. With costly returns, both matching supply and demand and shifting demand provide positive value.
7 Pricing

In this section, we endogenize pricing in our original model and address how the value of quick response is affected. We consider two types of pricing: fixed pricing (in which the retailer sets a single price for both periods) and dynamic pricing (in which the retailer may set different prices for each period).

7.1 Fixed Pricing

Unlike the inventory level, price is directly observed by consumers, and hence the firm acts as a Stackelberg leader in the price game. Thus, the model with fixed pricing entails a first stage in which the firm sets the (constant) selling price, and a second stage which behaves identically to the games analyzed in §§3–5. As a result, given a particular price, the previous results continue to hold (notably the equilibrium existence results) in the second stage of the game, and we need only analyze the firm’s choice of the selling price by comparing expected profits in the inventory/purchasing subgames using various price levels. The following theorem confirms that the result of Theorem 1–quick response may decrease firm profit–continues to hold even when the firm may set a (constant) price level.

**Theorem 3** Let \( \Delta_{fp} = \pi_{fp}^{qr} - \pi_{fp}^{sp} \) be the incremental equilibrium value of quick response with fixed pricing. \( \Delta_{fp} \) is strictly decreasing in the cost of quick response \( (c_2) \), and if \( c_2 = v \), \( \Delta_{fp} \leq 0 \).

**Proof.** Note that the existence of an equilibrium is immediate, due to the fact that we have already shown an equilibrium exists to the inventory/purchasing subgames and the firm’s expected payoffs are bounded (by 0 and \( \mathbb{E}N(v - c_1) \)) and its strategy space is a compact interval \([c_1, v]\) in the pricing supergame \(([c_2, v] \text{ when using quick response–if price is less than } c_2 \text{ but greater than } c_1, \text{ the firm will never use QR and reverts to the SP regime})\). Let \( \pi_{fp}^{qr}, p_{fp}^{qr}, \) and \( q_{fp}^{qr} \) be the equilibrium profit, price, and inventory of the firm with quick response and fixed pricing, and let \( \pi_{fp}^{sp} \) be the equilibrium profit without QR. Differentiating \( \pi_{fp}^{qr} \) with respect to \( c_2 \), we have, from the Envelope Theorem,

\[
\frac{d\pi_{fp}^{qr}}{dc_2} = \frac{\partial \pi_{fp}^{qr}}{\partial c_2} + \frac{\partial \pi_{fp}^{qr}}{\partial p} \frac{dp_{fp}^{qr}}{dc_2} + \frac{\partial \pi_{fp}^{qr}}{\partial \alpha} \frac{d\alpha_{fp}^{qr}}{dc_2} \leq \frac{\partial \pi_{fp}^{qr}}{\partial c_2}.
\]
Observe that either \( \frac{\partial \pi_{qr}^{sp}}{\partial p} = 0 \) (the firm prices at an interior optimum) or \( \frac{d\pi_{qr}^{sp}}{dc_2} = 0 \) (the firm prices on the boundary, i.e., \( c_2 \) or \( v \)). Unlike the case without pricing, \( \frac{d\alpha_{qr}^{sp}}{dc_2} \) in general does not equal zero. This is due to the fact that \( \frac{d\pi_{qr}^{sp}}{dc_2} \geq 0 \) and \( \frac{d\alpha_{qr}^{sp}}{dc_2} \geq 0 \)– in other words, higher costs of quick response lead to higher prices (a natural result) and higher prices lead to more consumers waiting until the late period, see equation (4). Because \( \frac{\partial \pi_{qr}^{sp}}{\partial c_2} \leq 0 \) (the more consumers that wait, the lower the firm’s profits), it follows that the \( \frac{\partial \pi_{qr}^{sp}}{\partial c_2} \frac{d\pi_{qr}^{sp}}{dc_2} \leq 0 \). Finally, since

\[
\frac{d\pi_{qr}^{sp}}{dc_2} \leq \frac{\partial \pi_{qr}^{sp}}{\partial c_2} = -Pr \left(D > q_{qr}^{sp}\right) = -1 + \frac{c_2 - c_1}{c_2} < 0,
\]

we find that profit is decreasing in \( c_2 \), precisely as in the case without pricing, and \( \Delta_{fp} \) is similarly decreasing in \( c_2 \). In the limit as \( c_2 \to v \), the firm’s optimal price with QR goes to \( v \), and margin on each unit sold that is procured via QR goes to zero. Hence, the firm’s profit effectively becomes the same as if it did not have QR capabilities, with two caveats: it is constrained to price at \( v \) (in the SP regime, the firm can price anywhere in the interval \([c_1, v]\)), and in equilibrium, more consumers will wait for the late period than if the firm did not have QR due to the fact that QR naturally shifts demand. In other words, if \( c_2 = v \),

\[
\Delta_{fp} = \pi_{fp}^{qr} - \pi_{fp}^{sp} = \pi_{fp}^{qr}\big|_{p=v} - \max_{p \in [c_1, v]} \pi_{fp}^{sp} \leq \pi_{fp}^{qr}\big|_{p=v} - \pi_{fp}^{sp}\big|_{p=v} \leq 0
\]

where the last inequality follows from Theorem 1.

The key to this result is the following: when prices are fixed across time, regardless of the optimal price level, adopting quick response increases the consumer incentive to wait and hence decreases advance selling and firm profit. The freedom to set the price is of little value in the quick response regime when \( c_2 \) is large, as the firm’s optimal price lies in the interval \([c_2, v]\)– if the the price is lower than \( c_2 \), then quick response is never used, hence the firm essentially moves to the single procurement regime. In the single procurement regime, the firm remains free to price anywhere in the interval \([c_1, v]\). When the cost of quick response is large, the quick response regime has two detrimental effects to the firm: pricing is constrained and more consumers delay purchasing due to higher availability. As a result, the single procurement regime becomes even more attractive than in the exogenous price case. Thus, Theorem 3 mirrors the result of Theorem 1: it is possible for
quick response to decrease profit ($\Delta f_p \leq 0$) even when the margin is positive ($c_2 \leq p \leq v$).

7.2 Dynamic Pricing

In the dynamic pricing case, we assume that the firm announces the first period price at the start of the first period and announces the second period price at the start of period two. Consumers develop rational expectations of future prices—that is, they correctly anticipate the price in period two. We first note that if the firm is free to set different prices in each period but is constrained only to mark prices down over time, Theorem 3 continues to hold. The reason is that it is never optimal in the current model to set a lower price in period two than in period one—lower late period prices would only encourage more consumers to delay purchasing and hence decrease the amount of advance selling. Thus, a firm constrained to mark down over time chooses to set a constant price, and the model reduces to the fixed pricing case analyzed above.

If the firm can raise prices over time, however, a different picture emerges. Let $p_1$ and $p_2$ be the selling price of the product in periods one and two, respectively. Note that the optimal selling price in the second period is $p_2 = v$; all consumers know their values, and possess values equal to $v$ or 0 for the product. Hence, the firm extracts all surplus from consumers purchasing in period two by charging the valuation of the high type consumers. Consequently, all consumers have zero surplus in period two (both high and low types, regardless of whether they successfully procure a unit), and all consumers with positive first period surplus purchase in that period. In general, the optimal first period price satisfies $p_1 \leq v$, i.e., the firm charges a lower first period price to induce some advance selling among consumers.

Because all consumers have identically zero surplus in period two, if the firm adopts quick response and raises the consumer expectation of availability in the late period ($\hat{\phi}$), the firm does not raise the expected surplus to any consumers from a second period purchase. Thus, quick response no longer shifts demand—the only effect remaining is matching supply and demand, hence quick response always has positive value. The following theorem summarizes this result.

**Theorem 4** Let $\Delta_{dp} = \pi_{dp}^{qr} - \pi_{dp}^{sp}$ be the incremental equilibrium value of quick response with dynamic pricing. $\Delta_{dp}$ is strictly decreasing in the cost of quick response ($c_2$), and if $c_2 = v$, $\Delta_{dp} \geq 0$. 

24
Proof. From the preceding discussion, \( p_2 = v \), and the existence of an equilibrium to the first period pricing supergame follows from a similar argument to Theorem 3. Combined with the rational expectations assumption of consumer beliefs concerning future pricing, this implies the critical \( \alpha \) is determined by the solution to \( \frac{\alpha \theta}{\alpha \theta + (1-\alpha)(1-\theta)} v - p_1 = 0 \), yielding

\[
\alpha^* = \frac{p_1 (1 - \theta)}{\theta (v - p_1) + p_1 (1 - \theta)}
\]

regardless of the firm’s operating regime. Note that because the second period price is equal to \( v \) the firm always makes a profit on a unit procured using quick response, and furthermore first period price may lie anywhere in the interval \([c_1, v]\). It is straightforward to see that \( \Delta_{dp} \) is decreasing in \( c_2 \). Next observe that, as a function of \( q \) and \( p_1 \), profit in the SP regime is \( \pi_{dp}^{sp}(q, p_1) = E[p_1 \min (\xi_1 N, q) - c_1 q] \), where \( \xi_1 \) is a function of \( p_1 \) (implicitly via \( \alpha^* \)). Similarly, profit in the QR regime is

\[
\pi_{dp}^{qr}(q, p_1) = E[p_1 \min (\xi_1 N, q) - c_1 q + v (\xi_2 N + (\xi_1 N - q)^+) - c_2 ((\xi_1 + \xi_2) N - q)^+] \cdot
\]

Because \( \xi_2 N + (\xi_1 N + -q)^+ \geq (\xi_1 N + \xi_2 N - q)^+ = ((\xi_1 + \xi_2) N - q)^+ \), it follows that \( \pi_{dp}^{qr}(q, p_1) \geq \pi_{dp}^{sp}(q, p_1) \) for any \( q \) and \( p_1 \) (and hence for optimal inventories and prices in the respective regimes). Thus, \( \Delta_{dp} \geq 0 \) for all \( c_2 \leq v \).

The key to Theorem 4 is that increasing prices over time provides consumers with greater incentive to purchase early, shifting demand from the later period to the early period. This effect counteracts the tendency of quick response to shift demand from the early period to the later period. Thus, dynamic pricing and quick response are complimentary in the sense that they enhance one another’s value: increasing prices reduces costly demand shifting due to quick response, and quick response eliminates costly supply/demand mismatches in the second period, mismatches which are particularly costly under dynamic pricing due to the higher price in period two.

We note that due to the assumption that consumer values follow a two point distribution, dynamic pricing in the present model completely eliminates strategic waiting in the sense that all consumers receive zero surplus in period two and hence consumers purchase in period one if and only if they have positive expected surplus. Should consumers have more than one positive
valuation in period two, in general dynamic pricing will not eliminate all strategic waiting, i.e., it will provide positive surplus to some consumers in period two. In that case, the adoption of quick response once again shifts demand from the early period to the late period and decreases advance selling; nevertheless, increasing prices over time continues to reduce the amount of strategic waiting that occurs and hence minimizes the negative aspects of demand shifting due to quick response.

8 A Manufacturer Selling to Many Retailers

The entirety of the discussion has focused on consumers as end users of the product; alternatively, we might think of the consumer population as a continuum of retailers, each with (potentially) unit demand for a product. In this case, the firm from our model is a manufacturer or supplier, and the retailers choose whether to stock the product based on their own private signals of whether the product will be in-demand at their location. This setting is described in the Sport Obermeyer case study, see Hammond and Raman (1994). In this section, we explore this alternative interpretation of the model, and consider how a different perception of a “customer” can produce new insights from the results.

Consider a single manufacturer introducing a new product to a continuous population of retailers. The manufacturer is uncertain as to how many retailers will consider purchasing the product, thus $N$ represents the (stochastic) number of retailers in the population and $F$ represents the manufacturer’s beliefs concerning $N$. Demand at the retailers follows independent draws from a two point distribution: with probability $\theta$ a retailer will face unit demand, and with probability $1 - \theta$ the retailer will face zero demand. All consumer demand occurs in period two; we may interpret period one purchases by the retailers as binding pre-orders or advance purchases from the manufacturer (as in the case of Sport Obermeyer). A consequence of this assumption is that $\delta = 1$ for the retailers (because all “consumption,” i.e. sales, occur at a single future date, there should be no discounting of period two consumption relative to period one consumption).

Each retailer sells the product to consumers for an identical price $v$ (e.g., a manufacturer suggested retail price). For simplicity, we assume that retailers cannot sell (or purchase) fractions...
of a unit, i.e., retailers purchase either one unit or zero units. The remainder of the model remains the same as that described in §3, with appropriate modifications made where necessary, e.g., each retailer receives a private signal $s \in \{l, h\}$ that is an indication of the eventual demand at her location—low (zero) or high (one).

Previous authors—for example, Cachon (2004), Taylor (2006), Dong and Zhu (2007), and Granot and Yin (2008)—have addressed the optimal timing of sales within the channel (before or after forecast updates) and the optimal allocation of inventory risk between a manufacturer and a retailer. Our model takes a different tack: given that retailers may purchase at any point in time (i.e., prior to or after learning demand), what are the manufacturer’s incentives to reduce leadtimes and create a flexible (upstream) supply chain that allows for additional production and replenishment in mid-season? In particular, the manufacturer’s decision to adopt quick response will alter retailer purchasing behavior, just as in the consumer oriented model discussed previously. The question is then: when does this change in retailer purchasing behavior help—or hurt—the manufacturer?

To answer this question, the following corollary summarizes and restates the results of Theorems 1–4 in a manner consistent with the manufacturer/retailer interpretation of the model. In that context, consumer returns correspond to manufacturer return or buy-back policies, and dynamic pricing (particularly increasing prices over time) corresponds to advance purchase contracts, while constant pricing corresponds to traditional static wholesale price contracts.

Corollary 1 When a manufacturer sells to a continuum of strategic retailers that are free to purchase before or after learning demand:

1. Under a wholesale price contract with exogenous, constant, or non-increasing prices, the manufacturer may not choose to adopt quick response even if $c_2$ is less than the wholesale price.

2. Under a wholesale price contract with unconstrained pricing in both periods (e.g., advance purchase discounts), the manufacturer always chooses to adopt quick response.

3. Under a buy-back contract, the manufacturer may not adopt quick response if the manufacturer’s margin on each unsold (returned) unit is positive.

4. Under a buy-back contract, the manufacturer will adopt quick response if the manufacturer’s margin on each unsold (returned) unit is negative.
As the corollary shows, whether the manufacturer wishes to adopt quick response depends on the type of supply contract it utilizes with the retailers: if, for instance, the contract allows for buy-backs (i.e., unsold products can be returned to the supplier, see Cachon 2003), then rapid production enables the manufacturer to mitigate strategic hoarding of inventory by the retailers (and subsequent costly returns). On the other hand, if the contract is of the wholesale price type (i.e., returns are not allowed), the manufacturer has less incentive to adopt quick response—it can benefit by advance selling to retailers due to the scarcity engendered by a slower supply chain. Furthermore, if the manufacturer can offer advance-purchase discounts to retailers, the incentive to adopt quick response increases; by raising prices over time (via the guise of “advance purchase discounts”) the manufacturer mitigates the degree of strategic waiting by retailers and hence induces advance purchasing.

9 Discussion

Quick response systems—or, more generally, leadtime reduction and rapid inventory replenishment—are often suggested as potential panaceas to the ill effects of supply and demand mismatches. Provided the fixed costs of implementing such systems are low enough, it is argued, the option to receive additional inventory after a forecast update can only increase a firm’s profit. We have demonstrated that this basic intuition may be incorrect once the consumer response to increased availability is taken into account. Though it is a commonly held belief that a faster, more responsive supply chain is a more profitable supply chain, we show that such responsiveness is not necessarily beneficial to a firm: when returns are forbidden or when prices are constant, the firm can exploit valuation uncertainty by advance selling, and quick response decreases the extent to which the firm can advance sell. By operating with rapid fulfillment capabilities, the firm loses its ability to credibly restrict inventory to create a stock-out risk, and thus may reduce its overall profitability.

The key to this result is that, when consumers learn about product value over time, there are two components of the value of quick response: matching supply and demand and shifting demand. The first component is well known to increase the profit of the firm by eliminating lost sales and reducing excess inventory. Shifting demand, on the other hand, can decrease firm profit by reducing the amount of advance selling and hence the overall demand. In some cases
(summarized in Table 1) demand shifting can be reduced (if prices increase over time) or beneficial to the firm (if consumer returns are allowed and are costly). In the former case, increasing prices mitigate the effects of strategic consumer waiting by reducing second period surplus and hence inducing more consumers to buy early. In the latter case, consumers are likely to hoard inventory—they would rather purchase a unit prior to learning their value and risk a return, than delay purchasing and risk a stock-out. Hoarding is costly for firms because of explicit costs of restocking returned inventory and implicit opportunity costs associated with reselling the unit. By adopting quick response, firms can mitigate consumer hoarding behavior by signaling high availability: if stock-outs are unlikely (or impossible) then consumers have little reason to hoard, which in turn reduces the number of costly returns for the firm. A consequence of quick response that was detrimental to the firm without consumer returns—shifting demand—becomes beneficial with returns.

We have also shown that the model of a firm selling to consumers is analogous to a manufacturer selling to strategic retailers. In that context, the manufacturer’s incentives to adopt a quick response production and fulfillment system depends upon the type of contract it employs with the retailers: quick response is beneficial when the contract consists of advance purchase discounts or costly buy-backs (unprofitable on a per-unit basis), while it may be detrimental to the manufacturer under a constant wholesale price or profitable buy-backs (profitable on a per-unit basis).

It is worth noting that quick response is an operational proxy for (more generically) information. Taken in that context, our results on the value of quick response are essentially results on the value of accurate information concerning product demand. The fact that quick response sometimes yields negative value supports the maxim that ignorance can be bliss; the lack of accurate information on demand can serve as a commitment mechanism to keep inventory scarce and increase advance selling. Taken together, our results provide insight into when a firm should adopt a fast supply
chain that allows action on improved demand information. Whether selling to consumers or retailers, the value of matching supply and demand depends not only on the reduction of lost sales and excess inventory, but also on the strategic response of the firm’s customers to increased product availability. This response can be harmful (if advance selling decreases as a result), beneficial (if costly returns are allowed and hoarding is an issue), and even diminished or eliminated by the appropriate pricing strategy (increasing prices over time in the optimal manner). Care must thus be taken when assessing the value of supply chain responsiveness in order to assess all consequences of such a strategy—both operational and behavioral.

Acknowledgements. Many thanks to Gérard Cachon, Marshall Fisher, Serguei Netessine, Senthil Veeraraghavan, Arvind Tripathi, and seminar participants at the University of Pennsylvania, the University of Texas at Dallas, the University of Rochester, the University of Washington, New York University, London Business School, Washington University in St. Louis, Northwestern University, the University of Chicago, Columbia University, Duke University, Stanford University, and the INFORMS Annual Meeting in Seattle for numerous comments and suggestions.

References


1 Second Period Fill Rate

With a random allocation rule, the actual second period fill rate is given by

\[ \phi(q, \alpha) = \mathbb{E}\left[ \min\left((q - \xi_1 N)^+, \xi_2 N\right) / \xi_2 N \right], \]

where \( \xi_1 \) and \( \xi_2 \) are functions of \( \alpha \). Note that

\[ \frac{d\xi_1}{d\alpha} = - (\theta \alpha + (1 - \theta) (1 - \alpha)) g(\alpha) < 0 \]

and

\[ \frac{d\xi_2}{d\alpha} = \theta \alpha g(\alpha) > 0, \]

and hence \( \frac{d(\xi_1 + \xi_2)}{d\alpha} = -(1 - \theta) (1 - \alpha) g(\alpha) < 0 \). Intuitively, higher \( \alpha \) implies more consumers wait until the late period to purchase, hence first period demand is decreasing in \( \alpha \) while second period demand is increasing in \( \alpha \). Furthermore, advance selling implies \( d(\xi_1 + \xi_2)/d\alpha < 0 \), i.e., if more consumers wait to purchase, less end up buying the product.

Note that if \( q < \xi_1 N \), no customers are served in period two, while if \( q > \xi_1 D_1 \) and \( q > (\xi_1 + \xi_2) D \),
all customers are served in period two. Consequently,

\[ \phi(q, \alpha) = \Pr(q/\xi_1 < N) \times 0 + \Pr(N < q/(\xi_1 + \xi_2)) \times 1 \]

\[ + \Pr(q/\xi_1 < N < q/(\xi_1 + \xi_2)) \times 1 \]

Thus, the derivative is given by

\[ \frac{d\phi(q, \alpha)}{d\alpha} = f\left(\frac{q}{\xi_1 + \xi_2}\right) \frac{d}{d\alpha}\left(\frac{q}{\xi_1 + \xi_2}\right) + \frac{d}{d\alpha}\int_{q/(\xi_1 + \xi_2)}^{q/\xi_1} \left(\frac{q - \xi_1 x}{\xi_2 x}\right) f(x) \, dx. \]

Calculating the derivative of the integral yields

\[ \int_{q/(\xi_1 + \xi_2)}^{q/\xi_1} \left(\frac{q - \xi_1 x}{\xi_2 x}\right) \frac{d\xi_1}{d\alpha} f(x) \, dx = f\left(\frac{q}{\xi_1 + \xi_2}\right) \frac{d}{d\alpha}\left(\frac{q}{\xi_1 + \xi_2}\right). \]

Thus, the derivative reduces to

\[ \frac{d\phi(q, \alpha)}{d\alpha} = -\frac{1}{\xi_2} \int_{q/(\xi_1 + \xi_2)}^{q/\xi_1} \left(\frac{q - \xi_1 x}{\xi_2 x}\right) \frac{d\xi_1}{d\alpha} f(x) \, dx. \]

Because \( d(\xi_1 + \xi_2)/d\alpha < 0 \) and \( \frac{q - \xi_1 x}{\xi_2 x} \frac{d\xi_2}{d\alpha} + \frac{d\xi_1}{d\alpha} < \frac{d\xi_2}{d\alpha} + \frac{d\xi_1}{d\alpha} \), due to the fact that the fill rate is less than one, it follows that \( \frac{d\phi(q, \alpha)}{d\alpha} > 0 \).