Intertemporal Pricing With Strategic Customer Behavior

Xuanming Su
University of Pennsylvania

Intertemporal Pricing With Strategic Customer Behavior

Abstract

This paper develops a model of dynamic pricing with endogenous intertemporal demand. In the model, there is a monopolist who sells a finite inventory over a finite time horizon. The seller adjusts prices dynamically to maximize revenue. Customers arrive continually over the duration of the selling season. At each point in time, customers may purchase the product at current prices, remain in the market at a cost to purchase later, or exit, and they wish to maximize individual utility. The customer population is heterogeneous along two dimensions: they may have different valuations for the product and different degrees of patience (waiting costs).

We demonstrate that heterogeneity in both valuation and patience is important because they jointly determine the structure of optimal pricing policies. In particular, when high-value customers are proportionately less patient, markdown pricing policies are effective because the high-value customers would buy early at high prices while the low-value customers are willing to wait (i.e., they are not lost). On the other hand, when the high-value customers are more patient than the low-value customers, prices should increase over time to discourage inefficient waiting. Contrary to intuition, we find that strategic waiting by customers may sometimes benefit the seller: when low-value customers wait, they compete for availability with high-value customers and thus increase their willingness to pay. Our results also shed light on how the composition of the customer population affects optimal revenue, consumer surplus, and social welfare. Finally, we consider the long-run problem of selecting the optimal initial stocking quantity.

Keywords

Disciplines

Other Business | Other Social and Behavioral Sciences | Sales and Merchandising

Inter-temporal Pricing with Strategic Customer Behavior

Xuanming Su

Haas School of Business, University of California, Berkeley, CA 94720

Abstract

This paper develops a model of dynamic pricing with endogenous customer behavior. In the model, there is a monopolist who sells a finite inventory over a finite time horizon. The seller adjusts prices dynamically in order to maximize revenue. Customers arrive continually over the duration of the selling season. At each point in time, customers may purchase the product at current prices, remain in the market at a cost in order to purchase later, or exit, and they wish to maximize individual utility. The customer population is heterogeneous along two dimensions: they may have different valuations for the product and different degrees of patience (waiting costs). We study this continuous-time game between the seller and the customers, show that it can be reduced into a single-variable nonlinear program, and characterize the equilibrium that maximizes revenue for the seller.

We demonstrate that heterogeneity in both valuation and patience is important because they jointly determine the structure of optimal pricing policies. In particular, when high-value customers are proportionately less patient, markdown pricing policies are effective because the high-value customers would still buy early at high prices while the low-value customers are willing to wait (i.e. they are not lost). On the other hand, when the high-value customers are more patient than the low-value customers, prices should increase over time in order to discourage inefficient waiting. Our results also shed light on how the composition of the customer population affects optimal revenue, consumer surplus, and social welfare. Finally, we consider the long run problem of selecting the optimal initial stocking quantity.

July 2005, revised February 2006, June 2006

1 Introduction

Pricing is one of the most fundamental but also most difficult decisions that firms have to make. An important reason is that businesses today are facing a generation of increasingly sophisticated customers. Whether firms adopt the most powerful category pricing software or the most data-intensive revenue management system, customers are becoming extremely adept at finding the "best deals." Readers of this article may even have some personal experiences that they are proud to share. According to *SmartMoney* magazine, there is constantly a "cat-and-mouse game between retailers, who hope to charge full price for everything, and shoppers, who wait for a sale" (see Kadet, 2004). The zero-sum nature of pricing makes this inevitable. Firms are constantly improving their pricing strategies in order to collect as much revenue as possible, and customers are constantly modifying their purchase plans in order to pay as little as possible.

Recently, a major electronics retailer, Best Buy, has expressed some strong opinions about this kind of strategic customer behavior. In an article that appeared on the front page of the *Wall Street Journal*, Best Buy Chief Executive Officer Brad Anderson openly labels some customers as "devils" (see McWilliams, 2004). According to Anderson, these are the customers who wait for markdowns, respond to promotions, and apply for rebates. In contrast, Anderson also describes the "angels" as the customers who snap up high-end gadgets without hesitation. Best Buy estimates that approximately 100 million out of its 500 million customer visits each year are "undesirable." Although these pejorative labels have attracted criticism (see Queenan, 2005), Best Buy has implemented customer relationship management programs to better distinguish the "angels" from the "devils" (see Arndorfer and Creamer, 2005). Apart from Best Buy, many other firms are also beginning to recognize that revenues are lost when customers wait for sales. Retailers, such as Bloomingdale's, Ann Taylor, Gap, and Home Depot, are turning to price optimization software instead of blindly slashing prices toward the end of the selling season (see Schlosser, 2004).

Although the importance of strategic customer behavior is recognized by many, its implications on inter-temporal pricing strategies has not been widely studied. This paper sets out with three main objectives. Our first goal is to formulate and solve the seller's dynamic pricing problem when facing strategic customers who may delay purchases and wait for sales. In this situation, should prices increase or decrease (or stay fixed) over time? What is the optimal timing and extent of the markups and/or markdowns? For the customers, how should they react to the seller's pricing strategies? Second, we would like to understand the main drivers behind the structure of the optimal policy characterized above. When prices change, what is the reason and what effect does this achieve? How should the seller respond when there are changes in the selling environment? Our final objective pertains to modeling. We would like to develop a comprehensive yet tractable framework to model dynamic pricing under

strategic customer behavior. The underlying problem is a dynamic principal-multi-agent problem between the seller and the customers. Although this class of problems is analytically complex, we would like to find an approach that captures a wide range of heterogeneous customer behavior, while retaining the seller's flexibility to control prices and ration inventory continuously over time.

In our model, there is a monopolist who sells a finite inventory over a finite time horizon. The seller may charge different prices over time, and may also practice rationing by fulfilling only a portion of market demand. There is a continuous inflow of customers arriving into the market. If they are unwilling to purchase the product immediately, they may leave the system, or may wait for more attractive purchase opportunities in the future. Although prices may fall, there is also the possibility that the product will become unavailable. Furthermore, customers incur waiting costs. Within this environment, the seller seeks to maximize revenue, and customers wish to maximize individual utility.

Heterogeneity plays a key role in our model. We allow the customers to vary along two dimensions: they may have different valuations for the firm's product and may have different degrees of patience (i.e. different waiting costs). Heterogeneous valuations imply that dynamic pricing is worthwhile because there is an opportunity to practice inter-temporal price discrimination. Heterogenous waiting costs allow us to capture a wide variety of customer behavior. At one extreme, when waiting costs are infinitely large, customers are myopic and make a one-time buy-or-exit decision upon arrival; on the other hand, customers with finite waiting costs may delay their purchases strategically. Existing models consider customer populations that are either purely myopic or purely strategic, whereas we allow for arbitrary combinations of both. With these two dimensions of heterogeneity, we believe that we can capture a good representation of reality.

This paper makes two main contributions. First, we demonstrate that strategic customer behavior is a main driver determining the *structure* of optimal pricing policies. Most existing models either explicitly impose structural restrictions (such as requiring monotonicity, or specifying the number of price changes), or do so implicitly; for instance, models based on stochastic customer arrivals tend to yield decreasing prices since the option value of unsold units decreases toward the end of the horizon, while models based on uncertain customer valuations tend to lead to increasing prices since customers who buy early need to be compensated for bearing additional risk. In contrast, our model does not impose any restrictions a priori. By endogenizing customer behavior, we find that a full spectrum of pricing policies may emerge at the optimal solution. This includes markups and markdowns, as well as other non-monotone price paths.

Our second contribution is to explain how optimal inter-temporal pricing strategies depend on the composition of the customer population. In particular, customer valuations, patience, as well as the interaction between these two dimensions of heterogeneity play an important role. Under any

price discrimination strategy, the seller has the choice between selling to low-valuation customers (at a low price) during the start or the end of the selling horizon. The latter implies holding "end-of-season sales," and this is preferred when low-valuation customers are sufficiently patient to wait for sales, while high-valuation customers are sufficiently impatient to buy early at higher prices. On the other hand, setting promotional low prices at the start is preferred when high-valuation customers are more patient than low-valuation customers: this discourages inefficient waiting and also captures surplus from high-valuation customers who miss the promotional prices. Finally, when high-valuation and low-valuation customers do not differ significantly in terms of patience, it may be optimal to have both promotional low prices as well as end-of-season sales, so the price schedule may not be monotone.

The remainder of this paper is organized as follows. Section 2 provides a literature review. Section 3 describes the model. Section 4 develops some structural properties, uses them to characterize an upper bound on seller revenues, and shows that this upper bound can be attained. Section 5 presents the main results, characterizing the seller's optimal policy. Section 6 examines the seller's revenue, consumer surplus, and social welfare under the optimal regime. The long run problem of selecting an optimal initial stocking quantity is analyzed in Section 7. Finally, Section 8 offers concluding remarks. All proofs are presented in the appendix.

2   Literature Review

The three main issues examined in this paper are: (i) strategic customer behavior, (ii) price dynamics, and (iii) limited capacity. There are several streams of related literature, each addressing different subsets of these issues.

The revenue management literature on dynamic pricing of finite inventories is closely related to our work. This stream of papers focuses on price dynamics and limited capacity (the primary question is how to set prices as a function of remaining inventory). However, strategic customer behavior is absent from the earlier models. The first papers were by Gallego and van Ryzin (1994, 1997). They model customer arrivals using Poisson processes and formulate the dynamic pricing problem as an intensity control problem; their 1997 paper generalizes the basic model to a network (multi-product) setting. Federgruen and Hetching (1999) combine pricing with inventory decisions. Feng and Gallego (1995, 2000) make a practically justified restriction: they consider a discrete menu of prices and policies involving at most one price change. Feng and Xiao (2000a, 2000b) extend this to policies involving multiple and reversible (non-monotonic) price changes. Similarly, Bitran and Mondschein (1997) consider periodic pricing policies that modify prices only at pre-specified times. Zhao and Zheng (2000) study dynamic pricing in more general situations with time inhomogeneous customer arrivals. For surveys of this literature, readers are referred to Bitran and Caldentey (2003),

Elmaghraby and Keskinocak (2003), and McAfee and te Velde (2005). For a comprehensive coverage of revenue management, readers are referred to the book by Talluri and van Ryzin (2005). A common approach in this literature is to determine optimal prices dynamically by considering the option value of unsold units. The result is that optimal price paths are decreasing over time (on average), because the option value of unsold units decreases as the deadline approaches. However, this result requires the assumption that demand is exogenous and independent across time. This no longer applies in the current work because strategic purchase delays in our model imply that demand may spill over into the future. As a result, we obtain optimal price schedules that may both increase or decrease over time.

Recent papers in revenue management have begun to examine customer behavior more closely. However, the focus is on how customers choose between substitute products offered by the firm (rather than on inter-temporal demand substitution). Talluri and van Ryzin (2004) use discrete choice models to describe how customers, in the context of airlines, choose among the set of fare classes offered; van Ryzin and Liu (2004) extends this analysis to the network setting. Shumsky and Zhang (2004) consider demand substitution, via upgrading, when inventory has been depleted. Netessine et al. (2004) consider cross-selling (i.e. offering customers a choice between their requested product and a package containing the requested product as well as another product). Cooper et al. (2004) show that neglecting substitution across products can lead to a spiral-down effect, in which the capacity allocation policy systematically performs worse and worse as the forecasting-optimization process continues. Zhang and Cooper (2005a) analyze a capacity allocation model with customer choice over parallel flights, and they extend this analysis (Zhang and Cooper, 2005b) to incorporate pricing decisions. Maglaras and Meissner (2006) show that the dynamic pricing problem when customers choose between multiple products can be reduced to an equivalent one-dimensional problem. In all these papers, customers choose what to buy, whereas in our work, customers choose when to buy. Both aspects of strategic customer behavior are important, and they are addressed using different modeling techniques.

The papers on dynamic pricing of finite inventories that are most closely related to ours involve inter-temporal demand. These papers explicitly model customers' decisions regarding when to buy. Aviv and Pazgal (2003) assume that there is a single price reduction and examine the optimal timing and extent of the discount in the presence of strategic customers. Elmaghraby et al. (2004) also focus on markdown mechanisms, and customers with multi-unit demands choose how many units to purchase at each price step. These two papers explicitly assume that prices should decrease over time, whereas we permit arbitrary price schedules. Next, using an infinite horizon model, Gallien (2004) shows that optimal prices should increase over time; for the simpler case where customers do not wait,

Arnold and Lippman (2001) and Das Varma and Vettas (2001) obtain similar conclusions. Unlike these papers, we consider a finite horizon, and we find that markups and markdowns may be optimal under different situations. Zhou et al. (2005) analyze the purchasing strategies of a single customer facing dynamic prices; our approach additionally considers the interaction between customers, in the sense that they are competing for the same pool of inventory. Rather than focusing on customers' decisions, Ovchinnikov and Milner (2005) focus on firms' pricing strategies when facing an exogenously specified profile of aggregate customer waiting behavior. This approach differs from our paper, in which we endogenously characterize customer waiting behavior as an equilibrium outcome. In another paper, Xu and Hopp (2004) show that prices should decrease when customers become increasingly price sensitive over time (and vice versa). However, they assume that customers commit to a purchasing time right from the start and do not wait in the market after observing current prices. Instead of pricing decisions, van Ryzin and Liu (2005) consider quantity decisions in a two-period capacity rationing model with strategic customers. They assume that the seller pre-commits to prices in both periods. Unlike the above two papers, we do not require commitment, either on the buyer side or on the seller side. Instead, we analyze a dynamic principal-agent game in which both the seller and the customers continuously optimize their decisions over time.

The aspect of strategic customer behavior analyzed in this paper (inter-temporal demand) first appeared in the economics literature on durable goods monopoly. The general approach is based on rational expectations: customers anticipate future price changes and adjust their purchase timing in response. However, unlike the papers reviewed above, capacity constraints and time deadlines are not considered, since the monopolist may sell as many units as desired over an infinite horizon. This literature has been inspired by the classic work of Nobel laureate Ronald Coase (1972): his main insight was that if customers strategically wait for price reductions, even a monopolist would be forced to price at marginal cost and earn zero profits. The earliest attempts to rigorously prove this result are by Stokey (1979, 1981) and Bulow (1982). Conlisk et al. (1984) introduce customer dynamics and show that the optimal price path involves periodic sales, with customers being willing to pay less as the next sale approaches. Gul et al. (1986), Ausubel and Deneckere (1989) and Sobel (1991) characterize a family of subgame perfect equilibria for the monopoly pricing game. Besanko and Winston (1990) present a dynamic programming procedure to compute the optimal price path. There are many variations of this basic setup; see Guth and Ritzberger (1998) for a survey. Unlike all these papers, we consider a fixed inventory and a finite time horizon. We show that these constraints influence customer expectations and thus have an important impact on optimal pricing strategies.

There are several other related papers that examine the relationship between capacity constraints (limited inventory) and pricing schemes. Harris and Raviv (1981) use a priority pricing mech-

anism to ration limited capacity. Wilson (1988) shows that when a monopolist sells a fixed quantity, it is optimal to post two prices and ration demand at the lower price. Lazear (1986) and Pashigian (1988) consider stochastic customer valuations and show how to use markdowns to extract high revenues during high-valuation realizations of demand. Desiraju and Shugan (1999) use a two-period model to investigate the profitability of various yield management practices, such as early discounting, overbooking and rationing. Dana (1998, 1999a) analyzes pricing and rationing decisions when customers face uncertainty over their own demands, and Dana (1999b, 2001) also allows for aggregate demand uncertainty. The common theme across all these papers is that by rationing demand at lower prices, firms can stimulate demand at higher prices, although precise implementation details may vary across different model setups. In our current study, we find that it may sometimes be profitable to generate scarcity, but we also identify situations when this is not necessary. Furthermore, all the papers above adopt a static mechanism design approach: the seller first establishes some mechanism, and then customers make a static buying decision. We hope to add to this research by studying the dynamics of pricing and rationing (on the seller side) as well as the dynamics of purchase and consumption (on the customer side).

3   Model

There is a monopolist seller who operates over a finite time horizon, the length of which is normalized to one (time unit). At the start of the time horizon, the seller is endowed with an inventory of Q units. By the end of the selling season, leftover units have zero value. This model is applicable to different industries, including travel (airplane seats and hotel rooms), retailing (fashion apparel and seasonal goods), and entertainment (concert and football tickets).

Customers are infinitesimally small and arrive continuously according to a deterministic flow of constant rate. This demand pattern is the same as that in the classical Economic Order Quantity (EOQ) model. We normalize the customer arrival rate to one (customer unit per time unit). Each customer demands a single unit of the seller's product. In other words, a mass of t customer units would have arrived by time t, and the aggregate demand of these customers is t units of the seller's product.

The customer population is heterogeneous along two dimensions. The first dimension is valuation. A fraction α of the customers value the product at V_H and the remaining $\overline{\alpha} \equiv 1 - \alpha$ value the product at V_L; we denote the difference using $\Delta \equiv V_H - V_L > 0$. We shall refer to these customers as "high-types" and "low-types" respectively. We assume that $V_L \geq \alpha V_H$; otherwise, it is optimal to sell only to high-types (because selling to low-types not only depletes inventory at a faster rate, but also earns lower revenue per unit time). The second dimension of heterogeneity is patience. Customers

7

who delay purchases incur different waiting costs, which may either be b_P or b_I per unit time, with $b_I > b_P \geq 0$. We shall adopt the terminology "patient" and "impatient" to distinguish these two cases. A fraction ϕ_H of the high-types are patient and the remaining $\overline{\phi}_H \equiv 1 - \phi_H$ are impatient; similarly, a fraction ϕ_L of the low-types are patient and the remaining $\overline{\phi}_L \equiv 1 - \phi_L$ are impatient. We shall refer to these four customer types as patient-high-types, impatient-high-types, patient-low-types, and impatient-low-types, denoted by $\theta \in \Theta \equiv \{PH, IH, PL, IL\}$. It is convenient to denote the proportion of each customer type using $f_{PH} \equiv \alpha\phi_H, f_{IH} \equiv \alpha\overline{\phi}_H, f_{PL} \equiv \overline{\alpha}\phi_L, f_{IL} \equiv \overline{\alpha}\overline{\phi}_L$, and to use V_θ and b_θ to denote the valuation and waiting cost of type-θ customers. We assume that customer types are unobservable to the seller, and that the customer composition is stationary over time.

The seller has to decide on pricing and rationing policies $\{p(t), r(t)\}$ and control policies $\{\mathbf{S}(t), \mathbf{D}(t)\}$. These choices, collectively referred to as the selling policy, are announced at the start of the time horizon. The price schedule $p(t)$ specifies the price charged at each time $t \in [0, 1]$. The rationing policy $r(t)$ specifies the fraction of current market demand that is fulfilled; we assume proportional rationing (see Tirole, 1988). The control policies represent cumulative sales processes $\mathbf{S}(t) \equiv \{S_\theta(t) : \theta \in \Theta\}$ and cumulative departure processes $\mathbf{D}(t) \equiv \{D_\theta(t) : \theta \in \Theta\}$ planned by the seller. That is, according to the seller's plans, by the end of time t, $S_\theta(t)$ units would have been sold to type-θ customers, and $D_\theta(t)$ type-θ customers would have departed from the market (without buying). These control processes also jointly determine cumulative market demand

$$Z_\theta(t) = A_\theta(t) - S_\theta(t) - D_\theta(t), \tag{1}$$

where $A_\theta(t) \equiv f_\theta t$ denotes cumulative arrival processes. Here, the market demand $Z_\theta(t)$ comprises of accumulated customers who have arrived and are waiting for a future purchase. When demand is rationed (i.e. $r(t) < 1$), all the $\sum_\theta Z_\theta(t)$ customers in the market have an equal $r(t)$ chance of getting the product. Customers who are rationed remain in the market if and only if their continuation utility is positive. Finally, we make the technical requirement that the seller's control policies $\{\mathbf{S}(t), \mathbf{D}(t)\}$ are right-continuous with left-limits (RCLL) and have a finite number of discontinuities. These regularity conditions ensure that the controls are implementable, in the sense that each individual customer's purchase or departure times (according to the controls) are well defined.

During the time horizon, based on the announced pricing and rationing policies $\{p(t), r(t)\}$, customers decide whether or not to purchase the product and whether or not to leave the market. Staying in the market incurs waiting costs at a constant rate, buying the product reaps an instantaneous surplus (valuation minus price paid) at the time of purchase, and leaving the market yields zero continuation payoff. For a type-θ customer with valuation V_θ and waiting cost b_θ, his total utility may be zero (if he leaves immediately upon arrival), or $V_\theta - p - b_\theta l$ (if he buys at price p after a time delay of length l), or $-b_\theta l$ (if he waits around for time l but ends up not buying). Therefore, each

customer faces a continuous time optimal stopping problem with two exit options. We assume that departed customers do not re-enter the market. Let $J_\theta(t)$ denote the optimal continuation value to a type-θ customer from staying in the market at the end of time t. Then, the customer is willing to buy at time t if $J_\theta(t) \leq V_\theta - p(t)$; if he does not buy, he is willing to leave if $J_\theta(t) \leq 0$.

There are two additional conditions that the selling policy must satisfy. First, the control policy $\{\mathbf{S}(t), \mathbf{D}(t)\}$ must be *incentive compatible*, because our customers are free to make their own utility maximizing choices. Specifically, cumulative sales $S_\theta(t)$ may increase if and only if customers are willing to purchase and cumulative departures $D_\theta(t)$ may increase if and only if customers are willing to leave. Second, the pricing and rationing policies $\{p(t), r(t)\}$ must be *credible*. In other words, there must not exist any time t when it is to the seller's advantage to deviate to a different price or rationing policy. Only then are customers willing to base their purchase decisions on these announcements.

Our goal is to determine the revenue-maximizing selling policy $\{p^*(t), r^*(t), \mathbf{S}^*(t), \mathbf{D}^*(t)\}$, subject to incentive compatibility and credibility. At first glance, this is a control problem in continuous time. However, most existing methods do not apply directly because verifying incentive compatibility involves solving additional control problems (one for each customer type). Furthermore, verifying credibility requires us to analyze the continuation-version of this problem at every point in time. Therefore, instead of tackling this problem directly, we shall first proceed to simplify it by exploiting some of its structural properties. This is done in the next section. Readers who prefer to first see the results and insights may proceed directly to Section 5, before returning to these technical and methodological details later.

4 Structural Properties and Upper Bounds

In this section, we show that the seller's continuous-time control problem can be reduced to a single-variable nonlinear program, which we solve explicitly. We proceed in four major steps. In Section 4.1, we identify some key features that parameterize feasible solutions. In Section 4.2, for each set of policy parameters, we establish an upper bound on revenues. In Section 4.3, we construct candidate policies attaining the upper bounds above. Finally, in Section 4.4, we optimize over policy parameters to find the largest upper bound, and conclude that the corresponding candidate policy must be optimal.

4.1 Policy Parameters

We begin by proving some structural properties that an optimal solution must satisfy. Based on these properties, we show that all feasible policies $\{p(t), r(t), \mathbf{S}(t), \mathbf{D}(t)\}$ are parameterized by several real-valued quantities. These parameters will be the focus of our analysis.

Lemma 1 *For any feasible policy, we must have $p(t) \geq V_L$.*

Intuitively, prices strictly below V_L are not credible because when the time comes to honor these prices, the seller will increase prices to V_L since all customers are still willing to buy at this price. At the time of purchase, waiting costs are sunk and customers will contend with any price that leaves them with non-negative surplus.

Proposition 1 *For any feasible policy, there exists parameters $\tau_0, \tau_1, \ldots, \tau_N, \pi_1, \ldots, \pi_N$, with $\sum_{i=0}^{N} \tau_i = 1$, such that at least the same revenues are earned under some policy satisfying*
(i) $p(t) = V_L$ and $r(t) = 1$ for every $t \in [0, \tau_0]$.
(ii) $p(t_k) = V_L$ and $r(t_k) = \pi_k$ for $t_k = \sum_{i=0}^{k} \tau_i$, for $k = 1, \ldots, N$.

This result illustrates that every feasible policy is associated with a set of parameters, namely $\tau_0, \vec{\tau} \equiv \{\tau_1, \ldots, \tau_N\}$, and $\vec{\pi} \equiv \{\pi_1, \ldots, \pi_N\}$. We may interpret τ_0 as the length of an initial time interval with price V_L. After this, "sale" prices of V_L occur only at discrete time points: each τ_k is the length of time that transpires before the next "sale" occurs, during which a fraction π_k of demand is fulfilled. We shall refer to the time interval with length τ_k as the k-th time segment, for $k = 0, 1, \ldots, N$.

4.2   Upper Bounds

Next, for each set of policy parameters $\tau_0, \vec{\tau} \equiv \{\tau_1, \ldots, \tau_N\}, \vec{\pi} \equiv \{\pi_1, \ldots, \pi_N\}$, we proceed to find an upper bound on revenues. In other words, any policy associated with the corresponding parameters can not earn more revenues than this upper bound. To obtain these bounds, we compute customers' willingness-to-pay (WTP) at each time, and aggregate these WTPs over all individuals. We first focus on the case $b_I > b_P = 0$; that is, patient customers face zero waiting cost. The other case with $b_I > b_P > 0$ will be examined later.

We begin by considering the stopping problem faced by low-valuation customers. First, if waiting costs are strictly positive, the customer will either buy immediately upon arrival (if price is V_L) or leave the market forever (if price exceeds V_L). These customers never wait because they know that they can at best enjoy zero surplus, since the lowest possible price is V_L. On the other hand, low-type customers with zero waiting cost will wait for a price of V_L before purchasing. Therefore, in the current case with $b_I > b_P = 0$, patient-low-types strategically wait for the sale, but impatient-low-types behave myopically and make a one-time buy-or-exit decision at the time of arrival.

Next, we consider the decision problem faced by high-valuation customers, given policy parameters $\tau_0, \vec{\tau} \equiv \{\tau_1, \ldots, \tau_N\}, \vec{\pi} \equiv \{\pi_1, \ldots, \pi_N\}$. First, we consider patient-high-types. Since waiting costs are $b_P = 0$, the WTP of patient-high-types arriving during the k-th time segment, denoted W_k^{PH}, can

be characterized using backward induction. Specifically, we have

$$W_N^{PH} = V_H - \pi_N \Delta, \tag{2}$$

$$W_k^{PH} = V_H - \pi_k \Delta - (1 - \pi_k)(V_H - W_{k+1}^{PH}), \quad k = 1, \ldots, N - 1. \tag{3}$$

These expressions imply the following upper bound on the WTP of patient-high-types:

$$W_k^{PH} \leq V_H - \pi_k \Delta. \tag{4}$$

By a similar logic, we can characterize the WTP of impatient-high-types. Since waiting costs $b_I > 0$ are now relevant, the WTP depend on the precise time of arrival within each time segment. Let $W_k^{IH}(s)$ denote the WTP of impatient-high-types arriving in the k-th time segment, at time $s \leq \tau_k$ before the end of the time segment. We have the following upper bound on WTP:

$$W_k^{IH}(s) \leq \min\{V_H - \pi_k \Delta + b_I s, V_H\}. \tag{5}$$

This is because impatient-high-types have an option to wait (at cost of $b_I s$) until the end of the time segment. If we assume that impatient-high-types are required to leave the market by the end of the current time segment, their WTP then would be $V_H - \pi_k \Delta$, so (5) would hold with equality. However, since exit is not mandatory, we only have an upper bound above. It is convenient to rewrite (5) as

$$W_k^{IH}(s) \leq V_H - [\pi_k \Delta - b_I s]^+. \tag{6}$$

Aggregating these WTP expressions yield an upper bound on seller revenues. In the following two lemmas, we characterize this upper bound and derive an useful condition that it satisfies.

Lemma 2 *Let $b_I > b_P = 0$. Consider any policy with parameters $\tau_0, \vec{\tau} \equiv \{\tau_1, \ldots, \tau_N\}, \vec{\pi} \equiv \{\pi_1, \ldots, \pi_N\}$. The revenue collected under this policy can not exceed the upper bound $UB(\tau_0, \vec{\tau}, \vec{\pi})$ given by*

$$\tau_0 V_L + \sum_{k=0}^{N} \left(f_{PL} \tau_k \pi_k V_L + f_{PH} \tau_k (V_H - \pi_k \Delta) + f_{IH} \int_0^{\tau_k} V_H - [\pi_k \Delta - b_I s]^+ ds \right). \tag{7}$$

Lemma 3 *Define $\bar{\tau} = \sum_{k=1}^{N} \tau_k$ and $\bar{\pi} = \frac{\sum_{k=1}^{N} \tau_k \pi_k}{\sum_{k=1}^{N} \tau_k}$. Then, we have $UB(\tau_0, \vec{\tau}, \vec{\pi}) \leq UB(\tau_0, \bar{\tau}, \bar{\pi})$.*

The preceding lemmas yield an important implication. Given any feasible policy with parameters $\tau_0, \vec{\tau}, \vec{\pi}$, we know that revenues do not exceed $UB(\tau_0, \vec{\tau}, \vec{\pi})$, which in turn is less than $UB(\tau_0, \bar{\tau}, \bar{\pi})$ with $\bar{\tau}, \bar{\pi}$ defined in Lemma 3. Therefore, we can start with using $UB(\tau_0, \bar{\tau}, \bar{\pi})$ as an upper bound on policy revenues. This allows us to make the following conclusion.

Proposition 2 *Let $b_I > b_P = 0$. Then, under any feasible policy, there exist parameters τ_0, τ_1, π_1 such that the revenues collected does not exceed the upper bound $UB(\tau_0, \tau_1, \pi_1)$.*

4.3 Candidate Policies

Next, for each set of parameters τ_0, τ_1, π_1, we proceed to construct a feasible policy attaining the upper bound given above. As we vary the parameters, this class of policies serves as possible candidates for the optimal policy. We will specify the price schedule $p(t)$, rationing function $r(t)$, as well as the controls for cumulative sales and departures $\{\mathbf{S}(t), \mathbf{D}(t)\}$. The policy described in the following lemma attains revenues that are arbitrarily close to the upper bound.

Proposition 3 *Let $\tau_0, \tau_1, \pi_1 \in [0,1]$ with $\tau_0 + \tau_1 = 1$. For some small $\epsilon > 0$, under the policy*

$$
p(t) \;=\; \begin{cases} V_L, & t \in [0, \tau_0], \\ V_H - [\pi_1 \Delta - b_I(1 - \epsilon - t)]^+, & t \in (\tau_0, 1 - \epsilon], \\ V_H - \pi_1 \Delta, & t \in (1 - \epsilon, 1), \\ V_L, & t = 1, \end{cases} \tag{8}
$$

$$
r(t) \;=\; \begin{cases} 1, & t \in [0, 1), \\ \pi_1, & t = 1, \end{cases} \tag{9}
$$

$$
S_\theta(t) \;=\; \begin{cases} f_\theta t, & \text{if } \theta = IH, \text{ or } (\theta = PH \text{ and } t \geq 1 - \epsilon), \\ f_\theta \min(t, \tau_0), & \text{if } \theta = IL, \text{ or } (\theta = PH \text{ and } t < 1 - \epsilon), \text{ or } (\theta = PL \text{ and } t < 1), \\ f_\theta(\tau_0 + \pi_1 \tau_1), & \text{if } (\theta = PL \text{ and } t = 1) \end{cases} \tag{10}
$$

$$
D_\theta(t) \;=\; \begin{cases} 0, & \text{if } \theta \in \{IH, PH\}, \text{ or } (\theta = PL \text{ and } t < 1), \text{ or } (\theta = IL \text{ and } t \leq \tau_0), \\ f_\theta \tau_1(1 - \pi_1), & \text{if } (\theta = PL \text{ and } t = 1), \\ f_\theta(t - \tau_0), & \text{if } (\theta = IL \text{ and } t > \tau_0), \end{cases} \tag{11}
$$

revenues collected are at least $UB(\tau_0, \tau_1, \pi_1) - \epsilon \Delta$.

The essence of this result is depicted in Figures 1 and 2, which plot the prices, rationing function, and controls over time. It is straightforward to verify that this policy attains the revenues as claimed. For convenience, we shall refer to this as the (τ_0, τ_1, π_1)-policy.

Next, observe that the seller must exhaust all available inventory Q by the end of the horizon. This is because any policy with leftover inventory violates the credibility condition, since the seller would always prefer to sell them to waiting customers at $t = 1$ rather than to discard them. Customers anticipate this move and these expectations would affect their WTP. Thus, the outcome would be as if the seller has planned on selling out right from the start. This implies that in order for the policy of Proposition 3 to be feasible (in particular, credible), we must have

$$
\tau_0 + \tau_1(f_{PH} + f_{IH} + \pi_1 f_{PL}) = Q, \tag{12}
$$

where the left-hand-side is the number of units sold. Together with the condition
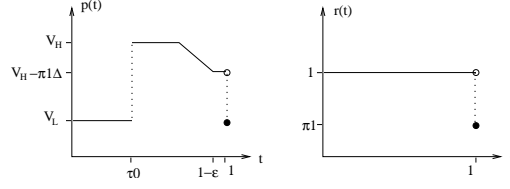
$$
\tau_0 + \tau_1 = 1, \tag{13}
$$

Figure 1: Price schedule and rationing function under (τ_0, τ_1, π_1)-policy

(a) (i) Sales to impatient–high–types   (a) (ii) Sales to patient–high–types   (a) (iii) Sales to patient–low–types   (a) (iv) Sales to impatient–low–types

(b) (i) Departures by impatient–high–types (b) (ii) Departures by patient–high–types (b) (iii) Departures by patient–low–types (b) (iv) Departures by impatient–low–types
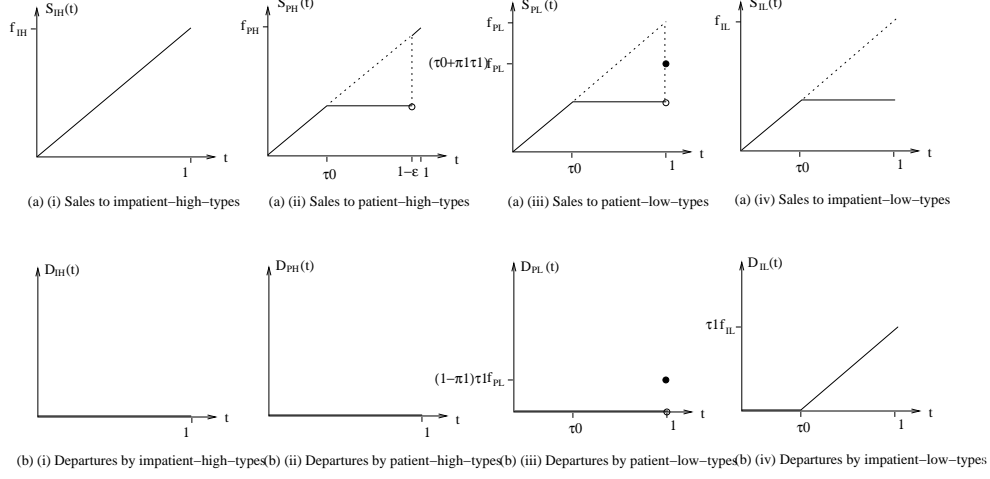
Figure 2: Cumulative sales and departures for each customer type under (τ_0, τ_1, π_1)-policy

we are left with only one degree of freedom among the parameters τ_0, τ_1, π_1. That is, specifying the value of one parameter pins down the value of the other two. In particular, we shall primarily use π_1 to uniquely characterize any (τ_0, τ_1, π_1)-policy. In this way, we have a continuum of candidate policies parameterized by π_1.

It is worthwhile to point out the two extreme cases in our continuum of candidate policies. These are shown on Figure 3. On one extreme, π_1 is minimized at $\pi_1 = 0$, with $\tau_0 = \frac{Q-\alpha}{1-\alpha}$ and $\tau_1 = \frac{1-Q}{1-\alpha}$ (recall that $f_{PH} + f_{IH} = \alpha$). This corresponds to a *pure markup* policy, since prices start at V_L and then increase to V_H at time τ_0. At the other extreme, π_1 is maximized at $\pi_1 = \bar{\pi}_1 \equiv \frac{Q-\alpha}{\overline{\alpha}\phi_L}$ (recall that $f_{PL} = \overline{\alpha}\phi_L$), and the other parameter values are $\tau_0 = 0, \tau_1 = 1$. This describes a *pure markdown* policy, since the zeroth time segment has length $\tau_0 = 0$ and prices decrease over all of the next time segment of length $\tau_1 = 1$ and culminate at V_L. At all intermediate policies with $\pi_1 \in (0, \bar{\pi}_1)$, the price schedule is not monotone: optimal prices start low at V_L, increase at time $t = \tau_0$, and then decrease again at time $t = 1$. We refer to these as *interior* policies.

The class of (τ_0, τ_1, π_1)-policies may be interpreted as follows. (Please refer to Figure 1.) There is initially a promotional pricing phase of length τ_0, at which all arriving customers purchase at price V_L. There is no rationing, and all demand is fulfilled. Then, the regular selling season of length τ_1
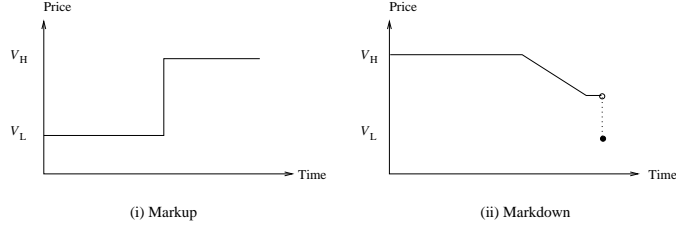
13

Figure 3: Pure markup and pure markdown policies

follows: prices decline steadily over the season and culminate in a sale at price V_L, when demand is rationed and there is only a π_1 chance of getting the product. Since customers balance the tradeoff between low prices and product availability, the price decrease is designed to induce high-types to buy instead of waiting for the final sale. During the season, impatient-high-types buy immediately, patient-high-types wait to buy right before the sale, patient-low-types wait to buy at the final sale price V_L (succeeding only if they are not rationed), and impatient-low-types are lost.

4.4 Attaining ϵ-Optimality

In this subsection, we maximize the upper bound $UB(\tau_0, \tau_1, \pi_1)$ over all $\tau_0, \tau_1, \pi_1 \in [0, 1]$ satisfying (12) and (13). From Propositions 1 and 2, we know that this maximum value must be an upper bound on optimal seller revenues. Based on the characterization given in Proposition 3, we can then construct an ϵ-optimal (τ_0, τ_1, π_1)-policy whose revenue is within $\epsilon\Delta$ of the optimal revenue. Since ϵ can be made arbitrarily small, we shall henceforth contend with ϵ-optimality and refer to it as "optimality" for brevity.

Before presenting the next result, let us define the constants $A \equiv \frac{\alpha(1-Q)}{1-\alpha} \cdot \left(1 - \frac{\phi_H}{\phi_L}\right) \cdot \Delta$, $B \equiv -\alpha \cdot \overline{\phi}_H \cdot \frac{\Delta^2}{b_I}$, and $K \equiv -\frac{A\phi_L}{B} = \frac{1-Q}{1-\alpha} \cdot \frac{\phi_L - \phi_H}{1-\phi_H} \cdot \frac{b_I}{\Delta}$. Also, let us define the function $G(u) \equiv u(1-\phi_L u)^2$ and notice that its maximum value over $u \in [0, 1]$ is $\frac{4}{27\phi_L}$. We are now ready to state our result.

Proposition 4 *Let $b_I > b_P = 0$. Then, the optimal policy is a (τ_0, τ_1, π_1)-policy, with π_1 given below.*
(i) When $K \le 0$, we have $\pi_1 = 0$.
(ii) When $K \in (0, \frac{4}{27\phi_L})$, we have either $\pi_1 = \bar{\pi}_1$ or $\pi_1 = \xi \in (0, \bar{\pi}_1)$, where ξ is the smallest solution of $G(\xi) = K$. In particular, if $K \le G(\bar{\pi}_1) \le \frac{4}{27\phi_L}$, we have $\pi_i = \xi$.
(iii) When $K \ge \frac{4}{27\phi_L}$, we have $\pi_1 = \bar{\pi}_1$.

Finally, we return to the case with $b_I > b_P > 0$. In this case, recall that all low-types, having a positive waiting cost, behave myopically, so that sales at the end of the length-τ_1 time segment benefit only high-types. Since all these high-types are willing to pay V_H, it is optimal to set prices at V_H after

14

time $t = \tau_0$, so $\pi_1 = 0$, which uniquely determines the optimal policy. We thus have the following result.

Proposition 5 *Let $b_I > b_P > 0$. Then, the optimal policy is a (τ_0, τ_1, π_1)-policy, with $\pi_1 = 0$.*

In this section, we have simplified a dynamic principal-agent control problem in continuous time into a nonlinear program in a single variable. The latter problem is then solved to characterize the seller's optimal policy. In the next section, we proceed to identify the main drivers that determine the structure of the optimal policy.

5 Optimal Policy

Should the seller still hold sales when customers strategically wait for them? Under the revenue-maximizing price discrimination strategy, should the seller have low-priced transactions at the start or at the end of the selling horizon? Or both? In this section, we show that the answers to these questions depend critically on customer valuations, waiting costs, as well as the interaction between these two dimensions of heterogeneity.

Theorem 1
(a) *The optimal policy is a pure markup policy if at least one of the following conditions hold:*
    (i) $\phi_L \leq \phi_H$,
    (ii) $b_P > 0$.
(b) *The optimal policy is a pure markdown policy if all of the following conditions hold:*
    (i) $\phi_L \geq \phi_H$,
    (ii) $b_P = 0$ and b_I is sufficiently large (that is, $b_I \geq \frac{4(1-\alpha)}{27(1-Q)} \cdot \frac{1-\phi_H}{\phi_L(\phi_L - \phi_H)} \cdot \Delta$).
(c) *The optimal policy is an interior policy if all of the following conditions hold:*
    (i) $\phi_L \geq \phi_H$,
    (ii) $b_P = 0$ and b_I is sufficiently small (that is, $b_I \leq \frac{(Q-\alpha)(1-Q)}{(1-\alpha)^2} \cdot \frac{1-\phi_H}{\phi_L(\phi_L - \phi_H)} \cdot \Delta$).

This theorem follows from Propositions 4 and 5. It indicates that strategic waiting on the part of customers is critical in determining the structure of optimal price schedules. Within our framework, we show that a full spectrum of pricing policies may be optimal, ranging from monotone increasing to monotone decreasing price paths. In case (i), pure markup policies are optimal, and the seller should concentrate all low-priced transactions at the start of the horizon, whereas in case (ii), the seller should use pure markdown policies and defer all sales to the end of the horizon. The intuition is as follows. Markups defend the seller against strategic delays since waiting does not pay off. This is important when the high-type population is patient and more likely to wait, compared to the low-types (i.e.

when $\phi_H \geq \phi_L$). Another situation in which markups are appropriate is when waiting costs are high in general (i.e. high b_P and b_I). In this case, it is important to use markups to discourage waiting, since waiting is inefficient and induces a deadweight loss. On the other hand, the success of price markdowns is contingent upon whether low-valuation customers will wait around to purchase at sales, after the (relatively more impatient) high-types have purchased at a higher price. This requires low waiting costs for patient-low-types and high waiting costs for impatient-high-types (i.e. low b_P and high b_I), as well as a relatively large segment of impatient customers among high-types (i.e. $\phi_L \geq \phi_H$). For other parameter values, neither of the opposing factors above dominate. Then, it may be optimal to sell low-priced units both during the start and the end of the horizon, using an interior policy. The promotional selling phase appeals to early customers who object to either paying high prices or waiting too long for a sale, and the regular selling season is short enough so that it is not too costly for self-selecting customers to wait for the sale, thereby achieving price discrimination.

Corollary 1 *Consider the limiting case with $b_P = 0$ and $b_I = \infty$. Then, the optimal policy is a pure markup policy if $\phi_H \geq \phi_L$ and a pure markdown policy when $\phi_L \geq \phi_H$.*

In this limiting case, Corollary 1 tells us that pure markups are optimal when high-types are relatively more patient, pure markdowns are optimal when high-types are relatively less patient, and interior policies are no longer needed. The same intuition above still applies. Since this limiting case allows us to derive the same insights, while being analytically more convenient, we shall focus on this case henceforth. Furthermore, since impatient customers with $b_I = \infty$ never wait, we may describe them as being myopic and the patient customers as being strategic.

The results presented thus far apply to the case where the seller's initial inventory Q lies between α and $\alpha + \phi_L \bar{\alpha}$. Equivalently, there is excess inventory after satisfying all high-type demand (of mass α), but this excess is not sufficient to supply to all patient-low-types (of mass $\phi_L \bar{\alpha}$). The other cases are not presented here because they do not yield any new insights. In the case where $Q < \alpha$, the seller simply sets a high price of V_H throughout, and in the case where $Q > \alpha + \phi_L \bar{\alpha}$, the seller initially sets promotional prices of V_L so that inventory can be depleted until it is within the range of our analysis, which is then directly applicable.

Anecdotal evidence suggests that markups are common in the travel industries whereas markdowns are common in fashion retailing. In fact, Feng and Gallego (1995, 2000) and Feng and Xiao (2000) use this observation to motivate their analysis of markup and markdown pricing mechanisms, but they do not explain whether markups or markdowns are more appropriate in any given situation. Bitran and Mondschein (1997) and Zhao and Zheng (2000) provide an explanation based on time-inhomogeneous demand patterns. They note that for fashion products, arrival intensities and

16

reservation prices are higher during the start of the selling season; however, in airlines and hotels, arrival intensities and reservation prices are higher during the end of the time horizon. They use this difference to build a model that explains why airlines and hotels use markups while fashion retailers use markdowns. It is important to note that time-inhomogeneity is crucial in these models: with stationary (and non-strategic) demand, inter-temporal price discrimination is not worthwhile because the seller should simply charge the monopoly price (on average) throughout the time horizon. However, in these models, time-inhomogeneity is exogenously specified.

Our results in Theorem 1 provide an alternative explanation based on strategic customer behavior. It is conceivable that in fashion retailing, high-valuation customers prefer to have the product earlier and are relatively less patient to wait for sales, whereas in the case of airlines, the high-types (e.g. business travelers) tend to be more patient and do not mind committing to travel schedules later. Based on our model, this difference would explain why markdowns are often seen in fashion retailing and markups are often seen in travel industries such as airlines. In fact, our explanation based on strategic customer behavior complements previous explanations based on time-inhomogeneity. This is because by incorporating strategic considerations, we show that time-inhomogeneity arises endogenously as an equilibrium outcome (e.g. when customers wait for markdowns). Although we assume stationary arrivals at the outset, we find that strategic customer behavior can generate the time-varying purchase patterns that have been shown to account for increasing or decreasing price paths.

Another observation is that each of the four customer segments in our model affects the seller in different ways. First, impatient-high-types benefit the seller because he is able to extract high revenues from these customers immediately when they arrive. Second, patient-high-types hurt the seller because strategic delays have an adverse effect on revenue. Third, patient-low-types benefit the seller because when they strategically delay purchase, they create competition with the other high-valuation customers for product availability at the end of the selling season; this discourages the high-types from waiting and increases their willingness to pay. Finally, impatient-low-types is of minimal value to the seller because it is not possible for these customers to generate high profits. Therefore, depending on the relative sizes of these four segments, the seller should adjust the price schedules accordingly as specified in Theorem 1.

In our model, the optimal price schedule may either increase or decrease over time, whereas most papers in the dynamic pricing literature generate price paths that, on average, only decrease over time. In a recent review paper on the airline industry, McAfee and te Velde (2005) write, "a remarkably robust prediction of theories ... (is that) prices are falling as takeoff approaches." This can be understood by considering the option value of unsold units. As the end of the horizon approaches, it becomes less likely that a given unit would be sold, so its option value declines. In

fact, Gallego and van Ryzin (1994) use their stochastic model with Poisson arrivals to establish the following structural properties: (i) for a fixed inventory level, the price should decrease over time, and (ii) at a fixed time, the price should increase as the remaining inventory decreases. This implies that along any sample path, the optimal price decreases continually, and jumps up whenever units are sold. McAfee and te Velde (2005) scrutinize the Gallego and van Ryzin (1994) model and conclude that the two structural properties above, when combined, lead to a decrease in expected prices over time. Therefore, these stochastic models and their corresponding option-value interpretations do not explain the use of markups (commonly observed in airlines); a notable exception is Zhao and Zheng (2000), whose model with time-inhomogeneous demand processes may lead to increasing price paths. Our game-theoretic model, in contrast, rationalizes both markups and markdowns by explicitly accounting for inter-temporal customer utility. This suggests that in some practical pricing contexts, strategic considerations are no less significant than stochastic influences.

Nevertheless, the option value interpretations from stochastic models can be combined with the insights from our game theoretic model. In the stochastic analogue of our model (e.g. customers arrive according to Poisson processes), we conjecture that our main findings will remain unchanged. The work of Gallego and van Ryzin (1994, 1997) lends support for this conjecture by showing that deterministic models provide good approximations because statistical fluctuations in demand average themselves out. In particular, they observed that compensating for demand uncertainty by adjusting prices dynamically achieves a minimal effect: their deterministic heuristic performs almost as well as the optimal dynamic pricing policy. In the same way, one can view the current work as a deterministic solution to the general yield management problem under strategic customer behavior. In general, although optimal prices should decrease along each sample path as the deadline approaches (because the option value decreases), these decreases only have a second-order effect. In contrast, first-order price changes are triggered by the "regime switches" captured in our deterministic model, in which the seller changes his target group of buyers. Specifically, markups reflect the seller's intention to restrict sales to a smaller group of high-valuation buyers, while markdowns reflect the opposite intention to include a larger set of potential buyers with lower valuations. In a stochastic setting, whether these first-order price changes should involve markups or markdowns depends on the composition of the customer pool in a similar fashion as before. However, all these statements are conjectures that remain to be verified by future research.

6   Seller Revenue, Consumer Surplus, and Social Welfare

We are interested in the seller's optimal revenue R, as well as the consumer surplus U_θ of type-θ customers (averaged over all customers of each type). The following proposition, proven in the

appendix, expresses these quantities of interest in terms of basic model parameters.

Proposition 6 *Let $b_P = 0$ and $b_I = \infty$. Under the optimal selling policy, the seller's revenue and customers' surplus are characterized below.*

(i) When $\phi_H \geq \phi_L$ (i.e. markups are used),

$$R = \alpha V_H + \frac{Q - \alpha}{1 - \alpha}(V_L - \alpha V_H), \tag{14}$$

$$U_{PH} = U_{IH} = \frac{Q - \alpha}{1 - \alpha}\Delta, \tag{15}$$

$$U_{PL} = U_{IL} = 0. \tag{16}$$

(ii) When $\phi_H \leq \phi_L$ (i.e. markdowns are used),

$$R = \alpha V_H + \frac{Q - \alpha}{\phi_L \overline{\alpha}}\left(\phi_L \overline{\alpha} V_L - \phi_H \alpha \Delta\right), \tag{17}$$

$$U_{PH} = \frac{Q - \alpha}{\overline{\alpha}\phi_L}\Delta, \tag{18}$$

$$U_{IH} = U_{PL} = U_{IL} = 0. \tag{19}$$

These expressions have intuitive interpretations. When the high-type population is relatively more patient ($\phi_H \geq \phi_L$), Theorem 1 tells us that the seller uses a markup. With a markup, the seller's revenue in Proposition 6(i) is the sum of the base revenue αV_H (obtained by charging V_H throughout), and the incremental revenue from charging the low introductory price of V_L for the first $\frac{Q-\alpha}{1-\alpha}$ time units. For high-types, the consumer surplus is Δ during the introductory period and zero at other times, so this yields $\frac{Q-\alpha}{1-\alpha}\Delta$ on average. For low-types, it is clear that the consumer surplus is always zero. In the other case where the high-type population is relatively less patient ($\phi_H \leq \phi_L$), Theorem 1 tells us that the seller uses a markdown. With a markdown, the seller's revenue in Proposition 6(ii) is the base revenue of αV_H (obtained if all high-types pay V_H), plus the incremental revenue from a fraction $\frac{Q-\alpha}{\overline{\alpha}\phi_L}$ of the patient-low-types who buy at the end of the horizon (the first term in parentheses), minus the revenue from patient-high-types that must be sacrificed in order to ensure that these customers will not choose the "deal" intended for the low-types (the second term in parentheses). This foregone revenue results in consumer surplus of $\frac{Q-\alpha}{\overline{\alpha}\phi_L}\Delta$ for all the patient-high-types, while all other customers earn zero surplus.

The expressions for seller revenue and high-type consumer surplus are plotted against ϕ_H and ϕ_L in Figure 4. (We omit the low-types because they receive zero surplus.) First, we look at the graphs on the left-hand-side. As the proportion of patient customers among the low-types increases (i.e. as ϕ_L increases), both the seller's revenue and consumers' surplus initially remain unchanged as long as $\phi_L \leq \phi_H$. As soon as ϕ_L increases beyond ϕ_H, there is a transfer of surplus from the
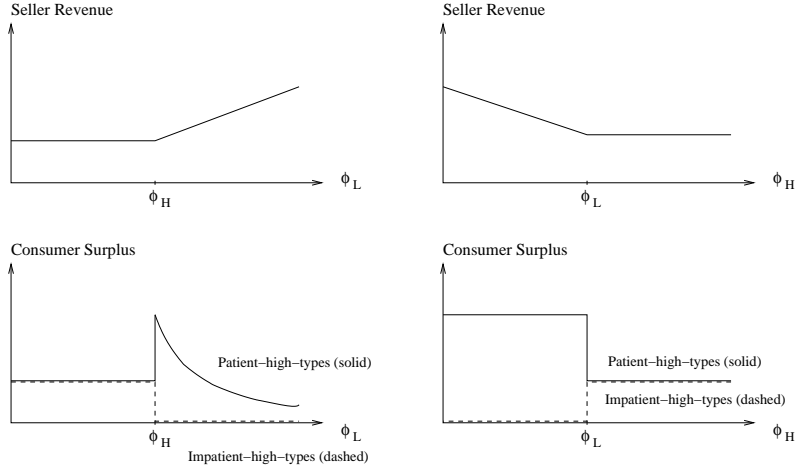
Figure 4: Seller's revenue and consumer surplus against ϕ_H and ϕ_L

impatient-high-types to the patient-high-types. This occurs due to the regime switch from markup pricing (positive surplus available to all high-types) to markdown pricing (positive surplus are exclusive to patient-high-types). As ϕ_L continues to increase, the surplus for impatient-high-types stays at zero and the surplus for patient-high-types decreases, while the seller's revenue increases. This transfer from patient-high-types to the seller occurs because a larger pool of patient-low-types (i.e. higher ϕ_L) increases the competition for availability at the final markdown price and helps the seller to extract a larger part of the patient-high-type consumer surplus. Next, we look at the graphs on the right-hand-side. As the proportion of patient customers among the high-types decreases (i.e. as ϕ_H decreases), all quantities initially remain unchanged. As ϕ_H falls below ϕ_L, the regime switch (from markups to markdowns) leads to a transfer from impatient-high-types to patient high-types, as discussed above. As ϕ_H continues to decrease, consumer surplus remains unchanged but the seller's revenue increases. The seller gains, even though each individual customer is unaffected, because there is an increased mass of impatient-high-types who pay the maximum price V_H.

In general, as illustrated in Figure 4, the seller is better off when either ϕ_L increases or ϕ_H decreases. When the high-valuation segment becomes proportionately less patient (i.e. ϕ_H decreases), there are more impatient-high-types who are willing to pay a high price immediately upon arrival, and there are less patient-high-types who use strategic delays to pay less than their individual valuation. Each of these two effects enhances revenues. Similarly, when the low-valuation segment becomes proportionately more patient (i.e. ϕ_L increases), there are more patient-low-types. These customers benefit the seller because they compete with other customers for end-of-season inventory, thus encouraging earlier purchases at higher prices.

Customers who arrive early during the selling season (early-birds) earn different levels of surplus from those who arrive later (late-comers). According to Theorem 1, when markups are used, early-birds (who arrive before time $t = \frac{Q-\alpha}{1-\alpha}$) with high valuations enjoy a surplus of Δ, but late-comers (who arrive after time $t = \frac{Q-\alpha}{1-\alpha}$) do not earn any surplus. On the other hand, when markdowns are used, individual surplus does not depend on arrival time: all patient-high-types have an equal chance of earning positive surplus (regardless of arrival time) and all other customers earn zero surplus. This suggests that in an environment with $\phi_H \geq \phi_L$, the use of markups will encourage customers (or the high-valuation customers, at least) to arrive earlier. On the other hand, in the opposite environment ($\phi_H \leq \phi_L$), the use of markdowns implies that arriving early is of no value. Therefore, if customers were able to choose when to arrive to the market, their arrival times would be influenced by the composition of the customer pool (which determines whether markups or markdowns are used). These choices of arrival times generate non-stationary intensities and valuations in the arrival process, which in turn influence the seller's optimal inter-temporal pricing strategies. We leave the task of endogenizing customers' choices of arrival times as a topic for future research. For a similar setting with endogenous arrivals to a queue, readers are referred to Lariviere and van Mieghem (2004).

Finally, observe that the optimal selling policy is socially efficient. Whether the seller uses a markup or a markdown, the limited inventory is always used up and allocated to all the high-valuation customers and a portion of the low-valuation customers. The only difference is in terms of low-type allocation, i.e., which low-type customers receive the product? When markups are used, early-bird low-types receive the product, but when markdowns are used, patient-low-types receive the product. In either case, social welfare (i.e. the sum of the seller's revenue and total consumer surplus) is maximized and equals $\alpha V_H + (Q - \alpha)V_L$.

7 Initial Inventory Choice

So far, the seller's inventory Q has been exogenously specified. Now, suppose that each unit can be procured at some cost, normalized to zero. If inventory planning were possible for the seller, how many units Q should he stock? The next proposition, proven in the appendix, provides the answer.

Proposition 7 *Let $b_P = 0$ and $b_I = \infty$. The seller's optimal stocking quantity and selling policy are characterized below.*

(i) *When $\alpha V_H \geq V_L$ and $\alpha \phi_H V_H \geq (\alpha \phi_H + \overline{\alpha} \phi_L)V_L$, the seller should stock $Q^* = \alpha$ units, charge the constant price $p^*(t) = V_H$, and satisfy all demand with $r^*(t) = 1$.*

(ii) *When $\alpha V_H \leq V_L$ and $\alpha \overline{\phi}_H V_H \leq (\alpha \overline{\phi}_H + \overline{\alpha} \overline{\phi}_L)V_L$, the seller should stock $Q^* = 1$ unit, charge the constant price $p^*(t) = V_L$, and satisfy all demand with $r^*(t) = 1$.*

(iii) In all other cases, the seller should stock $Q^* = \alpha + \overline{\alpha}\phi_L$ units, use the markdown price schedule

$$p^*(t) = \begin{cases} V_H, & t < 1, \\ V_L, & t = 1, \end{cases} \tag{20}$$

and satisfy all demand with $r^*(t) = 1$.

This proposition distinguishes between three cases. We refer to case (i) as the constant-high-price case, because the seller stocks α units, charges the high price V_H throughout the horizon, and satisfies all high-type demand. Next, we refer to (ii) as the constant-low-price case, because the seller stocks one unit, charges V_L throughout, and satisfies all demand at the low price. Finally, case (iii) involves a single markdown. The seller stocks $\alpha + \overline{\alpha}\phi_L$ units, sells to impatient-high-types (of mass $\alpha\overline{\phi}_H$) at the high-price V_H throughout the horizon, and sells to all patient customers (of mass $\alpha\phi_H + \overline{\alpha}\phi_L$) at the marked-down price V_L at the end.

High–types are valuable

(i) CONSTANT HIGH PRICE

(iii) MARKDOWN
PRICING

Low–types are patient

High–types are patient

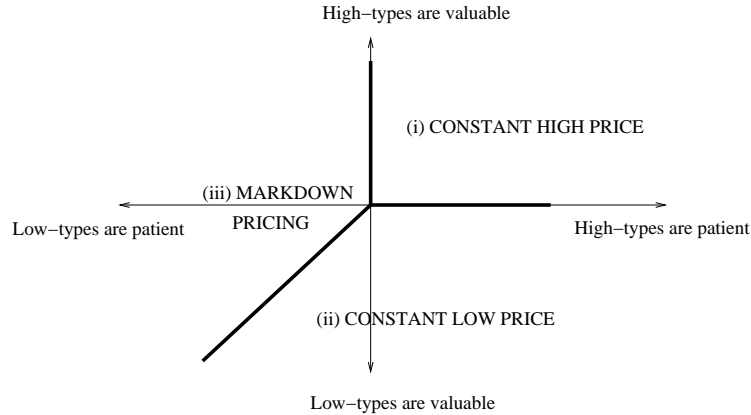(ii) CONSTANT LOW PRICE

Low–types are valuable

Figure 5: Regions for candidate pricing regimes

Figure 5 shows the regions (in parameter space) over which each of these three cases apply. On the x-axis, we have

$$x = \alpha\phi_H V_H - (\alpha\phi_H + \overline{\alpha}\phi_L)V_L, \tag{21}$$

which represents the profit differential between selling to the patient-high-types at the high price V_H and selling to all patient customers at the low price V_L. Observe that x is increasing in ϕ_H but decreasing in ϕ_L. Therefore, we can interpret an increase in x as the high-type population becoming proportionately more patient (i.e. an increase in ϕ_H), and we can interpret a decrease in x as the low-type population becoming proportionately more patient (i.e. an increase in ϕ_L). Next, on the y-axis, we have

$$y = \alpha V_H - V_L, \tag{22}$$

which represents the profit differential between selling to all the high-types at the high price V_H and selling to all customers at the low price V_L. Therefore, we can interpret an increase in y as the high-type population becoming more valuable (i.e. an increase in αV_H), and we can interpret a decrease in y as the low-type population becoming more valuable (i.e. an increase in V_L). According to Proposition 7, cases (i) and (iii) may apply in the upper half-plane (i.e. $y \geq 0$), and the other condition for (i) to hold here is $\alpha\phi_H V_H \geq (\alpha\phi_H + \overline{\alpha}\phi_L)V_L$, or $x \geq 0$. Similarly, in the lower half-plane (i.e. $y \leq 0$), cases (ii) and (iii) may apply. Here, the condition for (ii) to hold is $\alpha\overline{\phi}_H V_H \leq (\alpha\overline{\phi}_H + \overline{\alpha}\,\overline{\phi}_L)V_L$, or $x \geq y$. This yields the three regions shown in the diagram.

In a similar model, Wilson (1988) shows that uniform prices are optimal if the seller can choose initial inventory levels. In his model, all the customers are strategic (i.e. $\phi_H = \phi_L = 1$), so his results apply to situations on the $x = y$ line in Figure 5. In other words, our findings are consistent with earlier work. In fact, by considering a more general composition of the customer population, we show that apart from uniform pricing, markdowns may also be optimal when the initial inventory level is flexible.

Let us now draw some observations from the regions in Figure 5. First, when the low-type population is proportionately more patient than the high-type population, the seller should use markdown pricing. This agrees with our earlier findings when the initial inventory is fixed. On the other hand, when the high-type population is proportionately more patient, the seller should charge fixed prices: a constant high price should be used when there is a significant value differential between the high-type and low-type clienteles, otherwise, a constant low price should be used. This result differs from our earlier case where the initial inventory is fixed; in that case, markup pricing is appropriate because it provides the optimal time-mixture between the constant-low-price policy and the constant-high-price policy, subject to the inventory-constraint that all available units must be sold. However, when the initial inventory is flexible, such time-mixing is no longer needed, because the seller is free to choose the constant price (and stock the associated quantity) that maximizes profits.

These results suggest that price increases, which are quite common in practice, are driven by factors that extend beyond the realm of our stylized model. For example, consider uncertainty and non-stationarity in demand patterns. In environments with aggregate demand uncertainty, initially low prices may help the firm learn about market potential. Furthermore, with time-inhomogeneous demand, increasing prices may even be necessary if higher-value customers arrive later. A more complete investigation of the joint pricing-inventory problem with strategic customers should incorporate these factors, in order to better identify the forces favoring either markups or markdowns and understand their interplay with the initial inventory choice.

Our analysis has examined both short-term (when the initial inventory is fixed) and long-term (when the initial stock is flexible) solutions. Under some practical situations, the "short-term" solutions are also relevant in the long run. This may occur when there is uncertainty in demand during the inventory planning stage. The stock of Q units must be ordered/produced in the absence of perfect demand information. Subsequently, after demand is realized, the inventory has already been ordered/produced, and the seller must choose prices according to the "short-term" solutions. Alternatively, there may also be scenarios where the initial inventory is used to serve different streams of demand (for example, in airlines, the inventory of seats on the same plane is used on flights with different demand characteristics). In this case, the stocking quantity Q is not completely flexible even in the long run: there must be some situations when the seller faces a "fixed" inventory that has been optimized for some other stream of demand, and the "short-term" solutions must be used instead. These situations suggest that our "short-term" solutions in Theorem 1 may continue to persist in the long run.

8   Conclusion

The main contribution of this paper is to analyze the inter-temporal pricing problem when the customer pool consists of four distinct groups: patient-high-types, impatient-high-types, patient-low-types, and impatient-low-types. This model sheds light on how the composition of the customer pool influences the optimal pricing regime, as well as the division of surplus between the seller and his customers. In most existing work, customers either do not wait in the market or have homogeneous discount rates. In contrast, we find that heterogeneity in both valuations and waiting costs are crucial and they jointly determine the structure of optimal pricing policies.

When the seller has a fixed initial inventory, we find that prices should decrease over time when the market is dominated by either impatient-high-types or patient-low-types. On the other hand, when the market is dominated by patient-high-types or impatient-low-types, prices should increase over time. Figure 6 summarizes these conclusions. This provides a possible explanation for the prevalence of markdowns in fashion retailing, because high-valuation customers derive immediate consumption utility and are less willing to wait until the end of the season. On the other hand, in travel industries such as airlines, markup pricing could be justified because high-types (business travelers) tend to be more willing to wait.

In the long run, the seller is able to select an initial stocking quantity. We find that markdowns remain optimal when the high-valuation customer segment is relatively less patient. However, when the high-valuation segment is relatively more patient, the long run optimum is to charge a single price (which depends on market characteristics) throughout the selling season. The seller selects the price

	Patient	Impatient
High–value	INCREASING	DECREASING
Low–value	DECREASING	INCREASING

Figure 6: Structure of optimal pricing policy when each customer segment dominates the market

that maximizes his long run profit rate, and stocks the corresponding quantity.

The results in this paper can be extended in three broad directions. The first direction is to introduce inventory. In this paper, the seller is endowed with a fixed inventory. The long run stocking decision considered herein is also highly simplified because demand is deterministic. It would be interesting to consider the seller's (retailer's) inventory ordering decisions under demand uncertainty, in a newsvendor-type setting. How does the stocking decision (which determines availability) change when customers anticipate stockouts and markdowns? How should ordering and pricing decisions be made in conjunction? What are the implications of strategic customer behavior on decentralized decisions in a supply chain? These issues are examined in Su and Zhang (2005). These questions also extend to infinite horizon applications. Incorporating strategic customer behavior into classical inventory models is also a potential area of research; see Ahn et al. (2005). The second direction is to introduce competition. When there are multiple sellers, it is interesting to see how sellers react to each other's pricing strategies. For customers, a critical modeling component is to specify how they choose between sellers at different prices. Previous work involving competition (for example, Netessine and Shumsky, 2005) do not incorporate strategic customer behavior. Finally, the third direction is to extend our setup to a service setting. The service provider sets prices dynamically, and customers strategically choose when to seek service, taking into account future prices as well as negative congestion externalities. A similar kind of customer behavior has been examined by Lariviere and van Mieghem (2003) and Armony and Maglaras (2004a, 2004b), but these papers do not consider pricing. In my opinion, each of these directions present fruitful opportunities for future research.

Appendix

Proof of Lemma 1   Suppose there exists some time t at which positive sales occur at price $p(t) < V_L$. Let s be the supremum of this set of times. Then, at time $s - \epsilon$, consider increasing prices to $\max\{p(t), V_L\}$. Sales that were planned for the time interval $[s - \epsilon, s]$ are still incentive compatible because the new prices are still lower than all future prices after time s. This modification strictly increases revenues. Thus, the credibility condition is violated. ∎

Before presenting the proof of Proposition 1, we need the following two lemmas.

Lemma 4 *Let B be the set of times over which sales occur at price V_L. Then, $r(t) = 1$ over B, except possibly for a set of measure zero.*

Proof of Lemma 4   Suppose there exists some positive-measured subset $B' \subset B$ such that $r(t) < 1$ over B'. Then, we can find a closed interval $[a, b] \subset B'$, and let $\underline{r} = \min\{r(t) : a \le t \le b\} \in (0, 1)$. Let $c = (a + b)/2$. Since $r(t) < 1$ over $[a, c]$, the market size $Z_\theta(c) > 0$.

Fix $\epsilon > 0$. Since $Z_\theta(t)$ is RCLL, there must exist $\delta > 0$ such that $|Z_\theta(c + s) - Z_\theta(c)| \le \epsilon$ for every $s \le \delta$. However, for every $\delta > 0$, we can show that $Z_\theta(c + \delta)$ is arbitrarily close to zero, because for any integer k, $Z_\theta(c + \delta) \le Z_\theta(c) \cdot [1 - r(c + \delta/k)] \cdot [1 - r(c + 2\delta/k)] \cdots [1 - r(c + \delta)] \le (1 - \underline{r})^k Z_\theta(c)$. This contradicts the right continuity of $Z_\theta(t)$ at $t = c$. ∎

Lemma 5 *Let $I = (a, b)$ be some interval over which sales occur at price V_L. Then, revenues remain unchanged under the alternative price schedule*

$$
p'(t) = \begin{cases} V_L, & t \in [0, b - a], \\ p(t - (b - a)), & t \in (b - a, b], \\ p(t), & t \in (b, 1]. \end{cases} \tag{23}
$$

Proof of Lemma 5   By Lemma 4, we must have $r(t) = 1$ over all times with price V_L. By right-continuity, we must have $p(a) = V_L$ under the original policy. Therefore, revenues earned during time interval $(a, b]$ in the original policy are equal to revenues earned during time interval $(0, b - a]$ in the modified policy. Further, notice that the problem facing customers over time interval $(0, a]$ in the original policy is identical to the problem facing customers over time interval $(b - a, b]$ in the modified policy. Therefore, the modified policy (with control processes translated accordingly) must reap the same revenues. Finally, for the time interval $(b, 1]$, choices (and thus revenues) remain unchanged. ∎

With the two preceding lemmas, we are now ready to provide the proof for Proposition 1.

Proof of Proposition 1 By the repeated application of Lemma 5 to time intervals with price V_L, any feasible policy is reduced to some policy satisfying (i), with $r(t) = 1$ following from Lemma 4. After time τ_0, we have $p(t) = V_L$ only at isolated points. Since feasible controls are limited to a finite number of discontinuities, we can have only a finite number of subsequent time points with $p(t) = V_L$. Let there be M such time points. These time points, with rationing fractions denoted by π_k, punctuate the horizon into intervals of length τ_k. If $p(1) = V_L$, we are done, as we have $\sum_{i=0}^{M} \tau_i = 1$. Otherwise, we set $p(1) = V_L$ and $r(1) = 0$, so that this final $(M+1)$-th time segment has $\tau_{M+1} = 1 - \sum_{i=0}^{M} \tau_i$ and $\pi_{M+1} = 0$; we obtain $\sum_{i=0}^{M+1} \tau_i = 1$. ∎

Proof of Lemma 2 First, observe that in order to prove an upper bound on revenue, we may assume that sales to high-types are conducted during the time segment of arrival. To see this, suppose that a high-type who arrives during the k-th time segment were carried over to the next time segment. Then, even if the maximum revenue of V_H is anticipated in the next time segment, there is a π_k chance of this high-type securing the product at price V_L at the end of the k-th time segment, resulting in expected revenue of at most $V_H - \pi_k \Delta$. This does not exceed the WTP expressions (4) and (6) obtained for both patient and impatient high-types. Therefore, we may assume that each high-type contributes their WTP at time of arrival to the upper bound. Then, we can write down the desired upper bound. The leftmost term reflects revenue collected during $[0, \tau_0]$. Within the summation, the first term is revenue earned from patient-low-types in the k-th time segment, and the next two terms are revenue upper bounds based on the WTP expressions (4) and (6) of patient-high-types and impatient-high-types respectively. ∎

Proof of Lemma 3 We may assume that $\pi_1 \geq \pi_2$, which implies $\pi_1 \geq \hat{\pi} \geq \pi_2$. Define the function

$$L(\tau, \pi) \equiv \int_0^\tau [\pi \Delta - b_I u]^+ du = \begin{cases} \tau \pi \Delta - \frac{b_I \tau^2}{2}, & b_I \tau \leq \Delta \pi, \\ \frac{(\Delta \pi)^2}{2b_I}, & b_I \tau \geq \Delta \pi. \end{cases} \tag{24}$$

For any $\tau_1, \tau_2, \pi_1, \pi_2 \in (0, 1)$ such that $\tau_1 + \tau_2 \in (0, 1]$, let $\hat{\tau} = \tau_1 + \tau_2$ and $\hat{\pi} = (\tau_1 \pi_1 + \tau_2 \pi_2)/(\tau_1 + \tau_2)$. Denote $L_1 = L(\tau_1, \pi_1)$, $L_2 = L(\tau_2, \pi_2)$ and $\hat{L} = L(\hat{\tau}, \hat{\pi})$. To prove the lemma, it suffices to show that $L_1 + L_2 \geq \hat{L}$.

We consider three cases. First, suppose that $b_I \tau_1 \geq \Delta \pi_1$. Then, we must have $b_I \hat{\tau} \geq b_I \tau_1 \geq \Delta \pi_1 \geq \Delta \hat{\pi}$. This implies that $L_1 + L_2 \geq L_1 = \frac{(\Delta \pi_1)^2}{2b_I} \geq \frac{(\Delta \hat{\pi})^2}{2b_I} = \hat{L}$ as required.

Next, suppose $\Delta \pi_1 > b_I \tau_1 \geq \Delta(\pi_1 - \pi_2)$. Then we have, as required,

$$
\begin{aligned}
L_1 + L_2 &= \int_0^{\tau_1} [\pi_1 \Delta - b_I u]^+ du + \int_0^{\tau_2} [\pi_2 \Delta - b_I u]^+ du \\
&\geq \int_0^{\tau_1} [\hat{\pi} \Delta - b_I u]^+ du + \int_0^{\tau_2} [(\pi_1 \Delta - b_I \tau_1) - b_I u]^+ du
\end{aligned}
$$

$$= \int_0^{\tau_1} [\hat{\pi}\Delta - b_I u]^+ du + \int_{\tau_1}^{\tau_1 + \tau_2} [\pi_1 \Delta - b_I u]^+ du$$

$$\geq \int_0^{\tau_1} [\hat{\pi}\Delta - b_I u]^+ du + \int_{\tau_1}^{\tau_1 + \tau_2} [\hat{\pi}\Delta - b_I u]^+ du$$

$$= \int_0^{\hat{\tau}} [\hat{\pi}\Delta - b_I u]^+ du = \hat{L}.$$

Finally, consider the case $b_I \tau_1 < \Delta(\pi_1 - \pi_2)$. Define $\pi_1' = \pi_1 - \epsilon, \pi_2' = \pi_2 + \frac{\tau_1}{\tau_2}\epsilon$ for some $\epsilon > 0$. Notice that $\tau_1 \pi_1' + \tau_2 \pi_2' = \tau_1 \pi_1 + \tau_2 \pi_2$. Next, denote $L_1' = L(\tau_1, \pi_1')$ and $L_2' = L(\tau_2, \pi_2')$. Then,

$$(L_1' + L_2') - (L_1 + L_2) = (\tau_1 \pi_1' \Delta - b_I \tau_1^2/2) - (\tau_1 \pi_1 \Delta - b_I \tau_1^2/2) + \int_0^{\tau_2} [\pi_2' \Delta - b_I u]^+ - [\pi_2 \Delta - b_I u]^+ du$$

$$\leq -\tau_1 \Delta \epsilon + \tau_2 \Delta \frac{\tau_1}{\tau_2}\epsilon = 0.$$

Therefore, we can repeatedly apply this transformation to decrease τ_1 and increase τ_2 until the preceding cases applies. ∎

Proof of Proposition 4 Our goal is to maximize, subject to (12) and (13), the upper bound

$$UB(\tau_0, \tau_1, \pi_1) = \tau_0 V_L + f_{PL}\tau_1 \pi_1 V_L + f_{PH}\tau_1(V_H - \pi_1 \Delta) + f_{IH}\int_0^{\tau_1} V_H - [\pi_1 \Delta - b_I s]^+ ds \quad (25)$$

$$= \tau_0 V_L + f_{PL}\tau_1 \pi_1 V_L + (f_{PH} + f_{IH})\tau_1(V_H - \pi_1 \Delta) + H(\tau_1, \pi_1) \quad (26)$$

$$= C + \alpha\Delta(1 - 1/\phi_L)\tau_1 + H(\tau_1, \pi_1), \quad (27)$$

where $H(\tau_1, \pi_1) = f_{IH}\left[\tau_1 \pi_1 \Delta - (\pi_1 \Delta)^2/2b_I\right]$ if $\tau_1 b_I \geq \pi_1 \Delta$ and $H(\tau_1, \pi_1) = f_{IH} b_I \tau_1^2/2$ otherwise, C is some constant, and the last equality (27) follows from $\tau_1 \pi_1 = \frac{\tau_1}{\phi_L} - \frac{1-Q}{\alpha\phi_L}$ as implied by (12) and (13).

We begin by showing that $\tau_1 b_I \geq \pi_1 \Delta$ at the optimal solution. Suppose not. Then the objective in (27) above becomes $C + \alpha\Delta(1 - 1/\phi_L)\tau_1 + f_{IH} b_I \tau_1^2/2$. This is a convex function in τ_1, which must be maximized at an extreme point, contradicting the hypothesis $\tau_1 b_I < \pi_1 \Delta$.

The condition $\tau_1 b_I \geq \pi_1 \Delta$ allows us to rewrite the objective function (27) as

$$UB(\tau_0, \tau_1, \pi_1) = C + \alpha\Delta\left(1 - \frac{1}{\phi_L}\right)\tau_1 + f_{IH}\left[\tau_1 \pi_1 \Delta - \frac{(\pi_1 \Delta)^2}{2b_I}\right] \quad (28)$$

$$= C + C' + \alpha\Delta\left(1 - \frac{1}{\phi_L}\right)\tau_1 + f_{IH}\frac{\Delta}{\phi_L}\tau_1 - f_{IH}\frac{(\pi_1 \Delta)^2}{2b_I} \quad (29)$$

$$= C + C' + \alpha\Delta\left(1 - \frac{\phi_H}{\phi_L}\right)\tau_1 - f_{IH}\frac{(\pi_1 \Delta)^2}{2b_I} \quad (30)$$

$$= C + C' + \frac{A}{1 - \phi_L \pi_1} + \frac{B}{2}\pi_1^2, \quad (31)$$

where C' is some constants, (29) follows from $\tau_1 \pi_1 = \frac{\tau_1}{\phi_L} - \frac{1-Q}{\alpha\phi_L}$ and (31) follows from eliminating τ_1 using $\tau_1 = \frac{1-Q}{\alpha(1-\phi_L \pi_1)}$, and A, B are constants defined above.

We can ignore the constant terms and maximize $\overline{UB}(\pi_1) \equiv \frac{A}{1-\phi_L \pi_1} + \frac{B}{2}\pi_1^2$ over $\pi_1 \in [0, \bar{\pi}_1]$. Taking the derivative $\overline{UB}'(\pi_1)$, see that $\overline{UB}(\pi_1)$ is increasing if and only if $G(\pi_1) = \pi_1(1 - \phi_L \pi_1)^2 \leq K$,

keeping in mind that $B < 0$. Next, observe that $G(\pi_1) = 0$ at $\pi_1 = 0$ and $\pi_1 = 1/\phi_L > \bar{\pi}_1$; in between, $G(\pi_1)$ increases to a maximum at $G(\frac{1}{3\phi_L}) = \frac{4}{27\phi_L}$ and then decreases. Therefore, $\overline{UB}(\pi_1)$ is decreasing over $\pi_1 \in [0, \bar{\pi}_1]$ in case (i) and increasing in case (iii). In case (ii), $\overline{UB}(\pi_1)$ is increasing and then decreasing (and then possibly increasing again); however, if $K \leq G(\bar{\pi}_1)$, $\overline{UB}(\pi_1)$ must increase and then decrease (but can not increase again). The result thus follows. ∎

Proof of Theorem 1   Part (a) follows directly from Propositions 4(i) and 5. Part (b) follows from Proposition 4(iii). Part (c) follows from the second part of Proposition 4(ii), and from computing $G(\bar{\pi}_1) = G\left(\frac{Q-\alpha}{(1-\alpha)\phi_L}\right) = \frac{(Q-\alpha)(1-Q)^2}{(1-\alpha)^3} \cdot \frac{1}{\phi_L}$. ∎

Proof of Proposition 6   The expressions for consumer surplus $U_H^S, U_H^M, U_L^S, U_L^M$ follow directly from the optimal price schedule characterized in Theorem 1. To obtain the seller's revenue, we again use Theorem 1 to make the following calculations. For (i) we have

$$R = \frac{Q-\alpha}{1-\alpha}V_L + \left(1 - \frac{Q-\alpha}{1-\alpha}\right)\alpha V_H \tag{32}$$

$$= \alpha V_H + \frac{Q-\alpha}{1-\alpha}(V_L - \alpha V_H). \tag{33}$$

As for (ii), since all the myopic-high-types pay V_H, all the strategic-high-types pay $\left(V_H - \frac{Q-\alpha}{\phi_L\bar{\alpha}}\Delta\right)$, and a fraction $\frac{Q-\alpha}{\phi_L\bar{\alpha}}$ of the strategic-low-types pay V_L, we have

$$R = \bar{\phi}_H \alpha V_H + \phi_H \alpha \left(V_H - \frac{Q-\alpha}{\phi_L\bar{\alpha}}\Delta\right) + \frac{Q-\alpha}{\phi_L\bar{\alpha}}\phi_L\bar{\alpha}V_L \tag{34}$$

$$= \alpha V_H + \frac{Q-\alpha}{\phi_L\bar{\alpha}}\left(\phi_L\bar{\alpha}V_L - \phi_H \alpha\Delta\right), \tag{35}$$

which completes the proof. ∎

Proof of Proposition 7   We begin by showing that there are three possible candidates for the optimal stocking quantity Q^*, and then characterize them according to the theorem.

Notice that stocking $Q < \alpha$ is never optimal because the seller can always perform better by increasing Q slightly, while continuing to charge $p(t) = V_H$ and use $r(t) = 1$.

Next, for any stocking quantity Q between α and $\alpha + \bar{\alpha}\phi_L$, the seller's optimal revenue (from using the optimal selling policy) is characterized in (14) and (17) in Proposition 6. The markdown policy yields revenue (17), which is linear in Q; hence, the optimal stocking quantity Q^* must be an extreme solution, i.e. either α or $\alpha + \bar{\alpha}\phi_L$, and the corresponding selling policies are respectively the constant-high-price policy in case (i) and the single-markdown policy in case (iii). The markup policy, which might have been optimal when Q was fixed, is now suboptimal because it involves a

time-mixture of constant-low-price and constant-high-price policies. The optimal solution should use the policy generating a higher profit rate, so the optimal stocking quantity Q^* must either be α or 1, and the corresponding selling policies are respectively the constant-high-price policy in case (i) and the constant-low-price policy in case (ii).

Similarly, any stocking quantity strictly between $\alpha + \overline{\alpha}\phi_L$ and 1 is suboptimal. In order to sell out all $Q > \alpha + \overline{\alpha}\phi_L$ units, there must be a time-mixture between the constant-low-price policy (to sell to myopic-low-types) and some other policy already considered above. However, the optimal solution should use only one policy generating the highest profit rate. By the same argument above, the three candidates for the optimum are described in cases (i), (ii), and (iii) in the theorem.

It remains to distinguish between the situations when each of the three candidates are optimal. The constant-high-price and constant-low-price policies of (i) and (ii) generate revenue αV_H and V_L respectively, whereas the markdown policy in (iii) generates revenue $\alpha \overline{\phi}_H V_H + (\alpha \phi_H + \overline{\alpha}\phi_L)V_L$. When $\alpha V_H \geq V_L$, it suffices to check the seller's preferences between the constant-high-price policy in (i) and the markdown policy in (iii); the seller prefers the former when

$$\alpha V_H \geq \alpha \overline{\phi}_H V_H + (\alpha \phi_H + \overline{\alpha}\phi_L)V_L \tag{36}$$

$$\Leftrightarrow \quad \alpha \phi_H V_H \geq (\alpha \phi_H + \overline{\alpha}\phi_L)V_L. \tag{37}$$

When $\alpha V_H \leq V_L$, it suffices to check the seller's preferences between the constant-low-price policy in (i) and the markdown policy in (iii); the seller prefers the former when

$$\alpha \overline{\phi}_H V_H + (\alpha \phi_H + \overline{\alpha}\phi_L)V_L \leq V_L \tag{38}$$

$$\Leftrightarrow \quad \alpha \overline{\phi}_H V_H \leq (\alpha \overline{\phi}_H + \overline{\alpha}\overline{\phi}_L)V_L. \tag{39}$$

This completes the proof. ∎

References

[1] Ahn, H.S., M. Gümüş, P. Kaminsky. 2005. Pricing and manufacturing decisions when demand is a function of prices in multiple periods. Working paper.

[2] Armony, M., C. Maglaras. 2004a. On customer contact centers with a call-back option: customer decisions, routing rules and system design. *Oper. Res.* 52(2): 271-292.

[3] Armony, M., C. Maglaras. 2004b. Contact centers with a call-back option and real-time delay information. *Oper. Res.* 52(4): 527-545.

[4] Arndorfer, J.B., M. Creamer. 2005. Best Buy taps Rapp to weed out angels, devils. *Advertising Age.* 76(20): 8.

[5] Arnold, M.A., S.A. Lippman. 2001. The analytics of search with posted prices. *Econ. Theory.* 17: 447-466.

[6] Ausubel, L.M., R.J. Deneckere. 1989. Reputation in bargaining and durable goods monopoly. *Econometrica.* 57(3): 511-531.

[7] Aviv, Y., A. Pazgal. 2003. Optimal pricing of seasonal products in the presence of forward-looking consumers. Working paper.

[8] Besanko, D., W.L. Winston. 1990. Optimal price skimming by a monopolist facing rational consumers. *Mgmt. Sci.* 36(5): 555-567.

[9] Bitran, G.R., S.V. Mondschein. 1997. Periodic pricing of seasonal products in retailing. *Mgmt. Sci.* 43(1): 64-79.

[10] Bitran G., R. Caldentey. 2003. An overview of pricing models and revenue management. *Manufacturing & Service Oper. Mgmt.* 5(3): 203-229.

[11] Bulow, J.I. 1982. Durable-goods monopolists. *J. Polit. Econ.* 90(2): 314-332.

[12] Coase, R.H. 1972. Durability and monopoly. *J. Law & Econ.* 15(1): 143-149.

[13] Conlisk J., E. Gerstner, J. Sobel. 1984. Cyclic pricing by a durable goods monopolist. *Quar. J. Econ.* 99(3): 489-505.

[14] Cooper, W.L., T. Homem-de-Mello, A.J. Kleywegt. 2004. Models of the spiral-down effect in revenue management. Working paper.

[15] Dana, J.D., Jr. 1998. Advance-purchase discounts and price discrimination in competitive markets. *J. Polit. Econ.* 106(2): 395-422.

[16] Dana, J.D., Jr. 1999a. Using yield management to shift demand when the peak time is unknown. *RAND J. Econ.* 30(3): 456-474.

[17] Dana, J.D., Jr. 1999b. Equilibrium price dispersion under demand uncertainty: the roles of costly capacity and market structure. *RAND J. Econ.* 30(4): 632-660.

[18] Dana, J.D., Jr. 2001. Monopoly price dispersion under demand uncertainty. *International Econ. Rev.* 42(3): 649-670.

[19] Das Varma, G., N. Vettas. 2001. Optimal dynamic pricing with inventories. *Econ. Letters.* 72: 335-340.

[20] Desiraju, R., S.M. Shugan. 1999. Strategic service pricing and yield management. *J. Mktg.* 63(1):44-56.

[21] Elmaghraby, W., P. Keskinocak. 2003. Dynamic pricing in the presence of inventory considerations: research overview, current practices and future directions. *Mgmt. Sci* 49(10): 1287-1309.

[22] Elmaghraby, W., A. Gulcu, P. Keskinocak. 2004. Optimal markdown mechanisms in the presence of rational customers with multi-unit demands. Working paper.

[23] Federgruen, A., A. Hetching. 1999. Combined pricing and inventory control under uncertainty. *Oper. Res.* 47(3): 454-475.

[24] Feng, Y., G. Gallego. 1995. Optimal starting times for end-of-season sales and optimal stopping times for promotional fares. *Mgmt. Sci.* 41(8): 1371-1391.

[25] Feng, Y., G. Gallego. 2000. Perishable asset revenue management with Markovian time dependent demand intensities. *Mgmt. Sci.* 46(7): 941-956.

[26] Feng, Y., B. Xiao. 2000a. Optimal policies of yield management with multiple predetermined prices. *Oper. Res.* 48(2): 332-343.

[27] Feng, Y., B. Xiao. 2000b. A continuous-time yield management model with multiple prices and reversible price changes. *Mgmt. Sci.* 46(5): 644-657.

[28] Gallego, G., G. van Ryzin. 1994. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Mgmt. Sci.* 40(8): 999-1020.

[29] Gallego, G., G. van Ryzin. 1997. A multiproduct dynamic pricing problem and its applications to network yield management. *Oper. Res.* 45(1): 24-41.

[30] Gallien, J. 2004. Dynamic mechanism design for online commerce. Working paper.

[31] Gul, F., H. Sonnenschein, R. Wilson. 1986. Foundations of dynamic monopoly and the Coase conjecture. *J. Econ. Theory.* 39(1): 155-190.

[32] Guth, W., K. Ritzberger. 1998. On durable goods monopolies and the Coase-conjecture. *Rev. Econ. Design.* 3(3):215-236.

[33] Harris, M., A. Raviv. 1981. A theory of monopoly pricing schemes with demand uncertainty. *Amer. Econ. Rev.* 71(3): 347-365.

[34] Kadet, A. 2004. The price is right. *Smart Money.* Dec 2004: 90-94.

[35] Lariviere, M.A., J.A. van Mieghem. 2004. Strategically seeking service: how competition can generate Poisson arrivals. *Manufacturing & Service Oper. Mgmt.* 6(1): 23-40.

[36] Lazear, E.P. 1986. Retail pricing and clearance sales. *Amer. Econ. Rev.* 76(1): 14-32.

[37] Maglaras, C., J. Meissner. 2006. Dynamic pricing strategies for multi-product revenue management problems. *Manufacturing & Service Oper. Mgmt.* 8(2): 136-148.

[38] McAfee, R.P., V. te Velde. 2005. Dynamic pricing in the airline industry. Working paper.

[39] McWilliams, G. 2004. Minding the store: analyzing customers, Best Buy decides not all are welcome. *Wall Street Journal.* Nov 8: A1.

[40] Netessine, S., S. Savin, W.Q. Xiao. 2004. Revenue management through dynamic cross-selling in e-commerce retailing. Working paper.

[41] Netessine, S., R.A. Shumsky. 2005. Revenue management games: horizontal and vertical competition. *Mgmt. Sci.* 51(5): 813-831.

[42] Ovchinnikov, A., J.M. Milner. 2005. Strategic response to wait-or-buy: revenue management through last-minute deals in the presence of customer learning. Working paper.

[43] Pashigian, B.P. 1988. Demand uncertainty and sales: a study of fashion and markdown pricing. *Amer. Econ. Rev.* 78(5): 936-953.

[44] Queenan. J. 2005. The shoppers from hell. *Chief Executive.* 207: 70.

[45] Scholsser, J. 2004. Markdown lowdown. *Fortune.* 149(1): 40.

[46] Shumsky, R.A., F. Zhang. 2004. Dynamic capacity management with substitution. Working paper.

[47] Sobel, J. 1991. Durable goods monopoly with entry of new consumers. *Econometrica.* 59(5): 1455-1485.

[48] Stokey, N.L. 1979. Intertemporal price discrimination. *Quar. J.Econ.* 93(3): 355-371.

[49] Stokey, N.L. 1981. Rational expectations and durable goods pricing. *Bell J. Econ.* 12(1): 112-128.

[50] Su, X., F. Zhang. 2005. Strategic customer behavior, commitment, and supply chain performance. Working paper.

[51] Talluri, K.T., G.J. van Ryzin. 2004. Revenue management under a general discrete choice model of consumer behavior. *Mgmt. Sci.* 50(1): 15-33.

[52] Talluri, K.T., G.J. van Ryzin. 2005. *The Theory and Practice of Revenue Management.* Springer. New York, NY.

[53] Tirole, J. 1988. *The theory of industrial organization.* MIT Press. Cambridge, MA.

[54] van Ryzin, G., Q. Liu. 2004. On the choice-based linear programming model for network revenue management. Working paper.

[55] van Ryzin, G., Q. Liu. 2005. Strategic capacity rationing to induce early purchases. Working paper.

[56] Wilson, C.A. 1988. On the optimal pricing policy of a monopolist. *J. Polit. Econ.* 96(1): 164-176.

[57] Xu, X., W.J. Hopp. 2004. Customer heterogeneity and strategic behavior in revenue management: a martingale approach. Working paper.

[58] Zhang, D., W.L. Cooper. 2005a. Revenue management for parallel flights with customer-choice behavior. *Oper. Res.* 53(3): 415-431

[59] Zhang, D., W.L. Cooper. 2005b. Pricing substitutable flights in airline revenue management. Working paper.

[60] Zhao, W., Y-S Zheng. 2000. Optimal dynamic pricing for perishable assets with nonhomogeneous demand. *Mgmt. Sci.* 46(3): 375-388.

[61] Zhou, Y.P., M. Fan, M. Cho. 2005. On the threshold purchasing behavior of customers facing dynamically priced perishable products. Working paper.