COLUMBIA BUSINESS SCHOOL

Strategic Capacity Rationing to Induce Early Purchases

Q. Liu and G. van Ryzin

October 2005
DRO-2005-02

Decision, Risk & Operations Working Papers Series

Strategic Capacity Rationing to Induce Early Purchases

Qian Liu Garrett van Ryzin

October 15th, 2005

Abstract

Dynamic pricing offers the potential to increase revenues. At the same time, it creates an incentive for customers to strategize over the timing of their purchases. A firm should ideally account for this behavior when making its pricing and stocking decisions. In particular, we investigate whether it is optimal for a firm to create a rationing risk by deliberately under-stocking products. Then the resulting threat of shortages creates an incentive for customers to purchase early at higher prices. But when does such a strategy make sense? If it is profitable to create shortages, what is the optimal amount of rationing risk to create? We develop a stylized model to study this problem. In our model, customers have heterogeneous valuations for the firm's product and face declining prices over two periods. Customers are assumed to have identical risk preferences and know the price path and fill rate in each period. Via its capacity choice, the firm is able to control the fill rate and hence the rationing risk faced by customers. Customers behave strategically and weigh their payoff of immediate purchase against the expected payoff of delaying their purchases. We analyze the capacity choice that maximizes the firm's profits. First, we consider a monopoly market and characterize conditions under which rationing is optimal. We characterize how the optimal amount of rationing is affected by the magnitude of price changes over time and the degree of risk aversion among customers. We also analyze the case of aggregate demand uncertainty and show that demand uncertainty reduces a firm's optimal fill rates. Lastly, we analyze an oligopoly version of the model and show that competition reduces the firms' ability to profit from rationing. Indeed, there exists a critical number of firms beyond which a rationing equilibrium cannot be supported.

1 Introduction

Varying prices over time is a natural way for firms to increase revenue in response to uncertain and fluctuating market conditions. Apparel retailing is a canonical example. Demand for apparel is affected by factors such as weather, fashion trends and economic conditions, all of which are highly uncertain; hence, forecasting demand is inherently difficult. Moreover, lead times for design, production and distribution are normally longer than the selling season, so retailers must commit to their order quantities in advance of observing sales. As a result, they often end up with some popular products which sell out fast, while other unpopular products languish. In response, retailers dynamically change prices - maintaining full prices for their best-selling items while marking down slow sellers over time.

While most retailers still make such pricing decisions manually, many are now deploying sophisticated modeling and optimization software to help support pricing decisions. (See Talluri and van Ryzin [25] Chapter 5.) Such systems have proved quite effective; Ann Taylor, a U.S. women's apparel retailer with over 580 stores nationwide, reported a year-on-year increase in sales by 26% over the Christmas period of 2003 after it implemented markdown optimization software. Success stories like this have led to increased acceptance of model-based approaches to pricing among major retailers.

At its core, any dynamic pricing system is based on a model of demand; that is, a model of how demand responds to price changes. A typical one is to formulate demand at each point in time as only a function of the price charged at that time, the implicit assumption being that customers do not anticipate future prices; that is, they are *myopic* and buy if the current price is less than their reservation price. This model is pervasive; most commercial pricing software uses it, and it is common in the research literature too. For example, see Gallego and van Ryzin [15], Feng and Gallego [13], Federgruen and Heching [12] and Chen and Simchi-Levi [7]. The myopic customer assumption is reasonable when customers make impulse purchases and for consumable goods (food, beverages, etc.). In addition, it has the considerable practical advantage of leading to mathematically tractable models.

Yet it is arguably more realistic to model customers as behaving strategically. Faced with dynamic prices, customers may benefit from strategizing over the timing of their purchases. This is especially true in the case of *durable goods* - goods that customers buy once (or more realistically, infrequently) and use over an extended period of time. When purchasing durable goods, customers are more patient and may accelerate or postpone purchasing to obtain a lower price. In his classic work, Coase [8] gives real estate as an example of a durable good; Bulow [6] gives the example of diamonds. Even without such extreme cases of durability, goods with a finite life can also be categorized as "durable" provided they are purchased only infrequently. Many important retail categories - apparel, sporting goods, customer electronics and appliances - can be considered durable goods in this sense. Bulow [6] calls these *intermediate durable goods.*

Retailers certainly recognize the fact that customers often act strategically when purchasing such goods. Executives at Federated Department Stores, America's largest operator of department stores, have lamented to us that the department store industry's habit of running frequent promotional sales has "trained our customers to only buy on sale." Such behavior is evident in aggregate data too; for example, at Wal-Mart, the world's largest retailer, holiday season sales - including after-Thanksgiving and post-Christmas season sales events - account for close to 20% of total annual sales [23].

Some retailers have adopted strategies specifically aimed at thwarting such strategic behavior. For example, Zara, one of the largest Spanish apparel retailers, is known for deliberately setting low stock levels for its products to encourage customers to buy when they first see products they like, rather than waiting for sales [14]. Industry analysts estimate that unsold products at Zara represent less than 10% of stock, compared with the industry average of almost 20% as a result of this strategy [14].

Academic researchers too have investigated the mistakes that can occur if firms incorrectly model strategic customers as myopic. Aviv and Pazgal [2] report extensive numerical examples showing revenue losses of up to 30% by ignoring strategic behavior. Besanko and Winston [4] assess that the lost profit from ignoring customer strategic behavior could skyrocket to 60%.

1.1 Market and behaviorial assumptions

For all these reasons, understanding strategic customer behavior and its impact on pricing and quantity decisions is important. Yet there are a wide variety of assumptions one can make in modeling customer strategic behavior, each of which has important theoretical and practical implications. Here we review these assumptions in detail and relate them to the body of research on dynamic pricing in the presence of strategic customers. We also position our work relative to these assumptions and literature.

Credible commitment to prices - One important modeling issue is whether to assume a seller can credibly commit to its price schedule or whether a subgame perfect equilibrium (SPE) is a more plausible assumption. The earliest work on this issue dates back to the classic work of Coase [8], who considered a monopolist selling a durable good over time. Coase argued such a monopolist cannot implement a price-skimming strategy when customers rationally anticipate lower future prices and are willing to wait. In fact, with the possibility of continuous price adjustments, he showed that the only SPE price the monopolist can sustain is a uniform price equal to marginal cost. Discount factors play a significant role in the Coase outcome; if the firm is more patient than customers, it may be able to sustain price discrimination (c.f. von der Fehr and Kuhn [27]).

Besanko and Winston [4] show there exist SPE prices that decline over time when both the firm and customers discount utility over time; high value customers buy early at higher prices while low value customers delay their purchases to obtain lower prices. They show that the SPE price in each period is less than the single-period profit-maximizing price and the seller is worse-off as the number of periods increases. Both results show that the SPE outcome hurts the seller. Other papers that assume dynamic pricing strategies must follow

SPE solutions include Koenigsberg, Muller and Vilcassim [20], Gallien [16], Elmaghraby, G*ü*lc*ü* and Keskinocak [11] and Su [24].

While SPE behavior is an appropriate model is many settings, when a firm has repeated interactions with its customers SPE predictions may be overly pessimistic. While a firm may have an incentive to deviate from its announced decisions over time, it suffers in the long run because customers will then resort to the SPE and the resulting future losses could well outweigh any short-term gains. In such circumstances, a firm may be able to credibly commit to its decisions as a consequence of the folk theorem of infinitely repeated games. (See for example Osborne and Rubinstein [22].) To sustain such credible commitment requires an infinite time horizon and sufficiently low discount rates.

We assume that the selling firm can credibly commit to prices and quantities. The assumption is reasonable in many retail settings because of the repeated interaction argument given above; most stores sell products year after year, season after season and are therefore sensitive to future profits and the long-term impact of their pricing policies.

In addition, many sellers commit to prices simply because of advertising constraints or to simplify administration of prices. For example, Broadway theaters sell discounted tickets on the day of performance through TKTS outlets, where their "half-price tickets" policy is central to the concept. Filene's Basement has made a tradition of having automatic markdowns, in which products are marked down based on a preset schedule that begins with 25% off and drops to 75% off after four weeks. Standby airline fares at fixed discounts off full fares are yet another example of such price commitments.

In situations where price and quantity commitment is not a reasonable assumption, our results are, of course, less relevant. Our model also does not apply to consumable goods like gasoline, electricity and groceries, which are used up steadily and replaced periodically. But customers may still behave strategically in purchasing such goods. For example, they may stockpile them during promotions to take advantage of low prices. An extensive stream of research in the marketing science community covers dynamic pricing of consumer nondurables. See Ho, Tang and Bell [19], Assuncao and Meyer [1] and Bell and Iyer [3].

Strategic interaction among customers - A customer's decision may also be affected by the purchase behavior of other customers. This sort of strategic interaction among customers is a key feature of the theory of auction and optimal mechanism design. (See Myerson [21].) Similar behavior may occur in dynamic pricing problems. For example, a continuously descending dynamic pricing policy can be viewed as equivalent to a descending-price (Dutch) auction. (See Bulow and Klemperer [5].)

The important modeling question here is whether strategic interaction among customers should be taken into account when analyzing dynamic pricing. The answer ultimately depends on the market size. If the market consists of a small number of customers whose individual actions significantly affect each other, then modeling the strategic interaction among customers is vital for understanding their behavior. However, in markets consisting of a large number of customers, one customer's behavior has a negligible impact on the outcomes experienced by others. Hence, strategic interactions among customers can be reasonably ignored. It is then a matter of whether one is interested in analyzing a small or large market; our work assumes a large market.

Strategic interaction among customers is considered by Elmaghraby, G*ü*lc*ü* and Keskinocak [11], Gallien [16], Harris and Raviv [18], Aviv and Pazgal [2], Koenigsberg, Muller and Vilcassim [20]. In contrast, Besanko and Winston [4], Dana and Petruzzi [9], Su [24] - and our work - assume a market consisting of a large customer population and thus strategic interaction among customers is ignored. This assumption simplifies the analysis of customer behavior considerably, since one can then assume each customer reacts to prices and quantities without considering the effect their actions have on other customers. The assumption is also quite reasonable for most mainstream retail categories; as customers, most of us do not worry that our individual shopping decisions will alter the behavior of other "competing shoppers," even though we may account for the aggregate effect of other customers, for example when assessing the likelihood of future shortages. Such behavior is consistent with the large-market assumption.

Capacity constraints and rationing - Whether capacity is exogenously given or endogenously derived, it has a significant influence on pricing strategy. Harris and Raviv [18] find that the optimal pricing mechanism consists of a uniform price if capacity is not binding, while a differential, priority pricing scheme becomes optimal once capacity constraints are binding. In practice, capacity is frequently limited due to budget constraints, finite storage space, pre-commitments due to long procurement lead time, etc. It may also be the case that capacity is limited deliberately to induce high value customers to buy at higher prices to avoid rationing risk, as in the case of Zara mentioned above. That scarcity can increase value is well recognized, as illustrated by the following quote in FastCompany: [17]

> Scarcity, after all, is the cornerstone of our economy. The only way to make a profit is by trading in something that's scarce. This is why the music and movie industries are so terrified by the millions of people who download entertainment from the Internet every day. Downloading threatens to make supply virtually unlimited...

Scarcity also explains why customers are willing to buy broadway tickets at full price even though they are offered for half price at TKTS and why customers at Filene's Basement don't all wait for the lowest automatic markdown; waiting for a low price entails a risk of rationing.

But how best to balance the costs and benefits of creating scarcity is not well understood, and analyzing this trade-off is a central focus of our work. Su [24] also explicitly considers rationing effects and in his model it serves the same role of creating incentives for high-value customers to buy early at high prices. The main modeling difference is that in our paper rationing is implicitly determined by the firm's initial capacity choice, while Su introduces an exogenous rationing fraction as an explicit decision variable in his model.

Risk preferences - Risk aversion is a reasonable assumption when modeling customer behavior, especially when customers make "large" purchases - purchases whose cost represents a significant portion of a customer's wealth or budget. For example, computers, major household appliances and luxurious fashion items can be regarded as "large" purchases in this sense. Different assumptions of risk preferences can lead to different conclusions about pricing strategies, a fact illustrated by auction theory; a risk neutral firm prefers a first-price auction to a second-price auction in the face of risk averse bidders, while it is indifferent to these two formats with risk neutral buyers.

Our work compares stocking decisions in the presence of risk neutral and risk averse customers, respectively. Allowing customers to be risk averse is one distinct feature of our work and the assumption plays a key role in our analysis.

Price or quantity as a decision variable - Both prices and stocking quantities can affect the purchase behavior of strategic customers, and which is more appropriate to model depends on the context. Often firms must commit to price or capacity. Firms may commit to price for advertising or administrative reasons. They may commit to capacity because production is inflexible or lead times are long. Such commitments limit a firm's ability to adjust quantity or price decisions on a tactical level. For example, the airline, hotel and car rental industries traditionally manage quantity allocations and commit to prices in advance, while in retailing, due to long procurement lead times, firms normally commit to capacity and tactically change prices.

Our work considers the stocking quantity as the primary tactical decision variable. One motivation for this choice is to understand tactical rationing decisions in the case where firms commit to prices, as in the TKTS and Filene's Basement examples. On a theoretical level, our analysis also provides a complement to traditional inventory theory, which views stocking decisions as primarily driven by holding and fixed costs, overage and underage costs, demand uncertainty, service level constraints, etc.. In our model in contrast, the main motivation for stocking decisions is to control the rationing risk faced by customers in order to profitably influence their strategic behavior. In this sense, it provides a behavioral rather than cost-based explanation of stocking decisions.

To our knowledge, almost all papers on dynamic pricing with strategic customers use price as the decision variable. An exception is the work of Dana and Petruzzi [9], who extend the classical newsvendor model by considering demand to be affected by both price and inventory level (fill rate). In their model, strategic customers face a binary decision - to purchase or not. Because there is a cost to visit the store, their choices depend on both price and the anticipated fill rate. They find that higher fill rates (compared to the traditional newsvendor) are optimal when the firm internalizes the effect of its inventory on demand. Besides the exogenous price case, they explore the case where both price and stocking quantity are decision variables. Su [24] also considers optimizing over the initial capacity in his model. He shows the optimal strategy in this case falls into one of three cases, two of which correspond to selling at uniform prices (high or low), and one which corresponds to a single markdown (high-then-low-price) policy. When customers are risk neutral, we find that a similar strategy of selling to all customers at either a single high or low price is optimal.

Other assumptions - Besides the assumptions discussed thus far, the following are also worth noting: First, do customers arrive sequentially or simultaneously? Many paper - including ours - assume that all customers are present at the beginning of the sales season. Aviv and Pazgal [2], Gallien [16], Su [24] and Zhou et al. [28] in contrast consider sequential arrivals. Generally, achieving price discrimination is more difficult with simultaneous arrivals because customers have more flexibility in their purchase timing in this case.

Second, is utility constant or discounted over time? In reality, utility is of course discounted. But in order to simplify analysis, it is acceptable to regard utility as constant if the time horizon is not too long. In Aviv and Pazgal's work, utility is discounted in the sense that they assume customer valuations strictly decline over time. Gallien [16] studies an online commerce mechanism design in an infinite time horizon, in which case the use of discounting is natural. In Besanko and Winston's paper, discount factors play a key role in the SPE; in fact, in their model if utility is not discounted over time, all customers simply purchase in the last period at the lowest price as in the Coase problem. Our basic model does not consider discounting, though we include it as an extension. We ignore discounting partly as a practical approximation, but more importantly to isolate and study the effect that rationing risk alone has on customer behavior.

Lastly, is aggregate demand deterministic or stochastic? In the setting of uncertain demand, it becomes much harder to model customer strategic behavior. Still, quite a few papers including Aviv and Pazgal [2], Zhou et al. [28], Dana and Petruzzi [9], address dynamic pricing under uncertain aggregate demand. We also investigate the case of aggregate demand uncertainty, though our primary analysis is for the case of deterministic aggregate demand.

1.2 Overview of our paper

The remainder of the paper is organized as follows: We formulate our strategic capacity rationing model in section 1. In section 2, we model a single seller's stocking decision problem with deterministic aggregate demand. In section 3, we characterize the optimal stocking levels and rationing risk for the firm and provide comparative statics on how these decisions are affected by the magnitude of price discounting over time, the dispersion in customers' valuations and their degree of risk aversion. We then extend the model and analysis to allow for discounting, aggregate demand uncertainty and the case of symmetric competing firms in a market where customers assess the aggregate industry availability (rather than firm-level availability) in making their decision to buy early or wait. We characterize the equilibrium stocking decisions and examine the effect of competition on stocking quantities and the ability of firms to sustain price discrimination. Conclusions are given in Section 5.

2 The Model

A monopoly firm preannounces a single markdown pricing policy over two periods; the unit price in period 1, denoted p_1, is greater than the unit price in period 2, denoted p_2. We assume the seller is able to commit to this preannounced price path. We consider a market which consists of a large customer population. The market size, denoted N, is deterministic. (We will allow for a stochastic market size below.) Customers have heterogeneous valuations and unit demand for the good. All customers are present when sales begin and remain in the market until their requests are satisfied or the sales season is over. Customers behave strategically and take both the current and the future prices and availability into consideration when deciding to buy either in period 1 or 2. The firm seeks to maximize profit by choosing its stocking quantity (capacity) at the beginning of the sales season. Inventory is not replenishable once sales starts.

We assume the firm is risk neutral and customers are risk averse. (We study risk neutral customers as a limiting case.) Customers' valuations are independent and identically distributed with cumulative distribution function (c.d.f.) $F(v)$, which is common knowledge to both the firm and customers. Moreover, the distribution of customer valuations is constant over time. Customers have identical utility functions, denoted $u(\cdot)$, which are time-invariant, strictly increasing and concave.

All purchase requests in period 1 are filled, while customers may face a rationing risk in period 2 due to insufficient supply. (We justify these assumptions below.) Let q denote the probability of obtaining a unit in period 2 (the *fill rate*). Random (parallel) rationing is assumed; that is, each customer attempting to purchase in period 2 has an equal chance of obtaining a unit. The fill rate q is determined by the firm's capacity choice, as we show below. We assume customers can correctly anticipate the firm's fill rate; that is, customers have full information and rational expectations.

2.1 The Customer's Decision

As illustrated in Figure 1, a customer weighs the payoff of an immediate purchase at a high price against the expected payoff of a later purchase at a low price and buys one unit in period 1 if and only if: $u(v - p_1) \geq qu(v - p_2)$, and $v - p_1 \geq 0$. Intuitively, high-valuation customers are more likely to purchase immediately because they face a larger loss of utility if they are rationed out of the market. The following proposition formalizes this property:

Proposition 1 $\forall q \in [0, 1)$, *there exists the unique* $v(q) \geq p_1$, *such that customers with valuations greater than* $v(q)$ *buy in period 1 and those with valuations less than* $v(q)$ *wait to buy in period 2.*

At zero fill rate, all customers with valuations greater than p_1 buy at the high price in period 1; thus $v(0) = p_1$. When the fill rate is one (no rationing risk), no customers

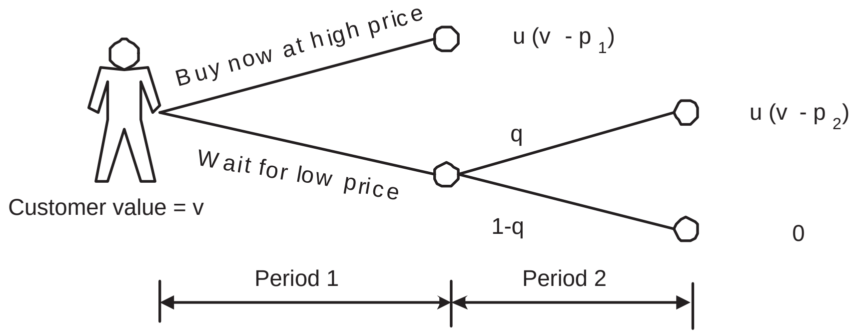

Figure 1: Two period decision model

purchase in period 1, and we define $v(1) = +\infty$. In the case of finite customer valuations, $v(1) = \bar{U}$ where $\bar{U}$ is the upper bound of valuation. Intuitively, with a higher fill rate, more customers are willing to risk purchasing later at low prices. This is verified by the following proposition:

Proposition 2 *The threshold function $v(q)$ is strictly increasing in q. Moreover, it is convex on q if the associated utility function $u(\cdot)$ has the nonnegative third derivative.* [1]

Proposition 2 implies that a lower fill rate (larger rationing risk) induces more customers to buy early at high prices. While a firm would like to induce customer to purchase early at high prices, this increase comes at the expense of lost sales due to rationing in period 2. This tradeoff between the benefits of inducing early purchases and the costs of lost sales in period 2 is key to understanding the firm's optimal stocking decisions.

2.2 The Firm's Stocking Decision

Without loss of generality, we normalize $p_1 = 1$ and $p_2 = \beta < 1$. Let C be the firm's stocking quantity before sales, and α_1 be the unit procurement cost, $\alpha_1 < \beta$. We assume the firm's cost function is linear, hence $\alpha_1 C$ is the cost of stocking C units.

The fill rate in period 2 is given by the ratio of residual capacity to residual demand in period 2, namely,

$$q = \min\left\{\left(\frac{C - N\bar{F}(v(q))}{N(F(v(q)) - F(\beta))}\right)^+, 1\right\} \tag{1}$$

where $(\cdot)^+ = \max\{\cdot, 0\}$. Equation (1) shows how the firm is able to influence the fill rate via its capacity choice. The fill rate in turn influences customer behavior through the threshold

[1] Nonnegativity of the third derivative of a utility function is called *prudence* in the economics literature, see Eeckhoudt *et al.* [10]. For example, the power function $u(x) = x^\gamma$, $0 < \gamma < 1$ and the exponential function $u(x) = 1 - e^{-kx}$, $k > 0$ are concave and have positive third derivatives.

function $v(q)$, defined (implicitly) by the indifference point:

$$u(v-1) = qu(v-\beta). \tag{2}$$

The firm would like to choose its capacity to induce the most profitable demand outcome.

An important fact to note is that, in general, a single capacity choice could induce multiple outcomes. An example is the following: suppose $u(x) = \sqrt{x}$, $F(x) = \frac{1}{4}x^2$, $\beta = 0.2$, $\alpha_1 = 0$ and $C = \frac{3}{4}N$. It is easy to verify both $q = 0$, $v(q) = 1$ and $q = 0.72$, $v(q) = 1.865$ are possible outcomes; that is, both pairs are solutions satisfying (1) and (2). Intuitively, the notion here is that of a "self-fulfilling prophecy"; in the case $q = 0$, $v(q) = 1$, customers expect a zero fill rate in period 2 ($q = 0$) and all those with valuations above one attempt to buy in period 1 as a result ($v(q) = 1$). The resulting demand entirely consumes the capacity in period 1, which indeed produces a fill rate of zero in period 2. On the other hand, if customers expect a high fill rate in period 2 ($q = 0.72$), many opt to postpone purchasing (all those with values below $v(q) = 1.865$). Demand in period 1 is therefore low, and hence there is excess product left over for period 2, which indeed produces a high fill rate of 0.72 in period 2. Just as with multiple equilibria in a game theory model, the predictions of our model are ambiguous in such cases. We will specialize the assumptions below to eliminate this ambiguity, though it is worth keeping in mind that multiple outcomes of this sort could very well be a realistic phenomenon.

We now specify the firm's optimization problem. First, we claim it is optimal for the firm to stock at least enough to meet potential demand at the high price, $N\bar{F}(1)$, and never more than required to satisfy the potential demand at the low price, $N\bar{F}(\beta)$. This follows because if the supply is not sufficient to meet demand at the high price, the fill rate in the second period is zero and there will even be rationing at the high price in period 1. Profits in this case are always dominated by choosing $C = N\bar{F}(1)$ and serving the entire market at the high price. Conversely, if the firm stocks more than demand at the low price, the fill rate in period 2 will be one, $N\bar{F}(\beta)$ customers will buy in period 2 and there will be positive left over stock. In this case, profits are always higher by stocking exactly $C = N\bar{F}(\beta)$. Therefore, the optimal capacity choice of the firm always satisfies $C \in [N\bar{F}(1), N\bar{F}(\beta)]$, and the fill rate q defined in (1) can therefore be simplified to

$$q = \frac{C - N\bar{F}(v(q))}{N(F(v(q)) - F(\beta))}. \tag{3}$$

We will assume customers' valuations are bounded above by $\bar{U}$. Then the firm's profit maximization problem can be expressed in terms of C and v as follows:

$$\begin{aligned} \max \quad & N(1-\beta)\bar{F}(v) + (\beta - \alpha_1)C \\ \text{s.t.} \quad & u(v-1) = \frac{C - N\bar{F}(v)}{N(F(v) - F(\beta))}u(v-\beta), \\ & 1 \le v \le \bar{U}, \quad N\bar{F}(1) \le C \le N\bar{F}(\beta), \end{aligned} \tag{4}$$

where the first constraint defines the threshold value v using (2) to define v and (3) to define the fill rate, and the second and third constraints limit the values of v and C to their relevant ranges.

Two things are worth mentioning about this formulation. First, the last constraint in (4) is implied by the first two constraints and is therefore redundant, and hence won't be explicitly considered hereafter. Second, when customers' valuations are bounded above by $\bar{U}$, the feasible range of fill rates in the model (4) is $[0, \bar{q}]$ where $\bar{q} = \frac{u(\bar{U}-1)}{u(\bar{U}-\beta)} < 1$. Once the fill rate exceeds $\bar{q}$, no customer buys at the high price. As a result, the firm stocks $N\bar{F}(\beta)$ exactly serving the entire low-price market. Therefore, the problem (4) only models the case of a segmented market. The case where $q = 1$ and the entire market is served at the low price must be analyzed separately. Let (v^0, C^0) denote the optimal solution to (4), and Π^0 be the associated optimal segmented profit. Let Π^{NS} denote the profit obtained by serving the entire market at a low price only, then $\Pi^{NS} = (\beta - \alpha_1)N\bar{F}(\beta)$. So, the firm's optimal stocking quantity corresponds to the one that achieves the maximum of Π^0 and Π^{NS}. We denote the optimal cutoff value by v^*, the optimal fill rate by q^* and the optimal stocking quantity by C^*.

3 Analysis of the Optimal Stocking Decision

A key question is whether it is optimal to create rationing risk or to simply serve the entire market at one price. If rationing is optimal, what level of rationing risk should be created? And how do these answers depend on the factors such as the price differences over time, the distribution of customer valuations and the level of risk aversion among customers? To answer these questions, we start with risk neutral customers, then proceed to the main result on risk averse customers.

To facilitate analysis, we assume in the remainder of this paper customers' valuations are uniformly distributed over $[0, \bar{U}]$. We also assume customers have a power utility function $u(x) = x^\gamma$ $(0 < \gamma < 1)$, which is a common form in the economics literature and corresponds to the case where customers have decreasing absolute risk aversion. Lower values of γ correspond to more risk aversion. These assumptions simplify the analysis and also eliminate the multiple-outcome problem mentioned above.

3.1 Risk Neutral Customers

For risk neutral customers, $v - p$ is the payoff given a customer's valuation v and the unit price p. It is easy to check that Proposition 1 still holds when $u(x) = x$. (The proof comes directly from the fact that $(v - p_1)/(v - p_2)$ strictly increases in v.) Hence, the risk neutral firm maximizes its profit as follows:

$$\max \quad \frac{N}{\bar{U}}(1-\beta)(\bar{U}-v) + (\beta - \alpha_1)C \tag{5}$$

$$\text{s.t.} \qquad v - 1 = \frac{C - \frac{N}{\bar{U}}(\bar{U} - v)}{\frac{N}{\bar{U}}(v - \beta)}(v - \beta), \quad \bar{U} \geq v \geq 1.$$

Facing risk neutral customers, the firm serves the market either at a high price only or at a low price only; this is established by the following proposition:

Proposition 3 *For risk neutral customers, 1) If $\bar{U} \geq 1 + \beta - \alpha_1$, $v^* = 1$, $q^* = 0$ and $C^* = \frac{N}{\bar{U}}(\bar{U} - 1)$; 2) If $\bar{U} < 1 + \beta - \alpha_1$, $v^* = \bar{U}$, $q^* = 1$ and $C^* = \frac{N}{\bar{U}}(\bar{U} - \beta)$.*

Intuitively, Proposition 3 says that when the market consists of a sufficiently large number of high-value customers ($\bar{U}$ is large), it is optimal for the firm to serve the market only at the high price in period 1; while when the number of high-value customers is small, the firm is better off serving the entire market at the low price only. But note rationing is never optimal.

3.2 Risk Averse Customers

Risk aversion among customers significantly changes the firm's optimal strategy. Using the power utility function $u(x) = x^\gamma$ $(0 < \gamma < 1)$, the firm's stocking decision is determined by solving:

$$\max \quad \frac{N}{\bar{U}}(1 - \beta)(\bar{U} - v) + (\beta - \alpha_1)C \tag{6}$$

$$\text{s.t.} \qquad \left(\frac{v - 1}{v - \beta}\right)^\gamma = \frac{\frac{\bar{U}}{N}C - \bar{U} + v}{v - \beta}, \quad \bar{U} \geq v \geq 1.$$

The profit can be expressed only in terms of v as follows:

$$\max_{\bar{U} \geq v \geq 1} \Pi(v) = \frac{N}{\bar{U}}\left((1 - \alpha_1)(\bar{U} - v) + (\beta - \alpha_1)(v - \beta)\left(\frac{v - 1}{v - \beta}\right)^\gamma\right). \tag{7}$$

The first order conditions yield:

$$\left(\frac{v - 1}{v - \beta}\right)^\gamma \left(1 + \frac{\gamma(1 - \beta)}{v - 1}\right) - \frac{1 - \alpha_1}{\beta - \alpha_1} = 0. \tag{8}$$

We then have the following lemma:

Lemma 1 *The profit function $\Pi(v)$ defined in (7) is strictly concave in $v \geq 1$. Furthermore, the maximizer of (7) is either the solution to (8) denoted by v^0 if $v^0 \leq \bar{U}$ or $\bar{U}$ otherwise.*

The proof of Lemma 1 establishes that the left hand side of (8) strictly decreases in $v > 1$, and has opposite signs at $v \to 1^+$ and $v \to +\infty$. Hence, there exists the unique root $v^0 > 1$ to (8). Notice also that v^0 is independent of $\bar{U}$. As discussed above, the firm's profit is maximized either at the segmented market optimum or at the low-price-only solution. The following proposition characterizes the optimal solution.

Proposition 4 *Suppose v^0 satisfies (8), and denote*

$$U_c = \frac{(\beta + \gamma(1-\alpha_1))v^0 - \beta(1+\gamma(\beta-\alpha_1))}{v^0 - 1 + \gamma(1-\beta)}.$$

If $\bar{U} \geq U_c$, the optimal strategy is to induce segmentation by creating rationing risk. The optimal solution in this case is $v^* = v^0$, $q^* = q^0 = \left(\frac{v^0-1}{v^0-\beta}\right)^\gamma$, and $C^* = C^0 = \frac{N}{\bar{U}}(\bar{U} - v^0 + (v^0-\beta)q^0)$. Otherwise, the optimal strategy is to serve the entire market at the low price; namely, $v^* = \bar{U}$, $q^* = 1$ and $C^* = \frac{N}{\bar{U}}(\bar{U} - \beta)$.

Proposition 4 shows that whether it is optimal to create rationing risk or not depends on the number of high-value customers in the market. When there are a large number of high-value customers ($\bar{U} \geq U_c$), the incremental demand induced in period 1 more than compensates for the lost sales cost of rationing in period 2. If there are relatively few high-value customers ($\bar{U} < U_c$), the opposite is true; the incremental demand induced in period 1 by creating rationing risk does not compensate for the lost sales in period 2. Also, note that, unlike in the risk neutral case, it is never optimal to serve the entire market only at the high price; the proof of Proposition 4 shows that the segmented solution (v^0, q^0, C^0) always dominates the high-price-only solution.

Sufficient conditions for creating rationing risk are provided in the following corollary, which directly follows with $1 + \gamma(\beta - \alpha_1) < U_c < 1 + \beta - \alpha_1$, as shown in the proof of Proposition 4.

Corollary 1 *When $\bar{U} \geq 1 + \beta - \alpha_1$, a strategy of creating rationing risk characterized by (v^0, q^0, C^0), is always optimal. When $\bar{U} \leq 1 + \gamma(\beta - \alpha_1)$, the optimal strategy is to serve the entire market only at the low price.*

Note the ratio of the potential demand at a high price to potential demand at a low price is $\frac{\bar{U}-1}{\bar{U}-\beta}$. That $\bar{U} \geq 1 + \beta - \alpha_1$ implies $\frac{N}{\bar{U}}(\bar{U}-1)(1-\alpha_1) \geq \frac{N}{\bar{U}}(\bar{U}-\beta)(\beta-\alpha_1)$ means the profit gained at the high-price-only solution is larger than that at the low-price-only solution. Since it is never optimal to serve customers only at the high price in the risk averse case, this implies when $\bar{U} \geq 1 + \beta - \alpha_1$ it is always optimal to create rationing risk. Recall the optimal strategy with risk neutral customers is to only serve the high-price market if $\bar{U} \geq 1 + \beta - \alpha_1$. Since the high-price-only profit values are the same in each case, this shows the firm is strictly better off with risk averse customers.

3.3 Comparative Statics

In this section, we examine how the optimal fill rate is affected by discount price and the degree of risk aversion. We first require the following result:

Proposition 5 *The capacity C satisfying (3) increases in the cutoff value v and fill rate q.*

Proposition 5 is intuitively obvious; to produce a higher fill rate (equivalently, a higher cutoff value) requires the firm to stock more. The total sales volume is then increased, but then customers' incentive to delay purchasing is increased. Again, the firm's decision is to create the "right" rationing risk, balancing the tradeoff between benefits of inducing segmentation and costs of lost sales. Proposition 6 and 7 below establish how the optimal amount of rationing risk is influenced by the price in period 2 and the degree of risk aversion among customers.

Proposition 6 *The optimal fill rate q^* is increasing in the discount price β.*

This result too is intuitive. As the ratio of the low price to high price, β, increases, the opportunity cost of rationing increases as well. On the other hand, a higher price in period 2 reduces customers' incentive to postpone their purchases. Both effects reduce the benefits of rationing in period 2, and drive fill rates up. Note that the optimal fill rate q^* is not necessarily continuous in β. Once a segmented market is no longer optimal, the optimal fill rate jumps to one.

Another question is how the degree of risk aversion impacts the optimal amount of rationing. Intuitively, the more risk averse customers are, the more incentive they have to purchase early. Hence, the firm should be able to use a higher fill rate. This is verified in the following Proposition:

Proposition 7 *The more risk averse customers are (smaller γ), the larger the optimal fill rate q^* as long as inducing segmentation via rationing is optimal.*

Proposition 7 implies more rationing risk is needed to induce segmentation as customers become less risk averse. However, at some point, the opportunity cost of rationing may become too great and it then becomes optimal to serve the entire market at the low price. Below we examine sufficient conditions under which segmentation using rationing is optimal. We first consider the limiting cases of risk aversion (γ approaching 0 or 1), and compare the results with risk neutral customers.

Proposition 8 *Suppose v^0 is the solution to (8), and $q^0 = \left(\frac{v^0-1}{v^0-\beta}\right)^\gamma$. Then $\lim_{\gamma\to 0^+} q^0 = 1^-$, $\lim_{\gamma\to 0^+} v^0 = 1^+$, $\lim_{\gamma\to 1^-} q^0 = 0^+$, and $\lim_{\gamma\to 1^-} v^0 = 1^+$.*

The above proposition says an extremely high rationing risk is required to induce segmentation as customers become much less risk averse ($\gamma \to 1$), and when customers are extremely risk averse ($\gamma \to 0$), a negligible rationing risk is all that is required to induce optimal segmentation. Again, however, whether rationing is optimal or not hinges on the market composition, specifically the number of high-value customers; the above limits only apply to the optimal rationing solution.

As γ becomes larger (customers become less risk averse), more rationing risk is needed to induce segmentation. At some point, it becomes too costly to create rationing and hence

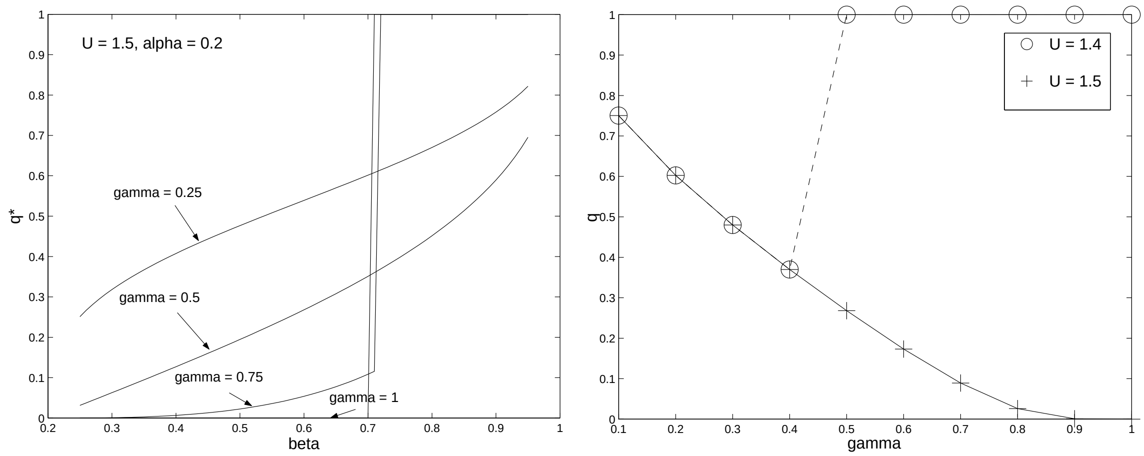

Figure 2: Optimal fill rate vs discount price and degree of risk aversion

a low-price-only solution becomes optimal. The precise condition under which the optimal stocking decision switches from a segmented market to a non-segmented market is provided in the following proposition.

Proposition 9 *v^0 is the solution to (8), and*

$$q^0 = \left(\frac{v^0 - 1}{v^0 - \beta}\right)^{\gamma}, \qquad \hat{q} = \left(\frac{\gamma(1 + \beta - \alpha_1 - \bar{U})}{(\bar{U} - \beta)(1 - \gamma)}\right)^{\gamma}.$$

When $1 + \gamma(\beta - \alpha_1) < \bar{U} < 1 + \beta - \alpha_1$, q^0 is the optimal fill rate if $q^0 \geq \hat{q}$; otherwise, the optimal fill rate is 1. Further, once the optimal fill rate jumps to 1, it will stay at 1 as γ continues to increase.

3.4 Numerical Examples

In this section, we conduct several examples to illustrate how the optimal stocking decision is affected by the difference in prices over time (β) and the degree of risk aversion (γ). We also give examples that illustrate the sensitivity of profits to errors in capacity and prices to get a sense of which variable most affects profits.

The left graph in Figure 2 shows that the optimal fill rate q^* increases in the discount price β. In this example, we set $\bar{U} = 1.5$, $\alpha_1 = 0.2$, and let γ takes values over $\{0.25, 0.5, 0.75, 1\}$. Observe that for risk neutral customers ($\gamma = 1$), the optimal strategy is either a high-price-only solution ($q^* = 0$ if $\beta \leq 0.7$ in this example) or a low-price-only solution ($q^* = 1$ if $\beta > 0.7$) as our theory predicts. Also, we note that the optimal fill rate is not necessarily continuous in β. At some point, the optimal strategy switches

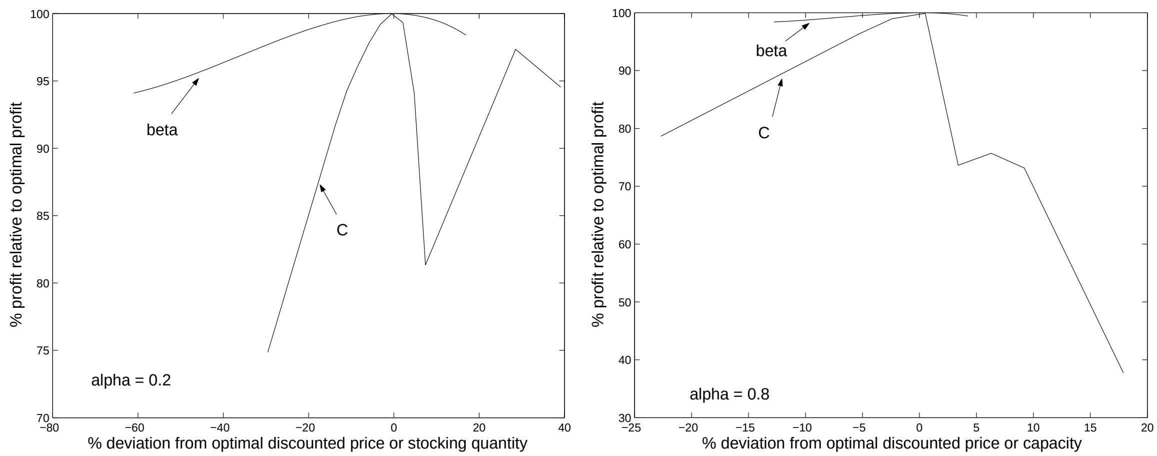

Figure 3: Sensitivity of profits to errors in discount price and capacity

from inducing segmentation to a low-price-only strategy, which, again, reflects the tradeoff between benefits of rationing and cost of lost sales.

The right graph in Figure 2 shows how the optimal fill rate changes with the degree of risk aversion. In this example, we test $\beta = 0.7$, $\alpha_1 = 0.2$, $\bar{U} = 1.4$ or 1.5. Observe that when customers are highly risk averse (i.e. γ is small), optimal fill rates are the same for different ranges of valuations. This is because the optimal segmenting point v^0 (or q^0) does not depend on the range of valuations, provided inducing segmentation is an optimal strategy. But, as customers become less risk averse (γ increases), differences emerge. For $\bar{U} = 1.5$, a larger rationing risk continues to be created as customers become less risk averse; in contrast, for $\bar{U} = 1.4$, the firm gives up inducing segmentation and switches to a low-price-only strategy. In the limiting case ($\gamma \to 1$), the former converges to a fill rate of one, while the latter converges to a fill rate of zero, exactly the results for risk neutral customers.

While price is not a variable in our analysis, one might question whether profits are more sensitive to the firm's pricing or capacity decisions. To investigate this question, we ran examples on different settings of parameters, and numerically optimized over prices as well as capacities. We first found the optimal discount price β^* numerically, then calculated the corresponding optimal stocking quantity and profit at β^*. Next, we computed relative profit losses under various deviations from β^* and from the associated optimal capacity at β^*, respectively.

Figure 3 illustrates the results for the case of $N = 1000$, $\gamma = 0.5$, $\bar{U} = 1.5$ and $\alpha_1 = \{0.2, 0.8\}$. In this example, the relative profit losses under 40% price deviations are within 5%, while the relative profit losses can be up to 25% under the same percentage of capacity deviations. Profits here are therefore much more sensitive to errors in capacity than errors in price. This suggests that inventory policies can be much more important than pricing when faced with strategic customers. Why? First, customers make decisions by comparing the current utility, $u(v-1)$, with their expected future utility, $qu(v-\beta)$. With a concave utility function, the threshold v may change slowly with changes in β; in contrast, a given percentage change in fill rate translates directly into the same percentage change in expected utility in period 2. Second, since fill rate is determined by residual capacity in period 2, a small percentage change in capacity choice can result in a large change in fill rate, thus leading to a large change in the threshold.

4 Extensions to the Basic Problem

We next examine several extensions of our basic problem. The first is to consider what happens when customers discount utility over time. In general, we show that discounting creates an exogenous incentive to buy early, which adds to the incentive created by rationing risk. Second, we consider the effect of aggregate uncertainty in demand. Our model of aggregate uncertainty is somewhat artificial but amenable to analysis. The general result here is that aggregate uncertainty makes it even more costly to serve customers in the second period and creates an incentive for the firm to lower its fill rate and serve even more customers in the high-price period. Lastly, we look at the effect of competition; specifically, the case where there are several stores in the market and customers assess the fill rate at the market level rather than the individual store/firm level. We analyze the equilibrium capacity strategies among the firms and show that, in general, competition reduces the industry's ability to profitably use rationing risk to segment customers.

4.1 Discounted Utility

Even without the risk of rationing, discounting of utilities over time creates an incentive for customers to buy early at high prices. When both discounted utilities and rationing risk are considered, we show that the results are qualitatively the same as those in our basic model without discounting, though characterizing the optimal strategies becomes more complex.

Let δ, $\frac{\alpha_1}{\beta} < \delta < 1$ ($\delta > \frac{\alpha_1}{\beta}$ guarantees positive margins at the low price), denote the discount factor applied to period two utilities and profits. We assume a perfect capital market, so that the firm and customers share the same discount factor. The firm maximizes its total profit, composed of a base profit and an extra profit margin via sales at the high price:

$$\begin{aligned} &\max \quad N(1-\beta\delta)\bar{F}(v) + (\beta\delta - \alpha_1)C \qquad (9) \\ \text{s.t.} \quad &u(v-1) = \delta \frac{C - N\bar{F}(v)}{N(F(v) - F(\beta))} u(v-\beta), \quad v \geq 1. \end{aligned}$$

Again, in what follows, customers' valuations are assumed to be uniformly distributed over $[0, \bar{U}]$.

4.1.1 Risk Neutral Customers

With risk neutral customers, (9) can again be rewritten only in terms of v:

$$\max_{1 \le v \le \bar{U}} \; -\frac{N}{\bar{U}}\Big(1-\beta+\frac{\alpha_1}{\delta}-\alpha_1\Big)v + N(1-\alpha_1) - \frac{N}{\bar{U}}\Big(\beta-\frac{\alpha_1}{\delta}\Big). \tag{10}$$

It is easy to see the above problem is maximized at $v = 1$, implying the optimal stocking decision is to serve the high-price market only, yielding a profit of $\frac{N}{\bar{U}}(\bar{U}-1)(1-\alpha_1)$.

Consider the situation of providing a fill rate of one in period 2, in contrast to the undiscounted case, customers may segment due to discounting even with a fill rate of one. In particular, the threshold value is $\hat{v} = \frac{1-\delta\beta}{1-\delta}$. If $\hat{v} < \bar{U}$, customers are segmented even though the fill rate in period 2 is one. However, the result of the profit-maximizing problem (10) indicates that a fill rate of one is never optimal; rather, the optimal strategy is to serve the entire market at the high price when considering both effects of discounting utility and rationing risk.

4.1.2 Risk Averse Customers

When customers are risk averse and have power utility functions, (9) can be written in terms of v as follows:

$$\max_{1 \le v \le \bar{U}} \Pi(v) = \frac{N}{\bar{U}}\Big\{(\bar{U}-v)(1-\alpha_1) + (\beta-\frac{\alpha_1}{\delta})(v-\beta)\Big(\frac{v-1}{v-\beta}\Big)^{\gamma}\Big\}. \tag{11}$$

As in Lemma 1 for the undiscounted case, one can show the profit function $\Pi(v)$ defined by (11) is strictly concave in $v \ge 1$. The first order conditions imply:

$$\Big(\frac{v-1}{v-\beta}\Big)^{\gamma}\Big(1+\frac{\gamma(1-\beta)}{v-1}\Big) - \frac{1-\alpha_1}{\beta-\frac{\alpha_1}{\delta}} = 0. \tag{12}$$

Applying the same argument as the undiscounted-utility case, one can show there exists a unique solution, denoted $v^D > 1$ to (12). Moreover, v^D is increasing in δ. This is intuitive; as the discount factor becomes smaller, future utilities and profits become less important and hence it is optimal to induce more customers to purchase early.

The precise characterization of optimal stocking decisions with discounts can be derived using the same approach as in Proposition 4. The results are algebraically complex and not presented here; instead, we only give sufficient conditions for creating rationing risk. Specifically, when $\bar{U} \ge \frac{1-\alpha_1+\alpha_1\beta-\delta\beta^2}{1-\delta\beta}$, it is optimal to create rationing risk; when $\bar{U} \le \frac{(1-\alpha_1)(1-\beta)(1+\gamma(\beta-\frac{\alpha_1}{\delta}))+\beta(\alpha_1+\beta-\frac{\alpha_1}{\delta}-\delta\beta)}{(1-\frac{\alpha_1}{\delta})(1-\delta\beta)}$, introducing rationing risk is not optimal.

Recall that there is a monotone relationship between fill rates and cutoff values in the undiscounted-utility case. However, when we take into account both discounted utility and rationing risk, this is no longer true. To see why, note that the optimal fill rate is

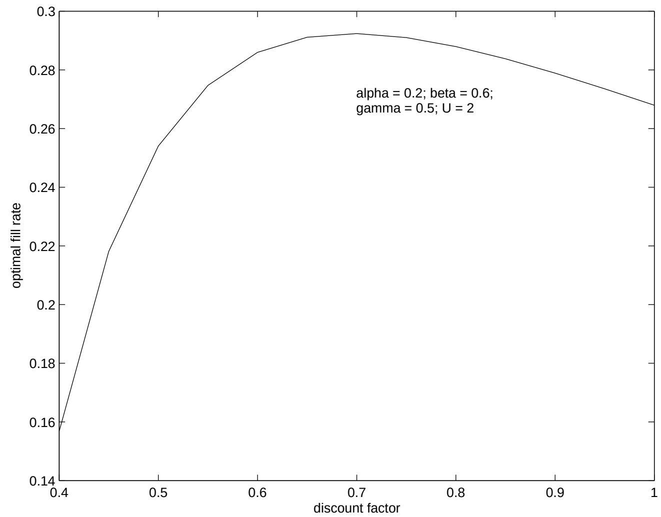

Figure 4: Optimal fill rate vs discount factor

determined by $q^* = \frac{1}{\delta}\left(\frac{v^*-1}{v^*-\beta}\right)^\gamma$. Hence, a smaller discount factor tends to produce a higher fill rate; conversely, smaller discount factors leads to lower threshold values because v^D is increasing in δ; this tends to reduce the optimal fill rate. It is hard to characterize the impact of these two offsetting effects on the optimal fill rate. Figure 4 gives a sense of how the optimal fill rate changes with discount factor when both rationing and discounting are taken into account.

4.2 Aggregate Demand Uncertainty

It is more realistic to assume that aggregate demand is uncertain due to fluctuating market conditions. In this section, we relax the assumption that the market size N is deterministic and instead assume N is a random variable following some known c.d.f. $G(N)$. We will assume $G(N)$ is uniformly distributed over $[\underline{D}, \bar{D}]$ for simplicity. Customers' valuations follow a c.d.f. $F(v)$, which is interpreted as the deterministic fraction of the market size N with given valuations. Recall the firm commits to its capacity upfront in the case of certain demand; shortages never happen at the high price in period 1, and no overage occurs at the low price in period 2. However, it this is no longer the case with aggregate demand uncertainty because different realizations of demand result in different ex-post fill rates. This complicates the analysis significantly. Hence, we will make assumptions that enable the firm to maintains its target fill rate for every realization of demand.

To do so, we assume the firm is able to reorder products in period 2 if realized demand is higher than expected and can salvage products in period 2 if demand turns out to be lower than expected. The unit purchase cost in period 2 is denoted α_2, and the unit salvage value in period 2 is s. Without loss of generality, we assume $0 < s < \alpha_1 < \alpha_2 < \beta < 1$. The firm makes a decision on its target fill rate and stocking quantity to maximize the expected profit as follows:

$$\begin{aligned}
\max_{q,C} \quad & E_N\Big\{N\bar{F}(v) + \beta N(F(v) - F(\beta))q - \alpha_1 C + s\Big[C - \Big(N\bar{F}(v) + N(F(v) - F(\beta))q\Big)\Big]^+ \\
& -\alpha_2\Big[C - \Big(N\bar{F}(v) + N(F(v) - F(\beta))q\Big)\Big]^-\Big\} \qquad (13)\\
\text{s.t.} \quad & u(v-1) = qu(v-\beta), \quad C \geq N\bar{F}(v) \;\; a.s.
\end{aligned}$$

Two features of this model are worthy of noting. First, as in the deterministic case, (13) models only the case in which market is segmented; that is, the target fill rate falls within $[0, \bar{q}]$ where $\bar{q} = \frac{u(\bar{U}-1)}{u(\bar{U}-\beta)}$. The non-segmented case (i.e. $q \in [\bar{q}, 1]$) needs to be analyzed separately, but turns out to be a standard newsvendor problem. Second, we assume there is no underage in period 1 for any realization of demand. This assumption is, of course, a bit artificial. It might be profitable for the firm to run out of stocks even in period 1. In such a situation, the model (13) does not apply and a much more complex model is required. We henceforth drop the last constraint of (13) in the following analysis, though keeping in mind that the resulting solution must be validated against this assumption. The problem (13) is solved by two-step solutions:

Step 1: For fixed $q \in [0, \bar{q}]$, solve for the capacity $C^*(q)$ that minimizes the total cost:

$$\begin{aligned}
& \min_C H(q) = \alpha_1 C + \\
& E_N\Big\{(-s)\Big[C - \Big(N\bar{F}(v) + N(F(v) - F(\beta))q\Big)\Big]^+ + \alpha_2\Big[C - \Big(N\bar{F}(v) + N(F(v) - F(\beta))q\Big)\Big]^-\Big\}\\
& \text{s.t.} \quad u(v-1) = qu(v-\beta).
\end{aligned}$$

Given q, v is uniquely determined, and hence, $H(q)$ is a standard newsvendor problem. Assuming $G(N)$ is uniformly distributed over $[\underline{D}, \bar{D}]$, then the optimal stocking quantity in period 1 has the closed form: $C^*(q) = \frac{(\alpha_2-\alpha_1)\bar{D}+(\alpha_1-s)\underline{D}}{\alpha_2-s}(\bar{F}(v(q)) + q(F(v(q)) - F(\beta)))$, and the optimal cost at $C^*(q)$ is $H(C^*(q)) = \frac{(2\alpha_1\alpha_2-\alpha_1^2-\alpha_2 s)\bar{D}+(\alpha_1^2+\alpha_2 s-2\alpha_1 s)\underline{D}}{2(\alpha_2-s)}(\bar{F}(v(q)) + q(F(v(q)) - F(\beta)))$. Note given q, the counterpart of capacity choice with certain demand, denoted $C^D(q)$, is equal to $\frac{\bar{D}+\underline{D}}{2}(\bar{F}(v(q)) + q(F(v(q)) - F(\beta)))$. It is easy to check that if $\alpha_2 - \alpha_1 \geq \alpha_1 - s$, $C^*(q) \geq C^D(q)$; otherwise, $C^*(q) < C^D(q)$. Note that $\alpha_2 - \alpha_1$ and $\alpha_1 - s$ are the unit underage cost and unit overage cost, respectively. This is intuitive; the firm stocks more inventory than in the deterministic case if underage cost is larger than overage cost, while it stocks less than in the deterministic case if underage cost is lower than overage cost.

Step 2: Substitute $H(C^*(q))$ into (13), and solve for q that maximizes the firm's profit:

$$\max_{0\leq q\leq\bar{q}} \quad \frac{\bar{D}+\underline{D}}{2}\Big((1-A)\bar{F}(v(q)) + (\beta - A)q(F(v(q)) - F(\beta))\Big) \qquad (14)$$

where $A = \frac{(\alpha_1(\alpha_2-\alpha_1)+\alpha_2(\alpha_1-s))\bar{D}+(s(\alpha_2-\alpha_1)+\alpha_1(\alpha_1-s))\underline{D}}{(\alpha_2-s)(\bar{D}+\underline{D})}$. It is easy to check $\beta - A > 0$. Therefore, the problem (14) has the same structure as the deterministic demand case if we take the deterministic market size as $\frac{\bar{D}+\underline{D}}{2}$ and the unit purchase cost in period 1 as A. Using the same approach as the deterministic demand case, we then get structurally similar results. Specifically, denote the optimal target fill rate in a segmented market by q^S, then $q^S = \left(\frac{v^S-1}{v^S-\beta}\right)^\gamma$ where v^S is the unique solution to:

$$\left(\frac{v-1}{v-\beta}\right)^\gamma \left(1 + \frac{\gamma(1-\beta)}{v-1}\right) - \frac{1-A}{\beta-A} = 0. \quad (15)$$

The case of $q \geq \bar{q}$, in which no segmentation is induced, follows a similar line of argument. In this case, it is easy to show the optimal fill rate is 1 and the optimal stocking quantity is $\frac{(\alpha_2-\alpha_1)\bar{D}+(\alpha_1-s)\underline{D}}{\alpha_2-s}\bar{F}(\beta)$, that is exactly the newsvendor solution without consideration of customer strategic behavior.

Combining the results of the segmented market and non-segmented market, the optimal stocking decision with uncertain aggregate demand is summarized in Proposition 10 below. Qualitatively, the results parallel the deterministic demand case given in Proposition 10 and Corollary 2.

Proposition 10 *Suppose v^S is the solution to (15), and denote*

$$U_c^S = \frac{1-A}{1-\beta} \cdot \frac{(\gamma(1-\beta)+\beta)v^S - \beta}{v^S - 1 + \gamma(1-\beta)} - \frac{\beta(\beta-A)}{1-\beta} \ .$$

1. *If $\bar{U} \geq U_c^S$, the optimal solution, denoted (v^*, q^*, C^*), is achieved at a segmented market: $v^* = v^S$, $q^* = \left(\frac{v^*-1}{v^*-\beta}\right)^\gamma$, and $C^* = \frac{(\alpha_2-\alpha_1)\bar{D}+(\alpha_1-s)\underline{D}}{\alpha_2-s}\Big(\bar{F}(v^*) + q^*(F(v^*) - F(\beta))\Big)$.*
2. *Otherwise, the entire market is served at the low price in period 2, that is, $v^* = \bar{U}$, $q^* = 1$, and $C^* = \frac{(\alpha_2-\alpha_1)\bar{D}+(\alpha_1-s)\underline{D}}{\alpha_2-s}\bar{F}(\beta)$.*

Remark: We need to check whether $C^* \geq \bar{D}\bar{F}(v^*)$ holds in case 1 to verify the assumption that no shortage ever occurs in period 1 for any realization of demand. If this is not true, our model no longer applies. A more complicated analysis is required.

Applying the same approach as in certain demand case, we have $1 + \gamma(\beta - A) < U_c^S < 1 + \beta - A$. Then a direct consequence of Proposition 10 is the following corollary, which establishes sufficient conditions for a rationing strategy to be optimal:

Corollary 2 *When $\bar{U} \geq 1 + \beta - A$, it is always optimal to create rationing risk; when $\bar{U} \leq 1 + \gamma(\beta - A)$, creating rationing risk is not optimal and it is best to serve the entire market at the low price.*

While qualitatively the outcome is similar to the deterministic case, in the uncertain demand case the optimal fill rate is lower than that in the certain demand case when both strategies involve rationing. This is established in Proposition 11.

Proposition 11 *If it is optimal to create rationing risk for both the cases of deterministic demand and stochastic aggregate demand, then the optimal fill rate with stochastic demand, denoted by q^S, is never more than the optimal fill rate with deterministic demand, denoted by q^D.*

This result at first looks counterintuitive. One would think that uncertainty in demand tends to create an extra incentive to purchase early due to an increased likelyhood of shortages in period 2. However, in our model, to sustain a target fill rate, the firm must bear underage or overage costs in period 2. Hence, serving customers in period 2 is more costly than in the case of deterministic demand, and the firm has an additional incentive to induce more customers to purchase early. It therefore selects a lower fill rate.

Also note that according to (15), v^S increases in the salvage value s, while it decreases in the second period ordering cost α_2. Intuitively, the larger the salvage value, the lower the overage cost and the lower the reordering cost, the lower the underage cost. Both lead to larger cutoff values and higher fill rates. In the limiting case, that is, $s \to \alpha_1^-$ or $\alpha_2 \to \alpha_1^+$, then $A \to \alpha_1$, that implies $v^S = v^D$. This makes a perfect sense, because when there are no overage or shortage costs, the problem is, of course, equivalent to the certain demand case.

4.3 Oligopolistic Competition

When the same product is carried by multiple retailers, it is unlikely that customers base their strategic behavior on the product's availability in any one store. Rather, it is more plausible that they assess availability across the entire market. In doing so, they will consider the aggregate supply and aggregate demand among all stores in the market. Here, we analyze a model of this situation.

Consider an oligopoly market of n firms providing the same product. We assume customers exhibit no preference over the source of supply and with equal probability, a customer buys a product from any firm as long as there is inventory. Capacity choice in the market is a vector, denoted by $C = (C_1, \cdots, C_i, \cdots, C_n)$. All the other notation is the same as in a monopoly case. Moreover, three assumptions are made to simply the analysis of stocking decisions under competition: i) Shortage never happens at the high price in period 1, and the potential demand exceeds the total supply, that is, $N\bar{F}(\beta) \geq \sum_{i=1}^{n} C_i \geq N\bar{F}(1)$; ii) Sales in period 1 are equally shared by all firms; and iii) A buyer will try other firms in period 2 if the firm he initially selects is out of stock until his request is accepted or the market supply is exhausted. The second and third assumptions are direct consequences of the assumption that customers randomly select suppliers and they have no preference of one specific supplier over another. These assumptions are reasonable in a commodity market.

Using the same argument as in the monopoly case, one can show that the aggregate equilibrium capacity is always greater than the potential demand at the high price and less than the potential demand at the low price. Hence, the aggregate fill rate is determined by $q = \frac{\sum_{i=1}^{n} C_i - N\bar{F}(v)}{N(F(v)-F(\beta))}$. Note that each firm's capacity choice contributes to aggregate fill rate and thus impacts not only its own market share, but also that of all its competitors. This creates a strategic interaction among the stocking decisions of the n firms in the market. Determining optimal stocking quantities under competition therefore becomes a problem of finding equilibria in their capacity choices. Given the perfect symmetry among firms, in what follows we focus only on symmetric equilibria.

Given the capacity choices of other firms', denoted $C_{-i} = (C_1, \cdots, C_{i-1}, C_{i+1}, \cdots, C_n)$, firm i maximizes its total profit, which consists of an equally-shared extra margin on first period sales and its own base margin of $\beta - \alpha_1$ on each unit stocked:

$$\begin{aligned} \max \quad & \Pi_i(C_i, C_{-i}) = \frac{N\bar{F}(v)}{n}(1-\beta) + (\beta - \alpha_1)C_i \qquad (16) \\ \text{s.t.} \quad & u(v-1) = \frac{\sum_{i=1}^{n} C_i - N\bar{F}(v)}{N(F(v) - F(\beta))} u(v-\beta), \quad v \geq 1. \end{aligned}$$

Again, to simplify the analysis we assume that customers' valuations are drawn from a uniform distribution and customers have power utility functions.

4.3.1 Risk Neutral Buyers

Analyzing (16) for the case of risk neutral customers we obtain the following proposition characterizing the symmetric equilibria:

Proposition 12 *If $\bar{U} \leq n(1-\alpha_1)+\beta$, a symmetric low-price-only Nash equilibrium exists; that is, no rationing risk is created and the entire market is served only at the low price. Namely, $v^* = \bar{U}$, $q^* = 1$ and $C_i^* = \frac{N}{n\bar{U}}(\bar{U} - \beta)$, $\forall\ i = 1, \cdots, n$. If $\bar{U} \geq 1 + n(\beta - \alpha_1)$, a symmetric high-price-only Nash equilibrium exists; that is, the market is served only at the high price. Namely, $v^* = 1$, $q^* = 0$ and $C_i^* = \frac{N}{n\bar{U}}(\bar{U} - 1)$, $\forall\ i = 1, \cdots, n$.*

Note in an oligopoly market ($n > 1$), it is always true that $\beta + n(1-\alpha_1) > 1 + n(\beta - \alpha_1)$. Hence, multiple equilibria may exist. In particular, when $\bar{U} > \beta + n(1-\alpha_1)$, the symmetric Nash equilibrium is uniquely attained at a high-price-only market; when $\bar{U} < 1 + n(\beta - \alpha_1)$, the unique symmetric Nash equilibrium is a low-price-only strategy; otherwise, both equilibria are attainable outcomes.

4.3.2 Risk Averse Buyers

Facing risk averse customers who have power utility functions, one can show that given C_{-i}, the optimization problem (16) can be expressed only in terms of v:

$$\max_{\bar{U}\geq v\geq 1} \quad \Pi_i(v) = \frac{N}{\bar{U}}\left(\beta-\alpha_1+\frac{1-\beta}{n}\right)(\bar{U}-v)+\frac{N}{\bar{U}}(\beta-\alpha_1)(v-\beta)\left(\frac{v-1}{v-\beta}\right)^{\gamma}-(\beta-\alpha_1)\sum_{j=1,j\neq i}^{n} C_j \, .$$

Again, it is easy to show that firm $i's$ profit function, $\Pi_i(v)$, is strictly concave in $v \geq 1$. The first order conditions yield:

$$\left(1+\frac{\gamma(1-\beta)}{v-1}\right)\left(\frac{v-1}{v-\beta}\right)^{\gamma} - \left(1+\frac{1-\beta}{n(\beta-\alpha_1)}\right) = 0. \tag{17}$$

Using the same argument as in (8), we conclude there exists a unique solution $v^0 > 1$ to (17).

A precise characterization of symmetric Nash equilibria is presented in Proposition 13 below. In a nutshell, as in the monopoly case, the equilibrium involves rationing and segmentation if the market consists of a sufficiently large high-value population; while the equilibrium has no rationing or segmentation if the market has a small number of high-value customers. However, both equilibria may be supportable.

Proposition 13 *v^0 is the solution to (17), and denote*

$$q^0 = \left(\frac{v^0-1}{v^0-\beta}\right)^{\gamma}, \qquad U_c = \frac{n(\beta-\alpha_1)(v^0-\beta)(1-q^0)}{1-\beta} + v^0 \, .$$

1. *If $\bar{U} \geq U_c$, there exists a symmetric segmented Nash equilibrium; namely, $v^* = v^0$, $q^* = q^0$ and $C_i^* = \frac{N}{n\bar{U}}(\bar{U} - v^0 + (v^0-\beta)q^0)$, $\forall\ i = 1, \cdots, n$.*
2. *If $\bar{U} \leq U_c$ or $\left(\frac{\bar{U}-1}{\bar{U}-\beta}\right)^{\gamma} \leq 1 - \frac{1}{n}$ or $q^0 < 1 - \frac{1}{n} < \left(\frac{\bar{U}-1}{\bar{U}-\beta}\right)^{\gamma}$ and $\bar{U} > U_c$, there exists a symmetric low-price-only Nash equilibrium; namely, $v^* = \bar{U}$, $q^* = 1$ and $C_i^* = \frac{N}{n\bar{U}}(\bar{U}-\beta)$, $\forall\ i = 1, \cdots, n$.*

Sufficient conditions of uniqueness of symmetric equilibrium is provided in the following Corollary:

Corollary 3 *If $\bar{U} \geq 1 + n(\beta-\alpha_1)$ and $q^0 \geq 1 - \frac{1}{n}$, the symmetric Nash equilibrium exists uniquely at a segmented market. If $\bar{U} \leq 1 + n\gamma(\beta-\alpha_1)$, the symmetric Nash equilibrium exists uniquely at a low-price-only solution.*

Intuitively, competition should make a segmentation strategy more difficult to sustain. This is because with large numbers of competitors, restricting supply has only a negligible impact on the overall market availability but the lost-sales cost of rationing is incurred entirely by firms that are restricting their supply. Indeed, the following corollary shows that there exists a critical number of firms beyond which creating rationing risk is never as sustainable equilibrium. This implies increased competition eventually eliminates the industry's ability to support segmentation via rationing.

Corollary 4 *When the number of firms n becomes sufficiently large, specifically,* $n \geq \frac{\bar{U}-1}{\gamma(\beta-\alpha_1)}$*, there exists the unique symmetric low-price-only Nash equilibrium.*

One can also compare the outcomes of the oligopoly market with those in the monopoly market. The first difference is that more competition leads to higher aggregate capacity and higher fill rates relative to the monopoly case and these differences increase in the level of competition. However, firms generate lower aggregate profits compared to the monopoly market under more competition, and the difference increases in the level of competition as well.

Proposition 14 *The optimal aggregate fill rate, cutoff value, and aggregate capacity in oligopoly market are larger than those in the monopoly market, while the aggregate profit obtained in oligopoly market is less than that in the monopoly market. Moreover, the optimal aggregate capacity, cutoff value and fill rate are all increasing in the number of competing firms n; however, the optimal aggregate profit is decreasing in n.*

4.3.3 Numerical Examples

We next show some numerical examples that illustrate the impact of competition over combinations of the following parameter ranges of $\alpha_1 \in \{0.2, 0.5, 0.8\}$, $\bar{U} \in \{1.25, 1.5, 1.75, 2\}$ and $\gamma = 0.5$. The examples show that a symmetric segmented Nash equilibrium is more likely to be attained when the marketplace has a small number of competing suppliers. Second, as the number of high-value customers increases (i.e. $\bar{U}$ becomes larger), the equilibrium tends to create rationing risk. For example, at $\alpha_1 = 0.8$ and $\bar{U} = 2$, a segmented Nash equlibrium exists uniquely even at $n = 10$. Figure 5 illustrates how the number of suppliers influences the optimal aggregate fill rate in the case of $\alpha_1 = 0.5$ and $\bar{U} = 2$.

5 Conclusions

Our model shows that rationing can be a profitable strategy to influence the strategic behavior of customers. It also provides a behavioral explanation for stocking and inventory service level decisions that are normally explained in terms of holding and lost-sales cost tradeoffs. In our case, the trade-off is between the benefits of inducing customers to purchase early at high prices and the cost of lost sales due to rationing. Our assumptions - a large market of customers who are strategic and risk averse, the ability of the firm to commit

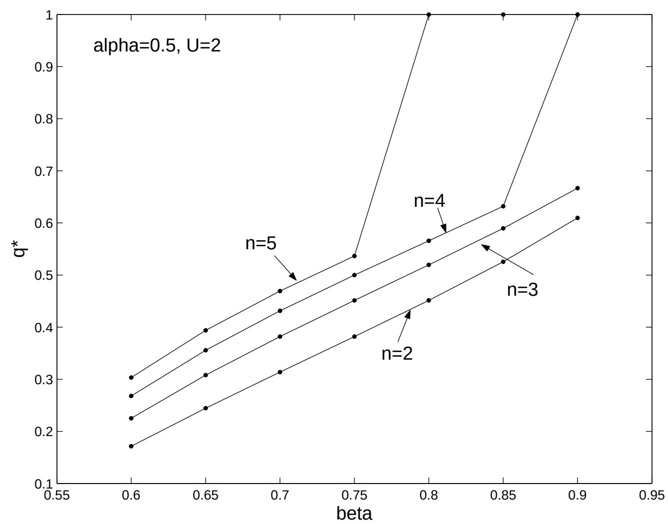

Figure 5: Optimal aggregate fill rates under competition

to prices and quantities, limited capacity and a finite selling season - are reasonable as a stylized model for big-ticket, seasonal and durable goods; such as autos, sporting goods, apparel and consumer electronics.

Under our assumptions, rationing is not profitable when customers are risk neutral. But even when customers are risk averse, rationing may not be optimal if the number of high-valuation customers is too small. In general, a large high-value customer segment, high levels of risk aversion and large differences in price over time all tend to favor rationing as an optimal strategy. Numerical examples suggest that rationing risk can be even more important than price in terms of influencing strategic customer behavior. Our analysis of aggregate uncertainty shows that demand uncertainty decreases the optimal fill rate because it makes serving customers in the markdown period more costly. Lastly, our oligopoly analysis shows that competition makes it more difficult to support segmentation using rationing, and when the number of competitors is sufficiently large, a low-price-only equilibrium is the only sustainable outcome. Thus, rationing is more likely to be used in cases where a firm has some reasonable degree of market power.

As for future work, we have a forthcoming paper investigating the case where customers do not perfectly anticipate fill rates but rather update their estimates over time based on experience. The issue here is understanding how the firm should respond to most profitably influence these expectations over time, whether the market converges to an equilibrium and, if so, how the equilibrium is related to the rational-expectations outcome derived here.

References

[1] J. Assuncao and R. J. Meyer. The rational effect of price promotions on sales and consumption. *Management Science*, 39:517–535, 1993.

[2] Y. Aviv and A. Pazgal. Optimal pricing of seasonal products in the presence of forward-looking consumers. Working Paper, Olin School of Business, Washington University, 2003.

[3] D. R. Bell, G. Iyer, and V. Padmanabhan. Price competition under stockpiling and flexible consumption. *Journal of Marketing Research*, 39:292–303, 2002.

[4] D. Bensanko and W. L. Winston. Optimal price skimming by a monopolist facing rational consumers. *Management Science*, 36:555–567, 1990.

[5] J. Bulow and P. Klemperer. Rational frenzies and crashes. *Journal of Political Economy*, 102:1–23, 1994.

[6] J. I. Bulow. Durable-goods monopolists. *Journal of Political Economy*, 90:314–332, 1982.

[7] X. Chen and D. Simchi-Levi. Coordinating inventory control and pricing strategies with random demand and fixed ordering costs: The finite horizon case. Technical report, MIT, Cambridge, MA, 2002. Working Paper.

[8] R. H. Coase. Durability and monopoly. *Journal of Law and Economics*, 15:143–149, 1972.

[9] J. D. Dana and N. C. Petruzzi. The newsvendor model with endogenous demand. Working Paper, Kellog Graduate School of Management, Northwestern University, 2001.

[10] L. Eeckhoudt, C. Gollier, and T. Schneider. Risk-aversion, rrudence and temperance: A unified approach. *Economics Letters*, 48:331–336, 1995.

[11] W. Elmaghraby, A. Gülcü, and P. Keskinocak. Optimal markdown mechanisms in the presence of rational customers with multi-unit demands. Working Paper, School of Industrial and Systems Engineering, Georgia Institute of Technology, 2004.

[12] A. Federgruen and A. Heching. Combined pricing and inventory control under uncertainty. *Operations Research*, 47:454–475, 1999.

[13] Y. Feng and G. Gallego. Optimal starting times for end-of-season sales and optimal stopping times for promotional fares. *Management Science*, 41:1371–1391, 1995.

[14] K. Ferdows, M. A. Lewis, and J. A. D. Machuca. Zara's secret for fast fashion. http://hbswk.hbs.edu/tools. **F**ebruary 21, 2005.

[15] G. Gallego and G. J. van Ryzin. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Science*, 40:999–1020, 1994.

[16] J. Gallien. Dynamic mechanism design for online commerce. Working Paper, Sloan School of Management, Massachusetts Institute of Technology, 2004.

[17] S. Goodin. The scarcity shortage. *FastCompany*, page 121, December 2003.

[18] M. Harris and A. Raviv. A theory of monopoly pricing schemes with demand uncertainity. *American Economic Review*, 71:347–365, 1981.

[19] T. H. Ho, C. S. Tang, and D. R. Bell. Rational shopping behavior and the option value of variable pricing. *Management Science*, 44:145–160, 1998.

[20] O. Koenigsberg, E. Muller, and N. J. Vilcassim. easyjet airlines: Small, lean, and with prices that increase over time. Working Paper, Graduate School of Business, Columbia University, 2004.

[21] R. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6:58–73, 1981.

[22] M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. The MIT Press, Cambridge, MA, 1995.

[23] T. Rozhon. Before christmas, wal-mart was stirring. http://www.nytimes.com/. **J**anuary 5, 2005.

[24] X. Su. Inter-temporal pricing with strategic customer behavior. Working Paper, Haas School of Business, University of California, 2005.

[25] K. T. Talluri and G. J. van Ryzin. *The Theory and Practice of Revenue Management*. Kluwer, New York, NY, 2004.

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

[27] N-H. M. von der Fehr and K-U. Kühn. Coase versus Pacman: Who eats whom in the durable-goods monopoly? *Journal of Political Economy*, 103:785–812, 1995.

[28] Y. Zhou, M. Fan, and M. Cho. Online purchasing strategies for dynamically priced perishable products. Working Paper, Department of Management Science, University of Washington Business School, 2004.

On-Line Appendix:
Strategic Capacity Rationing to Induce Early Purchases

Qian Liu Garrett van Ryzin

Proof of Proposition 1

Proof
Obviously, $v(q) \geq p_1$, $\forall\ q$, because only customers with valuations greater than p_1 possibly buy in period 1. Denote $h(v) = u(v-p_1)/u(v-p_2)$, $v \geq p_1$, then

$$h'(v) = \frac{u'(v-p_1)u(v-p_2) - u'(v-p_2)u(v-p_1)}{u^2(v-p_2)} > 0.$$

The inequality holds because $u(\cdot)$ is strict increasing and concave, thus $u(v-p_2) > u(v-p_1) \geq 0$ and $u'(v-p_1) \geq u'(v-p_2) > 0$ if $p_1 > p_2$. Furthermore, note that $h(p_1) = 0$ and $\lim_{v\to+\infty} h(v) = 1$, hence, $\forall\ 0 \leq q < 1$, $h(v)$ crosses q only once. □

Proof of Proposition 2

Proof
The threshold function $v(q)$ satisfies:

$$H(v(q), q) = \frac{u(v(q)-1)}{u(v(q)-\beta)} - q = 0.$$

By the Implicit Function Theorem,

$$\frac{dv}{dq} = -\frac{\frac{\partial H(v,q)}{\partial q}}{\frac{\partial H(v,q)}{\partial v}} = \frac{u^2(v-\beta)}{u'(v-1)u(v-\beta) - u(v-1)u'(v-\beta)} > 0.$$

The inequality holds because $u(\cdot)$ is a strictly increasing and concave function.

By the Inverse Function Theorem,

$$\frac{dq}{dv} = \left(\frac{dv}{dq}\right)^{-1} = \frac{u'(v-1)u(v-\beta) - u(v-1)u'(v-\beta)}{u^2(v-\beta)}.$$

Then,

$$\frac{d^2q}{dv^2} = \frac{u''(v-1)u(v-\beta) - u''(v-\beta)u(v-1)}{u^2(v-\beta)} + \frac{2u'(v-\beta)}{u^3(v-\beta)}\Big(u(v-1)u'(v-\beta) - u'(v-1)u(v-\beta)\Big). \quad (18)$$

Since the third derivative of $u(\cdot)$ is nonnegative, $0 \geq u''(v-\beta) \geq u''(v-1)$, and hence, the first term of L.H.S. of (18) is less than 0. Together with the nonnegative second term of

L.H.S. of (18), $\frac{d^2q}{dv^2} \leq 0$. It then follows that $v(q)$, the inverse function of $q(v)$, is convex. This is because the inverse function of a strictly increasing and concave function is strictly increasing and convex. [2] $\square$

Proof of Proposition 3

Proof

At $C = \frac{N}{\bar{U}}(\bar{U}-1)$, the first constraint in (5) holds for any $v \in [1, \bar{U}]$. Then regard to the profit function, the optimal cutoff value is equal to 1. That implies all potential customers at the high price buy in period 1. Correspondingly, the optimal fill rate is 0, and the associated profit is equal to $\frac{N}{\bar{U}}(\bar{U}-1)(1-\alpha_1)$.

Once $C > \frac{N}{\bar{U}}(\bar{U}-1)$, for any $v \in [1, \bar{U}]$, we have $v - 1 < \frac{C - \frac{N}{\bar{U}}(\bar{U}-v)}{\frac{N}{\bar{U}}(v-\beta)}(v-\beta)$. That means all rational customers wait to buy at the low price in period 2. Consequently, the firm stocks $\frac{N}{\bar{U}}(\bar{U}-\beta)$ units to serve the entire market. Then, the optimal fill rate is 1, the optimal cutoff value is $\bar{U}$, and the associated profit equals to $\frac{N}{\bar{U}}(\bar{U}-\beta)(\beta-\alpha_1)$.

Therefore, the profit is maximized at either zero fill rate or one fill rate. When $\bar{U} \geq 1+\beta-\alpha_1$, the profit earned at zero fill rate $\frac{N}{\bar{U}}(\bar{U}-1)(1-\alpha_1)$ is greater than the profit at one fill rate $\frac{N}{\bar{U}}(\bar{U}-\beta)(\beta-\alpha_1)$; otherwise, the profit at one fill rate is larger. $\square$

Proof of Lemma 1

Proof

The first and second derivatives of $\Pi(v)$ are the following:

$$\begin{aligned}
\Pi'(v) &= \frac{N}{\bar{U}}\left((\beta-\alpha_1)\left(1+\frac{\gamma(1-\beta)}{v-1}\right)\left(\frac{v-1}{v-\beta}\right)^{\gamma} - (1-\alpha_1)\right), \\
\Pi''(v) &= -\frac{N}{\bar{U}}(\beta-\alpha_1)\gamma(1-\gamma)(1-\beta)^2\frac{1}{(v-1)^2(v-\beta)}\left(\frac{v-1}{v-\beta}\right)^{\gamma} \leq 0.
\end{aligned}$$

Therefore, $\Pi(v)$ is strictly concave if $v > 1$. Note that $\Pi(v)$ is continuous at $v = 1$, hence, $\Pi(v)$ is strictly concave in $v \geq 1$.

Further, by the L-Hospital's rule, $\lim_{v\to 1^+}\Pi'(v) = +\infty$. So, the solution to the F.O.C. equation of $\Pi(v)$, denoted by v^0, must be strictly greater than 1. If $v^0 \leq \bar{U}$, v^0 maximizes $\Pi(v)$; otherwise, $\bar{U}$ is optimal. $\square$

[2] The brief reason is the following: suppose $h(\cdot)$ is strictly increasing and concave, and $g(\cdot)$ is the inverse function of $h(\cdot)$. Then, $g'(x) = \frac{1}{h'(g(x))} > 0$, and $g''(x) = \frac{-h''((g(x))g'(x)}{(h'(g(x)))^2} \geq 0$ due to $g'(x) > 0$ and $h''(g(x)) \leq 0$.

Proof of Proposition 4

Proof
Denote

$$\begin{aligned}\Pi^0 &= \frac{N}{\bar{U}}\Big((1-\alpha_1)(\bar{U}-v^0)+(\beta-\alpha_1)(v^0-\beta)\Big(\frac{v^0-1}{v^0-\beta}\Big)^{\gamma}\Big),\\ \Pi^{NS} &= \frac{N}{\bar{U}}(\beta-\alpha_1)(\bar{U}-\beta);\end{aligned}$$

Π^0 and Π^{NS} are the firm's profits at a segmented market if $v^0 \leq \bar{U}$ and at a low-price-only market, respectively.

Regarding the profit optimization problem for a segmented market defined by (6), it is maximized at (v^0, C^0) if $v^0 \leq \bar{U}$; otherwise, $(\bar{U}, C_s)$ maximizes (6) where $C_s = \frac{N}{\bar{U}}(\bar{U} - \beta)\Big(\frac{\bar{U}-1}{\bar{U}-\beta}\Big)^{\gamma}$.

Note the profit at $(\bar{U}, C_s)$ is strictly dominated by at $(\bar{U}, \bar{C})$ where $\bar{C} = \frac{N}{\bar{U}}(\bar{U}-\beta)$, hence, a segmented market, characterized by (v^0, C^0), is the optimal stocking decision if $v^0 \leq \bar{U}$ and $\Pi^0 \geq \Pi^{NS}$; otherwise, the firm stocks $\bar{C}$ to serve the entire market at a low price.

By some algebraic calculation,

$$\Pi^0 \geq \Pi^{NS} \quad \Leftrightarrow \quad \bar{U} \geq \frac{(\beta+\gamma(1-\alpha_1))v^0-\beta(1+\gamma(\beta-\alpha_1))}{v^0-1+\gamma(1-\beta)} = U_c\ .$$

Then the left is to show $\bar{U} \geq \max\{U_c, v_0\} = U_c$. We claim that the solution v_0 to (8) is strictly less than $1+\gamma(\beta-\alpha_1)$. This is because the L.H.S. of (8) is strictly decreasing in $v > 1$ and the value at $v = 1+\gamma(\beta-\alpha_1)$ is equal to $\frac{1-\alpha_1}{\beta-\alpha_1}\Big(\Big(\frac{\gamma(\beta-\alpha_1)}{\gamma(\beta-\alpha_1)+1-\beta}\Big)^{\gamma}-1\Big) < 0$. Furthermore, it is easy to check that U_c strictly declines in v^0. Then $1+\gamma(\beta-\alpha_1) < U_c < 1+\beta-\alpha_1$ follows with $1 < v^0 < 1+\gamma(\beta-\alpha_1)$. Therefore, $U_c > v_0$. □

Proof of Proposition 5

Proof
In a segmented market, the stocking quantity is determined by:

$$C = \frac{N}{\bar{U}}\Big(\bar{U} - v + \Big(\frac{v-1}{v-\beta}\Big)^{\gamma}(v-\beta)\Big). \tag{19}$$

Then,

$$\frac{dC}{dv} = \frac{N}{\bar{U}}\Big(\Big(1+\frac{\gamma(1-\beta)}{v-1}\Big)\Big(\frac{v-1}{v-\beta}\Big)^{\gamma}-1\Big),$$

$$\frac{d^2C}{dv^2} = -\frac{N}{\bar{U}}\gamma(1-\gamma)(1-\beta)^2(v-\beta)^{-1}(v-1)^{-2}\left(\frac{v-1}{v-\beta}\right)^{\gamma} \leq 0.$$

Therefore, $\frac{dC}{dv}$ decreases in v, together with $\lim_{v\to+\infty}\frac{dC}{dv} = 0$, we have $\frac{dC}{dv} \geq 0$, that is, C increases in v. Since v increases in q, so does C.

In a non-segmented market, $q \in [\bar{q}, 1]$, $v \equiv \bar{U}$, and $C = \frac{N}{\bar{U}}(\bar{U}-\beta)q$ where $\bar{q} = \left(\frac{\bar{U}-1}{\bar{U}-\beta}\right)^{\gamma}$. Obviously, C increases in v and q. Together with continuity of C at $\bar{q}$, C increases in both v and q over the entire range. □

Proof of Proposition 6

Proof

Substitute $v = \frac{1-\beta q^{\frac{1}{\gamma}}}{1-q^{\frac{1}{\gamma}}}$ into (8), we get

$$q^{1-\frac{1}{\gamma}} + \left(\frac{1}{\gamma}-1\right)q - \frac{1-\alpha_1}{\gamma(\beta-\alpha_1)} = 0.$$

The L.H.S. of the above equation, denoted by $G(q)$, is strictly decreasing in q when $0 < q < 1$ because

$$G'(q) = \left(\frac{1}{\gamma}-1\right)\left(1-q^{-\frac{1}{\gamma}}\right) < 0.$$

Furthermore, $G(0^+) \to +\infty$; $G(1^-) \to< 0$. Therefore, there exists the unique $q^0 \in (0,1)$ such that $G(q^0) = 0$. Because $G(q)$ strictly decreases in q, q^0 strictly increases in β.

Let Π^0 be the profit at fill rate q^0 if $q^0 \leq \bar{q}$ and $\bar{q} = \left(\frac{\bar{U}-1}{\bar{U}-\beta}\right)^{\gamma}$; and Π^{NS} be the profit at one fill rate. Then,

$$\Pi^0 = \frac{N}{\bar{U}\left(1-(q^0)^{\frac{1}{\gamma}}\right)}\left((1-\alpha_1)(\bar{U}-1) - (1-\alpha_1)(\bar{U}-\beta)(q^0)^{\frac{1}{\gamma}} + (\beta-\alpha_1)(1-\beta)q^0\right),$$

$$\Pi^{NS} = \frac{N}{\bar{U}}(\beta-\alpha_1)(\bar{U}-\beta).$$

The optimal fill rate q^* is equal to q^0 if $q^0 \leq \bar{q}$ and $\Pi^0 \geq \Pi^{NS}$; otherwise, $q^* = 1$.

Since q^0 strictly increases in β, that is, $\forall\ \beta_1 < \beta_2$, then $q^0(\beta_1) < q^0(\beta_2)$, the left work is to show if $q^0(\beta_1) < q^0(\beta_2) \leq \bar{q}$ and $q^*(\beta_1) = 1$, then it must be $q^*(\beta_2) = 1$ as well.

After some algebraic arrangements, $\Pi^0 \geq \Pi^{NS}$ is equivalent to

$$\frac{\bar{U}-\alpha_1}{1-\alpha_1} \geq \frac{1}{\gamma(q^0)^{1-\frac{1}{\gamma}} + (1-\gamma)q^0} + \frac{1}{(\frac{1}{\gamma}-1)(q^0)^{\frac{1}{\gamma}}+1}.$$

Denote the R.H.S. of the above inequality as $T(q)$ (we drop the superscript 0 of q to simplify notation), and

$$T'(q) = \frac{(1-\gamma)(1-q-q^{\frac{1}{\gamma}})}{q^{\frac{1}{\gamma}+2}(\gamma q^{-\frac{1}{\gamma}}+1-\gamma)^2} \; .$$

It is easy to check that $1-q-q^{\frac{1}{\gamma}}$ strictly decreases in $0 \le q \le 1$, and have different signs at $q=0$ and $q=1$. Therefore, there exists the unique $0 < \tilde{q} < 1$, such that $T'(q) \ge 0$ if $q \in [0,\tilde{q}]$ and $T'(q) < 0$ otherwise. It follows that $T(q)$ is a unimodal function maximized at $\tilde{q}$, that is, $T(q)$ increases in q when $q \in [0,\tilde{q}]$ while decreases in q afterwards.

Since $q^0(\beta_1) \le \bar{q}$ and $q^*(\beta_1) = 1$, $\frac{\bar{U}-\alpha_1}{1-\alpha_1} < T(q^0(\beta_1))$. Note also $\frac{\bar{U}-\alpha_1}{1-\alpha_1} < T(\bar{q})$ because the profit gained at $\bar{q}$ is strictly less than Π^{NS}. Together with the fact that $T(q)$ is unimodal in q, it must be $\frac{\bar{U}-\alpha_1}{1-\alpha_1} < T(q^0(\beta_2))$. Hence, $q^*(\beta_2) = 1$. □

Proof of Proposition 7

We construct an equivalent model and use the parametric supermodularity to show q^* is decreasing in γ if $q^* \le \bar{q}$ and $\bar{q} = \left(\frac{\bar{U}-1}{\bar{U}-\beta}\right)^{\gamma}$. We require the following lemma to prove Proposition 7:

Lemma 2 *The following problem (20) is equivalent to the one replacing the first constraint in (20) with the binding constraint.*

$$\begin{aligned} \max \quad & \frac{N}{\bar{U}}(1-\alpha_1)(\bar{U}-v) + \frac{N}{\bar{U}}(\beta-\alpha_1)(v-\beta)q \qquad (20)\\ s.t.& \\ (v-1)^{\gamma} &\ge q(v-\beta)^{\gamma}, \\ \bar{U} &\ge v \ge 1, \\ \bar{q} &\ge q \ge 0. \end{aligned}$$

<u>Proof</u>
Suppose (q^*, v^*) and $f(q^*, v^*)$ are the optimal solution and associated objective value to (20). If $q^* = \bar{q}$, then $v^* = \bar{U}$, and hence the first constraint in (20) holds with equality. If $q^* < \bar{q}$, suppose the first constraint in (20) is not binding at (q^*, v^*). Then there exists $\epsilon > 0$ such that the first constraint in (20) still holds at $(q^* + \epsilon, v^*)$. However, $f(q^* + \epsilon, v^*) > f(q^*, v^*)$, which contradicts the optimality of (q^*, v^*). □

Lemma 2 allows us to examine the equivalent model (20) to show the parametric supermodularity result as required.

Suppose (q^*, v^*) is the optimal solution to (20), we now prove that (q^*, v^*) decreases in γ. Define

$$
\begin{aligned}
X &= \{(q,v) \mid 0 \le q \le \bar{q},\ 1 \le v \le \bar{U},\ 0 < \gamma < 1\}, \\
\Theta &= \{\theta \mid \theta = \frac{1}{\gamma},\ 0 < \gamma < 1\}, \\
S_\theta &= \{(q,v) \mid (q,v) \in X,\ (v-1)^\gamma \ge q(v-\beta)^\gamma,\ 0 < \gamma < 1\}.
\end{aligned}
$$

Take the cross partial derivatives of the objective function in (20),

$$\frac{\partial^2 f(v,q)}{\partial q \partial v} = \frac{N}{\bar{U}}(\beta - \alpha_1) > 0.$$

In addition, X is a sublattice of R^2. Hence, $f(q,v)$ is supermodular in (q,v). Because $\left(\frac{v-1}{v-\beta}\right)^{\frac{1}{\theta}}$ increases in θ for every fixed v, the set

$$\left\{(q,v) \mid \left(\frac{v-1}{v-\beta}\right)^{\frac{1}{\theta}} \ge q\right\}$$

is enlarged as θ increases. Therefore, S_θ increases in θ. By Theorem 2.8.1 in Topkis [26], (q^*, v^*) decreases in γ.

Proof of Proposition 8

<u>Proof</u>
v^0 is the unique root of the equation:

$$\left(\frac{v-1}{v-\beta}\right)^\gamma \left(1 + \frac{\gamma(1-\beta)}{v-1}\right) - \frac{1-\alpha_1}{\beta-\alpha_1} = 0. \tag{21}$$

At $v = \gamma(\beta - \alpha_1) + 1$,

$$
\begin{aligned}
& \lim_{\gamma \to 0^+} \left(\frac{v-1}{v-\beta}\right)^\gamma \left(1 + \frac{\gamma(1-\beta)}{v-1}\right) \\
= & \frac{1-\alpha_1}{\beta-\alpha_1} \lim_{\gamma \to 0^+} \left(\frac{\gamma(\beta-\alpha_1)}{\gamma(\beta-\alpha_1)+1-\beta}\right)^\gamma \\
= & \frac{1-\alpha_1}{\beta-\alpha_1} \lim_{\gamma \to 0^+} \gamma^\gamma \\
= & \frac{1-\alpha_1}{\beta-\alpha_1}\ .
\end{aligned}
$$

Therefore, $v = \gamma(\beta - \alpha_1) + 1$ is indeed the solution to (21) as $\gamma \to 0^+$. Furthermore, $\lim_{\gamma \to 0^+} v^0 = 1^+$ and

$$\lim_{\gamma \to 0^+} q^0 = \lim_{\gamma \to 0^+} \left(\frac{v^0-1}{v^0-\beta}\right)^\gamma = \lim_{\gamma \to 0^+} \left(\frac{\gamma(\beta-\alpha_1)}{\gamma(\beta-\alpha_1)+1-\beta}\right)^\gamma = \lim_{\gamma \to 0^+} \gamma^\gamma = 1^-\ .$$

From the proof of Proposition 6, q^0 satisfies:

$$q^{1-\frac{1}{\gamma}} + \left(\frac{1}{\gamma} - 1\right)q - \frac{1-\alpha_1}{\gamma(\beta-\alpha_1)} = 0. \tag{22}$$

At $q = \left(\frac{1-\alpha_1}{\beta-\alpha_1}\right)^{-\frac{\gamma}{1-\gamma}}$, the L.H.S. of the above equation as $\gamma \to 1$ is the following:

$$\begin{aligned}
& \lim_{\gamma\to 1^-} q^{1-\frac{1}{\gamma}} + \left(\frac{1}{\gamma} - 1\right)q - \frac{1-\alpha_1}{\gamma(\beta-\alpha_1)} \\
= & \lim_{\gamma\to 1^-} \frac{1-\alpha_1}{\beta-\alpha_1} + \left(\frac{1}{\gamma} - 1\right)\left(\frac{1-\alpha_1}{\beta-\alpha_1}\right)^{-\frac{\gamma}{1-\gamma}} - \frac{1-\alpha_1}{\gamma(\beta-\alpha_1)} \\
= & \lim_{\gamma\to 1^-} \left(\frac{1}{\gamma} - 1\right)\left(\frac{1-\alpha_1}{\beta-\alpha_1}\right)^{-\frac{\gamma}{1-\gamma}} \\
= & \ 0.
\end{aligned}$$

Hence, $q = \left(\frac{1-\alpha_1}{\beta-\alpha_1}\right)^{-\frac{\gamma}{1-\gamma}}$ is the solution to (22) as $\gamma \to 1^-$. Moreover,

$$\begin{aligned}
\lim_{\gamma\to 1^-} q^0 &= \lim_{\gamma\to 1^-} \left(\frac{1-\alpha_1}{\beta-\alpha_1}\right)^{-\frac{\gamma}{1-\gamma}} = 0^+ \ , \\
\lim_{\gamma\to 1^-} v^0 &= \lim_{\gamma\to 1^-} \frac{1-\beta(q^0)^{\frac{1}{\gamma}}}{1-(q^0)^{\frac{1}{\gamma}}} = \lim_{\gamma\to 1^-} \frac{1-\beta\left(\frac{1-\alpha_1}{\beta-\alpha_1}\right)^{-\frac{1}{1-\gamma}}}{1-\left(\frac{1-\alpha_1}{\beta-\alpha_1}\right)^{-\frac{1}{1-\gamma}}} = 1^+ \ .
\end{aligned}$$

□

Proof of Proposition 9

Proof

By Proposition 4, v^0 is the optimal cutoff value if $\bar{U} \geq U_c$. Substitute $v^0 = \frac{1-\beta(q^0)^{\frac{1}{\gamma}}}{1-(q^0)^{\frac{1}{\gamma}}}$ into U_c,

$$\bar{U} \geq \beta + \frac{1-\alpha_1}{(\frac{1}{\gamma}-1)(q^0)^{\frac{1}{\gamma}}+1} \ , \tag{23}$$

which is equivalent to

$$q^0 \geq \left(\frac{\gamma(1+\beta-\alpha_1-\bar{U})}{(\bar{U}-\beta)(1-\gamma)}\right)^{\gamma} = \hat{q}.$$

So, $q^* = q^0$ if $q^0 \geq \hat{q}$; otherwise, $q^* = 1$.

From the proof of Proposition 7, v^0 decreases in γ. Then $(q^0)^{\frac{1}{\gamma}} = \frac{v^0-1}{v^0-\beta}$ decreases in γ as well. It follows with the R.H.S. of (23) increases in γ. Therefore, once there exists $\hat{\gamma}$

such that (23) is violated, (23) is also violated for any $\gamma > \hat{\gamma}$; namely, if $q^*(\hat{\gamma}) = 1$, then $q^*(\gamma) = 1$, $\forall\ \gamma \geq \hat{\gamma}$. □

Proof of Proposition 10

Proof
Denote

$$\begin{aligned}
\Pi^S &= \frac{\bar{D} + \underline{D}}{2}\Big((1-A)\bar{F}(v^S) + (\beta - A)q^S(F(v^S) - F(\beta))\Big), \\
\Pi^{NS} &= \frac{\bar{D} + \underline{D}}{2}(\beta - A)\bar{F}(\beta);
\end{aligned}$$

Π^S and Π^{NS} are optimal profits at a segmented market (if $v^S \leq \bar{U}$) and at a non-segmented market, respectively.

By (15), Π^S can be rewritten in terms of v^S only, that is,

$$\Pi^S = \frac{\bar{D} + \underline{D}}{2\bar{U}}\Big((1-A)(\bar{U} - v^S) + \frac{(1-A)(v^S - \beta)(v^S - 1)}{v^S - 1 + \gamma(1-\beta)}\Big).$$

It is easy to check

$$\Pi^S \geq \Pi^{NS} \quad \Leftrightarrow \quad \bar{U} \geq \frac{1-A}{1-\beta} \cdot \frac{(\gamma(1-\beta) + \beta)v^S - \beta}{v^S - 1 + \gamma(1-\beta)} - \frac{\beta(\beta - A)}{1-\beta} = U_c^S\ .$$

Therefore, if $\bar{U} \geq \{U_c^S, v^S\}$, a segmented market, characterized by (v^S, q^S), is an optimal strategy; otherwise, serving the entire market at a low price only is optimal.

We next show $\bar{U} \geq \{U_c^S, v^S\} = U_c^S$. We claim that the optimal solution v^S to (15) is strictly less than $1 + \gamma(\beta - A)$. This is because the L.H.S. of (15) strictly decreases in v and is negative at $1 + \gamma(\beta - A)$. Also, tt is easy to check U_c^S declines in v^S. Since $1 < v^S < 1 + \gamma(\beta - A)$, $1 + \gamma(\beta - A) < U_c^S < 1 + \beta - A$. Therefore, $U_c^S \geq v^S$. □

Proof of Proposition 11

Proof
Suppose v^D and v^S are solutions to (8) and (15), respectively. Since inducing segmentation is an optimal strategy for both cases, v^D and v^S are also optimal cutoff values corresponding to the deterministic and stochastic demand cases, respectively. Since $A > \alpha_1$, $v^S \leq v^D$. Then $q^S \leq q^D$ follows with the fact that v increases in q. □

Proof of Proposition 12

Proof

If $\sum_{i=1}^{n} C_i = N\bar{F}(1)$, the first constraint of (16) holds for any v. The profit function is then maximized at $v^* = 1$. The firm $i's$ symmetric capacity choice is $\frac{N\bar{F}(1)}{n}$. If $\sum_{i=1}^{n} C_i > N\bar{F}(1)$, the L.H.S of the first constraint of (16) is strictly less than the R.H.S for any v, implying all customers wait to buy in period 2; that is, $v^* = \bar{U}$, and the firm $i's$ symmetric equilibrium capacity choice is $\frac{N\bar{F}(\beta)}{n}$.

We first consider profitable deviations at $\frac{N\bar{F}(1)}{n}$. Given that all the other firms except firm i keep their capacities at $\frac{N\bar{F}(1)}{n}$, if firm i increases its capacity even just a little bit, all customers switch to buy at a low price in period 2. Then the firm i will deviate if and only if the profit it would gain at a low price exceeds its current share of profits at a high price, namely,

$$(\beta - \alpha_1)\Big(N\bar{F}(\beta) - \frac{n-1}{n}\bar{F}(1)\Big) > (1-\alpha_1)\frac{N\bar{F}(1)}{n} \ ,$$

which yields to $\bar{U} < 1 + n(\beta - \alpha_1)$. Therefore, $\frac{N\bar{F}(1)}{n}$ is a symmetric Nash equilibrium if $\bar{U} \geq 1 + n(\beta - \alpha_1)$.

A similar approach is used to examine profitable deviations at $\frac{N\bar{F}(\beta)}{n}$. Provided that all the other firms except itself keep their capacities at $\frac{N\bar{F}(\beta)}{n}$, the firm i deviates if and only if

$$\begin{aligned} N\bar{F}(1) - \frac{n-1}{n}N\bar{F}(\beta) \quad &> \quad 0, \\ (1-\alpha_1)\Big(N\bar{F}(1) - \frac{n-1}{n}N\bar{F}(\beta)\Big) \quad &> \quad (\beta - \alpha_1)\frac{N\bar{F}(\beta)}{n} \ . \end{aligned}$$

The first inequality guarantees the feasibility of such a deviation; the second one says firm $i's$ profit obtained by deviating to a high-price-only solution is greater than its profit share at a low price at present. Solving those two inequalities, we have $\bar{U} > \beta + n(1 - \alpha_1)$. Therefore, if $\bar{U} \leq \beta + n(1 - \alpha_1)$, a symmetric low-price-only Nash equilibrium exists. □

Proof of Proposition 13

Proof

In a segmented market, that is, $q \leq \bar{q}$ and $\bar{q} = \left(\frac{\bar{U}-1}{\bar{U}-\beta}\right)^{\gamma}$, given C_{-i}, the firm $i's$ profit function is strictly concave in $v \geq 1$. Hence, regarding the problem (16), the optimal cutoff value is either v^0 (if $v^0 \leq \bar{U}$) or $\bar{U}$ (if $v^0 > \bar{U}$). Once segmentation is not sustainable, that is, $\bar{q} \leq q \leq 1$, the firm simply serves the entire market at the low price. Therefore, the symmetric Nash equilibrium is either a segmented solution, namely,

$$v^* \quad = \quad v^0,$$

$$
\begin{aligned}
q^* &= q^0 = \left(\frac{v^0 - 1}{v^0 - \beta}\right)^{\gamma}, \\
C_i^* &= C_i^0 = \frac{N}{n\bar{U}}(\bar{U} - v^0 + (v^0 - \beta)q^0), \\
\Pi_i^* &= \Pi_i^0 = \frac{N}{n\bar{U}}(1-\beta)(\bar{U} - v^0) + (\beta - \alpha_1)C_i^0, \quad \forall\, i = 1, \cdots, n.
\end{aligned}
$$

or a low-price-only solution, that is, $q^* = 1$, $v^* = \bar{U}$, $C_i^* = \frac{N\bar{F}(\beta)}{n}$ and $\Pi_i^* = \frac{N\bar{F}(\beta)}{n}(\beta - \alpha_1)$, $\forall\, i = 1, \cdots, n$.

When $v^0 \geq \bar{U}$, each firm's profit function is strictly increasing its capacity choice. Hence, the market is served entirely at the low price and none of them have incentive to deviate. In what follows, we focus on the case of $v^0 < \bar{U}$.

We first consider profitable deviations at v^0. Given that all the other firms except the firm i keep their capacities at C_i^0, the firm i will deviate from C_i^0 if and only if

$$(\beta - \alpha_1)(N\bar{F}(\beta) - (n-1)C_i^0) > \Pi_i^0 \ .$$

Algebraic arrangements yield to

$$\bar{U} < \frac{n(\beta - \alpha_1)(v^0 - \beta)(1 - q^0)}{1 - \beta} + v^0 = U_c \ .$$

So, when $\bar{U} \geq \max\{v^0, U_c\}$, (v^0, q^0, C_i^0) is a symmetric Nash equilibrium solution. Note that $U_c \geq v^0$, then a segmented Nash equilibrium exists if $\bar{U} \geq U_c$.

We now consider profitable deviations at $\frac{N\bar{F}(\beta)}{n}$. Provided that all firms except the firm i keep their capacities at $\frac{N\bar{F}(\beta)}{n}$, the firm i will deviate if and only if one of the following two cases holds:

Case 1: It is feasible to achieve the optimal segmentation of customers by reducing the firm i's capacity, that requires:

$$
\begin{aligned}
nC_i^0 - \frac{n-1}{n}N\bar{F}(\beta) &\geq \frac{N\bar{F}(v^0)}{n}, \\
\frac{N\bar{F}(v^0)}{n}(1-\beta) + (\beta - \alpha_1)\left(nC_i^0 - \frac{n-1}{n}N\bar{F}(\beta)\right) &> (\beta - \alpha_1)\frac{N\bar{F}(\beta)}{n} \ .
\end{aligned}
$$

The first inequality guarantees such a deviation is feasible; the second one implies deviating to induce segmentation makes more profits. Solving those two inequalities, the firm will deviate if $\bar{U} > U_c$ and $q^0 \geq 1 - \frac{1}{n}$. That is, if $\bar{U} \leq U_c$ or $q^0 < 1 - \frac{1}{n}$, there exists a low-price-only Nash equilibrium.

Case 2: Although it is infeasible to achieve the optimal segmentation (The reduced capacity of firm i cannot meet its equally shared demand at the high price in period 1, that is one assumption made in our model), there exists some threshold $\hat{v} > v^0$ such that it is feasible to achieve the segmentation characterized by $\hat{v}$; and the firm i makes more profits by deviating its capacity to $\hat{C}$, which requires the following system of inequalities:

$$\left(\frac{\hat{v}-1}{\hat{v}-\beta}\right)^{\gamma} = \frac{\hat{C}+\frac{n-1}{n}N\bar{F}(\beta)-N\bar{F}(\hat{v})}{N(F(\hat{v})-F(\beta))}\ , \tag{24}$$

$$\frac{N\bar{F}(\hat{v})}{n}(1-\beta)+(\beta-\alpha_1)\hat{C} > \frac{N\bar{F}(\beta)}{n}(\beta-\alpha_1), \tag{25}$$

$$\hat{C} \geq \frac{N\bar{F}(\hat{v})}{n}\ , \tag{26}$$

$$\hat{v} > v^0. \tag{27}$$

(24), (26) and (27) require the feasibility of such a deviation; (25) means this deviation is profitable. Solve for this system of inequalities, then the firm i will deviate if and only if there exists $\bar{U} > \hat{v} \geq \max\{v^0, B\}$ such that $\bar{U} > H(\hat{v})$ where

$$\begin{aligned} B &= \frac{1-\beta\left(\frac{n-1}{n}\right)^{\frac{1}{\gamma}}}{1-\left(\frac{n-1}{n}\right)^{\frac{1}{\gamma}}}\ , \\ H(v) &= \frac{n(\beta-\alpha_1)(v-\beta)\left(1-\left(\frac{v-1}{v-\beta}\right)^{\gamma}\right)}{1-\beta} + v\ . \end{aligned}$$

Hence, the firm i stays at $\frac{N\bar{F}(\beta)}{n}$ if there does not exist $\bar{U} > \hat{v} \geq \max\{v^0, B\}$ such that $\bar{U} > H(\hat{v})$.

Note that $H(v)$ is strictly convex in $v \geq 1$ (The proof is similar as Lemma 1.); also, it has the same F.O.C. equation as (17). Therefore, $H(v)$ is minimized either at v^0 if $v^0 \geq B$, or at B if $v^0 < B$. That implies that the firm i will not deviate if $\bar{U} \leq U_c$ when $v^0 \geq B$, or $\bar{U} \leq H(B)$ when $v^0 < B$.

Combining the results in all cases yields Proposition 13. □

Proof of Corollary 3

Proof

The solution to (17) is always strictly less than $1+n\gamma(\beta-\alpha_1)$. This is because the L.H.S. of (17) decreases in v and has a negative value at $v = 1+n\gamma(\beta-\alpha_1)$.

On the other hand, by (17), U_c can be expressed in terms of v^0,

$$U_c = \frac{\left(n\gamma(\beta-\alpha_1)+\beta+\gamma-\beta\gamma\right)v^0 - n\beta\gamma(\beta-\alpha_1)-\beta}{v^0-1+\gamma(1-\beta)}\ ,$$

which is strictly decreasing in v^0. Since $1 < v^0 < 1 + n\gamma(\beta - \alpha_1)$, $1 + n\gamma(\beta - \alpha_1) < U_c < 1 + n(\beta - \alpha_1)$. Then Corollary 3 directly follows with Proposition 13. □

Proof of Proposition 14

Proof
Let $\bar{v}^0$, $(\bar{v}^*, \bar{q}^*)$, $\bar{C}^*$ and $\bar{\Pi}(v,q)$ denote the solution to (17), the pair of optimal cutoff value and aggregate fill rate, the optimal aggregate capacity and aggregate profit at (v,q) in an oligopoly market, respectively. Let v^0, (v^*, q^*), C^* and $\Pi(v,q)$ be their counterparts in the monopoly market. Then, $(\bar{v}^*, \bar{q}^*) = \arg\max\{\bar{\Pi}(\bar{v}^0, \bar{q}^0), \bar{\Pi}(\bar{U}, 1)\}$ and $(v^*, q^*) = \arg\max\{\Pi(v^0, q^0), \Pi(\bar{U}, 1)\}$.

According to (17), $\bar{v}^0$ strictly increases in n. Hence, $\bar{v}^0 > v^0$. To show $\bar{v}^* \geq v^*$, it suffices to show that $\bar{\Pi}(\bar{v}^0, \bar{q}^0) \geq \bar{\Pi}(\bar{U}, 1)$ implies $\Pi(v^0, q^0) \geq \Pi(\bar{U}, 1)$. Suppose it is not true, that is, $\bar{\Pi}(\bar{v}^0, \bar{q}^0) \geq \bar{\Pi}(\bar{U}, 1)$ and $\Pi(v^0, q^0) < \Pi(\bar{U}, 1)$. Since $\bar{\Pi}(\bar{U}, 1) = \Pi(\bar{U}, 1) = N\bar{F}(\beta)$, it must be $\bar{\Pi}(\bar{v}^0, \bar{q}^0) > \Pi(v^0, q^0)$. Note that $\bar{\Pi}(v,q)$ and $\Pi(v,q)$ have the same formulations in terms of v, namely,

$$\bar{\Pi}(v) = \Pi(v) = (\bar{U} - v)(1 - \alpha_1) + (\beta - \alpha_1)(v - \beta)\left(\frac{v-1}{v-\beta}\right)^{\gamma},$$

which is strictly concave in $v \geq 1$ and maximized at v^0 (see Lemma 1). Since $\bar{v}^0 > v^0$, it must be $\bar{\Pi}(\bar{v}^0) < \Pi(v^0)$, which contradicts to the assumption $\bar{\Pi}(\bar{v}^0, \bar{q}^0) > \Pi(v^0, q^0)$. Therefore, $\bar{v}^* \geq v^*$.

The optimal aggregate fill rate in an oligopoly market is higher than that in the monopoly market, that is, $\bar{q}^* \geq q^*$ directly follows with $\bar{v}^* \geq v^*$ because $q = \left(\frac{v-1}{v-\beta}\right)^{\gamma}$ if $v < \bar{U}$; otherwise, $q = 1$.

Note that capacity choices in an oligopoly market and in the monopoly market have the same expressions in terms of fill rates, thus $\bar{C}^* \geq C^*$ directly follows with $\bar{q}^* \geq q^*$ and Proposition 5.

The last is to show $\bar{\Pi}(\bar{v}^*, \bar{q}^*) \leq \Pi(v^*, q^*)$. Since $\bar{q}^* \geq q^*$, three cases are required to be discussed.

Case 1: $\bar{q}^* = q^* = 1$

Trivially, in this case, $\bar{\Pi}(\bar{U}, 1) = \Pi(\bar{U}, 1) = N\bar{F}(\beta)$.

Case 2: $\bar{q}^* < 1$ and $q^* < 1$

In this case, optimal stocking decisions for both markets are to induce segmentation. As in the proof of $\bar{v}^* \geq v^*$, we have shown $\Pi(\bar{v}^*, \bar{q}^*) \leq \Pi(v^*, q^*)$.

Case 3: $q^* < \bar{q}^* = 1$

$q^* < 1$ implies $\Pi(v^*, q^*) \geq \Pi(\bar{U}, 1) = \bar{\Pi}(\bar{U}, 1) = \bar{\Pi}(\bar{v}^*, \bar{q}^*)$.

We next show how the equilibrium depends on n. We first show the optimal cutoff value v^* increases in n. If $\bar{v}^0 < \bar{U}$ and $\bar{\Pi}(\bar{v}^0) \geq N\bar{F}(\beta)(\beta - \alpha_1)$, $v^* = \bar{v}^0$; otherwise, $v^* = \bar{U}$.

Since $\bar{v}^0$ strictly increases in n according to (17), thus $\forall\, n_1 < n_2$, $\bar{v}^0(n_1) < \bar{v}^0(n_2)$. Then it suffices to show if $\bar{v}^0(n_1) < \bar{v}^0(n_2)$ and $v^*(n_1) = \bar{U}$, it must be $v^*(n_2) = \bar{U}$ as well.

As shown in Lemma 1, $\bar{\Pi}(v)$ is strictly concave and maximized at v^0, and v^0 is the solution to (17) when $n = 1$. Note that $\bar{v}^0 > v^0$ $(n > 1)$, and $\bar{v}^0$ increases in n. Together with the fact that $\bar{\Pi}(v)$ is strictly decreasing when $v \geq v^0$, we conclude $\bar{\Pi}(\bar{v}^0)$ decreases in n, and hence, $\bar{\Pi}(\bar{v}^0(n_2)) \leq \bar{\Pi}(\bar{v}^0(n_1)) < N\bar{F}(\beta)(\beta - \alpha_1)$. It follows with $v^*(n_2) = \bar{U}$.

Similarly, we can show the optimal aggregate profit decreases in n.

Since q increase in v, and C increases in v by Proposition 5, the optimal aggregate capacity and fill rate increase in n as well. □