

# Optimal Multiperiod Pricing with Service Guarantees

Christian Borgs

Microsoft Research, New England Lab, Cambridge, Massachusetts 02142, borgs@microsoft.com

Ozan Candogan

Fuqua School of Business, Duke University, Durham, North Carolina 27708, ozan.candogan@duke.edu

Jennifer Chayes

Microsoft Research, New England Lab, Cambridge, Massachusetts 02142, jchayes@microsoft.com

Ilan Lobel

Stern School of Business, New York University, New York, New York 10012, ilobel@stern.nyu.edu

Hamid Nazerzadeh

Marshall School of Business, University of Southern California, Los Angeles, California 90089,  
hamidnz@marshall.usc.edu

We study the multiperiod pricing problem of a service firm with capacity levels that vary over time. Customers are heterogeneous in their arrival and departure periods as well as valuations, and are fully strategic with respect to their purchasing decisions. The firm's problem is to set a sequence of prices that maximizes its revenue while guaranteeing service to all paying customers. We provide a dynamic programming based algorithm that computes the optimal sequence of prices for this problem in polynomial time. We show that due to the presence of strategic customers, available service capacity at a time period may bind the price offered at another time period. This phenomenon leads the firm to utilize the same price in multiple periods, in effect limiting the number of different prices that the service firm utilizes in optimal price policies. Also, when customers become more strategic (patient for service), the firm offers higher prices. This may lead to the underutilization of capacity, lower revenues, and reduced customer welfare. We observe that the firm can combat this problem if it has an ability, beyond posted prices, to direct customers to different service periods.

*Keywords:* dynamic pricing; strategic customers; nonlinear programming; dynamic programming

*History:* Received June 17, 2011; accepted September 3, 2013, by Dimitris Bertsimas, optimization. Published online in *Articles in Advance* February 14, 2014.

## 1. Introduction

Dynamic pricing is one of the key tools available to a service firm trying to match time-varying supply with time-varying demand. It is, however, a delicate tool to use in the presence of customers who strategically time their purchases. As customers change the timing of their purchases, not only might the firm lose revenue, but it might also cause the firm's capacity to be strained in periods where a low price is offered.

We consider a general formulation of a multiperiod pricing problem of a service firm trying to maximize its revenue while selling service to strategic customers who arrive and depart over time. We assume the firm is constrained by its time-varying service capacity level and that it wishes to provide service guarantees to its customers. More specifically, the firm announces a sequence of prices in advance; each customer chooses the period with the lowest price between her arrival and departure. The sequence of prices is chosen to maximize the revenue of the firm

while guaranteeing that each customer who is willing to pay the price at a given period will obtain service. Our main contributions are to provide algorithms that compute the firm's optimal pricing policy and to characterize the properties of such optimal policy.

Service guarantees are an important contract feature that are often used when the customers themselves are businesses that rely on the service they purchase to run their own operations. An example of a setting with the aforementioned properties comes from the cloud computing market where firms sell computational services to their customers.<sup>1</sup>

In this market, despite demand for service varying quite significantly over time, customers typically demand reliability from their providers, in the sense

<sup>1</sup> Cloud computing is a large and quickly growing business. The combined revenues of this market were estimated to be more than \$22 billion in 2010 and are expected to reach \$55 billion by 2014 (see Lohr 2011).

that they should be able to purchase service whenever they need it and rationing is not tolerated. For instance, consider the case of Domino's Pizza, which is a client of Microsoft's Windows Azure cloud computing platform. As explained by Domino's director of eCommerce: "We have daily peaks for dinner rush, with Friday night being the biggest. Super Bowl, however, has a peak 50 percent larger than our busiest Friday night. Windows Azure allows me to focus on customer facing functionality, and not have to worry about whether or not I have enough hosting capacity to support it" (Vitek 2009). For Domino's and many other companies, the key managerial benefit of purchasing cloud computing services is that it permits them to completely ignore the hosting capacity needs of their online businesses. This is only the case because the cloud computing providers go to great lengths to ensure that their service is always available and, therefore, companies who rely on them need not worry about rationing risk.<sup>2</sup>

The clients of cloud computing services are also highly heterogeneous with respect to their willingness-to-wait for service. Although some companies utilize cloud computing to run on-demand services and websites, and thus always need immediate service, others use the cloud to run simulations or solve large-scale optimization problems such as the ones that arise in financial analysis, weather forecasting, and genome sequencing. Such clients will typically display more strategic behavior in their purchasing of cloud services and wait for lower prices before purchasing service (e.g., see the AWS case study DNAnexus (Amazon Web Services 2011)).

Currently, most cloud computing services are sold via static pricing or via a combination of long-term contracts and static pricing (e.g., Amazon Web Services 2012, Windows Azure 2012). More specifically, the customers can purchase computation capacity (starting at around 10 cents per hour) in a pay-as-you-go model where the per-hour price is constant over time; this hourly rate is reduced for customers who pay in advance, via yearly contracts, to reserve capacity. The largest players in this market have mostly shied away from selling their higher quality services via dynamic pricing;<sup>3</sup> this could potentially be

attributed to the difficulty of maintaining service level guarantees while customers are strategically timing their purchases. Our work provides the firm with techniques for using dynamic pricing in such a context and, therefore, giving them a tool to better manage their resources and revenues.

There are other examples of service firms that need to set profit-maximizing prices while guaranteeing service. For instance, the increasingly popular Uber online taxi service offers a flat pricing scheme on most days of the year, but utilizes dynamic pricing in high demand days (see Bilton 2012). As explained by its CEO, Uber is "aiming to provide a reliable ride to anybody who needs one, no matter how crazy demand is or what is going on in the city" (Kalanick 2011). Uber is able to provide such a service guarantee by, in times of higher demand, conserving resources by charging higher prices.

Another interesting application of our work is in the context of electricity markets. As smart meters (see Federal Energy Regulatory Commission 2008) are beginning to be widely deployed, allowing customers to immediately respond to price changes, the techniques we develop in this paper will become increasingly useful since electricity markets feature many of the elements of our model: capacity, demand, and prices are time varying, and rationing of service is highly undesirable.

In the industries discussed above, it is mainly the firm's responsibility to set prices that ensure that all service requests can be accommodated with a limited service capacity. This is in contrast to settings such as traditional retailing, where customers are exposed to rationing risk. In a traditional retail setting, strategic customers consider the risk of a stock-out and this incentivizes them to purchase the good earlier. This rationing risk mitigates the effect of strategic customer behavior on the firm's ability to set its own prices. In our setting, the firm's need to offer service guarantees places the entire burden of matching supply and demand over time on the firm. The intuition is that the firm should increase its prices when demand is high (or capacity is low) to shift some of the demand to the time periods with enough capacity. This is similar to the pricing scheme used by the major cellphone carriers that, in order to decongest their networks during business hours, often charge lower prices for making calls on nights and weekends (e.g., see AT&T 2012). The dynamic pricing tools we propose hold the promise of helping firms in such industries improve their resource utilization by better matching supply and demand over time.

### 1.1. Our Framework and Contributions

We consider a monopolist that offers service to customers over a finite horizon. The firm faces a (possibly

<sup>2</sup>"Organizations worry about whether utility computing services will have adequate availability, and this makes some wary of cloud computing. ...Google Search has a reputation for being highly available, to the point that even a small disruption is picked up by major news sources. Users expect similar availability from new services..." (Armbrust et al. 2010, p. 54).

<sup>3</sup>The main exception would be Amazon's spot market (Amazon Web Services 2012), which is a secondary market run by Amazon to sell the excess capacity of its main platform. Although the spot market prices fluctuate over time, the exact manner in which these prices are determined is not publicly available (see Ben-Yehuda et al. 2011).

time-varying) capacity constraint at each time period. The firm's objective is to implement a posted pricing scheme to maximize its revenue. At time 0, the monopolist declares—and commits to—a sequence of prices for its service, one for each time period. Given those preannounced prices, customers decide whether and when to purchase service. The firm needs to solve the constrained optimization problem of determining the prices that maximize revenue while still fulfilling all customer purchase requests.

Each customer is assumed to be infinitesimal and demands an (also infinitesimal) single unit of service. The valuation of a given customer for a unit of service is drawn from a known distribution. She is also associated with an arrival and a departure time. The arrival time corresponds to the time she enters the system and the departure time represents her deadline for obtaining service. All customers are fully strategic about whether and when they purchase service from the firm. That is, each customer either refuses to buy service (if her valuation is below any of the prices offered while she is present) or buys service at the period when she is offered the lowest price among all the periods in which she is present; if two periods have the same low price, the customer prefers the earlier one.

First, we consider a deterministic baseline model where the monopolist knows the total mass of customers that arrive at each given time period as well as their departure periods. The assumption of deterministic demand is justified when the number of customers is large and fairly predictable, which is often the case in the market for cloud computing. This modeling choice allows us to study the impact of strategic customers and time-varying demand and capacity on the optimal sequence of prices, but it deliberately removes the element of uncertainty from the model. The demand being deterministic also implies the optimality of using a sequence of preannounced prices. We show later in the paper that many of the insights obtained in this simple environment naturally extend to general settings that allow for uncertainty in the model.

Interestingly, even the solution of this baseline model is far from trivial. We show that because of the presence of strategic customers, the set of feasible price vectors is neither convex nor closed. This means no off-the-shelf software can be used to solve this problem efficiently. Despite these challenges, we are able to establish that the firm's price optimization problem is a tractable one and we provide an efficient polynomial-time algorithm for computing the optimal pricing policy. This result relies on two crucial ideas: the set of prices that the firm needs to consider is

not too large, and prices can be combined into a policy via dynamic programming because strategic customers never wait past a low price to purchase service at a future price that is higher.

We extend our results to models where the firm does not know its service capacity levels or the number of arriving customers. We do so in two distinct ways. First, we consider a robust optimization framework (see Ben-Tal and Nemirovski 2002, Bertsimas and Thiele 2006), where there is uncertainty about the firm's capacity and the size of the customer population at any given period, and the firm only knows that these parameters belong to given sets. In this setting, the firm tries to maximize its worst-case revenues while ensuring that the capacity constraints are not violated for any realization of demand and capacities. Second, we study the model in a stochastic setting where the seller knows the distribution of the uncertain parameters and is able to obtain additional capacity at a cost; the goal is to determine a sequence of (preannounced) prices that maximizes the expected profit. Additionally, this model allows for production costs and different value distributions for customers with differing patience levels. We establish that the insights from our baseline model carry on to these general environments. That is, using a dynamic programming approach similar to our baseline model, we are able to provide algorithms that compute the (near) optimal pricing policy in polynomial time.

We also consider a related setting where customers do not simply choose the earliest time period with the lowest price to receive service, but rather the firm chooses how customers should break ties between time periods with equal prices. In this setting, customers are guaranteed to receive service at the lowest price available to them, but the firm can more efficiently use its capacity by scheduling customers appropriately. Interestingly, our algorithms can be modified to account for this additional flexibility and solve the optimal pricing problem of the firm.

Finally, we conduct numerical studies using our algorithms and obtain further insights on the effect of strategic customers and service guarantees on both the firm and its customers. We show that even in settings with high volatility in service capacity and demand, the number of price levels that optimal pricing policy employs is small. For instance, in a 24-period model, the optimal price sequence includes four different price levels on average. This shows that even in complex multiperiod settings, the customers' strategic behavior severely constrains the firm's choice of price sequence. We also observe that if patient customers can wait longer for service, both the revenue of the firm and the aggregate customer welfare may decrease. This occurs because the firm is forced to

use higher prices to maintain its service guarantees, and consequently the service capacity is underutilized. Thus we conclude that in a phenomenon similar to Braess's paradox (Başar and Olsder 1999), when customers have additional freedom in choosing the time period they purchase service, the overall performance of the system may decrease.

## 1.2. Related Work

In this section, we present a brief overview of the literature on pricing mechanisms in the presence of customers who strategically time their purchases and discuss how the results in the literature relate to ours. There is also an extensive literature on dynamic pricing with myopic customers (see, e.g., Lazear 1986, Wang 1993, Gallego and Ryzin 1994, Feng and Gallego 1995, Bitran and Mondschein 1997, Federgruen and Heching 1999). We do not provide a summary of this line of literature here, but we refer the reader to excellent surveys by Talluri and van Ryzin (2004), Bitran and Caldentey (2003), Chan et al. (2004), Shen and Su (2007), and Aviv et al. (2009).

The study of monopoly pricing in the presence of strategic customers was pioneered by Coase (1972). Coase conjectured that in a setting in which a monopolist sells a durable good to patient customers, if the monopolist cannot commit to a sequence of posted prices, then the prices would converge to the production cost. Later, Stokey (1979, 1981), Gul et al. (1986), and Besanko and Winston (1990) showed that a decreasing sequence of prices is optimal for selling durable goods when customers face the trade-off of consuming right away versus the possibility of purchasing at the lower prices in the future. They observe that customers with high valuations buy in earlier periods and pay higher prices compared to the low valuation customers.

In the context of revenue management, Aviv and Pazgal (2008) study a model where a monopolist sells multiple items over a finite time horizon to strategic customers who arrive over time. The authors consider two classes of pricing strategies: contingent posted pricing, where the firm's prices may depend on the remaining inventory, and preannounced posted pricing. They observe that commitment (preannounced discount) can benefit the seller when customers are strategic. Also, ignoring the strategic customer behavior can lead to significant loss of revenue. Elmaghraby et al. (2008) and Dasu and Tong (2010) extend the analysis to a setting where the seller can reduce the prices multiple times over the time horizon. Other papers that consider commitment to a pricing policy include Arnold and Lippman (2001), Levin et al. (2010), and Cachon and Feldman (2013). These works mainly consider markdown pricing. Su (2007) shows that if the customers are heterogeneous regarding

their time sensitivity, then the optimal sequence of posted prices might also be increasing.

In the aforementioned works, the service provider uses the customers' fear of rationing to extract more revenue from strategic customers (see Liu and van Ryzin 2008). In contrast, in our model the firm ensures the customers does not face such risks and provides service guarantees. Su and Zhang (2009) consider the issue of rationing in the presence of strategic customers and find that sellers have an incentive to overinsure consumers against the risk of stockouts, thus showing that providing service guarantees can be in the firm's interest.

Another paper related to ours is the one by Ahn et al. (2007), which considers joint pricing and production decisions. Unlike us, they assume that customers are myopic with respect to prices, but, similarly to our model, they assume customers stay in the system for a number of periods unless they make a purchase. Interestingly, their analysis also relies on the notion that a low price effectively separates past and future through what they call regeneration points.

An altogether different approach to this problem is the one taken by the dynamic mechanism design literature. There, the firm offers a direct mechanism that allocates its service as a function of customers reports of their private valuations, entry period, and departure period. See Bergemann and Said (2011) for a survey. The paper closest to this one within this literature is Pai and Vohra (2013), where strategic customers arrive and depart over time, but the allocation problem they study is quite dissimilar to the one we consider.

The model we consider here differs from many papers in the literature in at least three key aspects: in our model, the firm guarantees service to all paying customers; therefore, the customers do not face rationing risk. Second, instead of having a fixed inventory at time 0, in our model, the firm has a time-varying service capacity, which is nonstorable. Finally, in the previous work, the customers are either present from the beginning of the time horizon or arrive over time but remain until the end of the horizon (or until they make a purchase). In our model, buyers arrive and depart the system over time.

## 2. The Baseline Model

In this section, we formulate the firm's revenue maximization problem, which will be studied in the subsequent sections. The firm sets a vector of prices over a finite horizon  $t = 1, \dots, T$ . The prices, denoted by  $\mathbf{p} = (p_1, \dots, p_T)$ , are announced up front, one price for each period  $t$ .<sup>4</sup> Customers arrive and depart over

<sup>4</sup>For instance, the horizon of the problem can be chosen as a day, with periods corresponding to different hours during the day, to



time and are infinitesimal. We denote the population of customers that arrive at period  $i$  and depart at period  $j$  by  $a_{i,j}$ . With slight abuse of notation, we also represent the mass of the population that arrives at period  $i$  and departs at period  $j$  by  $a_{i,j}$ .

Each customer wants one unit of service<sup>5</sup> from which she obtains a (nonnegative) value, and customers are strategic with respect to the timing of their purchases. Given the vector of prices  $\mathbf{p}$ , a customer from population  $a_{i,j}$ , with value  $v$  for the service, purchases the service at a time period with the lowest price between times  $i$  and  $j$ , if her value is larger than the lowest price, i.e., if  $v \geq \min_{\ell: i \leq \ell \leq j} \{p_\ell\}$ . If there is more than one period with the lowest price in  $\{i, \dots, j\}$ , the customer chooses the earliest period (with the minimum price) to obtain the service.<sup>6</sup>

Given a price vector  $\mathbf{p}$ , we can assign to each population  $a_{i,j}$  a service period, denoted by  $\pi_{i,j}(\mathbf{p})$ . This period has the lowest price among periods in  $\{i, \dots, j\}$  and is the earliest one (in  $\{i, \dots, j\}$ ) with this price. Each member of population  $a_{i,j}$  considers purchasing service at time  $\pi_{i,j}(\mathbf{p})$  and will purchase service if her value exceeds the price at that period. We call the mass of customers that, given prices  $\mathbf{p}$ , consider obtaining service at period  $t$  as the *potential demand* at time  $t$  and denote it by  $\bar{p}_t(\mathbf{p})$ . Formally, the potential demand is given by

$$\bar{p}_t(\mathbf{p}) = \sum_{i,j:1 \leq i \leq t \leq j \leq T} a_{i,j} \mathbf{1}\{t = \pi_{i,j}(\mathbf{p})\}, \quad (1)$$

where  $\mathbf{1}$  is an indicator function.

Each customer assigns a nonnegative value for obtaining service. The fraction of customers with value below  $v$  is given by  $F(v)$ . For simplicity of presentation, we assume that  $F$  is a continuous function and  $v \in [0, 1]$  for all customers. We also assume that customer valuations are independent of their arrival and departure periods, an assumption that we relax in §7. Hence, given price vector  $\mathbf{p}$ , the *demand* at time  $t$ , denoted by  $\bar{D}_t(\mathbf{p})$ , is equal to  $\bar{D}_t(\mathbf{p}) = (1 - F(p_t))\bar{p}_t(\mathbf{p})$ .

The firm's objective is to maximize its revenue, which is given by  $\sum_{t=1}^T p_t \bar{D}_t(\mathbf{p})$ . However, the firm is constrained by a service capacity level of  $c_t$  for each  $t \in \{1, \dots, T\}$ . The firm provides service guarantees to its customers, so it must set prices that ensure that the demand  $\bar{D}_t(\mathbf{p})$  does not violate the capacity  $c_t$  at

any period  $t$ . Thus, the firm's decision problem is given by

$$\begin{aligned} \sup_{\mathbf{p} \geq 0} \quad & \sum_{t=1}^T p_t \bar{D}_t(\mathbf{p}) \\ \text{s.t.} \quad & \bar{D}_t(\mathbf{p}) \leq c_t, \quad \text{for all } t \in \{1, \dots, T\}, \end{aligned} \quad (\text{OPT-1})$$

where  $\mathbf{p} \geq 0$  is a shorthand notation for  $p_t \geq 0$  for all  $t \in \{1, \dots, T\}$ . The above problem searches for the supremum of the objective function instead of the maximum, since the maximum of OPT-1 does not always exist. We demonstrate the nonexistence of an optimal solution in §3, where we also present our technique for handling this issue.

If there were no capacity constraints, the firm could use a single price  $p$  at all periods to maximize its revenue,<sup>7</sup> and this would result in a revenue equal to  $p(1 - F(p)) \sum_{i \leq j} a_{i,j}$ . Since  $\sum_{i \leq j} a_{i,j}$  is a constant, we call  $p(1 - F(p))$  the *uncapacitated revenue function*. We make the following regularity assumption to simplify our analysis.

**ASSUMPTION 1.** *The uncapacitated revenue function  $p(1 - F(p))$  is unimodal. That is, there exists some monopoly price  $p_M$  such that  $p(1 - F(p))$  is increasing for all  $p < p_M$  and decreasing for all  $p > p_M$ .*

Note that this assumption implies that  $p_M$  maximizes  $p(1 - F(p))$ , and it is satisfied for a wide range of distributions, including the uniform, normal, log-normal, and exponential distributions.

We now show, by means of an example, that the set of feasible prices of OPT-1 is nonconvex.

**EXAMPLE 1.** Let the time horizon be  $T = 3$  and assume that a single unit-mass of customers with uniform valuations in  $[0, 1]$  arrive at period 1 and depart at period 3. Assume that  $c_2 = 0$  and  $c_1, c_3 = 1$ . Then the price vectors  $(0, 0.1, 1)$  and  $(1, 0.1, 0)$  are both feasible. However, the average of these two price vectors,  $(0.5, 0.1, 0.5)$ , is infeasible since all customers with valuation above 0.1 seek service at period 2, violating the service capacity  $c_2 = 0$ . Therefore, the set of feasible prices of OPT-1 is nonconvex.

The above example illustrates that OPT-1 is a nonconvex optimization problem, and we cannot hope to solve it using off-the-shelf optimization tools. We show in §5 that despite being nonconvex, there exists a polynomial-time algorithm that solves this optimization problem. The construction of this algorithm relies on the structural properties of this pricing problem that are explored in §§3 and 4.

capture the problem of the firm selecting prices for its next business day. Such an approach would be reasonable when deciding day-ahead prices for cloud computing services or electricity markets.

<sup>5</sup> A customer that wants multiple units of service could be considered as multiple customers in our model.

<sup>6</sup> We relax this assumption in §8.

<sup>7</sup> Because customer valuations are independent of arrival and departure periods, it can be seen from OPT-1 that if there are no capacity constraints, setting  $p_t = \arg \max_p p(1 - F(p))$  for all  $t$  maximizes revenue.

### 3. Optimizing over Prices and Rankings

In this section, we show that there does not always exist a feasible solution achieving the supremum in the firm’s optimization problem. To address this issue, we construct a closely related optimization problem where the firm tries to maximize not only prices but also rankings of the prices. We show that this optimization problem always admits an optimal solution that can be used to obtain feasible solutions arbitrarily close to the supremum of the original problem.

We start with an example that shows that the supremum of OPT-1 may not be achieved by a feasible price vector. The main idea is that since the customers always seek the lowest price available, the potential demand function  $\bar{\rho}_t$  is a discontinuous function of  $\mathbf{p}$ ; thus, the feasible set of OPT-1 is open.

**EXAMPLE 2.** Consider a two-period model with customer valuations drawn uniformly from  $[0, 1]$ , capacity levels  $c_1 = \frac{1}{2}$  and  $c_2 = \infty$ , and customer populations  $a_{1,1} = a_{1,2} = 1$  (and  $a_{2,2} = 0$ ). Observe that solutions of the form  $(p_1, p_2) = (\frac{1}{2}, \frac{1}{2} - \epsilon)$  are feasible for any  $\epsilon > 0$ : the members of population  $a_{1,1}$  with value above  $\frac{1}{2}$  obtain service at time 1 and the members of population  $a_{1,2}$  with value above  $\frac{1}{2} - \epsilon$  are served at time 2. Hence,  $(p_1, p_2) = (\frac{1}{2}, \frac{1}{2} - \epsilon)$  yields the revenue of  $\frac{1}{2} \times \frac{1}{2} + (\frac{1}{2} - \epsilon) \times (\frac{1}{2} + \epsilon) = \frac{1}{2} - \epsilon^2$ . The revenue is decreasing in  $\epsilon$  and as  $\epsilon$  tends to 0 the revenue approaches  $\frac{1}{2}$ . The uncapacitated problem provides an upper bound on the revenue obtained, which is  $\frac{1}{2}$ . Therefore, the supremum of OPT-1 is equal to  $\frac{1}{2}$ . However,  $(p_1, p_2) = (\frac{1}{2}, \frac{1}{2})$  is not a feasible solution because under this price vector, both populations will choose the first period for service, and this violates the capacity constraints. Therefore, the feasible set of price vectors is open and the supremum of OPT-1 is not achieved by a feasible price vector.

The nonexistence of an optimal solution can be addressed by finding (feasible) solutions that are arbitrarily close to the (infeasible) supremum. In the remainder of this section, we introduce the notion of rankings and an alternative optimization formulation that allow us to obtain such solutions for OPT-1 (or the optimal solution itself in instances where the optimum is feasible).

We refer to permutations of  $\{1, \dots, T\}$  as *rankings*. We use the notation  $R_t$  to denote the rank of time period  $t$  under ranking  $R$ . We say that a ranking  $R$  is *consistent* with a price vector  $\mathbf{p}$  if periods with lower rank have lower prices. More precisely,  $R$  is consistent with  $\mathbf{p}$  if for all  $t$  and  $t'$ ,  $R_t < R_{t'}$  implies that  $p_t \leq p_{t'}$ .

We define the *customer-preferred ranking*, denoted by  $R^C(\mathbf{p})$ , as a ranking consistent with  $\mathbf{p}$ , such that when there are multiple periods with the same price, the

earlier periods are ranked lower. Namely, if  $p_t = p_{t'}$  and  $t < t'$  then  $R_t < R_{t'}$ . It can be seen from the definition of service period  $\pi_{i,j}(\mathbf{p})$  (introduced in §2) that in OPT-1 for a given price vector  $\mathbf{p}$ , each population  $a_{i,j}$  chooses the time period between  $i$  and  $j$ , with the lowest customer-preferred ranking to (potentially) receive service. Hence potential demand  $\bar{\rho}_t$  can be expressed as a function of this ranking.

More formally, for any period  $t$  and ranking of prices  $R$  we define the *R-induced potential demand*, denoted by  $\rho_t(R)$ , as

$$\rho_t(R) = \sum_{i \leq j} a_{i,j} \mathbf{1} \left\{ R_t = \min_{k: i \leq k \leq j} \{R_k\} \right\}. \quad (2)$$

Similarly, the *R-induced demand*, denoted by  $D_t(p_t, R)$ , is defined as

$$D_t(p_t, R) = (1 - F(p_t)) \rho_t(R). \quad (3)$$

It follows from (1), (2), and the definition of customer-preferred ranking that for any price vector  $\mathbf{p}$  and customer-preferred ranking  $R^C(\mathbf{p})$ , we have  $\rho_t(R^C(\mathbf{p})) = \bar{\rho}_t(\mathbf{p})$  and  $D_t(p_t, R^C(\mathbf{p})) = \bar{D}_t(\mathbf{p})$ . That is, it is possible to express demand ( $\bar{D}_t$ ) in terms of the  $R$ -induced demand function ( $D_t$ ) and customer-preferred ranking ( $R^C$ ).

Suppose that in OPT-1 the firm could select not only the vector of prices  $\mathbf{p}$  but also any ranking  $R$  consistent with  $\mathbf{p}$  (potentially different from the customer-preferred ranking), and customers decided when to obtain service according to this ranking; i.e., each customer chooses the period with the lowest ranking between her arrival and departure time. Then, the demand at any period is given by  $D_t(p_t, R)$ , and the corresponding revenue maximization problem can be formulated as

$$\begin{aligned} \max_{p \geq 0, R \in \mathcal{P}(T)} \quad & \sum_{t=1}^T p_t D_t(p_t, R) \\ \text{s.t.} \quad & D_t(p_t, R) \leq c_t \\ & \text{for all } t \in \{1, \dots, T\}, \\ & R_t < R_{t'} \Rightarrow p_t \leq p_{t'} \\ & \text{for all } t, t' \in \{1, \dots, T\}, \end{aligned} \quad (\text{OPT-2})$$

where  $\mathcal{P}(T)$  is the set of all possible rankings of  $\{1, \dots, T\}$ . Although this problem is different from OPT-1, the solutions of these problems are closely related, as we explain next.

Since  $D_t(p_t, R^C(\mathbf{p})) = \bar{D}_t(\mathbf{p})$ , it follows that any feasible solution  $\mathbf{p}$  of OPT-1 corresponds to a feasible solution of OPT-2 given by  $(\mathbf{p}, R^C(\mathbf{p}))$ , and these solutions lead to the same objective values. Additionally, it can be seen from (1) and (2) that for a given price

vector  $\mathbf{p}$  and any ranking  $R$  consistent with  $\mathbf{p}$ , the (potential) demand levels in OPT-1 and OPT-2 are equal except for periods where the price is equal to the price offered at another period. Intuitively, unlike OPT-1, in OPT-2 the firm can choose how the customers collectively break ties between time periods with equal price, by choosing the ranking  $R$  properly, and this may lead to a difference in demand levels only at such time periods. These observations can be used to show that OPT-2 always has an optimal solution, and this solution can be used to construct a solution of OPT-1 that is arbitrarily close to the supremum.

LEMMA 1. *The following claims hold:*

1. *The problem OPT-2 admits an optimal solution  $(\mathbf{p}^*, R^*)$ .*
2. *Let  $(\mathbf{p}^*, R^*)$  be an optimal solution of OPT-2. For any  $\epsilon > 0$ , the price vector  $\mathbf{p}^* + \epsilon R^*$  is a feasible solution of OPT-1 and the revenue it obtains converges to the supremum of OPT-1 as  $\epsilon$  tends to 0.*
3. *If  $\mathbf{p}$  is an optimal solution of OPT-1, then  $(\mathbf{p}, R^C(\mathbf{p}))$  is an optimal solution of OPT-2.*

The proof of this lemma can be found in the online appendix (available at <http://pages.stern.nyu.edu/~ilobel/multiperiod-pricing-online-app.pdf>). The idea behind this lemma is that the projection of the set of feasible solutions of OPT-2 onto the set of prices is the closure of the feasible set of OPT-1. Therefore, the optimal prices generated by OPT-2 can be perturbed in a way that maintains the ranking of prices, leading to a solution of OPT-1 that is arbitrarily close to the supremum. In the rest of the paper, we focus on the solution of OPT-2, keeping in mind that an optimal solution (or a solution arbitrarily close to optimal, if the optimal solution does not exist) of OPT-1 with (almost) the same prices can be constructed using this solution.

#### 4. Structure of the Optimal Prices

In this section, we explore the structure of optimal prices in OPT-2. We first show that at all periods the monopolist has incentive to keep the prices at least as high as the monopoly price  $p_M$ . Then we use this observation to study the optimality conditions in OPT-2. Exploiting these conditions, we construct a set of prices that contains all the possible candidate optimal prices. We show that the cardinality of this set is polynomial in the time horizon  $T$ , a result we later use in §5 to obtain a polynomial-time algorithm to solve OPT-2. Proofs of the results presented in this section can be found in the online appendix.

To gain some intuition, we first consider the optimal solution in a single period setting. By Assumption 1, choosing any price  $p < p_M$  is suboptimal, and the firm has incentive to increase its price to  $p_M$ .

If setting the price equal to  $p_M$  violates the capacity constraints, then the firm increases its price to the minimum price that respects the capacity constraints. Since customers' values are bounded by 1, such a price exists. Thus, it follows that an optimal price in  $[p_M, 1]$  can be found. The following proposition shows that this intuition extends to multiperiod settings.

PROPOSITION 1. *There exists an optimal solution  $(\mathbf{p}, R)$  of OPT-2 such that  $p_t \geq p_M$  for all  $t$ .*

To prove this result, we assume that a solution where  $p_t < p_M$  for some  $t$  is given, and we raise prices that are below the monopoly price  $p_M$  in a way that maintains the ranking of the prices. This ensures that as the prices increase to  $p_M$ , the revenue increases while the demand decreases. Thus, it is possible to obtain a feasible solution that (weakly) improves revenues and satisfies  $p_t \geq p_M$ .

Note that conditioned on prices being above the monopoly price  $p_M$ , by the assumption of unimodality of the uncapacitated revenue function, the incentives of the firm and the customers are aligned: both the firm and the customers prefer lower prices over higher ones. The firm never raises prices to obtain more revenue, only to satisfy capacity constraints.

We next provide a further characterization of the prices that are used at an optimal solution of OPT-2. This characterization significantly narrows down the set of prices that needs to be considered to find an optimal solution.

PROPOSITION 2. *There exists an optimal solution  $(\mathbf{p}, R)$  of the optimization problem OPT-2 such that for each period  $t$  one of the following three statements is true:  $p_t = p_M$ ,  $p_t = 1$ , or  $p_t = p_i$  for some  $i$ , such that  $c_i = D_i(p_i, R)$  and  $p_i \in [p_M, 1]$ .*

The proof of this proposition follows by showing that unless the conditions of the proposition hold, the monopolist can modify the prices in a way that increases its profits while maintaining the feasibility of the capacity constraints.<sup>8</sup> The third condition of the proposition suggests that the price at time  $t$  is either such that the capacity constraint at time  $t$  is tight or that this price is equal to the price offered at another time period, and the capacity constraint at this other time period is tight. Hence, because of the presence of strategic customers, capacity constraints at one period may bind the prices at another period, but this requires the prices to be identical at these two periods.

This proposition implies that for an optimal solution  $(\mathbf{p}, R)$  of OPT-2, each entry of the price vector  $\mathbf{p}$

<sup>8</sup> We note that if we strengthen Assumption 1 to impose concavity on the uncapacitated revenue function, then the proposition can be proved using the Karush–Kuhn–Tucker conditions.

either belongs to  $\{p_M, 1\}$  or is above  $p_M$  and satisfies the equation

$$c_i = D_i(p, R) = \rho_i(R)(1 - F(p)) \quad (4)$$

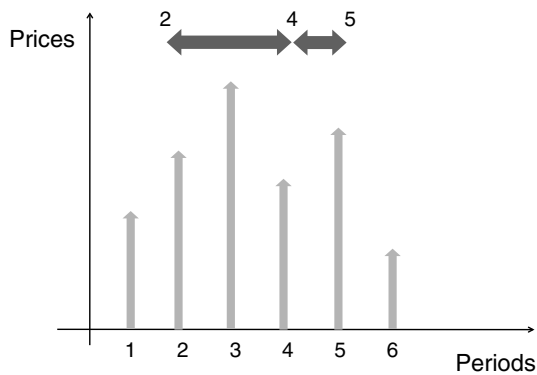
for some time period  $\hat{t}$ . However, to characterize the set of all prices that may appear at an optimal solution, we still need to consider all possible rankings. Although there are  $T!$  possible rankings  $R$ , there are a significantly smaller number of prices that satisfy equations of the form (4). To formalize this idea, we introduce the notion of attraction range, which is a representation of all the populations that choose the same period for service.

**DEFINITION 1 (ATTRACTION RANGE).** For a given consistent price-ranking pair  $(\mathbf{p}, R)$  the attraction range of a time period  $k$  is defined as the largest interval  $\{\underline{t}, \dots, \bar{t}\} \subseteq \{1, \dots, T\}$  containing  $k$  such that  $R_k = \min_{\ell \in \{\underline{t}, \dots, \bar{t}\}} R_\ell$ .

Assume that the attraction range of time period  $k$  for a consistent price-ranking pair  $(\mathbf{p}, R)$  is  $\{\underline{t}, \dots, \bar{t}\}$ . Since customers choose the time period with the lowest ranking available to them when purchasing service, customers who arrive at the system between periods  $\underline{t}$  and  $k$ , and who can wait until time period  $k$  but not beyond time period  $\bar{t}$ , are exactly the ones who will seek service at period  $k$ . Thus, the attraction range concept can be used to identify customers who are “attracted” to a particular time period for receiving service (see Example 3).

**EXAMPLE 3 (ATTRACTION RANGE).** Consider a problem instance with six time periods. Assume that a consistent price-ranking pair  $(\mathbf{p}, R)$  for this problem is given and the prices at different time periods are as in Figure 1. Since prices at all time periods are different, there is a unique ranking  $R$  consistent with these prices. The attraction range of time period 4 in this example is  $\{2, \dots, 5\}$ . Thus, customers who arrive between time periods 2 and 4 (inclusive) and who cannot wait beyond time period 5 are the ones who seek service at period 4.

**Figure 1** Attraction Range of Period 4 Is  $\{2, \dots, 5\}$  in This Six-Period Problem Instance



This example suggests that attraction ranges can be used to determine  $R$ -induced potential demand  $\rho_i(R)$ . Assume that  $(\mathbf{p}, R)$  is a consistent price-ranking pair and consider the attraction range of some time period  $k \in \{1, \dots, T\}$ , denoted by  $\{\underline{t}(k, R), \dots, \bar{t}(k, R)\}$ . As discussed earlier, customers who arrive at the system between  $\underline{t}(k, R)$  and  $k$  (inclusive) and who can wait until time  $k$  but not beyond time  $\bar{t}(k, R)$  are the only ones who can request service at time  $k$ . Thus, we obtain that  $\rho_k(R) = \sum_{i=\underline{t}(k, R)}^k \sum_{j=k}^{\bar{t}(k, R)} a_{ij}$ . From this equation it follows that  $\rho_k(R)$  can immediately be obtained by specifying the attraction range of time period  $k$ . By considering all the possible attraction ranges  $\{\underline{t}, \dots, \bar{t}\}$  corresponding to time period  $k$  we conclude that for any ranking  $R$ , we have  $\rho_k(R) \in \{\sum_{i=\underline{t}}^k \sum_{j=k}^{\bar{t}} a_{ij} \mid \underline{t} \leq k \leq \bar{t}\}$ . Using this observation, it follows that any  $p$  satisfying (4) for some  $R$  and  $\rho_k(R)$  belongs to the set

$$L_k \triangleq \left\{ \max \left\{ p_M, F^{-1} \left( 1 - \left( \frac{c_k}{\sum_{i=\underline{t}}^k \sum_{j=k}^{\bar{t}} a_{ij}} \right) \right) \right\} \mid c_k \leq \sum_{i=\underline{t}}^k \sum_{j=k}^{\bar{t}} a_{ij}, \text{ and } \underline{t} \leq k \leq \bar{t} \right\} \quad (5)$$

Here the condition  $c_k \leq \sum_{i=\underline{t}}^k \sum_{j=k}^{\bar{t}} a_{ij}$  is present since  $F^{-1}$  is defined over the domain  $[0, 1]$ . The maximum with  $p_M$  is taken to make sure that all the prices in  $L_k$  are at least equal to  $p_M$ , which follows from Proposition 2. By construction each element of  $L_k$  corresponds to an attraction range  $\{\underline{t}, \dots, \bar{t}\}$ . Since there are  $O(T^2)$  attraction ranges (there are  $O(T)$  values  $\underline{t}$  and  $\bar{t}$  can take), the cardinality of  $L_k$  is  $O(T^2)$ . Thus, we reach the following characterization of optimal prices, which is stated without proof because it immediately follows from Propositions 1 and 2 and the definition of  $L_k$  given in (5).

**PROPOSITION 3.** Let  $L$  be defined as  $L \triangleq (\bigcup_{k=1}^T L_k) \cup \{p_M\} \cup \{1\}$ . There exists an optimal solution  $(\mathbf{p}, R)$  of OPT-2 such that  $p_t \in L$  for all  $t \in \{1, \dots, T\}$ . Moreover, the cardinality of  $L$  is  $O(T^3)$ .

This proposition implies that without actually solving OPT-2, it is possible to characterize a superset of the prices that will be used at an optimal solution. Moreover, this set has polynomially-many elements, and it is sufficient for the monopolist to consider these prices when making its pricing decisions. However, finding the vector of optimal prices could still be a computationally intractable problem even if  $L$  has small cardinality. In the next section, we show that this is not case, and we develop a polynomial-time algorithm that determines the optimal sequence of prices.



## 5. A Polynomial-Time Algorithm

In this section, we use the characterization of the optimal prices obtained in §4 to design a polynomial-time algorithm for computing the optimal sequence of prices.

As shown in Proposition 3, an optimal solution of OPT-2 can be obtained by restricting attention to set of prices  $L$  given in Proposition 3. Thus, an optimal solution to OPT-2 can be obtained by restricting attention to prices in  $L$  and solving the following optimization problem:

$$\begin{aligned} \max_{\mathbf{p} \in L^T, R \in \mathcal{P}(T)} \quad & \sum_{t=1}^T p_t D_t(p_t, R) \\ \text{s.t.} \quad & D_t(p_t, R) \leq c_t \\ & \text{for all } t \in \{1, \dots, T\}, \\ & R_t < R_{t'} \Rightarrow p_t \leq p_{t'} \\ & \text{for all } t, t' \in \{1, \dots, T\}. \end{aligned} \quad (\text{OPT-3})$$

We next show that it is possible to find an optimal solution of OPT-3 by recursively solving problems that are essentially smaller instances of itself.

Consider an optimal solution of OPT-3, denoted by  $(\mathbf{p}^*, R^*)$ . Suppose time period  $k$  has the lowest ranking; i.e.,  $R_k^* = 1$ . In this case, the attraction range of  $k$  is  $\{1, \dots, T\}$ , and  $\rho_k(R^*) = \sum_{i=1}^k \sum_{j=k}^T a_{ij}$ . Hence, all customers who are present in the system at time  $k$  will seek service at time  $k$ . This implies that only populations  $a_{k_1, k_2}$ ,  $1 \leq k_1 \leq k_2 < k$  can receive service at time periods  $\{1, \dots, k-1\}$  (similarly, only populations  $a_{k_1, k_2}$ ,  $k < k_1 \leq k_2 \leq T$  can receive service at time periods  $\{k+1, \dots, T\}$ ). Therefore, if the monopolist knows  $p_k^*$  and that  $R_k^* = 1$ , it can solve for optimal prices at other time periods by solving two separate subproblems for time periods  $\{1, \dots, k-1\}$  and  $\{k+1, \dots, T\}$ : maximize the revenue obtained from time periods  $\{1, \dots, k-1\}$  assuming only populations  $a_{k_1, k_2}$  are present (with  $1 \leq k_1 \leq k_2 < k$ ) and similarly for time periods  $\{k+1, \dots, T\}$ . Note that in the solution of the subproblems, we need to impose the condition that prices are weakly larger than  $p_k^*$  because otherwise  $p_l^* < p_k^*$  for some  $l$ , and we obtain a contradiction to  $R_k^* = 1$ .

The above observation suggests that given the time period  $k$  with the lowest ranking, the pricing problem can be decomposed into two smaller pricing problems, where the prices that can be offered are lower bounded by the price offered at  $k$ . We next exploit this observation and obtain a dynamic programming algorithm for the solution of OPT-3.

Let  $\omega(i, j, \underline{p})$  denote the maximum revenue obtained from an instance of OPT-3 assuming (i)  $a_{k_1, k_2} = 0$  unless

$i < k_1 \leq k_2 < j$  and (ii) restricting prices to be weakly larger than  $\underline{p}$ . That is,

$$\begin{aligned} \omega(i, j, \underline{p}) = \max_{\mathbf{p} \in L^T, R \in \mathcal{P}(T)} \quad & \sum_{t=i+1}^{j-1} p_t D_t^{ij}(p_t, R) \\ \text{s.t.} \quad & D_t^{ij}(p_t, R) \leq c_t \\ & \text{for all } t \in \{1, \dots, T\}, \\ & R_t < R_{t'} \Rightarrow p_t \leq p_{t'} \\ & \text{for all } t, t' \in \{1, \dots, T\}, \\ & p_t \geq \underline{p} \text{ for all } t \in \{1, \dots, T\} \end{aligned} \quad (6)$$

where  $D_t^{ij}$  is defined similarly to (3) and denotes the demand at time  $t$ , assuming  $a_{k_1, k_2} = 0$  unless  $i < k_1 \leq k_2 < j$ . Observe that the optimal objective value of OPT-3 is equal to  $\omega(0, T+1, 0)$ .

For  $i+1 > j-1$ , we assume  $\omega(i, j, p)$  is equal to 0. On the other hand, for any  $i, j$  such that  $i+1 \leq j-1$ , we have

$$\omega(i, j, p) = \max_{k \in \{i+1, \dots, j-1\}} \left\{ \max_{p \in L: p \geq \underline{p}} \{ \omega(i, k, p) + \gamma_k^{ij}(p) + \omega(k, j, p) \} \right\}, \quad (7)$$

where  $\gamma_k^{ij}(p)$  is given by

$$\gamma_k^{ij}(p) = \begin{cases} \left( \sum_{l=i+1}^k \sum_{m=k}^{j-1} a_{lm} \right) (1 - F(p)) p \\ \text{if } \left( \sum_{l=i+1}^k \sum_{m=k}^{j-1} a_{lm} \right) (1 - F(p)) \leq c_k, \\ -\infty \text{ otherwise.} \end{cases} \quad (8)$$

To see why the recursion in (7) holds, consider a solution of (6) and assume that in this solution,  $k$  is the time period in  $\{i+1, \dots, j-1\}$  with the lowest ranking and  $p_k \geq \underline{p}$  is the corresponding price. Then all populations that are present in the system at  $k$  receive service at this time period. The total mass of these populations is  $\sum_{l=i+1}^k \sum_{m=k}^{j-1} a_{lm}$ , since  $a_{k_1, k_2} = 0$  unless  $i < k_1 \leq k_2 < j$ , as can be seen from the definition of  $\omega(i, j, \underline{p})$ . Since at the optimal solution of (6), the capacity constraints are satisfied, the revenue obtained from time period  $k$  is given by  $\gamma_k^{ij}(p_k)$ . Since  $k$  has the lowest ranking among  $\{i+1, \dots, j-1\}$ , only populations  $a_{k_1, k_2}$  such that  $i < k_1 \leq k_2 < k$  can receive service before time  $k$ , and the prices offered at those time periods should be weakly larger than  $p_k$ . It follows from the definition of  $\omega$  that the maximum revenue that can be obtained from these populations (with prices weakly larger than  $p_k$ ) is given by  $\omega(i, k, p_k)$ . Similarly, it follows that the maximum revenue that can be obtained from time periods

after  $k$  equals to  $\omega(k, j, p_k)$ . Thus, we conclude that  $\omega(i, j, p) = \omega(i, k, p_k) + \gamma_k^{ij}(p_k) + \omega(k, j, p_k)$ . The recursion in (7) follows since it searches for time period  $k$  with the lowest ranking and the corresponding price  $p_k$  that maximizes the objective of (6). Note that since  $\gamma_k^{ij}(p) = -\infty$  when a capacity constraint is violated, the solution obtained by solving this recursion also satisfies the capacity constraints.

The following theorem shows that a solution of OPT-2, or equivalently a solution of the alternative formulation in OPT-3, can be obtained by solving for prices using the dynamic programming recursion in (7) and constructing a ranking vector consistent with these prices.

**THEOREM 1.** *The optimal solution of OPT-2 can be computed in time  $O(T^6)$ .*

This theorem, which is proved in the appendix, suggests that firms that provide service guarantees can effectively implement optimal pricing policies by solving OPT-2. In the subsequent sections, we show that this result extends to other settings where there is uncertainty about the problem parameters and the firm has more general objectives.

## 6. A Robust Optimization Formulation

The baseline model we considered in previous sections assumes that both demand and service capacity available over the planning horizon can be fully anticipated. This assumption is motivated by our online services application, where the planning horizon is typically counted in hours or, at most, days. Over the next two sections, we show how to solve the firm's problem when this assumption is not valid, by either taking a robust or a stochastic view of the demand or capacity uncertainty. In particular, in this section we introduce a robust optimization formulation of the firm's pricing problem. We show that when there is uncertainty about either the service capacity levels or the size of the customer population, a variant of the algorithm of §5 can be used to obtain a solution that maximizes revenue while maintaining feasibility for all possible values of uncertain parameters. Furthermore, we can bound the firm's worst-case revenue loss as a function of the uncertainty in the problem parameters. The proofs of this section can be found in the online appendix.

Suppose the firm does not know its service capacity level  $c_t$  at a given period but only knows that it belongs to an interval  $\mathcal{C}_t = [c_t^L, c_t^U]$ . Similarly, the firm does not know the mass of customers in population  $a_{i,j}$ , but instead it knows only that  $a_{i,j} \in \mathcal{A}_{i,j} = [a_{i,j}^L, a_{i,j}^U]$ . We refer to a collection of population sizes  $A = \{a_{i,j}\}_{1 \leq i, j \leq T}$  as a population matrix and represent

the set of all possible capacity levels by  $\mathcal{C} = \prod_t \mathcal{C}_t$  and the set of all possible population matrices by  $\mathcal{A} = \prod_{i,j} \mathcal{A}_{i,j}$ . To make dependence of the demand (defined in Equation (3)) on population size explicit, in this section we denote the demand at period  $t$ , when population matrix is given by  $A \in \mathcal{A}$ , and the firm uses price  $p_t$  and ranking  $R \in \mathcal{P}(T)$  by  $D_t(p_t, R, A)$ .

The problem of selecting prices and ranking that are feasible for all capacity levels in  $\mathcal{C}$  and population matrices in  $\mathcal{A}$ , and that maximize the worst-case revenue, can be formulated as follows:

$$\begin{aligned} & \max_{p \geq 0, R \in \mathcal{P}(T), M} M \\ \text{s.t. } & M \leq \sum_{t=1}^T p_t D_t(p_t, R, A) \\ & \text{for all } A \in \mathcal{A}, \\ & D_t(p_t, R, A) \leq c_t \\ & \text{for all } t, c_t \in \mathcal{C}_t \text{ and } A \in \mathcal{A}, \\ & R_t < R_{t'} \Rightarrow p_t \leq p_{t'} \\ & \text{for all } t, t' \in \{1, \dots, T\}, \end{aligned} \tag{OPT-4}$$

Our next result shows that this robust optimization problem is tractable.

**PROPOSITION 4.** *The optimal solution of OPT-4 can be computed in time  $O(T^6)$ .*

The optimization problem in OPT-4 is not a special case of OPT-2 because it uses the population matrix  $A^L$  when computing revenue and a different population matrix  $A^U$  when computing feasibility. However, with minor modifications, the structural insights and the polynomial-time algorithm from §5 still apply, leading to the result in Proposition 4.

The robust optimization formulation finds a conservative solution that is feasible for all possible values of uncertain parameters. We next quantify the potential revenue loss due to the uncertainty, when the solution obtained from this formulation is used for pricing. Assume that a solution of OPT-4 is obtained using uncertainty sets  $\mathcal{C}$  and  $\mathcal{A}$ . Let  $V^{ROB}(\mathcal{C}, \mathcal{A}, \mathbf{c}, A)$  denote the revenue the firm achieves, using this solution, when realized parameter values are  $\mathbf{c} \in \mathcal{C}$  and  $A \in \mathcal{A}$ . We denote the revenues that could be obtained if we knew with certainty  $\mathbf{c}$  and  $A$  by  $V(\mathbf{c}, A)$ . The following proposition bounds the decrease in the revenues when there is uncertainty, and the solution of OPT-4 is used to obtain a robust pricing rule.

**PROPOSITION 5.** *Suppose  $c_t^U \leq (1 + \theta)c_t^L$  for all  $t \in \{1, \dots, T\}$  and  $a_{i,j}^U \leq (1 + \theta)a_{i,j}^L$  for all  $i, j \in \{1, \dots, T\}$ . Then,*

$$\sup_{\mathbf{c} \in \mathcal{C}, A \in \mathcal{A}} (V(\mathbf{c}, A) - V^{ROB}(\mathcal{C}, \mathcal{A}, \mathbf{c}, A)) \leq 3\theta \sum_{i,j} a_{i,j}^U.$$

This proposition implies that when the uncertainty in the problem parameters is small (i.e., when  $\theta$  is small), the revenue loss is also small, provided that a solution of OPT-4 is used for pricing. The proposition also suggests that if a nominal version of the problem with parameters  $(\mathbf{c}^{\text{NOM}}, A^{\text{NOM}})$  is known, and the realized parameters are between  $1 - \varepsilon$  and  $1 + \varepsilon$  times the nominal ones (i.e.,  $1 + \theta = 1 + \varepsilon/(1 - \varepsilon)$ ), then the maximum revenue loss due to uncertainty is equal to  $((6\varepsilon(1 + \varepsilon))/(1 - \varepsilon)) \sum_{i,j} a_{i,j}^{\text{NOM}}$ .

## 7. A General Framework and an Approximation Scheme

In this section, we generalize our baseline model by incorporating random customer arrivals and capacity levels, production costs, and customer valuations that are dependent on arrival and departure periods to the model. Namely, the distribution of the size and valuations of each population is known in advance and the firm determines a sequence of (preannounced) prices to maximize its expected profit. In order for the firm to be able to satisfy its service guarantees, we consider soft capacity constraints in the sense that the firm can exceed the allowable capacity by paying a penalty or purchasing more capacity.

At this level of generality, the characterization of the set of prices in an optimal solution, presented in §4, does not hold. However, we can extend our algorithm to obtain a fully polynomial-time approximation scheme (FPTAS) for the general model. An FPTAS is an approximation algorithm that for any  $\varepsilon > 0$  obtains a solution within a factor of  $1 - \varepsilon$  of the optimal solution and is polynomial in the size of the problem and in  $1/\varepsilon$ . Therefore, using an FPTAS, one can obtain a solution arbitrarily close to the optimal.

Intuitively, this problem is still tractable since, similar to our baseline model, if a time instant has the lowest price in the horizon, then all customers who are present in the system at this time instant prefer receiving service there. Consequently, our main divide-and-conquer approach is applicable, and we can reformulate the problem as a dynamic programming problem following the approach in §5. In this section, we formally explain this idea. The proofs of our results are presented in the online appendix.

We start by introducing an abstract problem that is the focus of this section:

$$\begin{aligned} \max_{\mathbf{p} \in [0, 1]^T, R \in \mathcal{P}(T)} \quad & \sum_{t=1}^T g_t(p_t, R) \\ \text{s.t.} \quad & h_t(p_t, R) \leq 0 \\ & \text{for all } t \in \{1, \dots, T\}, \\ & R_t < R_{t'} \Rightarrow p_t \leq p_{t'} \\ & \text{for all } t, t' \in \{1, \dots, T\}. \end{aligned} \tag{OPT-5}$$

We make the following assumption throughout this section:

**ASSUMPTION 2.** For any ranking  $R$  and period  $t$ , the functions  $g_t(\cdot, R)$  and  $h_t(\cdot, R)$  satisfy the following:

1. Each customer prefers the time period with the lowest rank (among those during which she is present) to (potentially) receive service. Hence, the dependence of functions  $g_t: \mathbb{R} \times \mathcal{P}(T) \rightarrow \mathbb{R}$  and  $h_t: \mathbb{R} \times \mathcal{P}(T) \rightarrow \mathbb{R}$  on  $R$  is through the attraction range of time period  $t$ . That is, there exist functions  $\hat{g}_t, \hat{h}_t$  such that

$$\begin{aligned} g_t(p_t, R) &= \hat{g}_t(p_t, b_t(R), e_t(R)) \quad \text{and} \\ h_t(p_t, R) &= \hat{h}_t(p_t, b_t(R), e_t(R)) \end{aligned} \tag{9}$$

where  $\{b_t(R), \dots, e_t(R)\}$  is the attraction range of time period  $t$ , when ranking  $R$  is chosen.

2.  $h_t(p_t, R)$  is decreasing in  $p_t$ , and  $g_t(p_t, R)$  is Lipschitz continuous in  $p_t$  with parameter  $l_t$ .

The first part of the assumption implies that the demand for service at period  $t$  and, therefore, the profit earned at period  $t$ , depends only on the price  $p_t$  and the attraction range generated by the ranking of prices  $R$ . That is, if we modify price  $p_{t+1}$  while leaving the ranking of prices  $R$  intact, thus leaving the attraction range intact, this change in the price vector  $\mathbf{p}$  will have no effect on the demand at period  $t$ . This excludes customer discounting, for example, but is a generalization of the customer behavior model assumed in earlier sections. The second part of the assumption just implies that demand in a given period decreases continuously in that period's price if we maintain a constant ranking of prices. Observe that OPT-2 is a special case of OPT-5, when  $g_t(p_t, R) = p_t D_t(p_t, R)$  and  $h_t(p_t, R) = D_t(p_t, R) - c_t$ , assuming that demand  $D_t(p_t, R)$  is Lipschitz continuous in  $p_t$ , for a fixed ranking  $R$ .

For a constant  $\varepsilon \in (0, 1)$ , consider the set of prices  $\mathbf{P}_\varepsilon = \{k\varepsilon \mid k \in \mathbb{Z}_+, k\varepsilon \leq 1\}$  and assume that we seek a solution to OPT-5 by restricting attention to the prices that belong to this set, i.e.,

$$\begin{aligned} \max_{\mathbf{p} \in \mathbf{P}_\varepsilon^T, R \in \mathcal{P}(T)} \quad & \sum_{t=1}^T g_t(p_t, R) \\ \text{s.t.} \quad & h_t(p_t, R) \leq 0 \\ & \text{for all } t \in \{1, \dots, T\}, \\ & R_t < R_{t'} \Rightarrow p_t \leq p_{t'} \\ & \text{for all } t, t' \in \{1, \dots, T\}. \end{aligned} \tag{OPT-6}$$

Note that any feasible solution of OPT-6 is feasible in OPT-5. We next show that for small  $\varepsilon$  the optimal objective values of these problems are also close. Hence, an optimal solution of OPT-6 can be used to provide a near-optimal solution of OPT-5.

LEMMA 2. Let the optimal solutions of OPT-5 and OPT-6 have objective values  $v$  and  $v_\epsilon$ , respectively. Then,  $v_\epsilon \geq v - \epsilon \sum_{t=1}^T l_t$ .

We next show that a modified version of the algorithm in §5 can be used to solve OPT-6. Our approach is again based on obtaining a solution by recursively solving smaller instances of the problem. For this purpose, we first define  $\hat{\omega}(i, j, \underline{p})$  to be the maximum utility that can be obtained assuming only populations  $a_{k_1, k_2}$ ,  $i < k_1 \leq k_2 < j$ , are present and the prices that can be used at periods  $\{t \mid i < t < j\}$  are (weakly) larger than  $\underline{p}$ . It can be seen that the optimal value of OPT-6 is equal to  $\hat{\omega}(0, T + 1, 0)$ .

We set  $\hat{\omega}(i, j, \underline{p}) = 0$ , for  $i + 1 > j - 1$ . Using the same argument given in §5 to justify the recursion in (7), it follows that for  $i + 1 \leq j - 1$ , the following dynamic programming recursion holds:

$$\hat{\omega}(i, j, \underline{p}) = \max_{k \in \{i+1, \dots, j-1\}} \left\{ \max_{p \in \mathbf{P}_\epsilon: p \geq \underline{p}} \{ \hat{\omega}(i, k, p) + \hat{\gamma}_k^{ij}(p) + \hat{\omega}(k, j, p) \} \right\}, \quad (10)$$

where  $\hat{\gamma}_k^{ij}(p)$  denotes the utility obtained at time  $k$  with price  $p$ , from all populations  $a_{k_1, k_2}$  that can receive service at this period and that satisfy  $i < k_1 \leq k_2 < j$ . That is,  $\hat{\gamma}_k^{ij}(p) = \hat{g}_k(p, i, j)$ , if  $\hat{h}_k(p, i, j) \leq 0$  and  $\hat{\gamma}_k^{ij}(p) = -\infty$  otherwise.

The intuition behind (10) is similar to the intuition of (7): to find  $\hat{\omega}(i, j, \underline{p})$ , we search for the time period with the lowest rank (maximization over  $k$  in (10)), and we search for the best possible price for this time period (maximization over  $p$ ). Since all populations that are present at the time period with the lowest ranking (say  $k$ ) receive service at this time period, the payoff obtained from this time period can be given by  $\hat{\gamma}_k^{ij}(p)$ . We then solve for prices of subproblems for time periods  $\{i + 1, \dots, k - 1\}$  and  $\{k + 1, \dots, j - 1\}$ . Since the time period with the lowest ranking also has the lowest price, we impose the prices for these subproblems to be weakly larger than  $\underline{p}$ . Thus, the payoffs of the subproblems are given by  $\hat{\omega}(i, k, \underline{p})$  and  $\hat{\omega}(k, j, \underline{p})$ . Hence, we obtain the recursion in (10) for computing optimal prices in OPT-6.

In the following lemma, we use this dynamic program to construct optimal prices and ranking for the solution of OPT-6 and characterize the computational complexity of the solution. Note that since we are dealing with general functions  $g_t$  and  $h_t$ , our result depends on the computational complexity of evaluating these functions.

LEMMA 3. Assume that for any given  $t, p, R$ , computation of  $g_t(p, R)$  and  $h_t(p, R)$  takes  $O(s(T))$  time. An optimal solution of OPT-6 can be found in  $O((T^3 s(T))/\epsilon^2)$  time.

Lemmas 2 and 3 imply that an approximate solution to OPT-5 can be found in polynomial time provided that  $g_t(p, R)$  and  $h_t(p, R)$  can be evaluated in polynomial time.

THEOREM 2. Assume that for any given  $t, p, R$ , computation of  $g_t(p, R)$  and  $h_t(p, R)$  takes  $O(s(T))$  time. An  $\epsilon$ -optimal solution of OPT-5 can be found in  $O((T^3 s(T))/\epsilon^2)$  time.

The proof immediately follows from Lemmas 2 and 3 and is omitted. In many of the relevant cases (such as revenue maximization subject to capacity constraints as introduced in §§2 and 3), for given prices and rankings, evaluating constraints and the objective function ( $h_t$  and  $g_t$ ) can be completed in  $O(1)$  time. In such settings, Theorem 2 implies that an approximate solution can be obtained in  $O(T^3/\epsilon^2)$  time.

We conclude this section by showing that this general framework allows us to find approximately optimal prices in polynomial time for problem instances that involve random arrivals and capacity levels, production costs, and a richer class of customer valuations.

*Correlated Valuations.* Here, we relax the assumption made before that the customers' valuations are independent of their arrival and departure periods. Let  $F_{i,j}(v)$  represent the fraction of the  $a_{i,j}$  population that values service at most  $v$ . We assume that all valuations are in  $[0, 1]$  and that  $F_{i,j}$  is differentiable, but we no longer suppose that Assumption 1 holds; i.e., the corresponding uncapacitated revenue function need not be single peaked. Customer demand at period  $t$  as a function of price  $p_t$  and the ranking of prices  $R$  is now given by

$$D_t(p_t, R) = \sum_{i \leq j} a_{i,j} (1 - F_{i,j}(p_t)) \mathbf{1}\{R_t \leq R_k \text{ for all } i \leq k \leq j\}.$$

By choosing  $g_t(p_t, R) = p_t D_t(p_t, R)$  and  $h_t(p_t, R) = D_t(p_t, R) - c_t$ , the corresponding revenue maximization problem is an instance of OPT-5. Note that for a fixed  $R$ , denoting  $f_{i,j}(p) = dF_{i,j}(p)/dp$  and assuming  $f_{i,j}$  is bounded by  $l_{i,j}$ , we conclude

$$\begin{aligned} \left| \frac{\partial D_t(p_t, R)}{\partial p_t} \right| &= \sum_{i \leq j} a_{i,j} f_{i,j}(p_t) \mathbf{1}\{R_t \leq R_k \text{ for all } i \leq k \leq j\} \\ &\leq \sum_{i \leq j} a_{i,j} l_{i,j}. \end{aligned}$$

Thus, it follows that when  $f_{i,j}$  is bounded for all  $i, j$ ,  $D_t(\cdot, R)$  is Lipschitz continuous. This implies that  $g_t(\cdot, R)$  is also Lipschitz continuous for all  $t$  and  $R$ . Moreover,  $h_t$  is decreasing in  $p_t$  (since demand is decreasing in  $p_t$ ). Furthermore, for any  $t, p, R$  evaluating  $D_t(p_t, R)$  and in turn  $g_t(p, R)$  and  $h_t(p, R)$  takes  $O(T^2)$  time. Thus, Theorem 2 applies, and we conclude that the approximate revenue maximization problem can be solved in  $O(T^5/\epsilon^2)$ .



*Production Costs and Soft Capacity Constraints.* We also incorporate production costs into the model. We assume that it costs the firm  $\mu_t(d)$  to provide service to mass  $d$  of customers at period  $t$ . We make the following regularity assumption on the production costs:

ASSUMPTION 3. Assume that for any period  $t$ , the production cost  $\mu_t$  is a nonnegative, nondecreasing,  $\lambda$ -Lipschitz continuous function.

Along with capturing the cost of producing the service to be delivered, the function  $\mu_t$  can also capture a soft capacity constraint: if  $\bar{c}_t$  represents a capacity level above which any unit produced costs  $\bar{\mu}$ , then we can capture this by setting  $\mu_t(d) = \max\{0, \bar{\mu}(d - \bar{c}_t)\}$ . Even with soft capacity constraints, we still assume that the firm provides service guarantees. Whenever the firm is incapable of providing the purchased service itself, it contracts service delivery out to a third party with a unit cost of  $\bar{\mu}$ .

By letting  $g_t(p_t, R) = p_t D_t(p_t, R) - \mu_t(D_t(p_t, R))$  and  $h_t(p_t, R) = D_t(p_t, R) - c_t$ , we obtain an instance of OPT-5. From Assumption 3, it follows that  $g_t(p_t, R)$  is Lipschitz continuous in  $p_t$ . Since demand is decreasing with price, we observe that  $h_t(p_t, R)$  decreases with price. Thus, Theorem 2 applies, and since  $D_t(p, R)$  (and hence  $g_t(p_t, R)$ ,  $h_t(p_t, R)$ ) can be evaluated in polynomial time for any given  $p$  and  $R$ , it follows that an approximate solution of the problems with production costs and soft capacity constraints can be obtained in polynomial time.

*Stochastic Arrival and Capacity Processes.* Assume that population sizes  $\{a_{i,j}\}_{i,j}$  and capacities  $\{c_t\}_t$  are random variables with known distributions. Let  $E[a_{i,j}] = \hat{a}_{i,j}$  and  $E[c_t] = \hat{c}_t$ . In this setting, if a monopolist wants to guarantee that the total service request does not exceed the capacity for any realization of the parameters, it can use the robust optimization framework in §6. On the other hand, if the firm has the capability to contract service delivery out whenever the capacity is exceeded (hence it has soft capacity constraints), it can solve the following expected revenue maximization problem:

$$\begin{aligned} \max_{p \in \mathbf{P}_e, R \in \mathcal{P}(T)} \quad & \sum_{t=1}^T E[p_t D_t(p_t, R) - \mu_t(D_t(p_t, R))] \\ \text{s.t.} \quad & R_t < R_{t'} \Rightarrow p_t \leq p_{t'} \\ & \text{for all } t, t' \in \{1, \dots, T\}. \end{aligned} \tag{OPT-7}$$

By choosing

$$\begin{aligned} h_t(p_t, R) &= 0 \text{ and} \\ g_t(p_t, R) &= E[p_t D_t(p_t, R) - \mu_t(D_t(p_t, R))], \end{aligned}$$

we obtain an instance of OPT-5. Note that if  $\mu_t$  is Lipschitz continuous, then so is  $g_t(p_t, R) = E[p_t D_t(p_t, R) - \mu_t(D_t(p_t, R))]$ . Thus, under Assumption 3, provided

that the expectation in  $E[p_t D_t(p_t, R) - \mu_t(D_t(p_t, R))]$  can be evaluated in polynomial time, OPT-7 can be solved using the dynamic programming recursion in (10) in polynomial time.

## 8. Multiperiod Pricing with Customer Scheduling

In the earlier sections, we studied the pricing problem of a firm in a setting where the customers choose the earliest time instant with the lowest price to receive service. Consider an example where all customers arrive at the initial period and can wait until the end of the horizon to receive service, but service capacity is spread over many periods. For any pricing rule, in this example all customers receive service at the same time instant (with the lowest price). However, this results in inefficient use of capacity and reduced revenues.

Motivated by this example, in this section, we consider the pricing problem in a setting where the firm can choose how customers should break ties between time instants with equal prices. That is, the customers still receive service at a time instant with the lowest price, but if there are multiple such time instants, the firm schedules how customers should be served. Observe that in the example described above, such a scheduling of customers would avoid the inefficiency created by serving all customers at the same period (and wasting the capacity in the remaining periods).

Note that it may not always be in the power of the firm to schedule its customers as described above since at the least this requires knowledge of the arrival and departure times (or deadlines) of each individual customer, which is not always available. However, in this section we establish that if the firm has the necessary means to schedule its customers, then it can decide on the optimal prices and schedule by following a dynamic programming approach similar to the one discussed in §5.

In this setting, the firm's optimization problem can be formulated as follows:

$$\begin{aligned} \max_{p \in \mathbf{P}^T, x} \quad & \sum_{t=1}^T p_t \left( \sum_{i,j: i \leq t \leq j} x_{i,j}^t \right) \\ \text{s.t.} \quad & \sum_{i,j: i \leq t \leq j} x_{i,j}^t \leq c_t \\ & \text{for all } t \in \{1, \dots, T\}, \\ & x_{i,j}^t p_t = x_{i,j}^t \min_{k: i \leq k \leq j} \{p_k\} \\ & \text{for all } t \in \{1, \dots, T\}, \\ & \sum_{t: i \leq t \leq j, F(p_t) < 1} \frac{1}{(1 - F(p_t))} x_{i,j}^t = a_{ij} \\ & \text{for all } i, j \in \{1, \dots, T\}. \end{aligned} \tag{OPT-8}$$

Here,  $x_{i,j}^t$  corresponds to the mass of customers that belong to population  $a_{i,j}$  and receive service at time  $t$ , and  $\mathbf{P}$  is the set of prices that can be used by the firm. For simplicity, in this section we assume that  $\mathbf{P}$  is a finite set. The first constraint suggests that the capacity constraint is satisfied at all time instants, and the second one ensures that if a fraction of population  $a_{i,j}$  is scheduled to receive service at time  $t$  (i.e.,  $x_{i,j}^t > 0$ ), then the price at time  $t$  should be equal to the minimum price offered from time  $i$  up to (and including) time  $j$ . The final constraint guarantees that all customers that belong to some population  $a_{i,j}$  and have valuation larger than the lowest price between  $i$  and  $j$  will receive service.

It is possible to interpret our paper as a two-stage game where the firm moves first and selects a sequence of prices and the consumers respond in the second stage. This two-stage game might have multiple equilibria. One equilibrium is the one we call the baseline model—consumers break ties in favor of buying service early. Another potential equilibrium is the one where consumers break ties in favor of buying at the most favorable period for the firm. The second equilibrium is the one studied in this section.

Observe that our baseline model can be viewed as a version of OPT-8 where the firm is restricted to breaking ties by assigning the customers to the earliest time instant with the lowest price. Hence, if the firm has the ability to break the ties favorably, it may obtain a higher revenue. We next show that in this case the optimal pricing policy (i.e., the solution of OPT-8) can be obtained by using a dynamic programming approach that generalizes the one used for the solution of our baseline model. In particular, given a set of prices that can be used by the seller  $\mathbf{P}$ , this algorithm finds the optimal solution of OPT-8 in time polynomial in  $T$  and size of  $\mathbf{P}$ .

To compute the optimal revenue in the OPT-8, similar to §5, define  $\hat{\omega}(i, j, \underline{p})$  as the maximum revenue that can be obtained from all populations arriving between periods  $i$  and  $j$ , using prices weakly greater than  $\underline{p}$ , with the boundary condition that all customers who can wait until or after time  $j$  receive service at a later time instant. In our original model, we recursively calculated  $\omega(i, j, \underline{p})$  by finding a period  $k$  where it gets the lowest price and highest ranking (see §5). In the new model, since a population can be split into different periods, instead of finding a single period that takes the lowest price, we find a set of such periods. Before presenting the algorithm, we need a couple of definitions.

Let  $\mathcal{S}(i, j)$  be the set of all feasible price vectors for subproblem  $(i, j)$ . Namely,  $\mathcal{S}(i, j)$  is the set of all feasible solutions of problem OPT-8 where the mass of all populations except those arriving between time  $i$  and  $j$  is assumed to be zero. Define  $\mathcal{S}(i, j, \underline{p}, k)$  to be

the set of price vectors in  $\mathcal{S}(i, j)$  such that all prices are (weakly) greater than  $\underline{p}$  and  $k$  is the latest period that has price  $\underline{p}$  (i.e., all periods between time  $k$  and  $j$  have a price (strictly) larger than  $\underline{p}$ ). If no such feasible set exists, then let  $\mathcal{S}(i, j, \underline{p}, k) = \emptyset$ . Additionally, we define  $\hat{L}(i, j, \underline{p}, k) = \{t \mid \exists S \in \mathcal{S}(i, j, \underline{p}, k) \text{ and } p_t(S) = \underline{p}\}$ , where  $p_t(S)$  denotes the price at time  $t$  in price vector  $S$ . This definition suggests that  $\hat{L}(i, j, \underline{p}, k)$  is the set of all periods that take price  $\underline{p}$  for some price vector in  $\mathcal{S}(i, j, \underline{p}, k)$ . Using these definitions, we first provide a characterization of the optimal pricing policy when we restrict attention to price vectors in  $\mathcal{S}(i, j, \underline{p}, k)$ .

**LEMMA 4.** *Suppose  $\mathcal{S}(i, j, \underline{p}, k)$  is nonempty and  $\underline{p}$  is (weakly) larger than the monopoly price  $p_M$ . Then there exists a price vector  $S^* \in \mathcal{S}(i, j, \underline{p}, k)$  such that  $p_t(S^*) = \underline{p}$  for all  $t \in \hat{L}(i, j, \underline{p}, k)$  and  $S^*$  maximizes the revenue (objective of OPT-8) among all the price vectors  $\mathcal{S}(i, j, \underline{p}, k)$ .*

Now, suppose  $\mathcal{S}(i, j, \underline{p}, k)$  is nonempty, and  $S^* \in \mathcal{S}(i, j, \underline{p}, k)$ . Observe that if we remove periods in  $\hat{L}(i, j, \underline{p}, k)$  from the set of periods between  $i$  and  $j$ , we will end up with some intervals (consecutive time periods belong to the same interval and we may have intervals with only one period). We denote the  $\ell$ th such interval by  $[I_{i,j,p,k,0}^\ell, I_{i,j,p,k,1}^\ell]$  and note that there are fewer than  $(j-i)$  intervals. We also observe that in  $S^*$ , only populations that arrive after time  $(I_{i,j,p,k,0}^\ell - 1)$  and leave before period  $(I_{i,j,p,k,1}^\ell + 1)$  receive service at the periods in  $[I_{i,j,p,k,0}^\ell, I_{i,j,p,k,1}^\ell]$  because the price at periods  $I_{i,j,p,k,0}^\ell - 1$  and  $I_{i,j,p,k,1}^\ell + 1$  is equal to  $\underline{p}$ . Now we can state the recursion for computing  $\hat{\omega}(i, j, \underline{p})$ :

$$\hat{\omega}(i, j, \underline{p}) = \max_{k: i < k < j, p \geq \underline{p}} \max \left\{ \hat{\gamma}_k^{ij}(p) + \sum_{\ell} \hat{\omega}(I_{i,j,p,k,0}^\ell - 1, I_{i,j,p,k,1}^\ell + 1, p) \right\}, \quad (11)$$

where, similar to §5,  $\hat{\gamma}$  is given by

$$\hat{\gamma}_k^{ij}(p) = \begin{cases} \left( \sum_{l, m: \exists t \in \hat{L}(i, j, p, k), l \leq t \leq m} a_{lm} \right) (1 - F(p)) p & \text{if } \hat{L}(i, j, p, k) \neq \emptyset, \\ -\infty & \text{if } \hat{L}(i, j, p, k) = \emptyset. \end{cases} \quad (12)$$

Namely,  $\hat{\gamma}_k^{ij}(p)$  denotes revenue obtained from the periods in  $\hat{L}(i, j, p, k)$ , by setting their prices equal to  $p \geq \underline{p}$ , assuming only the populations between periods  $i$  and  $j$  are present in the system. Following the same argument in §5, and using Lemma 4, it is easy to show that the recursion calculates the optimal value for  $\hat{\omega}(i, j, \underline{p})$ . Note that (12) suggests that if  $\hat{L}(i, j, p, k)$

can be found in polynomial time, then  $\hat{\gamma}$  can be computed efficiently. Our next result, which is proved in the online appendix, shows that this is the case.

LEMMA 5. Set  $\hat{L}(i, j, p, k)$  can be found in polynomial time.

Observe that we have  $O(T^2|\mathbf{P}|)$  subproblems  $\hat{\omega}(i, j, p)$ . Therefore, by the lemma given above and the recursion in (11), we immediately obtain the following theorem.

THEOREM 3. The optimal solution of OPT-8 can be found in polynomial time.

## 9. Numerical Insights

In this section, we consider generic instances of the firm's pricing problem and obtain qualitative insights about the optimal pricing scheme introduced in §2. We first investigate the effect of available capacity on the optimal prices used by the firm and establish that the prices closely track the capacities; e.g., decreasing capacities induce increasing prices and vice versa. Then, we focus on how patience level of players affect the outcome and show that as customers become more patient, the firm offers higher prices that leads to underutilization of capacity and lowered revenues and customer welfare. In addition, we observe that when customers are patient, the firm ends up using only a few different prices, thereby considerably decreasing the complexity of the pricing policy. Finally, we compare the pricing schemes we introduce in this paper with static pricing that is commonly employed in practice and establish that it is possible to significantly improve the revenues using our algorithms. We conclude by testing the run time of our algorithms and showing that for realistic scenarios optimal prices can be computed in a few minutes.

Unless noted otherwise, we will always consider problem instances with 36 time periods, where we focus on the middle 24 periods to avoid potential boundary effects.<sup>9</sup> We let customer valuations be uniformly distributed between 0 and 1. We assume that there are two types of populations arriving at each time period: (i) impatient (or myopic) customers (who are only interested in purchasing service at the period they arrived), and (ii) strategic (or  $s$ -patient) customers, who are willing to wait up to  $s$  periods to purchase service. This is captured by setting all  $a_{i,j}$  equal to 0 unless  $j = i$  (myopic customers) or  $j = i + s$  (strategic ones). For each  $i$ ,  $a_{i,i}$  is generated at random from a uniform distribution between 0 and  $m_1$ ,

whereas  $a_{i,i+s}$  is generated from a uniform distribution between 0 and  $m_2$ , where  $m_1$  and  $m_2$  are simulation parameters.

### 9.1. Effects of Capacity Constraints

We first consider different capacity regimes and try to understand their impact on pricing rules. We consider capacity vectors that satisfy one of the following cases: (case 1)  $c_t = 1$  for all  $t \in \{1, \dots, T\}$ , i.e., constant capacity; (case 2)  $c_t = 1.25 - 0.5(t-1)/(T-1)$  for all  $t \in \{1, \dots, T\}$ , i.e., capacity decreasing from 1.25 to 0.75 over the horizon; (case 3)  $c_t = 0.75 + 0.5(t-1)/(T-1)$  for all  $t \in \{1, \dots, T\}$ , i.e., capacity increasing from 0.75 to 1.25 over the horizon; (case 4)  $c_t = 1.25 - (t-1)/(T-1)$  for all  $t \leq T/2$  and  $c_t = 0.75 + (t-1 - T/2)/(T-1)$  for all  $t > T/2$ , i.e., capacity first decreasing from 1.25 to 0.75 and then increasing back to the original level, with a midday minimum. The first three cases capture the constant, decreasing, and increasing capacity settings, respectively. The last one captures a phenomenon that is typical in cloud computing markets: during peak business hours, part of the service capacity is usually unavailable because of high demand on the servers due to other contracts and obligations.

We next plot the average price vector over 100 randomly generated problem instances (Figure 2). We consider three settings in which (i) the entire population is impatient ( $[m_1, m_2] = [6, 0]$ ), (ii) half of the population is patient ( $[m_1, m_2] = [3, 3]$ ), (iii) the entire population is patient ( $[m_1, m_2] = [0, 6]$ ).

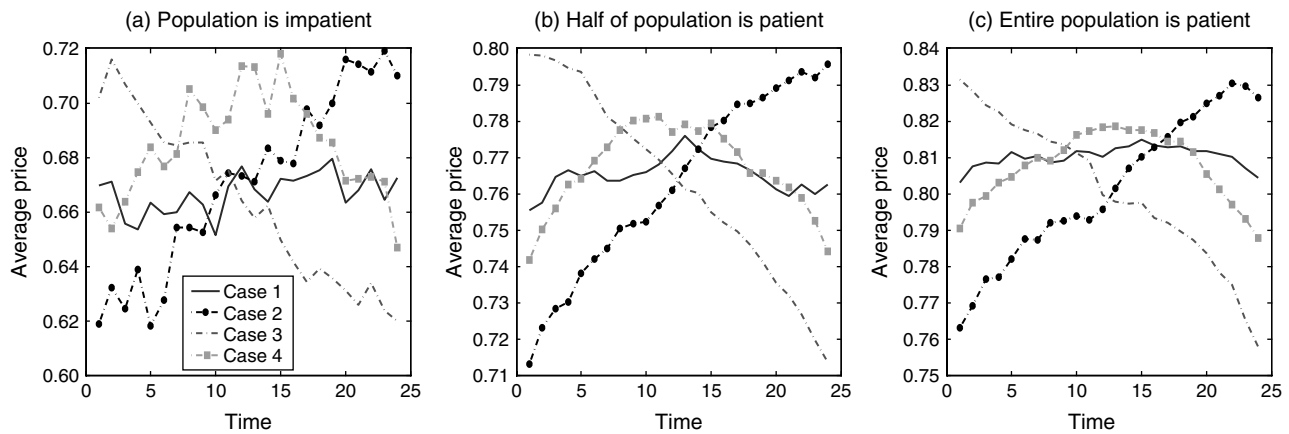
Figure 2 shows that in all four cases (and for all population structures), prices track service capacities. Prices are lower when service capacities are higher and vice versa (note that the ranges of the vertical axes in Figure 2 are not the same). Case 2 is to some extent analogous to a typical revenue management setting. As capacity dwindles toward the end of the horizon, prices rise accordingly. An interesting phenomenon occurs as we move from the graph with impatient customers, on the left, toward the one with patient customers, on the right: Prices become both smoother and higher as customers become more patient. We explore this observation more in the next subsection.

### 9.2. Effects of Strategic Behavior

We now investigate how strategic behavior of customers affects revenues, capacity usage, and customer welfare as the parameters  $s$  (willingness to wait for strategic customers) and  $m_2/(m_1 + m_2)$  (fraction of customers who are strategic) change. Our results indicate that as customers become more patient (and strategically time their purchases), the monopolist uses fewer different prices that are on average higher. This leads to inefficient use of the available capacity and

<sup>9</sup> We note that no significant changes are observed in our results in any of the subsections of this section, when the entire time horizon is used for the analysis.

Figure 2 Average Price over 24 Time Periods of Interest



reduces both the revenue of the firm and the welfare of customers.

We assume that willingness to wait for strategic customers  $s$  belongs to set  $\{0, 1, \dots, 8\}$ . The capacities are generated independently and uniformly between 0.5 and 1.5 for each time period. Additionally, the sizes of impatient and patient populations are characterized by the following cases: (case 1)  $[m_1, m_2] = [5, 1]$ ; (case 2)  $[m_1, m_2] = [3, 3]$ ; (case 3)  $[m_1, m_2] = [1, 5]$ ; (case 4)  $[m_1, m_2] = [0, 6]$ . That is, case 1 captures the scenario, where most of the population is impatient, whereas, case 4 captures the one where all buyers are patient. Since parameters are generated randomly, we present our results by averaging them over 100 problem instances.

We first consider the average price (over the horizon) offered by the monopolist for different cases and  $s$  parameters. We also plot average number of different prices used by the monopolist for the optimal solution of the pricing problems. Proposition 2 stated

that the optimal price in a given time period must either belong to  $\{p_M, 1\}$  or be equal to the price of a time period where the capacity is tight. This implies that the total number of price levels used might be smaller than the total number of periods. Our simulation in Figure 3(a) shows that this is indeed the case. In particular, it shows that the average number of price levels over the 24-period horizon drops both when a higher fraction of the population is willing to wait for service and when the customers who are willing to wait become more patient. For example, in case 2, while roughly 14 prices are needed when customers are willing to wait only up to one period, this number drops to 8 if they are willing to wait for two periods and 5 if they are willing to wait for three periods. Moreover, this drop in the number of prices is more significant if a larger proportion of the population is patient. We note that when an optimal solution for the original pricing problem OPT-1 does not exist, Lemma 1 suggests using perturbed prices  $\{p_t + \epsilon R_t\}_t$

Figure 3 Average Number of Different Price Levels (a) and Average Prices (b)

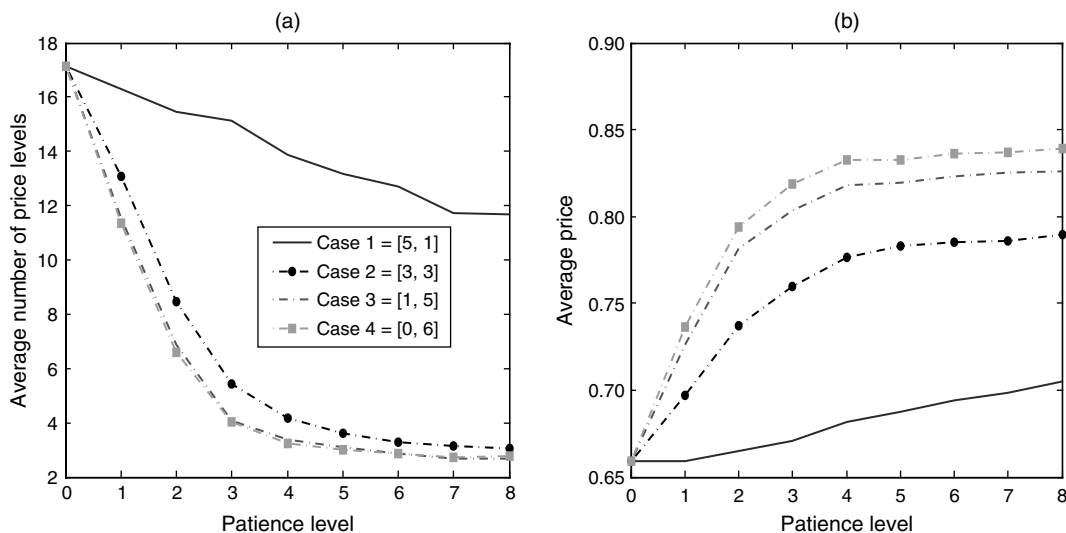
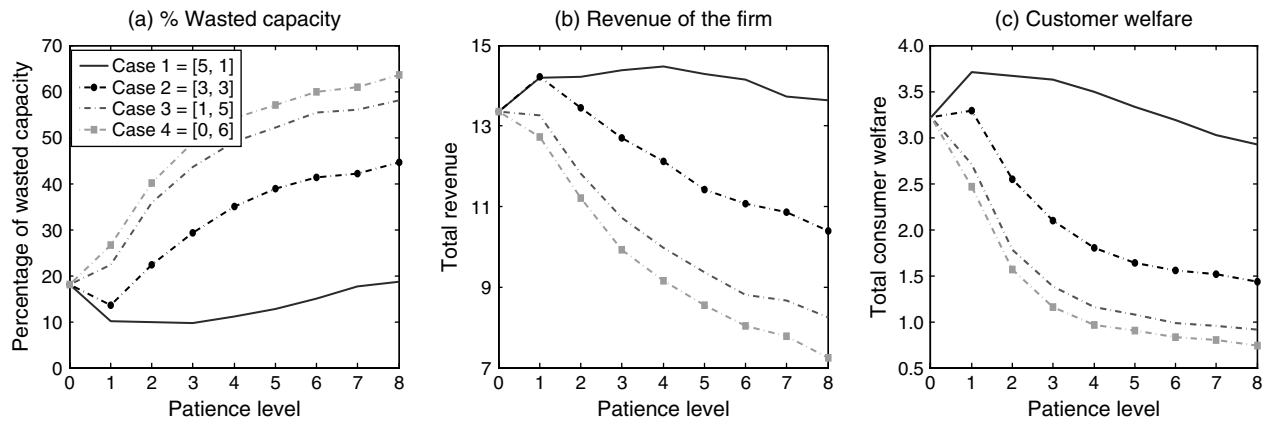




Figure 4 Wasted Capacity, Revenue, and Customer Welfare over 24 Time Periods



for an arbitrarily small  $\epsilon$  to obtain solutions arbitrarily close to the optimal. In such cases, our results indicate that the firm uses “essentially” few prices. Our first conclusion in this set of simulations is that patient customers lead to fewer price levels.

As customers become more patient, the firm becomes more constrained in the prices it can offer. Consequently, to maintain feasibility with fewer prices and sustain its service guarantees, it may need to increase the prices at some periods. Recall that the prices were already at or above monopoly price to begin with, so both the firm and the customers lose as the prices go up. Even a small increase in customer patience causes a fairly large increase in average prices (see Figure 3(b)) and, as expected, the effect is more pronounced when a larger fraction of the population is strategic. Thus, we also conclude that patient customers lead to higher prices.

We next focus on the effect of these higher prices on capacity usage, revenue, and customer welfare. At first glance, the presence of customers that are more patient would seem to lead to better use of resources. After all, high demand and low supply in one period followed by low demand and high supply in the next period could be properly matched if customers are willing to wait. Indeed this phenomenon does show up in our numerical analysis to a small extent, when customers switch from being completely impatient to willing to wait for one period (see cases 1 and 2 in Figure 4(a)). However, we mainly observe the opposite effect. As customers become more patient (for  $s \geq 1$ ), the firm is forced to use fewer and higher prices. These prices lead to inefficient use of the firm’s resources. The inefficiency is higher when a larger fraction of customers is impatient (Figure 4(a)). This phenomenon lowers the firm’s revenue and simultaneously reduces customer welfare (i.e., total surplus of the customers who purchase the service, where surplus is defined as the difference between the value a

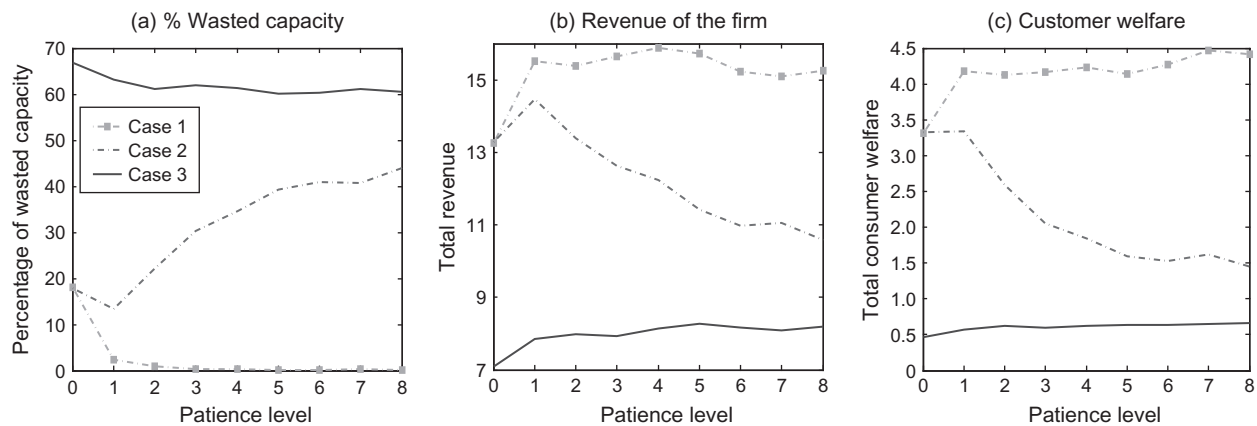
customer has for the service and her payment). In Figures 4(b) and 4(c), respectively, we plot the average revenue and customer welfare for different cases and  $s$  parameters. We establish that as customers become more patient, both revenue loss and customer welfare reduction become quite significant. For instance, it can be seen that for the case  $m_1 = 1, m_2 = 5$ , the revenue and welfare for  $s = 8$  are, respectively, 35% and 75% lower compared to a scenario with  $s = 0$ . This phenomenon analogous to Braess’s paradox that arises in transportation problems (where opening a new road may lead to higher overall congestion in the transportation network). However, the mechanism at work here is different from the one at Braess’s paradox since in our setting the lower welfare is a consequence of the firm’s price adjustment (raising prices to maintain feasibility of solution).

### 9.3. Static Pricing and Multiperiod Pricing

In this subsection, we compare the performance of our multiperiod pricing algorithms with the static pricing that is commonly used in practice. The three scenarios considered here are the following: (case 1) monopolist can use multiple prices and selects the service time of the customers, as in §8; (case 2) monopolist can use multiple prices, but customers receive service in their preferred (earliest) period, as in our base model; and (case 3) monopolist is restricted to using a single price and customers receive service in their preferred period, which is the case closest to current practice in the cloud computing industry. We assume that half the population is impatient and half is patient with willingness-to-wait  $s \in \{0, \dots, 8\}$ .

Figure 5 shows our result. The cases where the firm is most flexible (case 1) and least flexible (case 3) in terms of its pricing (and scheduling) policies, respectively, lead to the best and worst cases in terms of capacity management and firm profits. The efficiency gains from better pricing strategies are large, and thus the customers are better off when the firm has the

Figure 5 Wasted Capacity, Revenue, and Customer Welfare over 24 Time Periods of Interest



most tools in its arsenal. Interestingly, our base algorithm that uses only pricing to direct customers to periods performs well for low values of patience  $s$ , but its performance degrades as  $s$  increases. If the firm has some mechanism aside from pricing for scheduling customers to periods and customers are fairly patient, then the pricing scheme with scheduling should be used because the performance of this alternative algorithm improves with customer patience.

#### 9.4. Running Time

We conclude this section by discussing the running time of our base algorithm. In this section, we consider problem instances where the horizon length is given by  $T \in \{24, 48, 96\}$ . We still assume that the capacities are drawn from  $[0.5, 1.5]$  uniformly at random. We first focus on problems where there are two populations impatient and  $s$ -patient, and  $s$ -patient players can wait for  $s \in \{0, 1, \dots, 8\}$  time instants to receive service. Moreover, the size of impatient and patient populations are drawn from  $[0, m_1]$  and  $[0, m_2]$ , respectively, uniformly at random. We then consider a *heterogeneous* setting, where all population groups (characterized by an arrival and departure period) are present and their sizes are drawn at random. For each of these randomly generated problem instances, we run our simulations 100 times and report the average running time in *seconds* in Table 1.

Running times increase as either  $s$  or  $T$  grows, but our results do indicate that in instances of reasonable size (up to  $T = 96$ ), the optimal prices can be found in a few minutes using a standard laptop

(with 3.06 GHz Intel Core 2 Duo processor and 4 GB 1,067 MHz DDR3 memory). Therefore, the algorithm is sufficiently efficient to be implemented in practice.

## 10. Conclusions

We study a service firm’s multiperiod pricing problem in the presence of time-varying capacities and heterogeneous customers that are strategic with respect to their purchasing decisions. A distinct feature of our model is the service guarantees provided by the firm that ensure that any customer willing to pay the announced service price will be able to receive service. Such guarantees are quite appealing to customers because they allow them to ignore rationing risk, tremendously simplifying the consumers’ decision-making process. However, providing such guarantees requires the firm to use prices that ensure the firm has sufficient service capacity in every period. We propose an efficient algorithm to compute revenue-maximizing prices while maintaining service guarantees. We show, via numerical simulation, that in a typical instance the optimal pricing policy involves only a few prices and it enables the firm to obtain significantly more revenues than the static pricing schemes that are common in practice. We show that such algorithms and insights generalize to complex versions of the problem with random arrivals, departures and capacity levels, production costs, and customer valuations that depend on arrival and departure periods. We also construct an algorithm that the firm can use if it is capable of scheduling customer service times in addition to

Table 1 Average Running Times for Random Problem Instances (in Seconds)

	$s=0$	$s=1$	$s=2$	$s=3$	$s=4$	$s=5$	$s=6$	$s=7$	$s=8$	Heterogeneous
$T=24$	0.99	3.55	5.88	8.48	10.24	13.23	14.23	18.77	20.10	32.74
$T=48$	14.9	51.4	85.0	105.6	141.8	167.3	168.4	165.0	190.7	233.6
$T=96$	193	655	868	1,102	1,232	1,304	1,501	1,489	1,550	1,802

using dynamic pricing and demonstrate numerically that such an algorithm could yield even higher revenues and better resource utilization for the firm.

### Acknowledgments

The authors thank Ishai Menache (Microsoft Research), Georgia Perakis (MIT), Rakesh Vohra (University of Pennsylvania), and the anonymous associate editor and referees for their fruitful comments and suggestions.

### Appendix

**PROOF OF PROPOSITION 2.** Note that if for a ranking  $R$ , we have  $\rho_t(R) = 0$ , then, without loss of generality, we can let  $p_t = 1$ . Now, by Proposition 1, we can assume that  $(\mathbf{p}, R)$  is an optimal solution of OPT-2 such that  $p_M \leq p_t \leq 1$  and  $p_t = 1$  if  $\rho_t(R)(1 - F(p_t)) = 0$  for all  $t \in \{1, \dots, T\}$ . We will prove that  $(\mathbf{p}, R)$  is such that  $p_t$  necessarily satisfies one of the conditions 1–3 of the proposition for all  $t \in \{1, \dots, T\}$ . By contradiction, assume that for  $(\mathbf{p}, R)$  none of the conditions 1–3 hold at time  $t_0$ . Since conditions 1–2 do not hold  $1 > p_{t_0} > p_M$ . Note that  $p_{t_0} \neq 1$  implies that  $\rho_{t_0}(R)(1 - F(p_{t_0})) \neq 0$ , and hence  $\rho_{t_0}(R) > 0$ . Since condition 3 does not hold, for  $\tilde{t} \in S \triangleq \{t \in \{1, \dots, T\} \mid p_t = p_{t_0}\}$  we have that  $c_{\tilde{t}}$  is not tight; i.e.,

$$\rho_{\tilde{t}}(R)(1 - F(p_{\tilde{t}})) < c_{\tilde{t}} \quad \text{for all } \tilde{t} \in S. \quad (13)$$

Let  $\delta$  be a constant such that

$$\delta = \begin{cases} p_{t_0} - p_M & \text{if } p_{t_0} \leq p_k \text{ for all } k \in \{1, \dots, T\}, \\ p_{t_0} - \max_{\{t_1 \mid p_{t_1} < p_{t_0}\}} p_{t_1} & \text{otherwise.} \end{cases}$$

Consider the price vector  $\hat{\mathbf{p}}$ , for which  $p_k = \hat{p}_k$  for  $k \notin S$ , and  $\hat{p}_k = p_k - \epsilon$  otherwise, for some  $0 < \epsilon < \delta$ . It follows from the definition of  $\delta$  that if  $p_i \leq p_j$  for some  $i, j \in \{1, \dots, T\}$ , then  $\hat{p}_i \leq \hat{p}_j$ . Hence, the price vector  $\hat{\mathbf{p}}$  is also consistent with ranking  $R$ . Moreover, since  $(1 - F(p))$  is a continuous function, by (13) we conclude that  $\epsilon$  can be chosen small enough to guarantee that for time periods  $t \in S$ ,  $\rho_t(R)(1 - F(\hat{p}_t)) < c_t$ . Since  $(\mathbf{p}, R)$  is feasible and  $p_t = \hat{p}_t$  for  $t \notin S$ , it also follows that for  $t \notin S$ , we have  $\rho_t(R)(1 - F(\hat{p}_t)) = \rho_t(R)(1 - F(p_t)) \leq c_t$ . Consequently,  $(\hat{\mathbf{p}}, R)$  is feasible in OPT-2. The definition of  $\delta$  also suggests that  $p_M < \hat{p}_t < p_t = p_{t_0}$  for  $t \in S$ . It follows by the definition of  $p_M$  and the unimodality of the uncapacitated revenue function that  $p_t(1 - F(p_t)) < \hat{p}_t(1 - F(\hat{p}_t))$  for  $t \in S$ . Thus, since  $\rho_{t_0}(R) \neq 0$  we conclude that the revenue obtained from time periods  $t \in S$  increases under  $\hat{\mathbf{p}}$ ; i.e.,

$$\sum_{t \in S} \hat{p}_t(1 - F(\hat{p}_t))\rho_t(R) > \sum_{t \in S} p_t(1 - F(p_t))\rho_t(R). \quad (14)$$

Since  $p_t = \hat{p}_t$  for  $t \notin S$ , it also follows that  $\sum_{t \notin S} \hat{p}_t(1 - F(\hat{p}_t))\rho_t(R) = \sum_{t \notin S} p_t(1 - F(p_t))\rho_t(R)$ . Hence, we conclude that the overall revenue improves when  $(\hat{\mathbf{p}}, R)$  is used. Therefore, we reach a contradiction and  $(\mathbf{p}, R)$  has to satisfy one of the conditions 1–3 of the proposition.

**PROOF OF THEOREM 1.** We first describe how the optimal prices and ranking in OPT-2 are obtained, and then we consider the computational complexity of the solution. As explained in the text, given the set  $L$ , the optimal solution of OPT-2 can be obtained by solving OPT-3. The solution of the latter problem is identical to that of (6), with  $i = 0$ ,  $j = T + 1$ ,

$p = 0$ , and hence the optimal value is equal to  $\omega(0, T + 1, 0)$ . Given  $\omega(i, j, p)$ , for  $0 \leq i \leq j \leq T + 1$ , one can construct the optimal sequence of prices in this problem using the recursion in (7): We say that  $k$  is the solution for  $\omega(i, j, p)$  if the right-hand side of Equation (7) takes its maximum at  $k$  and  $k$  is the earliest time period that achieves the maximum. Let  $(k^*, p_{k^*})$  be the optimal solution of  $\omega(0, T + 1, 0)$  in (7). Then the price of time period  $k^*$  in the optimal solution of (6) is  $p_{k^*}$ , and prices for time periods earlier and later than  $k^*$  can be obtained by solving for the prices in the subproblems  $\omega(0, k^*, p_{k^*})$  and  $\omega(k^*, T + 1, p_{k^*})$ .

We assume that at each step the leftmost subproblem is solved first. We say that the time period  $k^*$ , which solves the  $i$ th subproblem, has priority  $i$  (hence the time period that solves  $\omega(0, T + 1, 0)$  has priority 1). Using these priorities together with prices, we next construct the ranking vector (consistent with the already obtained prices) that appears in the solution of OPT-3 (or equivalently to (7) with  $i = 0$ ,  $j = T + 1$ ,  $p = 0$ ). Consider time periods  $k_1$  and  $k_2$ . If  $p_{k_1} \neq p_{k_2}$ , it is clear how to rank them: the lower price will have a smaller rank. Now suppose  $p_{k_1} = p_{k_2}$ ; then the time period with lower priority receives lower ranking. Note that under this ranking, the ranking vector is consistent with prices. Moreover, when there are multiple time periods with the same price, the time period that has lower ranking is the one that is used by the algorithm to solve an earlier subproblem. This implies that the ranking is consistent with the time period each population receives service in the solution of the recursion (7).

We next characterize the computational complexity of providing a solution to OPT-2. Note that by Proposition 3, there exists an optimal solution for OPT-2 with prices that belong to set  $L$ . It can be seen from (5) that to compute the prices in this set we need quantities of the form  $z_{ijk} = \sum_{t_1=i}^k \sum_{t_2=k}^j a_{t_1, t_2}$  for all  $i \leq j \leq k$ . Note that there are  $O(T^3)$  values  $z_{ijk}$  can take, and each value takes at most  $O(T^2)$  to compute. Thus, all values of  $z_{ijk}$  and the set  $L$  can be computed in  $O(T^5)$  time (the computation time can be further reduced by exploiting the relation between different values of  $z_{ijk}$ ; this is omitted because it does not affect our final complexity result). Thus, in  $O(T^5)$  time we can reduce OPT-2 to OPT-3. We characterize the computational complexity of the latter problem.

Observe that the algorithm relies on characterizing  $\omega(i, j, p)$  for all time periods  $i \leq j$  and  $p \in L$ . Since cardinality of  $L$  is  $O(T^3)$ , there are  $O(T^5)$  values of  $\omega(i, j, p)$  that need to be characterized. These can be computed using the condition  $\omega(i, j, p) = 0$  if  $i + 1 \geq j - 1$  and the recursion in (7). At each step of the recursion, there are  $O(T)$  different values  $k$  can take. On the other hand, for a given value of  $k$ , the corresponding optimal  $p_k$  can be computed in  $O(1)$ : Since for all  $p \in L$  we have  $p \geq p_M$ , it follows that  $\gamma_k^{ij}(p)$  is decreasing in  $p$ , provided that  $p \in L$ . Moreover,  $\omega(i, j, p)$  is also decreasing in  $p$  for all  $i, j$  since larger  $p$  corresponds to tighter constraints in (6). Thus, the  $p_k$  that solves (7) is the smallest  $p \geq p_M$  that makes the capacity constraint feasible. Therefore, it follows that  $p_k = \max\{p_M, F^{-1}(1 - c_k / \sum_{l=i+1}^k \sum_{m=k}^{j-1} a_{lm})\}$ , where the latter is the price that makes the capacity at time  $k$  tight. Since by construction both these prices belong to  $L$ , and elements of  $L$  were computed earlier, it follows that given  $k$ ,  $p_k$  can be constructed in  $O(1)$ .

Thus, we conclude that each step of the recursion in (7) can be computed in  $O(T)$ . Thus, the overall complexity of computing all  $\omega(i, j, p)$  is  $O(T^6)$ . Finally, given all values of  $\omega(i, j, p)$ , the construction of the prices that solve (6) takes  $O(T^2)$  following the procedure described in the beginning of the proof: to solve for each  $p_k$ , an instance of the recursion (7) needs to be solved. This takes  $O(T)$  time, and there are  $O(T)$  prices to be solved for. Similarly constructing priorities and rankings consistent with these prices takes another  $O(T)$ . Thus, the overall complexity of the algorithm is  $O(T^5 + T^6 + T^2 + T) = O(T^6)$ .

## References

- Ahn H, Gümüş M, Kaminsky P (2007) Pricing and manufacturing decisions when demand is a function of prices in multiple periods. *Oper. Res.* 55(6):1039–1057.
- Amazon Web Services (2011) AWS Case Study: DNAnexus. Accessed June 1, 2013, <http://aws.amazon.com/solutions/case-studies/dnanexus/>.
- Amazon Web Services (2012) Amazon EC2 Pricing. Accessed June 1, 2013, <http://aws.amazon.com/ec2/pricing/>.
- Armbrust M, Fox A, Griffith R, Joseph AD, Katz R, Konwinski A, Lee G, et al. (2010) A view of cloud computing. *Comm. ACM* 53(4):50–58.
- Arnold MA, Lippman SA (2001) The analytics of search with posted prices. *Econom. Theory* 17(2):447–466.
- AT&T (2012) Wireless Home Phone Plans. Accessed June 1, 2013, <http://www.att.com/shop/wireless/plans/individualplans.html>.
- Aviv Y, Pazgal A (2008) Optimal pricing of seasonal products in the presence of forward-looking consumers. *Manufacturing Service Oper. Management* 10(3):339–359.
- Aviv Y, Levin Y, Nediak M (2009) Counteracting strategic consumer behavior in dynamic pricing systems. Netessine S, Tang CS, eds. *Consumer-Driven Demand and Operations Management Models* (Springer, New York), 323–352.
- Başar T, Olsder GJ (1999) *Dynamic Noncooperative Game Theory*, 2nd ed. (Society for Industrial Mathematics, Philadelphia).
- Ben-Tal A, Nemirovski A (2002) Robust optimization methodology and applications. *Math. Programming* 92(3):453–480.
- Ben-Yehuda OA, Ben-Yehuda M, Schuster A, Tsafir D (2011) Deconstructing Amazon EC2 spot instance pricing. *Proc. 2011 Third Internat. Conf. Cloud Comput. Tech. Sci.* (IEEE Computer Society, Washington, DC), 304–311.
- Bergemann D, Said M (2011) Dynamic auctions: A survey. *Wiley Encyclopedia of Operations Research and Management Science* (John Wiley & Sons, Hoboken, NJ).
- Bertsekas D (1999) *Nonlinear Programming* (Athena Scientific, Nashua, NH).
- Bertsimas D, Thiele A (2006) A robust optimization approach to inventory theory. *Oper. Res.* 54(1):150–168.
- Besanko D, Winston WL (1990) Optimal price skimming by a monopolist facing rational consumers. *Management Sci.* 36(5):555–567.
- Bilton N (2012) Disruptions: Taxi supply and demand, priced by the mile. *New York Times* (January 8), <http://bits.blogs.nytimes.com/2012/01/08/disruptions-taxi-supply-and-demand-priced-by-the-mile/>.
- Bitran G, Caldentey R (2003) An overview of pricing models for revenue management. *Manufacturing Service Oper. Management* 5(3):203–229.
- Bitran GR, Mondschein SV (1997) Periodic pricing of seasonal products in retailing. *Management Sci.* 43(1):64–79.
- Cachon GP, Feldman P (2013) Price commitments with strategic consumers: Why it can be optimal to discount more frequently than optimal. Working paper, University of Pennsylvania, Philadelphia.
- Chan LMA, Shen ZJM, Simchi-Levi D, Swann JL (2004) Coordination of pricing and inventory decisions: A survey and classification. Simchi-Levi D, Wu SD, Shen Z-J, eds. *Handbook of Quantitative Supply Chain Analysis*, International Series in Operations Research & Management Science, Vol. 74 (Springer, New York), 335–392.
- Coase RH (1972) Durability and monopoly. *J. Law Econom.* 15(1):143–149.
- Dasu S, Tong C (2010) Dynamic pricing when consumers are strategic: Analysis of posted and contingent pricing schemes. *Eur. J. Oper. Res.* 204(3):662–671.
- Elmaghraby W, Gülcü A, Keskinocak P (2008) Designing optimal preannounced markdowns in the presence of rational customers with multiunit demands. *Manufacturing Service Oper. Management* 10(1):126–148.
- Federgruen A, Heching A (1999) Combined pricing and inventory control under uncertainty. *Oper. Res.* 47(3):454–475.
- Feng Y, Gallego G (1995) Optimal starting times for end-of-season sales and optimal stopping times for promotional fares. *Management Sci.* 41(8):1371–1391.
- Federal Energy Regulatory Commission (2008) Assessment of demand response and advanced metering. Staff report, Federal Energy Regulatory Commission, Washington, DC. <http://www.ferc.gov/legal/staff-reports/12-08-demand-response.pdf>.
- Gallego G, Ryzin GV (1994) Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Sci.* 40(8):999–1020.
- Gul F, Sonnenschein H, Wilson R (1986) Foundations of dynamic monopoly and the Coase conjecture. *J. Econom. Theory* 39(1):155–190.
- Kalanick T (2011) NYE surge pricing explained. *Uber.com* (blog) (December 31), <http://blog.uber.com/2011/12/31/nye-surge-pricing-explained/>.
- Kimball WS (1952) *Calculus of Variations by Parallel Displacement* (Butterworth, London).
- Lazear EP (1986) Retail pricing and clearance sales. *Amer. Econom. Rev.* 76(1):14–32.
- Levin Y, McGill J, Nediak M (2010) Optimal dynamic pricing of perishable items by a monopolist facing strategic consumers. *Production Oper. Management* 19(1):40–60.
- Liu Q, van Ryzin GJ (2008) Strategic capacity rationing to induce early purchases. *Management Sci.* 54(6):1115–1131.
- Lohr S (2011) The business market plays cloud computing catch-up. *New York Times* (April 15), <http://www.nytimes.com/2011/04/15/business/15cloud.html>.
- Pai MM, Vohra R (2013) Optimal dynamic auctions and simple index rules. *Math. Oper. Res.* 38(4):682–697.
- Shen ZJM, Su X (2007) Customer behavior modeling in revenue management and auctions: A review and new research opportunities. *Production Oper. Management* 16(6):713–728.
- Stokey NL (1979) Intertemporal price discrimination. *Quart. J. Econom.* 93(3):355–371.
- Stokey NL (1981) Rational expectations and durable goods pricing. *Bell J. Econom.* 12(1):112–128.
- Su X (2007) Intertemporal pricing with strategic customer behavior. *Management Sci.* 53(5):726–741.
- Su X, Zhang F (2009) On the value of commitment and availability guarantees when selling to strategic consumers. *Management Sci.* 55(5):713–726.
- Talluri K, van Ryzin GV (2004) Revenue management under a general discrete choice model of consumer behavior. *Management Sci.* 50(1):15–33.
- Vitek J (2009) Windows Azure lessons learned: Domino's Pizza. *MSDN Blog* (November 17), <http://blogs.msdn.com/b/jonbox/archive/2009/11/17/domino-s-demonstrates-tomcat-site-on-windows-azure.aspx>.
- Wang R (1993) Auctions versus posted-price selling. *Amer. Econom. Rev.* 83(4):838–851.
- Windows Azure (2012) Windows Azure Pricing. Accessed June 1, 2013. <http://www.windowsazure.com/en-us/pricing/calculator/>.