Publication information

Title	Strategic Capacity Management When Customers Have Boundedly Rational Expectations
Author(s)	Huang, Tingliang; Liu, Qian
Source	Production and Operations Management , v. 24, (12), December 2015, p. 1852-1869
Version	Pre-Published version
DOI	https://doi.org/10.1111/poms.12420
Publisher	Wiley-Blackwell

Copyright information

This is the peer reviewed version of the following article: [Huang, T., Q. Liu. Strategic Capacity Management When Customers Have Boundedly Rational Expectations. Production and Operations Management (2015)], which has been published in final form at [\[https://doi.org/10.1111/poms.12420\]](https://doi.org/10.1111/poms.12420). This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

Notice

This version is available at HKUST Institutional Repository via

<http://hdl.handle.net/1783.1/75896>

If it is the author's pre-published version, changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published version.

Strategic Capacity Management When Customers Use Anecdotal Reasoning

Tingliang Huang

Department of Management Science and Innovation, University College London, London, WC1E 6BT, United Kingdom
t.huang@ucl.ac.uk

Qian Liu

Department of Industrial Engineering and Logistics Management, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
qianliu@ust.hk

In retailing industries, such as apparel, sporting goods, customer electronics and appliances, many firms deploy sophisticated modeling and optimization software to conduct dynamic pricing in response to uncertain and fluctuating market conditions. However, the possibility of markdown pricing creates an incentive for customers to strategize over the timing of their purchases. How should a retailing firm optimally account for customer behavior when making its pricing and stocking/capacity decisions? For example, is it optimal for a firm to create rationing risk by deliberately understocking products? In this paper, we develop a stylized modeling framework to answer these questions. In our model, customers strategize over the timing of their purchases. However, customers have boundedly rational expectations in the sense of anecdotal reasoning about the firm's fill rate, i.e., they have to rely on anecdotes, past experiences, or word-of-mouth to infer the firm's fill rate. In our modeling framework, we distinguish two settings: (i) capacity commitment, where the firm commits to its capacity level in the long run, or (ii) the firm dynamically changes it in each season. For both settings, within the simplest form of anecdotal reasoning, we prove that strategic capacity rationing is not optimal independent of customer risk preferences. Then, using a general form of anecdotal reasoning, we provide sufficient conditions for capacity rationing to be optimal for both settings respectively. We show that the result of strategic capacity rationing being suboptimal is fairly robust to different valuation distributions and utility functions, heterogeneous sample size, and price optimization.

Key words: strategic capacity management; bounded rationality; capacity rationing; anecdotal reasoning; revenue management

1. Motivation & Introduction

Dynamic pricing and capacity management have gained an increasing popularity in revenue management practice, and have engendered a significant body of academic research (see, e.g., Bitran and Caldentey 2003, and Talluri and van Ryzin 2004, Chapter 5). Inter alia, much attention has been focused on how customers react to firm strategies: historically, customers are assumed not to anticipate future prices or availability; they are *myopic* and purchase if their current utility from purchasing is positive. Recognizing that dynamic pricing creates an incentive for customers to

strategize over the timing of their purchases, the recent body of literature has started to investigate such *forward-looking* or *strategic* customers who may intentionally delay their purchases in anticipation of future discounts (Su 2007, Gallego et al. 2008, Zhang and Cooper 2008, Aviv and Pazgal 2008, Liu and van Ryzin 2008, Yin et al. 2009, Cachon and Swinney 2009, Lai et al. 2010, Mersereau and Zhang 2012, Li et al. 2011 and references therein). A key common assumption made in this extensive literature is that customers are fully rational: they are perfect utility-maximizers and form “rational expectations” about the firm’s strategies; that is, they perfectly anticipate the firm’s strategies in equilibrium.¹ Undoubtedly, this might be a strong assumption in some real settings. Then a natural and important question for both academics and practitioners is how robust the existing findings and managerial insights are with respect to this assumption. This paper attempts to scratch the surface of this general question in the specific context of strategic capacity management. In other words, it focuses on determining the optimal size of capacity that should be stocked to maximize a firm’s profit, given that customers behave strategically.

Recently, Liu and van Ryzin (2008) show that strategic capacity rationing (i.e., deliberately understocking products) can be an optimal strategy to influence the strategic behavior of customers when they can perfectly anticipate the product availability. Despite the fact that strategic customer behavior has been well recognized by both academics and practitioners, how customers can actually form rational expectations that are commonly assumed in the existing literature remains unclear. Some have explained that rational expectations can be formed through repeated purchasing experiences. However, one may question whether the setting of selling durable goods might be at odds with the long-run learning from repeated purchases. Purchasing cars as the examples of “General Motors” and “Saturn” in Liu and van Ryzin (2011, p. 98) may happen just a few times in a person’s life, then a focus exclusively on “post-convergence” behavior that results from long-run learning might not be always warranted (Camerer and Lowenstein 2003). Hence, in some situations, it also appears reasonable to assume that customers only have scarce opportunities to learn (Spiegler 2006b). To quote Camerer and Lowenstein (2003, p. 8-9), “many important aspects of economic life are like the first few periods of an experiment rather than the last.”

In this study, we shall extend and complement this stream of literature by investigating the settings where long-run learning from repeated purchases is not possible, and it is difficult for customers to perfectly learn the firm’s capacity choice. Therefore, customers naturally resort to

¹ Notice that rational expectations are imbedded in the Nash equilibrium, subgame perfection, and its other refinements (cf. Aumann and Dreze 2008, Osborne and Rubinstein 1998, and Spiegler 2006b for more discussions). We use the term in the sense of Aumann and Dreze (2008), but not in the sense in macroeconomics. See Liu and van Ryzin (2011) for the same interpretation.

simple heuristics or short-cuts and infer from a limited number of samples or anecdotes, i.e., to use *anecdotal reasoning*. Consequently, customers have *boundedly rational expectations* about the firm's strategy. However, they are still forward-looking since they still take the future into account when making their current purchasing decisions. Hence, this study isolates and highlights the role of boundedly rational expectations (in the sense of anecdotal reasoning) on the equilibrium outcomes. Our main research question is: would strategic capacity rationing be optimal when customers have boundedly rational expectations in the sense of anecdotal reasoning?

Our model of boundedly rational expectations is inspired by the recent economics literature on bounded rationality in general (see the textbooks, Rubinstein 1998 and Spiegler 2011, for comprehensive surveys), and the $S(1)$ and $S(K)$ procedures proposed by Osborne and Rubinstein (1998) in particular. To model limited learning and anecdotal reasoning that results in boundedly rational expectations, we first assume that each customer purchases once or rarely (as is the case with cars and other durable products of similar nature) and before deciding which period to purchase, he obtains a single sample or "anecdote" ($S(1)$ procedure). For example, he asks one of his friends (who has interacted with the firm before) about his former experience with the firm (e.g., whether he faced stockouts in the markdown period). Based on the limited information from others (i.e., word of mouth), he can only rely on anecdotal reasoning to make his own purchasing decision. Hence, customers rely on random, casual stories regarding the service level (fill rate) of the firm in the markdown period, and react to these stories as if they are fully informative of the actual service level. In situations when learning from repeated purchases is unavailable, customers are more likely to rely on anecdotes such as "a friend of mine has bought a Chevrolet at a discount in a clearing sale," "my colleague has waited for a cheap flight ticket from Chicago to Bali but didn't get it." We also generalize our model to capture that a customer may obtain multiple samples or anecdotes ($S(K)$ procedure) before making his purchase decision. Based on this model, we are able to investigate how the *level of rationality in customer expectations* affects the firm's optimal capacity decision.

The fundamental contribution of the paper is that we develop a capacity-commitment model and a dynamic capacity management model, allowing us to study the impact of customer boundedly rational expectations, in the sense of anecdotal reasoning, on the firm's strategic capacity management decisions. Our main findings are as follows. First, when customers use a single sample or anecdote to make purchasing decisions ($S(1)$ model), we obtain a somewhat unexpected result that it is not optimal for a firm to strategically ration capacity, for two distinct settings of capacity commitment (where the firm commits to a capacity level) and dynamic capacity management (where

the firm can adjust its capacity level in each period). This result does not depend on customer risk preference (i.e., the functional forms of customer utility). Second, we generalize the $S(1)$ model to a general $S(K)$ model that allows customers to combine a collection of samples to estimate the firm's service level (fill rate). The advantage of our model is that the single parameter K can be treated as the *level of rationality in customer expectations*. Notably, the rational-expectations model in the existing literature becomes a special case of this $S(K)$ model when the level of rationality K becomes infinity. The numerical results show that the finding of capacity rationing being suboptimal appears fairly robust to different valuation distributions and utility functions, heterogeneous sample size, price optimization, and discounted utility. The intuition behind the results is that each individual customer's belief of the fill rate becomes more homogeneous and close to the true one, as customers become more rational (K increases). In other words, when customers use very limited samples (small K) to infer the fill rate, each individual customer's estimates are quite diverse, thus weakening the firm's advantage of using capacity rationing to optimally influence customer strategic purchase behavior. In summary, our findings are complementary to the existing capacity management literature, and unearth the importance of accounting for customer bounded rationality in capacity management that has been somewhat overlooked in the operations and revenue management literature.

The remainder of this paper is organized as follows. Section 2 reviews related literature. Section 3 presents the basic framework of modeling boundedly rational expectations in the sense of anecdotal reasoning. Section 4 studies a variety of extensions to the basic model. Discussions and conclusions are provided in the last section. Proofs are relegated to the appendix.

2. Related Literature

Our model follows the recent economics literature on modeling boundedly rational expectations. Osborne and Rubinstein (1998) formulate the $S(1)$ and $S(K)$ procedure in the context of strategic-form games, in which all players behave according to this procedure. They focus on developing a novel equilibrium concept. In contrast, in our study, the firm is fully rational, and only the customers employ the $S(K)$ procedure. The $S(1)$ procedure prescribes that each agent chooses his action after randomly sampling each alternative from his strategy set, picking the action that yields the highest payoff. Osborne and Rubinstein (2003) also apply a variant on " $S(1)$ -equilibrium" in a strategic voting setting. The $S(1)$ procedure has been applied in a variety of settings. For example, Spiegler (2006a) studies a setting where patients do not have enough opportunities to learn doctors' "quality," and choose doctors according to the $S(1)$ procedure which reflects "anecdotal reasoning."

He finds that the element of boundedly rational expectations has significant implications, e.g., the “market for quacks” is active. In another paper, Spiegler (2006b) studies a market model in which profit-maximizing firms compete in multidimensional pricing strategies for a consumer, who is limited in his ability to grasp such complicated objects and therefore uses a sampling procedure (i.e., the $S(1)$ procedure) to evaluate them. He shows that firms respond to increased competition with an increased effort to obfuscate, rather than with more competitive pricing.

As Spiegler (2006a,b, 2011) pointed out, the $S(1)$ procedure is related to and consistent with other departures from standard probabilistic reasoning. Tversky and Kahneman (1971) demonstrate experimentally that people over-infer from small samples, which they refer to as “the law of small numbers.” They explain this phenomenon as a consequence of the “representativeness” heuristic: people expect a small sample to mirror the probability distribution from which it is drawn. Rabin (2002) proposes a formal model of inference by “believers in the law of small numbers,” an extreme version of which the $S(1)$ procedure reflects. The idea behind the $S(1)$ rule is to capture a simple way of *anecdotal reasoning* (as opposed to *probabilistic reasoning*): in our basic model to be presented in Sections 3, 4.1 and 5.1, each customer independently asks some former client of the firm and infers his own utility from purchasing in each period based on what he heard. Our model captures scarce learning and word of mouth in inferring the firm’s capacity decision for individual heterogeneous customers, which reflects customer boundedly rational expectations. The $S(1)$ procedure is also closely related to the model of “case-based reasoning” due to Gilboa and Schmeidler (2001), where decision makers evaluate an action by recalling its performance in past “cases.” We refer the reader to Szech (2011a,b), Spiegler (2011, Chapter 6-7), and references therein for more discussions and applications of the $S(1)$ procedure. To the best of our knowledge, this type of customer bounded rationality has not been explored in any capacity management settings. Furthermore, due to the analytical complexity of the $S(K)$ model, the economics literature works only with the $S(1)$ model. Our study appears to be among the first to endow the parameter K with a concrete meaning, the level of rationality in customer expectations, and investigate its impact on the firm’s capacity decision in an important operations and revenue management setting.²

While full rationality is modeled in a unique way by economists, there are naturally a variety for approaches of modeling bounded rationality. For a general and comprehensive discussion (e.g., which approach is more appropriate in different situations, whether purely incomplete information is equivalent to bounded rationality, etc.), we refer the reader to the classic textbook Rubinstein

² We refer the reader to Huang and Yu (2014) for an application of the $S(K)$ model in a marketing setting of probabilistic selling.

(1998) and the most recent one, Spiegler (2011). There is an emerging literature on behavioral operations management; we refer the readers to Bendoly et al. (2006) for this stream of research. Su (2008) applies the quantal response framework to the classic newsvendor setting. Kalkanci and Perakis (2013) investigate the impact of bounded rationality in a two-tier supply chain in which a supplier offers a wholesale price contract and sells to a boundedly rational retailer facing a newsvendor setting. Veeraraghavan and Debo (2011) and Kremer and Debo (2012) study herding behavior in service systems both theoretically and experimentally. Huang et al. (2013) and Huang and Chen (2014) examine customer incapability of estimating expected waiting time in service settings. Our work contributes to this stream of literature by studying the impact of customer boundedly rational expectations on the firm's capacity management decisions in retail settings.

Inspired by and based on the work of Liu and van Ryzin (2008), our work is quite related to the fast growing research stream of revenue management with strategic customer behavior. See Netessine and Tang (2009) for more comprehensive reviews. In the theoretical literature, authors typically assume rational expectations as a starting point, or else they assume long-run repeated interactions which facilitate learning of fill rates. In the empirical operations literature, for instance, the recent work by Li et al. (2011) investigates whether customers are really forward-looking; it also assumes "rational expectations" given that testing the assumption of rational expectations is extremely challenging (if not impossible). The challenge of empirically testing this strong assumption of rational expectations actually justifies the importance of theoretically investigating what happens if customers have boundedly rational expectations. This argument also provides motivations for our study. There are several papers that have examined bounded rationality in capacity and revenue management settings. Nasiry and Popescu (2011) investigate dynamic pricing with loss-averse consumers and peak-end anchoring. Recently, Özer and Zheng (2014) study a seller's optimal pricing and inventory strategies when non-pecuniary motives affect consumers' purchase decisions. They investigate the impact of two salient behavioral regularities - anticipated regret and misperception of product availability - on seller's profit. Our work shares a similar theme as the two studies, although our approach of modeling customer behavior differs. It appears that anecdotal reasoning consumers have not received much attention in the operations and capacity management literature.

3. Basic Model

We adopt the model framework in Liu and van Ryzin (2008) and highlight the differences in modeling customer behavior between their setup and ours. This facilitates our focus on the role of customer boundedly rational expectations in the sense of anecdotal reasoning.

3.1. Model Setup

Consider a firm that sells a product over two periods. The price p_1 in period 1 and p_2 in period 2 are pre-determined and exogenous with $p_1 > p_2$.³ The assumption of fixed prices or price commitment is maintained in the basic model. There are several key reasons for this assumption. First, it allows us to isolate and focus on our research question on the firm's capacity decision, i.e., whether capacity rationing is optimal or not. The same assumption is also made in Liu and van Ryzin (2008, 2011) and Ovchinnikov and Milner (2012). Second, it may be reasonable in many retail settings because of reputation considerations, price stickiness (Blinder et al. 1998), etc.

The market consists of a large population with a deterministic market size N . Each individual customer is assumed to be infinitesimally small so that strategic interactions among customers can be reasonably ignored. Customer valuations are heterogeneous, and each customer demands a single unit. Customers decide to purchase in period 1 or wait for a lower price in period 2. The firm is a rational profit-maximizer and chooses its capacity (stocking quantity) C at the beginning of period 1, with procurement cost $\alpha < p_2$ per unit. All purchases in period 1 are guaranteed to be satisfied, but demand in period 2 may not be all unmet. The reason is the following: if the supply is not enough to meet demand even at the high price, there will be no supply left for meeting the demand in the second period. This profit is strictly increased by choosing a higher supply that precisely meets all the demand at the high price. The firm does not have any incentive to stock more than the demand at the low price either, because there will be no rationing in period 2 and there will be positive leftover inventory. The profit is strictly increased by choosing a lower supply that precisely meets all the demand at the low price. Let q be the probability of obtaining a unit in period 2, which is called the fill rate or service level and is determined by both the firm's capacity decision C and customer purchasing behavior (to be specified below in equation (3)). We assume that a uniform allocation rule applies; that is, each customer has equal chance to obtain the product when the remaining capacity in period 2 is less than the demand.

Customer valuations are independently and identically distributed with cumulative distribution function $F(v)$, which is known to both the firm. Each customer has the same utility function $u(\cdot)$, which is time invariant. Different from Liu and van Ryzin (2008), we only require u to be strictly increasing. Hence, we allow customers to be risk neutral, risk averse or risk seeking. ***TL: want to confirm with you that the customers do have to know F , right?***

Liu and van Ryzin (2008) make a key assumption that customers have rational expectations about the firm's capacity choice: customers correctly anticipate the firm's optimal stocking quantity

³ We shall consider price optimization as an extension of the basic model in Section 4.

decision. As argued earlier, this may not be the real case in practice. In many situations, customers do not have enough opportunities to learn the firm's capacity choice under the repeated purchasing environment, and hence can only rely on *anecdotal reasoning* to make their purchasing decisions. In particular, customers obtain samples/anecdotes from different sources. For concreteness, let us imagine the following setting: Each customer "samples" the firm by randomly asking a few former clients of the firm about their past experiences in the markdown period, which we may call *word of mouth*. He relies on the anecdotes in order to judge the pros and cons of purchasing in period 1: a stockout in period 2 he heard about makes a customer more willing to purchase in period 1; an availability in period 2 he heard about makes the customer more likely to postpone his purchase. After he decides in which period to purchase, his own realized *actual* utility would be driven by a *new, independent* draw from the random demand fulfillment. This sampling process is motivated by the $S(K)$ procedure (K is the number of samples a customer obtains before making his purchase decision) proposed by Osborne and Rubinstein (1998).

We now analytically formalize the $S(K)$ model. Suppose a customer observes K independent samples or anecdotes. Let 1_{A_k} be a binary random variable to denote whether the k -th sample indicates availability ($1_{A_k} = 1$) or stockout ($1_{A_k} = 0$) in the markdown period, $k = 1, 2, \dots, K$. We assume that each customer's sample size K does not depend on his valuation v or endogenously determined by the customer. In other words, each customer has access to the same "size" (or amount) of information independent of his valuation. This assumption is reasonable in many situations, as how much information customers can collect typically depends on factors that are out of their control, e.g., market conditions, technological capability, etc. Traditionally, the K may be extremely small. With the prevalence of the advanced web-based technology (e.g., Internet, facebook, twitter, etc.), the K presumably becomes larger.⁴ We further assume that each customer combines these samples by simply taking the average⁵, i.e., he purchases the product in period 1 if and only if

$$u(v - p_1) \geq \frac{1}{K} \sum_{k=1}^K [1_{A_k} u(v - p_2) + (1 - 1_{A_k}) u(0)] \quad (1)$$

and $v - p_1 \geq 0$, where $u(0)$ is normalized to zero.

According to the $S(K)$ model, customers fully rely on the signals they obtain from others' experiences. They over-infer from their samples, especially when K is small. They attribute the occasional

⁴ This is analogous to the interpretation of the level of bounded rationality β in Huang et al. (2013) where β is not controlled by the customers themselves (but may be influenced by the firm).

⁵ How a customer combines different samples to make his purchasing decision is an interesting question. For example, he may adopt a conservative rule by replying on the "worst" sample or an optimistic rule by using the "best" sample only. In this paper, we adopt the commonly-used statistical rule of taking the sample average.

availability (stockout) to ample (limited) capacity rather than luck. This non-parametric approach is in contrast to the well-known (parametric) Bayesian method, where any new information is combined with a prior belief of stockout using the Bayesian formula. Hence, a Bayesian customer partially relies on her anecdotes (and partially relies on her prior belief). As $K \rightarrow \infty$, the sample average $\frac{1}{K} \sum_{k=1}^K 1_{A_k}$ becomes the population expectation, hence, the *true* fill rate q . Therefore, the model of Liu and van Ryzin (2008) is a special case of this $S(K)$ model when $K = \infty$, which is formally stated in Lemma 1 below.

LEMMA 1. *If $K = \infty$, the $S(K)$ model reduces to the model in Liu and van Ryzin (2008).*

We thus use the parameter K to measure the *level of rationality in customer expectations*. The larger the K is, the more rational the customers are. In the extreme case of $K = \infty$, the customers have (fully) rational expectations. Any finite K indicates how boundedly rational the customers are in their expectations of the firm's fill rate.

3.2. Capacity Commitment

The model with capacity commitment is technically simple, and serves as a starting point for us to investigate the impact of customer anecdotal reasoning behavior on the firm's strategic capacity decision. This assumption may be justified in settings where changing capacity is costly.

We start from the basic model where each customer obtains only one sample/anecdote. When $K = 1$, the customer purchase decision condition (1) reduces to the following: a customer with valuation v purchases in period 1 if and only if

$$u(v - p_1) \geq u(v - p_2)1_A + u(0)(1 - 1_A), \quad (2)$$

and $v - p_1 \geq 0$, where 1_A is a binary random variable to denote whether the customer being consulted (i.e., the former client of the firm) obtained a product or not in period 2. Therefore, a customer purchases in period 1 if and only if $1_A = 0$ and $v \geq p_1$. Notice that the word of mouth received by each customer is independent of others', and thus heterogeneity among customers is captured. In other words, different customers may well receive *different* word of mouth. Note that, under the $S(1)$ model, customers' purchase decision does not depend on the magnitude of the markdown price p_2 . This is a byproduct of the assumption that each customer only has a single anecdote. In Section 4.1, we shall relax this assumption by allowing for multiple anecdotes, in which case the magnitude of the markdown price p_2 does play a role.

Sequence of Events. The firm first *commits* to a capacity choice C at the initialization season $t = 0$ for all the future seasons. Each season $t = 0, 1, 2, \dots$ consists of two periods: period 1 with

Figure 1 Sequence of Events of Capacity Commitment and Consumer Anecdotal Learning

price p_1 and period 2 with price p_2 . At each season, a constant number N of customers enter the market, make a *one-shot* purchasing decision and then *exit* the market. Hence, learning from “repeat experiences” for each individual customer is not allowed. At the initial season $t=0$, each customer does not have any information about the fill rate. Hence, it is reasonable to assume that customers with valuations greater than p_1 purchase in period 1 and the remaining customers wait for the markdown. Each customer at season $t=1, 2, \dots$, obtains an anecdote from a random customer in the previous season who has waited for the markdown. The anecdote indicates whether the customer in season $t-1$ obtained a product or faced a stockout. Based on the anecdote, each customer chooses which period to purchase according to the $S(1)$ rule specified in (2). Figure 1 depicts the sequence of events and consumer dynamic learning through sampling. We are interested in the long-run steady state of the dynamic learning system.

3.2.1. Equilibrium Fill Rate Let $\varphi(q, p_1)$ denote the *fraction* of the customers who purchase in period 1, then $\varphi(q, p_1) = (1 - q)\bar{F}(p_1)$ since $P(1_A = 0) = 1 - q$ and we assume that customer valuation and demand fulfillment are independent random variables. The fill rate in period 2 is then given by

$$q = \min \left\{ \left(\frac{C - N\varphi(q, p_1)}{N(1 - \varphi(q, p_1) - F(p_2))} \right)^+, 1 \right\}, \quad (3)$$

where the denominator $N(1 - \varphi(q, p_1) - F(p_2))$ is the number of customers who purchase in period 2 (i.e., those customers who not only obtained positive anecdotes but also have valuations higher than p_2). As argued earlier, the firm will never choose a capacity less than the demand in period 1 since it can always improve the profit by increasing capacity to exactly satisfy the demand in period

1. The firm will never stock more than the demand at the lower price p_2 as there is even leftover inventory at the end of period 2. Hence, the optimal capacity must satisfy $C \in [N\varphi(q, p_1), N\bar{F}(p_2)]$. The fill rate expression (3) is then reduced to

$$q = \frac{C - N\varphi(q, p_1)}{N(1 - \varphi(q, p_1) - F(p_2))}. \quad (4)$$

The fill rate is required to satisfy this fixed-point type of equation because of the consistency or stability of the “system,” and thus called the equilibrium fill rate.

DEFINITION 1. The long-run **Equilibrium Fill Rate** q of the market is the fill rate that satisfies the fixed-point equation (4).

In the Online Supplement, we demonstrate that equation (4) has a unique solution, and the market always converges to the unique equilibrium.

The novelty of our model vis-à-vis the economics literature is that the markdown-period fill rate q is *endogenously* determined by customer purchasing behavior, whereas there the random service quality level (e.g., the likelihood that a customer obtains satisfactory/successful service/treatment from the firm) is exogenously given (cf. Spiegler 2011, Chapter 6-7). Hence, our model has the fixed-point equation which does not arise in the existing economics literature.

Note that, customers in our model are not myopic, rather they are still strategic or forward-looking since they do take the future price and availability into account when making their current purchasing decisions. However, they form *boundedly rational expectations* about the firm’s fill rate due to their anecdotal learning about the firm. Therefore, our model is able to exclusively highlight the role of boundedly rational expectations while still retaining the strategic behavior.

3.2.2. Firm’s Capacity Decision We are now ready to formulate the firm’s profit optimization problem. By the fixed-point equation (4), for any given $q \in [0, 1]$, the capacity level C is a quadratic and convex function of q :

$$C(q) \equiv C = N\bar{F}(p_1)q^2 + N(\bar{F}(p_2) - 2\bar{F}(p_1))q + N\bar{F}(p_1) \quad (5)$$

The relationship of $C(q)$ and q suggests that finding the optimal capacity level is transformed to solving for the optimal fill rate. Interestingly, $C(q)$ is not necessarily monotonically increasing in q in our setting, while the capacity C always increases in the fill rate q when customers have rational expectations (see Proposition 5 of Liu and van Ryzin 2008). The reason is as follows: Increasing the fill rate q leads to a decrease in first-period sales $S_1(q) \equiv N(1 - q)\bar{F}(p_1)$, and such a decrease is linear and hence first-order. Meanwhile, it leads to an increase in second-period sales $S_2(q) \equiv N[1 - (1 - q)\bar{F}(p_1) - F(p_2)]q$. However, such an increase is quadratic and hence

second-order. It could be the case that the total sales $C(q) = S_1(q) + S_2(q)$ decreases since the decrease in the first-period sales may dominate. This surprising result is due to customer bounded rationality in the sense of anecdotal reasoning. The fill rate perceived by each customer through anecdotal reasoning may be different, depending on his random sampling outcome. This customer misperception provides the firm an opportunity to stock less yet yielding a higher fill rate. However, in the rational-expectations model of Liu and van Ryzin, customers perfectly anticipate the fill rate. As expected, to achieve a higher fill rate requires the firm to stock more.

Since the second-period fill rate q uniquely determines the firm's capacity level C , we can transform the firm's capacity management problem to the one with the fill rate as the decision variable as follows:

$$\max_{q \in [0,1]} \Pi(q) := N(p_1 - p_2) \bar{F}(p_1)(1 - q) + N(p_2 - \alpha) [\bar{F}(p_1)q^2 + (\bar{F}(p_2) - 2\bar{F}(p_1))q + \bar{F}(p_1)], \quad (6)$$

where the profit function is derived from the summation of the revenues from the two periods, $p_1 S_1(q)$, and $p_2 S_2(q)$, minus the capacity cost $\alpha C(q)$.

The key question is whether the firm should deliberately create rationing risk for customers. The following proposition answers this question.

PROPOSITION 1. *Strategic capacity rationing is never optimal: 1) If $\frac{p_1 - p_2}{p_2 - \alpha} \geq \frac{\bar{F}(p_2) - \bar{F}(p_1)}{\bar{F}(p_1)}$, then $q^* = 0$ and $C^* = N\bar{F}(p_1)$; 2) If $\frac{p_1 - p_2}{p_2 - \alpha} < \frac{\bar{F}(p_2) - \bar{F}(p_1)}{\bar{F}(p_1)}$, then $q^* = 1$ and $C^* = N\bar{F}(p_2)$.*

If we further assume that customer valuations are uniformly distributed over $[0, \bar{U}]$ as Liu and van Ryzin (2008) do, we have the immediate corollary of Proposition 1:

COROLLARY 1. *Suppose that customer valuations are uniformly distributed over $[0, \bar{U}]$. When $\bar{U} \geq p_1 + p_2 - \alpha$, serving the market only at the high price in period 1 is optimal, $C^* = N\bar{F}(p_1)$. When $\bar{U} < p_1 + p_2 - \alpha$, the optimal strategy is to serve the entire market only at the low price in period 2, $C^* = N\bar{F}(p_2)$.*

This result is new vis-à-vis Corollary 1 and Proposition 4 in Liu and van Ryzin (2008, p. 1122), where they find that when $\bar{U} \geq p_1 + p_2 - \alpha$ and customers are risk averse, it is always optimal to create rationing risk by deliberately understocking. We show that, in the presence of boundedly rational expectations following the $S(1)$ procedure, however, creating rationing risk can *never* be optimal, which is independent of the customers' risk preference.

Note that the customers are still strategic, but are boundedly rational. The key message we want to deliver is that customer rationality behavior, rather than customer strategic behavior, has significant implications on the firm's optimal strategy and profitability in that bounded rationality

can even reverse the policy recommendations. Mathematically, the key driver of our results is the non-concavity of the profit function, whereas the profit function is concave in Liu and van Ryzin (2008, 2011). Indeed, the profit function is convex in our model, yielding the result of rationing being suboptimal. What are the intuitions behind these contrasting results?

Let's first recall the benchmark case when customers have rational expectations, i.e., the fill rate can be *perfectly* anticipated by customers. The key result under rational expectations is that there exists a unique threshold value $v(q)$ for a given fill rate q such that the customers with valuations greater than $v(q)$ purchase in period 1, while those with valuations less than $v(q)$ wait to buy in period 2. This implies that there is a *strict ranking* among all the customers in terms of who would purchase at the regular price versus who would wait for the markdown (which we call *customer segmentation*). When making its capacity decision, the firm trades off between benefits of inducing segmentation and costs of lost sales: there is incremental demand induced in period 1 by lowering q ; however, there is also a lost sales cost of rationing in period 2. When the market consists of a large number of high-value customers, the incremental demand induced in period 1 can more than compensate for the lost sales cost of rationing in period 2. Indeed, Liu and van Ryzin show that for risk-averse customers (i.e., their utility function is concave), the threshold value function $v(q)$ is convex in q . This leads to concavity of the firm's profit function when there are sufficiently large number of high-value customers, i.e., when $\bar{U}$ is large enough.

When customers have boundedly rational expectations (in the sense of anecdotal reasoning), i.e., they cannot perfectly anticipate the fill rate and can only use anecdotes to estimate the true fill rate, the key result that there exists a strict ranking regarding customer segmentation would *no longer* hold here. It is possible that a high-valuation customer would wait if he gets a sample indicating that the fill rate is high in the second period, as illustrated in the right panel of Figure 2. Indeed, when each customer follows the $S(1)$ procedure, regardless how high a customer valuation is, he will wait if his anecdote does not indicate a stock-out. Hence, the firm cannot use the capacity rationing tool to effectively segment the market so that all the high-valuation customers are induced to purchase early while the low-valuation customers wait for markdowns. As a result, when reducing the fill rate in period 2, the benefit of the incremental demand induced in period 1 does not necessarily compensate for the cost of lost sales in period 2. To be precise, as discussed earlier, the incremental demand induced in period 1 is first-order (proportional in the fill rate); while the lost sales cost of rationing in period 2 is second-order, i.e., increases in the fill rate at an increasing rate. The combined *rate of change* in firm's profit due to fill rate change will be dominated by the second-order effect, and thus the incremental demand induced in period 1 *cannot* compensate for the lost sales cost of rationing in period 2.

4. Extensions to the Basic Model

Having shown that it is never optimal to ration capacity in our basic model, we shall investigate the robustness of this finding in the remainder of the paper. In particular, we consider several extensions to the basic model: allowing for multiple samples, including endogenous pricing, and dynamic capacity management over multiple seasons.

4.1. Multiple Samples

The $S(1)$ model suffers from the limitation that a customer has access to only a single sample before making his purchase decision. This extreme assumption may not be appropriate for some real situations where multiple independent samples are available to a customer. In this section, we extend the $S(1)$ model to a general $S(K)$ model where a customer makes his decision based on K independent samples, where $K = 1, 2, 3, \dots$. These samples can be obtained through a variety of channels. For brevity, we discuss only a few below.

Anecdotes. In Section 3, we assumed that one customer only obtains one anecdote, although we allow for the customers' anecdotes to be heterogeneous. Obviously, it is well possible that a customer gets several anecdotes from different persons or sources. Using the same examples as in the introduction, we can think of the situation that "one friend of mine has bought a Chevrolet at a discount in a clearing sale, while the other friends have not." In the setting of airline tickets, the anecdotes can be the following: "both of my two colleagues have waited for cheap flight tickets from Chicago to Bali but didn't get any." With the ubiquity of the online social networking service and microblogging service (e.g., facebook, twitter, etc.), a customer may easily obtain more than a few anecdotes before making his purchase decision. Note that, in the extreme, if every customer obtains an *infinite* number of independent anecdotes, he would be able to *perfectly* estimate the fill rate by simply taking the sample average. That extreme situation corresponds to Liu and van Ryzin (2008). However, the main premise and focus of this paper is on the scenario when K is small, which best approximates anecdotal reasoning in the real life.

Denote $Q_K \equiv \sum_{k=1}^K 1_{A_k}$. Then Q_K follows the binomial distribution with parameters K and q_K , where $q_K = P(1_{A_k} = 1)$ is the fill rate. Denote φ_K as the *fraction* of the customers who purchase in period 1, then

$$\varphi_K = \mathbb{E}P\left(Q_K \leq \frac{Ku(v - p_1)}{u(v - p_2)}\right), \quad (7)$$

where the expectation is taken with respect to the random valuation v and $v - p_1 \geq 0$. **TL: not sure if i understand R1's comments 9 and 12.**

Using the same argument as in the basic model, for any given K , the fill rate q_K solves the equation:

$$q_K = \frac{C - N\varphi_K}{N(1 - \varphi_K - F(p_2))}. \quad (8)$$

The firm chooses capacity C to maximize its expected profit

$$N(p_1 - p_2)\varphi_K + (p_2 - \alpha)C \quad (9)$$

constrained to equation (8). From equation (8), we can express the capacity $C(q_K)$ as a function of the fill rate q_K . We then substitute C and obtain the firm's problem as follows:

$$\max_{q_K \in [0,1]} \Pi(q_K) := N(p_1 - p_2)\varphi_K + (p_2 - \alpha)[N(1 - \varphi_K - F(p_2))q_K + N\varphi_K],$$

where φ_K is given by equation (7).

Following Liu and van Ryzin (2008), we assume that customers' valuations are uniformly distributed over the interval $[0, \bar{U}]$, and customers have the power utility function $u(x) = x^\gamma$ ($0 < \gamma < 1$). For coherence and concreteness, we maintain these two assumptions for the remainder of this section. (However, our qualitative results do not change if we use alternative valuation distributions and utility functions, as confirmed in Section 4.1.2.)

Under these assumptions, the firm's profit maximization problem becomes

$$\begin{aligned} \max_{q_K \in [0,1]} \Pi(q_K) = & N(p_1 - p_2) \int_{p_1}^{\bar{U}} \sum_{n=0}^{\lfloor \frac{K(v-p_1)^\gamma}{(v-p_2)^\gamma} \rfloor} B(n; K, q_K) \frac{1}{\bar{U}} dv + \\ & (p_2 - \alpha) \left[N \left(1 - \int_{p_1}^{\bar{U}} \sum_{n=0}^{\lfloor \frac{K(v-p_1)^\gamma}{(v-p_2)^\gamma} \rfloor} B(n; K, q_K) \frac{1}{\bar{U}} dv - F(p_2) \right) q_K + N \int_{p_1}^{\bar{U}} \sum_{n=0}^{\lfloor \frac{K(v-p_1)^\gamma}{(v-p_2)^\gamma} \rfloor} B(n; K, q_K) \frac{1}{\bar{U}} dv \right], \end{aligned} \quad (10)$$

where $B(n; K, q_K)$ denotes the probability mass function of the binomial distribution with parameters K and q_K . We are interested in whether and when capacity rationing is optimal, i.e., whether the optimal fill rate q_K^* lies in the interval $(0, 1)$. We denote $\varphi(q_K) \equiv \varphi_K = \int_{p_1}^{\bar{U}} \sum_{n=0}^{\lfloor \frac{K(v-p_1)^\gamma}{(v-p_2)^\gamma} \rfloor} B(n; K, q_K) \frac{1}{\bar{U}} dv$, and $\bar{n}_K(v) \equiv \lfloor \frac{K(v-p_1)^\gamma}{(v-p_2)^\gamma} \rfloor$. The following proposition provides sufficient conditions for capacity rationing to be optimal:

PROPOSITION 2. For the problem (10), when $\bar{U} \geq p_1 + p_2 - \alpha$ and

$$\bar{U} > \frac{p_1 - \alpha}{p_2 - \alpha} \int_{p_1}^{\bar{U}} \sum_{n=0}^{\bar{n}_K(v)} \binom{K}{\bar{n}_K(v) + 1} dv + \int_{p_1}^{\bar{U}} \sum_{n=0}^{\bar{n}_K(v)} \binom{K}{n} dv + p_2 \quad (11)$$

hold, capacity rationing is optimal, i.e., $q_K^* \in (0, 1)$, and q_K^* satisfies the first-order condition

$$(p_1 - p_2)\varphi'(q_K^*) + (p_2 - \alpha)[\bar{F}(p_2) + \varphi'(q_K^*)(1 - q_K^*) - \varphi(q_K^*)] = 0,$$

and the optimal capacity level is $C_K^* \equiv N\bar{F}(p_1)(q_K^*)^2 + N(\bar{F}(p_2) - 2\bar{F}(p_1))q_K^* + N\bar{F}(p_1)$.

Note that, in the rational-expectations setting of Liu and van Ryzin (2008), the single condition $\bar{U} \geq p_1 + p_2 - \alpha$ is sufficient to guarantee that capacity rationing is optimal. However, in our setting with boundedly rational expectations, that condition is not sufficient. We find out an additional sufficient condition (11) that depends on the level of rationality K in this proposition. This additional condition requires that there should be enough number of high-value customers in the market for a given K . It is of interest to investigate when this additional condition is satisfied. We carried out an extensive set of numerical studies, in which the parameters are set so that capacity rationing is always optimal if customers have rational expectations (i.e., the condition $\bar{U} \geq p_1 + p_2 - \alpha$ is satisfied). We here provide a few representative examples to illustrate our numerical observations (and refer the reader to the Online Supplement for more numerical studies): (1) we observed that, the finding that capacity rationing is suboptimal is fairly robust to K . Figure 3 shows an example where the parameters are $\bar{U} = 1.8$, $\alpha = 0.35$, $\gamma = 0.75$, $p_1 = 1.2$, $p_2 = 0.7$, and $N = 2000$. (2) We did observe cases where the sufficient conditions in Proposition 2 are satisfied; hence, capacity rationing is optimal. These cases typically happen when K is sufficiently large. For example, Figure 4 shows that capacity rationing is optimal when K is as large as 1000. In those cases, the assumption that customers can access to a large number of “anecdotes” and that they can compute their average appears strong. Hence, the interest of our study focuses on small K , which presumably best approximates customer anecdotal reasoning in real life.

TL: As suggested by R1, we may remove Figure 3, but i think we need to update Figure 4 adding the case of $K=20$.

4.1.1. Heterogeneous Sample Size In this section, we investigate the case in which different customers may have access to different sizes of samples. In other words, customers have heterogeneous sample size K . We adopt a simple approach that captures the heterogeneity in K . For each customer in the population, suppose K is a random variable that follows a two-point distribution over K_1 and K_2 with $P(K = K_1) = \theta$ and $P(K = K_2) = 1 - \theta$. We are interested in how this heterogeneity affects our main finding from our basic model. In the Online Supplement, we present representative examples which show that our main finding from the basic model that capacity rationing is suboptimal with anecdotal reasoning consumers is robust to heterogeneous sample size K .

4.1.2. Other Valuation Distributions and Utility Functions We have also investigated whether our result is robust to different distributions of customer valuations. In the Online Supplement, we present a representative example when the customer valuation is exponentially distributed $F(v) = 1 - e^{-\lambda v}$, $v \geq 0$. The numerical study provides good evidence that our basic finding

Figure 2 Firm Profit $\Pi(q_K)$ as a Function of Fill Rate q_K for Different K ($q_K^* = 0$, $\bar{U} = 1.8$, $\alpha = 0.35$, $\gamma = 0.75$, $p_1 = 1.2$, $p_2 = 0.7$, and $N = 2000$)

Figure 3 Firm Profit $\Pi(q_K)$ as a Function of Fill Rate q_K for Different K ($\bar{U} = 1.8$, $\alpha = 0.35$, $\gamma = 0.65$, $p_1 = 1.3$, $p_2 = 0.8$, and $N = 2000$)

of capacity rationing being suboptimal is robust to the valuation distribution specifications.

Is our result robust to the functional form of customer utility function? What if customers are risk-seeking? To address these questions, we also use other utility functions that allow for

risk-seeking behavior in the Online Supplement. The numerical examples show that our main finding that capacity rationing is suboptimal with anecdotal reasoning consumers is robust to the specification of valuation distributions and utility functions.

4.2. Price Optimization

In the basic model, the prices are fixed. In this section, we investigate scenarios where the firm has the ability to optimize over both prices and capacity. We are interested in whether the finding from the basic model is robust to endogenous pricing, i.e., prices are optimally set and committed. In this case, the firm solves the following optimization problem:

$$\max_{\bar{U} \geq p_1 \geq p_2 \geq \alpha; q \in [0,1]} \Pi(p_1, p_2, q) \equiv N(p_1 - p_2)\varphi + N(p_2 - \alpha)q [\bar{F}(p_2) - \varphi] + N(p_2 - \alpha)\varphi, \quad (12)$$

where $\varphi = (1 - q)\bar{F}(p_1)$.

Following the economics literature (see, e.g., Fudenberg and Villas-Boas 2006), we assume that $(p - \alpha)\bar{F}(p)$ is strictly quasi-concave in p , which is the condition necessary for the existence of a unique local maximum. For expositional convenience, we define $p^* \equiv \arg \max_p (p - \alpha)\bar{F}(p)$. Proposition 3 shows that capacity rationing is not optimal.

PROPOSITION 3. *For the problem (12), the firm's optimal pricing policy is to charge a uniform price equal to p^* , and the optimal capacity level is $N\bar{F}(p^*)$.*

To obtain sharper results and compare with the rational-expectations case in Liu and van Ryzin (2008), we continue to assume a uniform valuation distribution F over $[0, \bar{U}]$. The firm's optimal pricing policy is then to charge a uniform price equal to $\frac{\bar{U} + \alpha}{2}$, and the optimal capacity level is $\frac{N(\bar{U} - \alpha)}{2\bar{U}}$.

Recall that, when customers have rational expectations, Liu and van Ryzin (2008) show that the firm solves the following optimization problem:

$$\max_{\bar{U} \geq v \geq p_1 \geq p_2 \geq \alpha} \frac{N}{\bar{U}} \left((p_1 - \alpha)(\bar{U} - v) + (p_2 - \alpha)(v - p_2) \left(\frac{v - p_1}{v - p_2} \right)^\gamma \right),$$

where $0 < \gamma < 1$, and v is the threshold valuation such that customers with valuations greater than v buy in period 1 and those with valuations less than v wait to buy in period 2. While an explicit solution cannot be obtained for this optimization problem, they show that the optimal pricing strategy is a markdown strategy ($p_1 > p_2$) and the optimal capacity decision is to create a fill rate strictly less than 1 in period 2. Moreover, the optimal capacity level with fully rational customers is strictly higher than the optimal capacity level $N\bar{F}(p^*)$ with anecdotal reasoning customers, and

the firm can earn a higher profit by charging the high-low prices strictly lower than the uniform price in our setting, $p^* > p_1 > p_2$.

Proposition 3 demonstrates that our finding from the basic model is robust to endogenous pricing; capacity rationing is not optimal when customers use anecdotal reasoning and the firm also optimizes over prices. Customer bounded rationality in the sense of anecdotal reasoning weakens the effectiveness of rationing capacity to segment customers, and thus hurts the firm's profit.

What if the firm cannot commit to the prices at the beginning of the regular period? In that case, we look for subgame perfect equilibria by backward induction. First, notice that, for any first period p_1 set, the optimal markdown period price would be $p_2 = p_1 - \epsilon$ for an arbitrarily small ϵ . The reason is that this price extracts the maximum consumer surplus for the firm from the customers who obtained the sample ($1_A = 1$) given $p_2 < p_1$. Hence, the firm would effectively sell in one period only, and the result becomes identical to the outcome characterized in Proposition 3.

In the Online Supplement, we also investigate whether customer utility discounting in the second period affects our finding from the basic model. We find that the result of capacity rationing being suboptimal remains.

4.3. Dynamic Capacity Management

In many situations, it may be more appropriate to allow the firm to dynamically adjust its capacity level for each season. In this section, we shall investigate how this flexibility for the firm affects the result of the optimality of capacity rationing. *A priori*, it appears intuitively possible that the firm may use a cycling strategy to take advantage of consumers. For example, the firm may first select a low fill rate, which sends a signal to customers that they should buy early. However, in the next period, the firm chooses a higher capacity level to satisfy more demand. Then, customers get new signals indicating that they could wait, but the firm reduces capacity again in the next period. Such a “low-high” cycling strategy might be optimal in a dynamic setting. To understand the pattern of the optimal capacity strategy, we have to resort to a dynamic programming model where the firm is allowed to dynamically adjust its capacity level. As before, we first consider the basic model where customers follow the $S(1)$ procedure, we then generalize it to the $S(K)$ framework.

4.3.1. A Single Sample Consider a firm that sells products over repeated sales seasons to different generations of new customers. Each season consists of two selling periods: a full-price (high-price) period and a markdown (low-price) period. There are N customers in each season, who are all *new* and called the generation of customers in that season. At the beginning of each season, the firm makes its capacity decision. Resupply is not possible within a season. As in the capacity

Figure 4 Sequence of Events of Dynamic Capacity Management and Consumer Anecdotal Learning

commitment model discussed earlier, the assumptions regarding customer valuation and behavior are the same. The only difference here is that, the firm can *dynamically* adjust its capacity (or fill rate) for each season. Figure 5 depicts the sequence of events.

During each season t , new customers decide when to buy. They either buy early at the high price p_1 and obtain one unit of product for sure, or they wait for a markdown at price p_2 , but may not obtain the product. We assume that each generation- t customer obtains one anecdote from a customer in the previous generation $t-1$. How each customer makes her purchasing decision is the same as what we described in the capacity commitment model, i.e., the $S(1)$ decision rule.

We denote q_{t-1} as the fill rate of period 2 in season $t-1$. Then the fraction of the generation- t customers who purchase in period 1 during season t can be written as

$$\varphi_t \equiv (1 - q_{t-1})\bar{F}(p_1). \quad (13)$$

Consequently, the fill rate in period 2 of season t can be determined by

$$q_t \equiv \frac{C_t - N\varphi_t}{N(1 - \varphi_t - F(p_2))}, \quad (14)$$

where C_t is the firm's capacity level set for season t .

Let $\pi(q_{t-1}, q_t)$ be the firm's one-stage profit, given that the fill rate in the previous period is q_{t-1} and the firm's capacity decision induces the fill rate q_t for season t . We have

$$\pi(q_{t-1}, q_t) = N(p_1 - p_2)\varphi_t + (p_2 - \alpha)C_t. \quad (15)$$

Let $V(q_{t-1})$ denote the maximum discounted profit, given that customers experience the true fill rate q_{t-1} in the previous period (q_{t-1} is the state in season t). Future profit is discounted by a discount factor of δ ($0 < \delta < 1$) per season. We have $V(q_{t-1})$ satisfying the following Bellman equation:

$$V(q_{t-1}) = \max_{q_t \in [0,1]} \{\pi(q_{t-1}, q_t) + \delta V(q_t)\}. \quad (16)$$

Combining equations (13), (14), (15) and (16), we obtain the Bellman equation

$$V(q_{t-1}) = \max_{q_t \in [0,1]} \{N(p_1 - p_2)(1 - q_{t-1})\bar{F}(p_1) + N(p_2 - \alpha) [(1 - q_{t-1})\bar{F}(p_1) + q_t (1 - (1 - q_{t-1})\bar{F}(p_1) - F(p_2))] + \delta V(q_t)\}. \quad (17)$$

We are first interested in whether there exists an optimal stationary policy.

PROPOSITION 4. *For the problem (17), given any initial state $q_1 \in [0, 1]$, there exists a monotone optimal state path.*

By Proposition 4, the optimal fill rate policy converges to a constant value (called stationary or steady state) as a bounded monotone sequence must converge. The immediate question to ask is whether capacity rationing can be the optimal stationary policy. The following proposition shows that it is never optimal to ration capacity.

PROPOSITION 5. *For the problem (17), the optimal fill rate eventually converges to either 0 or 1, i.e., the optimal stationary capacity level is $N\bar{F}(p_1)$ or $N\bar{F}(p_2)$.*

This proposition shows that capacity rationing cannot be the optimal stationary policy even with dynamic capacity management.

4.3.2. Multiple Samples In this section, we extend the $S(1)$ model to a general $S(K)$ model where a customer makes his decision based on K ($K \geq 1$) independent samples. Suppose a generation- t customer observes K independent samples or anecdotes from generation- $(t-1)$ customers. Let $1_{A_k, t-1}$ be a binary random variable to denote whether the sample/signal indicates availability or stockout in period 2 of season $t-1$, $k = 1, 2, \dots, K$. Again, we assume each customer combines these samples by simply taking the sample average, i.e., he purchases the product in period 1 at season t if and only if

$$u(v - p_1) \geq \frac{1}{K} \sum_{k=1}^K 1_{A_k, t-1} u(v - p_2) \quad (18)$$

and $v - p_1 \geq 0$. Note that, as $K \rightarrow \infty$, the sample average $\frac{1}{K} \sum_{k=1}^K 1_{A_k, t-1}$ becomes the firm's true fill rate q_{t-1} .

Denote $Q_{K,t-1} \equiv \sum_{k=1}^K 1_{A_{k,t-1}}$. Then $Q_{K,t-1}$ follows the binomial distribution with parameters K and $q_{K,t-1}$, where $q_{K,t-1} = P(1_{A_{k,t-1}} = 1)$ is the fill rate at season $t-1$. Denote $\varphi_{K,t}$ as the fraction of the customers who purchase in period 1 of season t , then

$$\varphi_{K,t} = \mathbb{E}P\left(Q_{K,t-1} \leq \frac{Ku(v-p_1)}{u(v-p_2)}\right), \quad (19)$$

where the expectation is taken with respect to the random valuation v .

Following the same argument as before, for any given K , the fill rate $q_{K,t}$ solves the equation:

$$q_{K,t} = \frac{C_{K,t} - N\varphi_{K,t}}{N(1 - \varphi_{K,t} - F(p_2))}, \quad (20)$$

where $C_{K,t}$ is the firm's capacity level at season t .

Let $\pi(q_{K,t-1}, q_{K,t})$ be the firm's one-stage profit, given that the fill rate in the previous period is $q_{K,t-1}$ and the firm's capacity decision induces the fill rate $q_{K,t}$ for season t (i.e, the state is $q_{K,t-1}$ and the decision variable is $q_{K,t}$ in season t). We have

$$\pi(q_{K,t-1}, q_{K,t}) = N(p_1 - p_2)\varphi_{K,t} + (p_2 - \alpha)C_{K,t}. \quad (21)$$

Recall we have assumed that customers' valuations are uniformly distributed over the interval $[0, \bar{U}]$. Customers have the power utility function $u(x) = x^\gamma$ ($0 < \gamma < 1$). Under these assumptions, equation (19) can be further expressed as

$$\varphi_{K,t} = \varphi(q_{K,t-1}) = \int_{p_1}^{\bar{U}} \sum_{n=0}^{\lfloor \frac{K(v-p_1)^\gamma}{(v-p_2)^\gamma} \rfloor} B(n; K, q_{K,t-1}) \frac{1}{\bar{U}} dv. \quad (22)$$

Let $V(q_{K,t-1})$ denote the maximum discounted profit, given that customers experience the fill rate $q_{K,t-1}$ in the previous period. Future value is discounted by a discount factor of δ per season. Then $V(q_{K,t-1})$ satisfies the following Bellman equation:

$$V(q_{K,t-1}) = \max_{q_{K,t} \in [0,1]} \{\pi(q_{K,t-1}, q_{K,t}) + \delta V(q_{K,t})\}. \quad (23)$$

Combining equations (20) - (23), we obtain the Bellman equation:

$$V(q_{K,t-1}) = \max_{q_{K,t} \in [0,1]} \{N(p_1 - p_2)\varphi(q_{K,t-1}) + N(p_2 - \alpha)[(1 - \varphi(q_{K,t-1}) - F(p_2))q_{K,t} + \varphi(q_{K,t-1})] + \delta V(q_{K,t})\}. \quad (24)$$

We aim to characterize the optimal stationary policy, if it exists, for this dynamic programming problem.

PROPOSITION 6. *For the problem (24), given any initial state $q_{K,1} \in [0, 1]$, there exists a monotone optimal state path.*

By Proposition 6, the optimal fill rate policy converges to a stationary state as a bounded monotone sequence must converge. We now provide a sufficient condition under which capacity rationing is the optimal stationary policy.

PROPOSITION 7. *For the problem (24), when $\bar{U} \geq p_1 + p_2 - \alpha$ and*

$$\bar{U} > \frac{1}{(p_2 - \alpha)\bar{F}(p_2)} \left[(1 + \delta(p_1 - p_2)) \int_{p_1}^{\bar{U}} \sum_{n=0}^{\bar{n}_K(v)} \binom{K}{\bar{n}_K(v) + 1} dv + (p_2 - \alpha) \int_{p_1}^{\bar{U}} \sum_{n=0}^{\bar{n}_K(v)} \binom{K}{n} dv \right] \quad (25)$$

hold, capacity rationing is an optimal stationary policy, denoted by $q_K^0 \in (0, 1)$, and q_K^0 satisfies the Euler equation

$$\delta(p_1 - p_2)\varphi'(q_K^0) + (p_2 - \alpha) [\bar{F}(p_2) + \varphi'(q_K^0)(1 - q_K^0) - \varphi(q_K^0)] = 0,$$

and the optimal stationary capacity level is $C_K^0 \equiv N\bar{F}(p_1)(q_K^0)^2 + N(\bar{F}(p_2) - 2\bar{F}(p_1))q_K^0 + N\bar{F}(p_1)$.

When is the condition (25) more likely to hold? We conduct an extensive numerical study to answer this question. We use the following parameters: $\bar{U} = 2$, $\alpha = 0.2$, $\gamma = 0.75$, $p_1 = 1$, $p_2 = 0.8$, and $N = 1000$ (used in Liu and van Ryzin 2011). In these numerical examples, we increase the level of consumer rationality (i.e., the number of samples) K from 1 up to 1000, which we believe is a large enough number for any practical setting. We report our numerical results in Table 1, for various combinations of the discount factor and the level of consumer rationality. For example, for the discount factor lower than 0.7, capacity rationing is not optimal even for K up to 1000. When $\delta = 0.8$, capacity rationing becomes optimal when K is 89 or larger. When $\delta = 0.9$, capacity rationing becomes optimal when K is 80 or larger.

For the same set of data as that for the capacity commitment setting, we also carried out numerical investigations. In Table 2, we report the results for the data: $\bar{U} = 1.8$, $\alpha = 0.35$, $\gamma = 0.65$, $p_1 = 1.3$, $p_2 = 0.8$, and $N = 2000$.

From these numerical examples, we have the following observations: First, whether capacity rationing is an optimal stationary policy crucially depends on the level of consumer rationality K . In majority of the cases in our numerical study, the level of consumer rationality K has to be fairly large for capacity rationing being optimal. This indicates some robustness of our previous results. Second, note that for cases when capacity rationing being optimal, the firm's discount factor is relatively large. It appears that, as the firm's discount factor δ becomes sufficiently close to one, capacity rationing is more likely to an optimal stationary policy. Interestingly, this result shares a similar theme as Liu and van Ryzin (2011) who show that the equilibrium under adaptive learning converges to that under rational expectations as the discount factor approaches one.

Table 1 Optimal stationary policies for different discount factor δ and level of rationality K ($\bar{U} = 2$, $\alpha = 0.2$, $\gamma = 0.75$, $p_1 = 1$, $p_2 = 0.8$, and $N = 1000$)

Discount factor δ	Level of rationality K	Capacity rationing?	Optimal fill rate q_K^0
0.4	$\{1, 2, \dots, 1000\}$	No	1
0.5	$\{1, 2, \dots, 1000\}$	No	1
0.6	$\{1, 2, \dots, 1000\}$	No	1
0.7	$\{1, 2, \dots, 1000\}$	No	1
0.8	$\{1, 2, \dots, 88\}$	No	1
0.8	$\{89, 90, \dots, 1000\}$	Yes	$\{0.7150, 0.7150\}$
0.9	$\{1, 2, \dots, 79\}$	No	1
0.9	$\{80, 81, \dots, 1000\}$	Yes	$\{0.6825, 0.6825\}$

Table 2 Optimal stationary policies for different discount factor δ and level of rationality K ($\bar{U} = 1.8$, $\alpha = 0.35$, $\gamma = 0.65$, $p_1 = 1.3$, $p_2 = 0.8$, and $N = 2000$)

Discount factor δ	Level of rationality K	Capacity rationing?	Optimal fill rate q_K^0
0.4	$\{1, 2, \dots, 2000\}$	No	1
0.5	$\{1, 2, \dots, 2000\}$	No	1
0.6	$\{1, 2, \dots, 2000\}$	No	1
0.7	$\{1, 2, \dots, 2000\}$	No	1
0.8	$\{1, 2, \dots, 2000\}$	No	1
0.9	$\{1, 2, \dots, 2000\}$	No	1
0.99	$\{1, 2, \dots, 9\}$	No	1
0.99	$\{10, 11, \dots, 1000\}$	Yes	$\{0.2133, 0.2133, \dots, 0.2733\}$

5. Concluding Remarks

In this paper, we investigated how customer boundedly rational expectations, in the sense of anecdotal reasoning, impact the firm's optimal capacity decision in the concrete contexts of both capacity commitment and dynamic capacity management. With anecdotal reasoning customers using the $S(1)$ procedure, we proved that it is never optimal to strategically ration capacity, although customers are still strategic, regardless of whether the firm commits to its capacity or dynamically changes it. This result does not depend on customer risk preference. We then extended the restrictive $S(1)$ model to the general $S(K)$ model where the parameter K denotes the number of independent samples capturing the level of rationality in customer expectations. The advantage of this model is that it includes the rational-expectations model as a special case. Based on this analytical framework, we were able to provide sufficient conditions for capacity rationing to be optimal for a given level of rationality K , for both capacity commitment and dynamic adjustment settings. From our extensive numerical studies, we found that, for a large range of K , capacity rationing is not optimal. We believe that this range represents many (if not all) realistic settings when customers make their purchasing decisions, especially in the settings of selling durable goods. We show that the result of strategic capacity rationing being suboptimal is fairly robust to different

valuation distributions and utility functions, heterogeneous sample size, price optimization, and discounted utility. Our results thus demonstrate the peril of strategic capacity rationing when customers have boundedly rational expectations, and are complementary to the extant capacity management literature. For managers and practitioners, the results suggest the importance of carefully investigating customer behavior when a firm designs its dynamic pricing and capacity management systems and implements corresponding strategies.

Admittedly, while being among the first in the operations literature, this study only scratches the surface of how to appropriately incorporate customer bounded rationality in practical operations management settings. As we have argued, anecdotal reasoning is a commonly and naturally observed behavioral phenomenon in the real life. However, we acknowledge that there are several other alternatives regarding how to model boundedly rational expectations, e.g., the “level- k thinking” proposed by Stahl and Wilson (1995), and the “experienced-weighted attraction learning” model proposed by Camerer and Ho (1999). Compared to the $S(K)$ model, these models may not be easily applied to the setting we studied here. First, it is unclear how to model a fully rational firm and boundedly rational customers using these alternative models. Second, these models do not possess a clean structure that captures the level of rationality in customer expectations that simply includes the rational-expectations as a special case. Regardless of these disadvantages, they remain a good potential for future exploration. We hope that this study stimulates more future research in this area.

6. Acknowledgement

The work of Qian Liu was supported by the Hong Kong Research Grant Council (Grants RGC615911).

References

- Aumann, R. J., J.H. Dreze. 2008. Rational expectations in games. *American Economic Review* 98(1) 72-86.
- Aviv, Y, A. Pazgal. 2008. Optimal pricing of seasonal products in the presence of forward-looking consumers. *Manufacturing Service Oper. Management* 10(3) 339-359.
- Bendoly, E., K. Donohue, K. Schultz. 2006. Behavioral operations management: Assessing recent findings and revisiting old assumptions. *J. Oper. Management* 24(6) 737-752.
- Bitran, R.G., R. Caldentey. 2003. Pricing models for revenue management. *Manufacturing Service Oper. Management* 5(3) 203-229.
- Blinder, A., E. Canetti, D. Lebow, J. Rudd. 1998. *Asking About Prices: A New Approach to Understanding Price Stickiness*. Russell Sage Foundation, New York.

- Camerer, C.F., G. Lowenstein. 2003. Behavioral economics: Past, present, future. In *Advances in Behavioral Economics* (Colin F. Camerer, George Lowenstein, and Matthew Rabin, eds.), 3-51, Princeton University Press, Princeton, New Jersey.
- Cachon, G. P., R. Swinney. 2009. Purchasing, pricing and quick response in the presence of strategic customers. *Management Science* 55(3) 497-511.
- Camerer, C., T.-H. Ho. 1999. Experienced-weighted attraction learning in normal form games. *Econometrica* 67(4) 827-874.
- Fraiman, N., M. Singh, L. Arrington, C. Paris. 2008. ZARA. Columbia Business School Caseworks, Columbia University, New York.
- Fudenberg, D., J. M. Villas-Boas. 2006. Behavior-based price discrimination and customer recognition. *Handbook on Economics and Information Systems* 1 377-436.
- Gallego, G., R. Phillips, Ö. Şahin. 2008. Strategic management of distressed inventory. *Production and Operations Management* 17(4) 402-415.
- Gilboa, I., D. Schmeidler. 2001. *A Theory of Case-Based Decisions*. Cambridge: Cambridge University Press.
- Huang, T., G. Allon, A. Bassamboo. 2013. Bounded rationality in service systems. *Manufacturing Service Oper. Management* 15(2) 263-279.
- Huang, T., Y.-J. Chen. 2014. Service systems with experience-based anecdotal reasoning consumers. *Production and Operations Management*, Published online in *Articles in Advance*.
- Huang, T., Y. Yu. 2014. Sell probabilistic goods? A behavioral explanation for opaque selling. *Marketing Science*, Published online in *Articles in Advance*.
- Kalkanci, B., G. Perakis. 2013. Bounded rationality in supply chain interactions. Working paper, Available at SSRN: <http://ssrn.com/abstract=2361999>.
- Kremer, M., L. G. Debo. 2012. Herding in a queue: A laboratory experiment. Chicago Booth Research Paper.
- Lai, G., L. G. Debo, K. Sycara. 2010. Buy now and match later: The impact of posterior price matching on profit with strategic consumers. *Manufacturing Service Oper. Management* 12(1) 33-55.
- Li, J., N. Granados, S. Netessine. 2011. Are consumers strategic? Structural estimation from the air-travel industry. Working paper, The Wharton School, University of Pennsylvania, Philadelphia, PA.
- Liu, Q., G. van Ryzin. 2008. Strategic capacity rationing to induce early purchases. *Management Sci.* 54(6) 1115-1131.
- Liu, Q., G. van Ryzin. 2011. Strategic capacity rationing when customers learn. *Manufacturing Service Oper. Management* 13(1) 89-107.
- Mersereau, A., D. Zhang. 2012. Markdown pricing with unknown fraction of strategic customers. *Manufacturing Service Oper. Management* 14(3) 355-370.

- Nasiry, J., I. Popescu. 2011. Dynamic pricing with loss-averse consumers and peak-end anchoring. *Operations Research* 59(6) 1361-1368.
- Netessine, S., C. Tang, eds. 2009. *Consumer-driven Demand and Operations Management Models*. Springer Publisher, New York, NY.
- Osborne, M. J., A. Rubinstein. 1998. Games with procedurally rational players. *American Economic Review* 88 834-849.
- Osborne, M. J., A. Rubinstein. 2003. Sampling equilibrium with an application to strategic voting. *Games and Economic Behavior* 45 434-441.
- Ovchinnikov, A., J. M. Milner. 2012. Revenue management with end-of-period discounts in the presence of customer learning. *Production and Operations Management* 21(1) 69-84.
- Özer, Ö., Y. Zheng. 2014. Markdown or everyday-low-price? The role of consumer regret and availability misperception. *Management Science*, Forthcoming.
- Rabin, M. 2002. Inference by believers in the law of small numbers. *Quarterly Journal of Economics* 117 775-816.
- Rubinstein, A. 1998. *Modeling Bounded Rationality*. MIT Press, Cambridge, MA.
- Spiegler, R. 2006a. The market for quacks. *Review of Economic Studies* 73(4) 1113-1131.
- Spiegler, R. 2006b. Competition over agents with boundedly rational expectations. *Theoretical Economics* 1(2) 207-231.
- Spiegler, R. 2011. *Bounded Rationality and Industrial Organization*. Oxford University Press, New York, NY.
- Stahl, D., P. Wilson. 1995. On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior* 10 218-254.
- Su, X. 2007. Inter-temporal pricing with strategic customer behavior. *Management Science* 53(5) 726-741.
- Su, X. 2008. Bounded rationality in newsvendor models. *Manufacturing Service Oper. Management* 10(4) 566-589.
- Sundaram, R. K. 1996. *A First Course in Optimization Theory*. Cambridge university press.
- Szech, N. 2011a. Becoming a bad doctor. *Journal of Economic Behavior and Organization* 80 244-257.
- Szech, N. 2011b. Welfare in markets where consumers rely on word of mouth. Working paper, University of Bonn.
- Talluri, K., G. van Ryzin. 2004. *The Theory and Practice of Revenue Management*, Springer-Verlag/Kluwer Academic Publishers.
- Tversky, A., D. Kahneman. 1971. Belief in the law of small numbers. *Psychological Bulletin* 2 105-110.
- Van Mieghem, J. A. 2003. Commissioned paper: Capacity management, investment, and hedging: Review and recent developments. *Manufacturing & Service Operations Management* 5(4) 269-302.

- Veeraraghavan, S. K., L. G. Debo. 2011. Herding in queues with waiting costs: Rationality and regret. *Manufacturing Service Oper. Management* 13(3) 329-346.
- Yin, R., Y. Aviv, A. Pazgal, C. S. Tang. 2009. Optimal markdown pricing: Implications of inventory display formats in the presence of strategic customers. *Management Sci.* 55(8) 1391-1408.
- Zhang, D., W. L. Cooper. 2008. Managing clearance sales in the presence of strategic customers. *Production and Operations Management* 17 416-431. doi: 10.3401/poms.1080.0039

Appendix. Proofs

Proof of Lemma 1. Taking the limit of inequality (1) as K goes to ∞ , we have

$$u(v - p_1) \geq \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K 1_{A_k} u(v - p_2) = qu(v - p_2),$$

which becomes identical to the customer decision problem in Section 2.1 of Liu and van Ryzin (2008). $\square$

Proof of Proposition 1. We solve the firm's profit maximization problem (6). It is clear that $\Pi(q)$ is strictly convex in q . The maximum can only be achieved at the boundary $q = 0$ or $q = 1$, which correspond to the boundary points for the capacity levels. Therefore, $C^* \in \{N\bar{F}(p_1), N\bar{F}(p_2)\}$.

If $\frac{p_1 - p_2}{p_2 - \alpha} \geq \frac{\bar{F}(p_2) - \bar{F}(p_1)}{\bar{F}(p_1)}$, then $\Pi(N\bar{F}(p_1)) \geq \Pi(N\bar{F}(p_2))$. Therefore, $q^* = 0$ and $C^* = N\bar{F}(p_1)$. Otherwise, the reverse holds. $\square$

Lemma 2 characterizes the impact of the fill rate change in period 2 on the fraction of customers who purchase in period 1, which will be used in proving subsequent propositions.

$$\text{LEMMA 2. } \frac{d\varphi(q_K)}{dq_K} = -K \int_{p_1}^{\bar{U}} B(\bar{n}_K(v); K-1, q_K) \frac{1}{\bar{U}} dv.$$

Proof of Lemma 2. Taking the first-order derivative, we obtain:

$$\frac{d\varphi(q_K)}{dq_K} = \int_{p_1}^{\bar{U}} \sum_{n=0}^{\bar{n}_K(v)} \binom{K}{n} q_K^{n-1} (1 - q_K)^{K-n-1} (n - Kq_K) \frac{1}{\bar{U}} dv.$$

For ease of exposition, we introduce some notations: Let $A(q_K) \equiv \sum_{n=1}^{\bar{n}_K(v)} n \binom{K}{n} q_K^{n-1} (1 - q_K)^{K-n-1}$ and $B(q_K) \equiv K \sum_{n=0}^{\bar{n}_K(v)} \binom{K}{n} q_K^n (1 - q_K)^{K-n-1}$. Some algebra yields

$$A(q_K) = \frac{K}{1 - q_K} F_B(\bar{n}_K(v) - 1; K - 1, q_K),$$

where $F_B(\bar{n}_K(v) - 1; K - 1, q_K)$ is the cumulative distribution function of the binomial distribution with parameters $K - 1$ and q_K . Similarly, we have $B(q_K) = \frac{K}{1 - q_K} F_B(\bar{n}_K(v); K, q_K)$. We thus obtain

$$A(q_K) - B(q_K) = \frac{K}{1 - q_K} [F_B(\bar{n}_K(v) - 1; K - 1, q_K) - F_B(\bar{n}_K(v); K, q_K)].$$

Using the Pascal's rule

$$\binom{K}{n} = \binom{K-1}{n-1} + \binom{K-1}{n}$$

for $n \geq 1$ and some algebra, we have

$$A(q_K) - B(q_K) = -\frac{K}{1-q_K} \binom{K-1}{\bar{n}_K(v)} q_K^{\bar{n}_K(v)} (1-q_K)^{K-\bar{n}_K(v)} = -KB(\bar{n}_K(v); K-1, q_K).$$

Substituting it back into the equation for $\frac{d\varphi(q_K)}{dq_K}$, we have proved the lemma. $\square$

Proof of Proposition 2. The condition $\bar{U} \geq p_1 + p_2 - \alpha$ implies that $\Pi(0) = N(p_1 - \alpha)\bar{F}(p_1) \geq \Pi(1) = N(p_2 - \alpha)\bar{F}(p_2)$, i.e., the profit at the high-price-only solution is higher than that at the low-price-only solution.

Taking the first-order derivative of the profit function, we obtain

$$\Pi'(q_K) = N(p_1 - p_2)\varphi'(q_K) + N(p_2 - \alpha) [\bar{F}(p_2) + \varphi'(q_K)(1 - q_K) - \varphi(q_K)].$$

It is clear that, if $\Pi'(0) > 0$, then there exists some $q_K^* \in (0, 1)$ such that $\Pi(q_K^*) > \Pi(0)$. Notice that

$$\varphi'(0) = -K \int_{p_1}^{\bar{U}} \binom{K-1}{\bar{n}_K(v)} \frac{1}{\bar{U}} dv$$

and

$$\varphi(0) = \int_{p_1}^{\bar{U}} \sum_{n=0}^{\bar{n}_K(v)} \binom{K}{n} \frac{1}{\bar{U}} dv.$$

Substituting them into the expression of $\Pi'(0)$, we obtain the condition in this proposition after simplifying algebraic operations. It is clear that, the necessary first-order condition is satisfied at the optimal fill rate. $\square$

Proof of Proposition 3. Recall that we have shown that for any fixed prices p_1 and p_2 , capacity rationing cannot be optimal. Hence, $q^* = 0$ or 1 . We only need to find the optimal price in each of the two cases. If $q^* = 0$, then $\Pi(p_1, p_2, q) = N(p_1 - \alpha)\bar{F}(p_1)$. Clearly, p^* is the profit-maximizing price. If we further assume F is a uniform distribution, we have the following sharper results: $\Pi(p_1, p_2, q) = \frac{N}{\bar{U}}(p_1 - \alpha)(\bar{U} - p_1)$. First-order condition yields $p_1^* = \frac{\bar{U} + \alpha}{2}$. Similarly, if $q^* = 1$, we obtain $p_2^* = \frac{\bar{U} + \alpha}{2}$. Under either case, we have the same optimal profit $\frac{N(\bar{U} - \alpha)^2}{4\bar{U}}$. $\square$

Proof of Proposition 4. Because

$$\frac{\partial^2 \pi}{\partial q_{t-1} \partial q_t} = \bar{F}(p_1) > 0,$$

the firm's one-stage profit $\pi(q_{t-1}, q_t)$ is supermodular in (q_{t-1}, q_t) . Hence, the inner maximization term in the problem (17) is supermodular in (q_{t-1}, q_t) . By Theorem 10.6 in Sundaram (1996), the

optimal fill rate decision q_t^* satisfies $q_t^*(q_{t-1}^A) \geq q_t^*(q_{t-1}^B)$ for any given $q_{t-1}^A > q_{t-1}^B$. This implies that the optimal state path is monotone. $\square$

Proof of Proposition 5. Suppose the stationary state, denoted by q^0 , is an interior solution; that is, $0 < q^0 < 1$. Then q^0 satisfies the following Euler equation:

$$\left. \frac{\partial \pi(q_{t-1}, q_t)}{\partial q_t} \right|_{(q^0, q^0)} + \delta \left. \frac{\partial \pi(q_t, q_{t+1})}{\partial q_t} \right|_{(q^0, q^0)} = 0.$$

Hence, we have

$$q^0 = \frac{\delta \bar{F}(p_1)(p_1 - \alpha) - (p_2 - \alpha)(F(p_1) - F(p_2))}{(1 + \delta)\bar{F}(p_1)(p_2 - \alpha)}. \quad (26)$$

That $0 < q^0 < 1$ requires both

$$\frac{\bar{F}(p_1)}{F(p_1) - F(p_2)} > \frac{p_2 - \alpha}{\delta(p_1 - \alpha)}, \quad (27)$$

and

$$\frac{\bar{F}(p_1)}{\bar{F}(p_2)} < \frac{p_2 - \alpha}{\delta(p_1 - p_2)}.$$

When q^0 is the stationary policy, we can easily derive the maximum discounted profit for a given state q^0 :

$$V(q^0) = \frac{N}{1 - \delta} [(p_1 - \alpha)\bar{F}(p_1)(1 - q^0) + (p_2 - \alpha)q^0(q^0\bar{F}(p_1) + F(p_1) - F(p_2))]. \quad (28)$$

We now claim that an interior solution $0 < q^0 < 1$ cannot be the stationary policy. The reason is the following: for a given state q^0 , consider the policy $\{0, 0, \dots\}$ over time (i.e., the firm sets a zero fill rate in all the following seasons). We denote the discounted profit under this policy by $\hat{V}(q^0)$, which can be determined below.

$$\hat{V}(q^0) = N(p_1 - p_2)(1 - q^0)\bar{F}(p_1) + N(p_2 - \alpha)(1 - q^0)\bar{F}(p_1) + \frac{\delta}{1 - \delta} N\bar{F}(p_1)(p_1 - \alpha). \quad (29)$$

Comparing (28) and (29), we have

$$\begin{aligned} \hat{V}(q^0) - V(q^0) &= \frac{Nq^0}{1 - \delta} [\delta\bar{F}(p_1)(p_1 - \alpha) - (p_2 - \alpha)(q^0\bar{F}(p_1) + F(p_1) - F(p_2))] \\ &= \frac{Nq^0}{1 - \delta} \left[\delta\bar{F}(p_1)(p_1 - \alpha) - (p_2 - \alpha) \left(\frac{\delta\bar{F}(p_1)(p_1 - \alpha)}{(1 + \delta)(p_2 - \alpha)} + \frac{\delta}{1 + \delta} (F(p_1) - F(p_2)) \right) \right] \\ &= \frac{\delta Nq^0}{1 - \delta^2} [\bar{F}(p_1)(p_1 - \alpha) - (p_2 - \alpha)(F(p_1) - F(p_2))] \\ &> 0. \end{aligned}$$

In above, the second equality comes from (26), and the last inequality follows from (27). Thus we can conclude that a rationing policy $0 < q^0 < 1$ cannot be sustained as the stationary policy. $\square$

Proof of Proposition 6. Because

$$\frac{\partial^2 \pi}{\partial q_{K,t-1} \partial q_{K,t}} = -N(p_2 - \alpha) \frac{d\varphi(q_{K,t-1})}{dq_{K,t-1}} = N(p_2 - \alpha) K \int_{p_1}^{\bar{U}} B(\bar{n}_K(v); K-1, q_{K,t-1}) \frac{1}{\bar{U}} dv > 0,$$

where the second equality comes from the same argument as Lemma 2. Thus, the firm's one-stage profit $\pi(q_{K,t-1}, q_{K,t})$ is supermodular in $(q_{K,t-1}, q_{K,t})$. Hence, the inner maximization term in the problem (24) is supermodular in $(q_{K,t-1}, q_{K,t})$ as well. This implies that the state path is monotone.

□

Proof of Proposition 7. The condition $\bar{U} \geq p_1 + p_2 - \alpha$ implies that the one-stage profit at the high price only (i.e., $q = 0$) is greater than that at the low price only (i.e., $q = 1$) in the stationary state.

Suppose that the stationary state, denoted by q_K^0 , is an interior optimal solution, i.e., $0 < q_K^0 < 1$. Then q_K^0 satisfies the following Euler equation:

$$e(q_{K,t-1}, q_{K,t}) \Big|_{(q_K^0, q_K^0)} \equiv \frac{\partial \pi(q_{K,t-1}, q_{K,t})}{\partial q_{K,t}} \Big|_{(q_K^0, q_K^0)} + \delta \frac{\partial \pi(q_{K,t}, q_{K,t+1})}{\partial q_{K,t}} \Big|_{(q_K^0, q_K^0)} = 0.$$

If at point $(0, 0)$, we have

$$e(q_{K,t-1}, q_{K,t}) \Big|_{(0,0)} > 0, \quad (30)$$

then there exists a stationary policy $\tilde{q}_K \in (0, 1)$ such that the one-stage profit at this state is higher than that at the high price only. Together with the earlier result that the one-stage profit at $q = 0$ is higher than that at $q = 1$, this indicates that capacity rationing is the optimal stationary policy.

Inequality (30) is equivalent to

$$\delta(p_1 - p_2)\varphi'(0) + (p_2 - \alpha) [\bar{F}(p_2) + \varphi'(0) - \varphi(0)] > 0. \quad (31)$$

Recall that

$$\varphi'(0) = -K \int_{p_1}^{\bar{U}} \binom{K-1}{\bar{n}_K(v)} \frac{1}{\bar{U}} dv = - \int_{p_1}^{\bar{U}} \sum_{n=0}^{\bar{n}_K(v)} \binom{K}{\bar{n}_K(v) + 1} dv,$$

where $\bar{n}_K(v) \equiv \lfloor \frac{K(v-p_1)^\gamma}{(v-p_2)^\gamma} \rfloor$, and

$$\varphi(0) = \int_{p_1}^{\bar{U}} \sum_{n=0}^{\bar{n}_K(v)} \binom{K}{n} \frac{1}{\bar{U}} dv.$$

Substituting them into inequality (31) and simplifying, we obtain the condition (25). □