

# Appointment Scheduling under Patient Preference and No-Show Behavior

Jacob Feldman

School of Operations Research and Information Engineering,  
Cornell University, Ithaca, NY 14853  
jbf232@cornell.edu

Nan Liu

Department of Health Policy and Management, Mailman School of Public Health,  
Columbia University, New York, NY 10032  
nl2320@columbia.edu

Huseyin Topaloglu

School of Operations Research and Information Engineering,  
Cornell University, Ithaca, NY 14853  
topaloglu@orie.cornell.edu

Serhan Ziya

Department of Statistics and Operations Research,  
University of North Carolina, Chapel Hill, NC 27599  
ziya@unc.edu

April 9, 2014

## Abstract

Motivated by the rising popularity of electronic appointment booking systems, we develop appointment scheduling models that take into account the patient preferences regarding when they would like to be seen. The service provider dynamically decides which appointment days to make available for the patients. Patients arriving with appointment requests may choose one of the days offered to them or leave without an appointment. Patients with scheduled appointments may cancel or not show up for the service. The service provider collects a “revenue” from each patient who shows up and incurs a “service cost” that depends on the number of scheduled appointments. The objective is to maximize the expected net “profit” per day. We begin by developing a static model that does not consider the current state of the scheduled appointments. We give a characterization of the optimal policy under the static model and bound its optimality gap. Building on the static model, we develop a dynamic model that considers the current state of the scheduled appointments and propose a heuristic solution procedure. In our computational experiments, we test the performance of our models under the patient preferences estimated through a discrete choice experiment that we conduct in a large community health center. Our computational experiments reveal that the policies we propose perform well under a variety of conditions.

# 1 Introduction

Enhancing patient experience of care has been set as one of the “triple aims” to improve the healthcare system in many developed countries including the United States, Canada, and the United Kingdom. This aim is considered as equally, if not more, important as the other aims of improving the health of the population and managing per capita cost of care; see Berwick et al. (2008) and Institute for Healthcare Improvement (2012). An important component of enhancing patient experience of care is to provide more flexibility to patients regarding how, when and where to receive treatment. In pursuit of this objective, the National Health Service in the United Kingdom launched its electronic booking system, called Choose and Book, for outpatient appointments in January 2006; see Green et al. (2008). In the United States, with the recent Electronic Health Records and Meaningful Use Initiative, which calls for more and better use of health information technology, online scheduling is being adopted by an increasingly larger percentage of patients and clinics; see US Department of Health and Human Services (2011), Weiner et al. (2009), Silvestre et al. (2009), Wang and Gupta (2011) and Zocdoc (2012).

In contrast to the traditional appointment scheduling systems where patients are more or less told by the clinic when to come and whom to see or are given limited options on the phone, electronic appointment booking practices make it possible to better accommodate patient preferences by providing patients with more options. Giving patients more flexibility when scheduling their appointments has benefits that can go beyond simply having more satisfied patients. More satisfied patients lead to higher patient retention rates, which potentially allow providers to negotiate better reimbursement rates with payers; see Rau (2011). More satisfied patients can also lead to reduced no-show rates, helping maintain the continuity of care and improve patient health outcomes; see Bowser et al. (2010) and Schectman et al. (2008). An important issue when providing flexibility to patients is that of managing the operational challenges posed by giving more options. In particular, one needs to carefully choose the level of flexibility offered to the patients while taking into account the operational consequences. It is not difficult to imagine that giving patients complete flexibility in choosing their appointment times would lead to high variability in the daily load of a clinic. Thus, options provided to the patients need to be restricted in a way that balances the benefits with the costs. While such decisions have been studied in some industries, such as airlines, hospitality and manufacturing, scheduling decisions with consideration of patient preferences has largely been ignored in the appointment scheduling literature; see Cayirli and Veral (2003), Gupta and Denton (2008) and Rohleder and Klassen (2000). The goal of this paper is to develop models that can aid in the appointment scheduling process while considering the patient preferences.

In this paper, with electronic appointment booking systems in mind, we develop models, which can be used to decide which appointment days to offer in response to incoming appointment requests. Specifically, we consider a single service provider receiving appointment requests every day. The service provider offers a menu of appointment dates within the scheduling window for patients to choose from. During the day, patients arrive with the intention of scheduling

appointments and they either choose to book an appointment on one of the days made available to them or leave without scheduling any. We assume that patient choice behavior is governed by a multinomial logit choice model; see McFadden (1974). In the meantime, patients with appointments may decide not to show up and those with appointments on a future day may cancel. The service provider generates a “revenue” from each patient who shows up for her appointment and incurs a “cost” that depends on the number of patients scheduled to be seen on a day. The objective is to maximize the expected net “profit” per day by choosing the offer set, the set of days offered to patients who demand appointments. We begin by developing a static model, which can be used to make decisions without using the information about the currently booked appointments. Then, we build on the static model and develop a dynamic formulation, which can be used to take the current state of the booked appointments into consideration when making scheduling decisions.

Our static model is a mathematical program in which the decision variables are the probabilities with which different subsets of appointment days will be offered to the patients, independent of the state of the booked appointments. One difficulty with this optimization problem is that the number of decision variables increases exponentially with the length of the scheduling window. To overcome this, we exploit the special structure of the multinomial logit model to reformulate the static model in a more compact form. The number of decision variables in the compact formulation increases only linearly with the number of days in the scheduling window, making it tractable to solve. We show that if the no-show probability, conditional on the event that the patient has not canceled before her appointment, does not depend on patient appointment delays, then there exists an easily implementable optimal policy from the static model, which randomizes between only two adjacent offer sets. To assess the potential performance loss as a result of the static nature of the model, we provide a bound on the optimality gap of the static model. The bound on the optimality gap approaches zero as the average patient load per day increases, indicating that the static model may work well in systems handling high patient demand with respect to their capacity. In addition to randomized scheduling policies, we discuss how our approach can also be used to calculate the optimal deterministic policy, which may be even easier to implement.

Our dynamic model improves on the static one by taking the state of the booked appointments into consideration when making its decisions. The starting point for the dynamic model is the Markov decision process formulation of the appointment scheduling operations. Unfortunately, solving this formulation to optimality using standard dynamic programming tools is not practical since the state space is too large for realistically sized instances. Therefore, we propose an approximation method based on the Markov decision process formulation. This approximation method can be seen as applying a single step of the standard policy improvement algorithm on an initial “good” policy. For the initial “good” policy, we employ the policy provided by the static model. Implementing the policy provided by the dynamic model allows us to make decisions in an online fashion by solving a mathematical program that uses the current state of the booked appointments. The structure of this mathematical program is similar to the one we solve for the static model. Thus, the structural results obtained for the static model, at least partially, apply

to the dynamic model as well. We carry out a simulation study to investigate how our proposed methods perform. In order to create a clinic environment that is as realistic as possible, we estimate patient preferences using data from a discrete choice experiment conducted in a large community health center. We also generate numerous problem instances by varying model parameters so that we can compare the performance of our policies with benchmarks over a large set of possible settings regarding the clinic capacity and the patient demand. These benchmarks are carefully chosen to mimic the policies used in practice. Our simulation study suggests that our proposed dynamic policy outperforms benchmarks, in particular when capacity is tight compared to demand and overtime cost is high.

The studies in Ryan and Farrar (2000), Rubin et al. (2006), Cheraghi-Sohi et al. (2008), Gerard et al. (2008) and Hole (2008) all point out that patients do have preferences regarding when to visit the clinic and which physician to see. In general, patients prefer appointments that are sooner than later, but they may prefer a later day over an earlier one if the appointment on the later day is at a more convenient time; see Sampson et al. (2008). Capturing these preferences in their full complexity while keeping the formulation sufficiently simple can be a challenging task, but our static and dynamic models yield tractable policies to implement in practice. The models we propose can be particularly useful for clinics that are interested in a somewhat flexible implementation of open access, the policy of seeing today’s patient today; see Murray and Tantau (2000). While same-day scheduling reduces the access time for patients and keeps no-show rates low, it is a feasible practice only if demand and service capacity are in balance; see Green and Savin (2008). Furthermore, some patients highly value the flexibility of scheduling appointments for future days. A recent study in Sampson et al. (2008) found that a 10% increase in the proportion of same-day appointments was associated with an 8% reduction in the proportion of patients satisfied. This is a somewhat surprising finding, attributed to the decreased flexibility in booking appointments that comes with restricting access to same-day appointments. Our models help find the “right” balance among competing objectives of providing more choices to patients, reducing appointment delays, increasing system efficiency, and reducing no-shows.

While past studies have found that patients have preferences regarding when they would like to be seen, to our knowledge, there has been limited attempt in quantifying the relative preferences of patients using actual data. To understand patient preferences in practice and to use this insight in populating the model parameters in our computational study, we collected data from a large urban health center in New York City and used these data to estimate the parameters of the patient choice model. For the patients of this particular clinic, the choice model estimates how patients choose one day over the other. The choice model confirms that when patients have a health condition that does not require an emergency but still needs fast attention, they do prefer to be seen earlier, even after taking their work and other social obligations into account.

In addition to capturing the preferences of the patients on when they would like to be seen, a potentially useful feature of our models is that they capture the patient no-show process in a quite

general fashion. In particular, the probability that a customer cancels her appointment on a given day depends on when the appointment was scheduled and how many days there are left until the appointment day arrives. In this way, we can model the dependence of cancellation probabilities on appointment delays. Similarly, we allow the probability that a patient shows up for her appointment to depend on how much in advance the appointment was booked. Past studies are split on how cancellation and no-show probabilities depend on appointment delays. A number of articles find that patients with longer appointment delays are more likely to not show up for their appointments; see Grunebaum et al. (1996), Gallucci et al. (2005), Dreier et al. (2008), Bean and Talaga (1995) and Liu et al. (2010). However, other studies found no such relationship; see Wang and Gupta (2011) and Gupta and Wang (2011). In this paper, we keep our formulation general so that no-show and cancellation probabilities can possibly depend on the delays experienced by patients.

The rest of the paper is organized as follows. Section 2 reviews the relevant literature. Section 3 describes the model primitives we use. Section 4 presents our static model, gives structural results regarding the policy provided by this model, provides a bound on the optimality gap, and derives solutions for the optimal deterministic policies. Section 5 develops our dynamic model. Section 6 discusses our discrete choice experiment on patient preferences, explains our simulation setup and presents our numerical findings. Section 7 provides concluding remarks. In the Online Appendix, we discuss a number of extensions that can accommodate more general patient choice models, and present the proof of the results that are omitted in the paper.

## 2 Literature Review

The appointment scheduling literature has been growing rapidly in recent years. Cayirli and Veral (2003) and Gupta and Denton (2008) provide a broad and coherent overview of this literature. Most work in this area focuses on intra-day scheduling and is typically concerned with timing and sequencing patients with the objective of finding the “right” balance between in-clinic patient waiting times and physician utilization. In contrast, our work focuses on inter-day scheduling and does not explicitly address how intra-day scheduling needs to be done. In that regard, our models can be viewed as daily capacity management models for a medical facility. Similar to our work, Gerchak et al. (1996), Patrick et al. (2008), Liu et al. (2010), Ayvaz and Huh (2010) and Huh et al. (2013) deal with the allocation and management of daily service capacity. However, none of these articles take patient preferences and choice behavior into account.

To our knowledge, Rohleder and Klassen (2000), Gupta and Wang (2008) and Wang and Gupta (2011) are the only three articles that consider patient preferences in the context of appointment scheduling. These three articles deal with appointment scheduling within a single day, whereas our work focuses on decisions over multiple days. By focusing on a single day, Rohleder and Klassen (2000), Gupta and Wang (2008) and Wang and Gupta (2011) develop policies regarding the specific timing of the appointments, but they do not incorporate the fact that the appointment preferences

of patients may not be restricted within a single day. Furthermore, their proposed policies do not smooth out the daily load of the clinic. In contrast, since our policies capture preferences of patients over a number of days into the future, they can improve the efficiency in the use of the clinic’s service capacity by distributing demand over multiple days. Another interesting point of departure between our work and the previous literature is the assumptions regarding how patients interact with the clinic while scheduling their appointments. Specifically, Gupta and Wang (2008) and Wang and Gupta (2011) assume that patients first reveal a set of their preferred appointment slots to the clinic, which then decides whether to offer the patient an appointment in the set or to reject the patient. In our model, the clinic offers a set of appointment dates to the patients and the patients either choose one of the offered days or decline to make an appointment.

While there is limited work on customer choice behavior in healthcare appointment scheduling, there are numerous related papers in the broader operations literature, particularly in revenue management and assortment planning. Typically, these papers deal with problems where a firm chooses a set of products to offer its customers based on inventory levels and remaining time in the selling horizon. In response to the offered set of products, customers make a choice within the offered set according to a certain choice model. Talluri and van Ryzin (2004) consider a revenue management model on a single flight leg, where customers make a choice among the fare classes made available to them. Gallego et al. (2004), Liu and van Ryzin (2008), Kunnumkal and Topaloglu (2008) and Zhang and Adelman (2009) extend this model to a flight network. Bront et al. (2009) and Rusmevichientong et al. (2010) study product offer decisions when there are multiple customer types, each type choosing according to a multinomial logit model with a different set of parameters. The authors show that the corresponding optimization problem is NP-hard. In contrast, Gallego et al. (2011) show that the product offer decisions under the multinomial logit model can be formulated as a linear program when there is a single customer type. We refer the reader to Talluri and van Ryzin (2004) and Kk et al. (2009) for detailed overview of the literature on revenue management and product offer decisions.

Two papers that are particularly relevant to ours are Liu et al. (2010) and Topaloglu (2013). Liu et al. (2010) consider an appointment scheduling problem. Their model keeps track of appointments over a number of days and uses the idea of applying a single step of the policy improvement algorithm to develop a heuristic method. Liu et al. (2010) do not take patient preferences into account in any way and assume that patients always accept the appointment day offered to them. This assumption makes the formulation and analysis in Liu et al. (2010) significantly simpler than ours. Furthermore, we are able to characterize the structure of the optimal state-independent policy and provide a bound on the optimality gap of this policy, whereas Liu et al. (2010) do neither. On the other hand, Topaloglu (2013) considers a joint stocking and product offer model over a single period. The model in Topaloglu (2013) is essentially a newsvendor model with multiple products, where the decision variables are the set of offered products and their corresponding stocking quantities. Customers choose among the set of offered products according to a certain choice model and the objective is to maximize the total expected profit. This model is a single

period model with newsvendor features. In contrast, we take on a dynamic appointment scheduling problem, where patients arrive randomly over time and we need to adjust the set of offer days over time as a function of the state of the system. Furthermore, our model has to explicitly deal with cancellations and no-shows. This dimension results in a high-dimensional state vector in a dynamic programming formulation and we need to come up with tractable approximate policies. Such an issue does not occur at all in a newsvendor setting.

### 3 Model Primitives

We consider a single care provider who schedules patient appointments over time and assume that the following sequence of events occur on each day.<sup>1</sup> First, we observe the state of the appointments that were already scheduled and decide which subset of days in the future to make available for the patients making appointment requests on the current day. Second, patients arrive with appointment requests and choose an appointment day among the days that are made available to them. Third, some of the patients with appointments scheduled for future days may cancel their appointments and we observe the cancellations. Also, some of the patients with appointments scheduled on the current day may not show up and we observe those who do. In reality, appointment requests, cancellations and show-ups occur throughout the day with no particular time ordering among them. Thus, it is important to note that our assumption regarding the sequence of events is simply a modeling choice and our policies continue to work as long as the set of days made available for appointments are chosen at the beginning of each day. The expected revenue the clinic generates on a day is determined by the number of patients that it serves on that day. The expected cost the clinic incurs is determined by the number of appointments scheduled for that day before no-show realizations. This is mainly because staffing is the primary cost driver and staffing decisions have to be made beforehand.

The number of appointment requests on each day has Poisson distribution with mean  $\lambda$ . Each patient calling in for an appointment can potentially be scheduled either for the current day or for one of the  $\tau$  future days. Therefore, the scheduling horizon is  $\mathcal{T} = \{0, 1, \dots, \tau\}$ , where day 0 corresponds to the current day and the other days correspond to a day in the future. The decision we make on each day is the subset of days in the scheduling horizon that we make available for appointments. If we make the subset  $S \subset \mathcal{T}$  of days available for appointments, then a patient schedules an appointment  $j$  days into the future with probability  $P_j(S)$ . We assume that the choice probability  $P_j(S)$  is governed by the multinomial logit choice model; see McFadden (1974). Under this choice model, each patient associates the preference weight of  $v_j$  with the option of scheduling an appointment  $j$  days into the future. Furthermore, each patient associates the nominal preference weight of 1 with the option of not scheduling an appointment at all. In this case, if we offer the

---

<sup>1</sup>Some physicians work as a member of a physician team and reserves a certain number of appointment slots in each day for urgent patients (e.g., walk-ins) who can be seen by any one of the team members, but use the rest of their service capacity for scheduling their own patients only. In such a team practice, our focus is on managing the capacity that is reserved for scheduled appointments only.

subset  $S$  of days available for appointments, then the probability that an incoming patient schedules an appointment  $j \in S$  days into the future is given by

$$P_j(S) = \frac{v_j}{1 + \sum_{k \in S} v_k}. \quad (1)$$

With the remaining probability,  $N(S) = 1 - \sum_{j \in S} P_j(S) = 1/(1 + \sum_{k \in S} v_k)$ , a patient leaves the system without scheduling an appointment.

If a patient called in  $i$  days ago and scheduled an appointment  $j$  days into the future, then this patient cancels her appointment on the current day with probability  $r'_{ij}$ , independently of the other appointments scheduled. For example, if the current day is October 15, then  $r'_{13}$  is the probability a patient who called in on October 14 and scheduled an appointment for October 17 cancels her appointment on the current day. We let  $r_{ij} = 1 - r'_{ij}$  so that  $r_{ij}$  is the probability that we retain a patient who called in  $i$  days ago and scheduled an appointment  $j$  days into the future. If a patient called in  $i$  days ago and scheduled an appointment on the current day, then this patient shows up for her appointment with probability  $s_i$ , conditional on the event that she has not canceled until the current day. We assume that  $r_{ij}$  is decreasing in  $j$  for a fixed value of  $i$  so that the patients scheduling appointments further into the future are less likely to be retained.

For each patient served on a particular day, we generate a nominal revenue of 1. We have a regular capacity of serving  $C$  patients per day. After observing the cancellations, if the number of appointments scheduled for the current day exceeds  $C$ , then we incur an additional staffing and overtime cost of  $\theta$  per patient above the capacity. The cost of the regular capacity of  $C$  patients is assumed to be sunk and we do not explicitly account for this cost in our model. With this setup, the profit per day is linear in the number of patients that show up and piecewise-linear and concave in the number of appointments that we retain for the current day, but our results are not tied to the structure of this cost function and it is straightforward to extend our development to cover the case where the profit per day is a general concave function of the number of patients that show up and the number of appointments that we retain. Using such a general cost structure, one can easily enforce a capacity limit beyond which no patients can be scheduled by setting the cost rate beyond this capacity limit sufficiently high. Note that we also choose not to incorporate any cost associated with the patients' appointment delay mainly because in our model patients choose their appointment day based on their preferences anyway. Still, it is important to note that one can in fact incorporate such costs into our formulation rather easily since our formulation already keeps track of the appointment delay each patient experiences.

We can formulate the problem as a dynamic program by using the status of the appointments at the beginning of a day as the state variable. Given that we are at the beginning of day  $t$ , we can let  $Y_{ij}(t)$  be the number of patients that called in  $i$  days ago (on day  $t - i$ ), scheduled an appointment  $j$  days into the future (for day  $t - i + j$ ) and are retained without cancellation until the current day  $t$ . In this case, the vector  $Y(t) = \{Y_{ij}(t) : 1 \leq i \leq j \leq \tau\}$  describes the state of the scheduled appointments at the beginning of day  $t$ . This state description has  $\tau(\tau + 1)/2$  dimensions and the



state space can be very large even if we bound the number of scheduled appointments on a day by a finite number. In the next section, we develop a static model that makes its decisions without considering the current state of the booked appointments. Later on, we build on the static model to construct a dynamic model that indeed considers the state of the booked appointments.

## 4 Static Model

In this section, we consider a static model that makes each subset of days in the scheduling horizon available for appointments with a fixed probability. Since the probability of offering a particular subset of days is fixed, this model does not account for the current state of appointments when making its decisions. We solve a mathematical program to find the best probability with which each subset of days should be made available.

To formulate the static model, we let  $h(S)$  be the probability with which we make the subset  $S \subset \mathcal{T}$  of days available. If we make the subset  $S$  of days available with probability  $h(S)$ , then the probability that a patient schedules an appointment  $j$  days into the future is given by  $\sum_{S \subset \mathcal{T}} P_j(S) h(S)$ . Given that a patient schedules an appointment  $j$  days into the future, we retain this patient until the day of the appointment with probability  $\bar{r}_j = r_{0j} r_{1j} \dots r_{jj}$ . Similarly, this patient shows up for her appointment with probability  $\bar{s}_j = r_{0j} r_{1j} \dots r_{jj} s_j$ . Thus, noting that the number of appointment requests on a day has Poisson distribution with parameter  $\lambda$ , if we make the subset  $S$  of days available with probability  $h(S)$ , then the number of patients that schedule an appointment  $j$  days into future and that are retained until the day of the appointment is given by a Poisson random variable with mean  $\lambda \bar{r}_j \sum_{S \subset \mathcal{T}} P_j(S) h(S)$ . Similarly, we can use a Poisson random variable with mean  $\lambda \bar{s}_j \sum_{S \subset \mathcal{T}} P_j(S) h(S)$  to capture the number of patients that schedule an appointment  $j$  days into the future and that show up for their appointments. In this case, using  $\text{Pois}(\alpha)$  to denote a Poisson random variable with mean  $\alpha$ , on each day, the total number of patients whose appointments we retain until the day of the appointment is given by  $\text{Pois}(\sum_{j \in \mathcal{T}} \sum_{S \subset \mathcal{T}} \lambda \bar{r}_j P_j(S) h(S))$  and the total number of patients that show up for their appointments is given by  $\text{Pois}(\sum_{j \in \mathcal{T}} \sum_{S \subset \mathcal{T}} \lambda \bar{s}_j P_j(S) h(S))$ . To find the subset offer probabilities that maximize the expected profit per day, we can solve the problem

$$\max \quad \sum_{j \in \mathcal{T}} \sum_{S \subset \mathcal{T}} \lambda \bar{s}_j P_j(S) h(S) - \theta \mathbb{E} \left\{ \left[ \text{Pois} \left( \sum_{j \in \mathcal{T}} \sum_{S \subset \mathcal{T}} \lambda \bar{r}_j P_j(S) h(S) \right) - C \right]^+ \right\} \quad (2)$$

$$\text{subject to} \quad \sum_{S \subset \mathcal{T}} h(S) = 1 \quad (3)$$

$$h(S) \geq 0 \quad S \subset \mathcal{T}, \quad (4)$$

where we use  $[\cdot]^+ = \max\{\cdot, 0\}$ . In the problem above, the two terms in the objective function correspond to the expected revenue and the expected cost per day. The constraint ensures that the total probability with which we offer a subset of days is equal to 1. Noting  $\emptyset \subset \mathcal{T}$ , the problem above allows not offering any appointment slots to an arriving patient.

## 4.1 Reformulation

Problem (2)-(4) has  $2^{|\mathcal{T}|}$  decision variables, which can be too many in practical applications. For example, if we have a scheduling horizon of a month, then the number of decision variables in this problem exceeds a billion. However, it turns out that we can give an equivalent formulation for problem (2)-(4) that has only  $|\mathcal{T}|+1$  decision variables, which makes this problem quite tractable. In particular, Proposition 1 below shows that problem (2)-(4) is equivalent to the problem

$$\max \quad \sum_{j \in \mathcal{T}} \lambda \bar{s}_j x_j - \theta \mathbb{E} \left\{ \left[ \text{Pois} \left( \sum_{j \in \mathcal{T}} \lambda \bar{r}_j x_j \right) - C \right]^+ \right\} \quad (5)$$

$$\text{subject to} \quad \sum_{j \in \mathcal{T}} x_j + u = 1 \quad (6)$$

$$\frac{x_j}{v_j} - u \leq 0 \quad j \in \mathcal{T} \quad (7)$$

$$x_j, u \geq 0 \quad j \in \mathcal{T}. \quad (8)$$

In the problem above, we interpret the decision variable  $x_j$  as the probability that a patient schedules an appointment  $j$  days into the future. The decision variable  $u$  corresponds to the probability that a patient does not schedule an appointment. The objective function accounts for the expected profit per day as a function of the scheduling probabilities. Constraint (6) captures the fact that each patient either schedules an appointment on one of the future days or does not. To see an interpretation of constraints (7), we note that if we offer the subset  $S$  of days, then the probability that a patient schedules an appointment  $j$  days into the future is given by  $P_j(S) = v_j / (1 + \sum_{k \in S} v_k)$  if  $j \in S$  and 0 otherwise. On the other hand, the probability that a patient does not schedule an appointment is given by  $N(S) = 1 / (1 + \sum_{k \in S} v_k)$ . Therefore, we have  $P_j(S)/v_j - N(S) \leq 0$ . Surprisingly constraints (7) are the only place where the parameters of the multinomial logit choice model appears in problem (5)-(8) and these constraints turn out to be adequate to capture the choices of the patients as stipulated by the multinomial logit choice model. In the next proposition, we show that problems (2)-(4) and (5)-(8) are equivalent to each other. The proof uses the approach followed by Topaloglu (2013). We give the main ideas of the proof here, but defer the details to the appendix.

**Proposition 1.** *Problems (2)-(4) is equivalent to problem (5)-(8) in the sense that given an optimal solution to one problem, we can generate an optimal solution to the other one.*

*Proof.* Assume that  $h^* = \{h^*(S) : S \subset \mathcal{T}\}$  is an optimal solution to problem (2)-(4). Letting  $x_j^* = \sum_{S \subset \mathcal{T}} P_j(S) h^*(S)$  and  $u^* = \sum_{S \subset \mathcal{T}} N(S) h^*(S)$ , we show in the appendix that  $(x^*, u^*)$  with  $x^* = (x_0^*, \dots, x_\tau^*)$  is a feasible solution to problem (5)-(8) providing the same objective value as the solution  $h^*$ . On the other hand, assume that  $(x^*, u^*)$  with  $x^* = (x_0^*, \dots, x_\tau^*)$  is an optimal solution to problem (5)-(8). We reorder and reindex the days in the scheduling horizon so that we have  $x_0^*/v_0 \geq x_1^*/v_1 \geq \dots \geq x_\tau^*/v_\tau^*$ . Constraints (7) in problem (5)-(8) also ensure that  $u^* \geq x_0^*/v_0 \geq x_1^*/v_1 \geq \dots \geq x_\tau^*/v_\tau^*$ . We define the subsets  $S_0, S_1, \dots, S_\tau$  as  $S_j = \{0, 1, \dots, j\}$ . For

notational convenience, we define  $x_{\tau+1}^* = 0$ . In this case, letting

$$h^*(\emptyset) = u^* - \frac{x_0^*}{v_0} \quad \text{and} \quad h^*(S_j) = \left[1 + \sum_{k \in S_j} v_k\right] \left[\frac{x_j^*}{v_j} - \frac{x_{j+1}^*}{v_{j+1}}\right] \quad (9)$$

for all  $j = 0, 1, \dots, \tau$  and letting  $h^*(S) = 0$  for all other subsets of  $\mathcal{T}$ , we show in the appendix that  $\{h^*(S) : S \subset \mathcal{T}\}$  ends up being a feasible solution to problem (2)-(4) providing the same objective value as the solution  $(x^*, u^*)$ .  $\square$

The proof above indicates that we can obtain an optimal solution to problem (2)-(4) by solving problem (5)-(8) and using (9). Furthermore, (9) shows that there are at most  $|\mathcal{T}| + 1$  subsets for which the decision variables  $\{h(S) : S \subset \mathcal{T}\}$  take strictly positive values in an optimal solution. Problem (5)-(8) has a manageable number of decision variables and constraints, but its objective function is nonlinear. Nevertheless, the nonlinear term in the objective function is of the form  $F(\alpha) = \mathbb{E}\{[\text{Pois}(\alpha) - C]^+\}$  for  $\alpha \in \mathfrak{R}_+$  and Lemma 7 in the appendix shows that  $F(\alpha)$  is differentiable and convex in  $\alpha$ . Given that the objective function of problem (5)-(8) is convex and its constraints are linear, we can use a variety of approaches to solve this problem.

One approach to solve problem (5)-(8) is to use a cutting plane method for convex optimization, which represents the objective function of problem (5)-(8) with a number of cuts; see Ruszczyński (2006). A more direct approach for solving this problem is to observe that the function  $F(\cdot)$  as defined in the previous paragraph is a scalar and convex function. So, it is simple to build an accurate piecewise linear and convex approximation to  $F(\cdot)$  by evaluating this function at a finite number of grid points. We denote the approximation constructed in this fashion by  $\hat{F}(\cdot)$ . In this case, to obtain an approximate solution to problem (5)-(8), we can maximize the objective function  $\sum_{j \in \mathcal{T}} \lambda \bar{s}_j x_j - \theta \hat{F}(\alpha)$  subject to the constraint that  $\alpha = \sum_{j \in \mathcal{T}} \lambda \bar{r}_j x_j$  and constraints (6)-(8). Since this optimization problem has a piecewise linear objective function and linear constraints, it can be solved as a linear program. Furthermore, if we choose  $\hat{F}(\cdot)$  as a lower bound or an upper bound on  $F(\cdot)$ , then we obtain a lower or upper bound on the optimal objective value of problem (5)-(8). When the lower and upper bounds are close to each other, we can be confident about the quality of the solution we obtain from our approximation approach.

Another algorithmic strategy that we can use to solve problem (5)-(8) is based on dynamic programming. For a fixed value of  $u$ , we observe that the constraints in this problem correspond to the constraints of a knapsack problem. In particular, the items correspond to the days in the scheduling horizon, the capacity of the knapsack is  $1 - u$  and we can put at most  $v_j \times u$  units of item  $j$  into the knapsack. So, for a fixed value of  $u$ , we can solve problem (5)-(8) by using dynamic programming. In this dynamic program, the decision epochs correspond to days in the scheduling horizon  $\mathcal{T}$ . The state at decision epoch  $j \in \mathcal{T}$  has two components. The first component corresponds to the value of  $\sum_{k=0}^{j-1} x_k$ , capturing the portion of the knapsack capacity consumed by the decisions in the earlier decision epochs. The second component corresponds to the value of  $\sum_{k=0}^{j-1} \lambda \bar{r}_k x_k$ , capturing the accumulated value of the argument of the second term in the objective

function of problem (5)-(8). Thus, for a fixed value of  $u$ , we can solve problem (5)-(8) by computing the value functions  $\{\Theta_j(\cdot, \cdot | u) : j \in \mathcal{T}\}$  through the optimality equation

$$\Theta_j(b, c | u) = \max \quad \lambda \bar{s}_j x_j + \Theta_{j+1}(b + x_j, c + \lambda \bar{r}_j x_j | u) \quad (10)$$

$$\text{subject to} \quad 0 \leq x_j/v_j \leq u, \quad (11)$$

with the boundary condition that  $\Theta_{\tau+1}(b, c | u) = -\theta \mathbb{E}\{\text{Pois}(c) - C\}^+$  when  $b = 1 - u$  and  $\Theta_{\tau+1}(b, c | u) = -\infty$  when  $b \neq 1 - u$ . In the dynamic program above, we accumulate the first component of the objective function of problem (5)-(8) during the course of the decision epochs, but the second component of the objective function through the boundary condition. Having  $\Theta_{\tau+1}(b, c | u) = -\infty$  when  $b \neq 1 - u$  ensures that we always consume the total capacity availability of  $1 - u$  in the knapsack. It is straightforward to accurately solve the dynamic program for all  $u \in \mathbb{R}_+$  by discretizing the state variable  $(b, c)$  and  $u$  over a fine grid in  $\mathbb{R}_+^3$ . Solving the dynamic program for all  $u \in \mathbb{R}_+$ ,  $\max_{u \in \mathbb{R}_+} \Theta_1(0, 0 | u)$  gives optimal objective value of problem (5)-(8).

## 4.2 Static Model under Delay-Independent Show-up Probabilities

As we describe at the end of the previous section, it is not difficult to obtain good solutions to problem (5)-(8). Nevertheless, although we can obtain good solutions to problem (5)-(8), the structure of the optimal subset offer probabilities obtained from problem (2)-(4) is still not obvious. Furthermore, many of the subset offer probabilities  $\{h(S) : S \subset \mathcal{T}\}$  may take positive values in the optimal solution, indicating that the static policy may be highly randomized, which may be undesirable for practical implementation. In this section, we consider the special case where the show-up probability of an appointment, conditional on it having been retained until the day of service, does not depend on how many days ago the patient called in. We prove two results regarding the structure of the optimal subset availability probabilities. First, we show that the subsets of days that we make available always consists of a certain number of consecutive days. Second, we show that the optimal subset availability probabilities randomize between only two subsets and these two subsets differ only in one day. Thus, in this special case, the optimal solution turns out to have a very simple structure even though no explicit restriction on the set of allowable policies is made. The assumption of conditional independence of no-shows on appointment delays somewhat limits the general appeal of our formulation. However, some studies have indeed found that once cancellations are factored in, dependence of no-shows to appointment delays either disappears or weakens significantly. For example, Gupta and Wang (2011) analyzed data sets from three different health organizations and found that the association between patient no-shows, after accounting for cancellations, and appointment delays is either statistically insignificant or statistically significant but very weak. Furthermore, if the assumption of conditional independence of no-shows on appointment delays does not hold, then we can use the algorithm strategies described at the end of the previous section to obtain an optimal solution to problem (5)-(8).

Throughout this section, we assume that the conditional show-up probability  $s_j$  given an

appointment delay of  $j$  days is independent of the value of  $j$  and we use  $s_0$  to denote the common value of  $\{s_j : j \in \mathcal{T}\}$ . Thus, noting the definitions  $\bar{r}_j = r_{0j} r_{1j} \dots r_{jj}$  and  $\bar{s}_j = r_{0j} r_{1j} \dots r_{jj} s_j$ , we obtain  $\bar{s}_j = \bar{r}_j s_0$ . We emphasize that although the conditional show-up probabilities are assumed to be independent of how many days ago an appointment was scheduled, the probability of retaining an appointment  $r_{ij}$  can still be arbitrary. Define the scalar function  $R(\cdot)$  as  $R(\alpha) = s_0 \alpha - \theta \mathbb{E}\{\text{Pois}(\alpha) - C\}^+$ . Using the fact that  $\bar{s}_j = \bar{r}_j s_0$ , we can write the objective function of problem (5)-(8) succinctly as  $R(\sum_{j \in \mathcal{T}} \lambda \bar{r}_j x_j)$  and problem (5)-(8) becomes

$$\max \quad R\left(\sum_{j \in \mathcal{T}} \lambda \bar{r}_j x_j\right) \quad (12)$$

$$\text{subject to} \quad \sum_{j \in \mathcal{T}} x_j + u = 1 \quad (13)$$

$$\frac{x_j}{v_j} - u \leq 0 \quad j \in \mathcal{T} \quad (14)$$

$$x_j, u \geq 0 \quad j \in \mathcal{T}. \quad (15)$$

In the next proposition, we show that there exists an optimal solution to the problem above where at most one of the decision variables  $(x_0, x_1, \dots, x_\tau)$  satisfy  $0 < x_j/v_j < u$ . The other decision variables satisfy either  $x_j/v_j = u$  or  $x_j/v_j = 0$ . This result ultimately becomes useful to characterize the structure of the optimal subset offer probabilities. Before giving this result, we make one observation that becomes useful in the proof. In particular, noting the assumption that  $r_{ij}$  is decreasing in  $j$  for fixed  $i$ , the definition of  $\bar{r}_j$  implies that  $\bar{r}_j = r_{0j} r_{1j} \dots r_{jj} \geq r_{0,j+1} r_{1,j+1} \dots r_{j,j+1} r_{j+1,j+1} = \bar{r}_{j+1}$ , establishing that  $\bar{r}_0 \geq \bar{r}_1 \geq \dots \geq \bar{r}_\tau$ .

**Proposition 2.** *Assume that  $\{s_j : j \in \mathcal{T}\}$  share a common value. Then, there exists an optimal solution  $(x^*, u^*)$  with  $x^* = (x_0^*, \dots, x_\tau^*)$  to problem (12)-(15) that satisfies*

$$u^* = \frac{x_0^*}{v_0} = \dots = \frac{x_{k-1}^*}{v_{k-1}} \geq \frac{x_k^*}{v_k} \geq \frac{x_{k+1}^*}{v_{k+1}} = \dots = \frac{x_\tau^*}{v_\tau} = 0 \quad (16)$$

for some  $k \in \mathcal{T}$ .

*Proof.* We let  $(x^*, u^*)$  be an optimal solution to problem (12)-(15) and  $y^* = \sum_{j \in \mathcal{T}} \lambda \bar{r}_j x_j^*$  so that the optimal objective value of the problem is  $R(y^*)$ . For a fixed value of  $u$ , consider the problem

$$\zeta(u) = \min \quad \sum_{j \in \mathcal{T}} x_j \quad (17)$$

$$\text{subject to} \quad \sum_{j \in \mathcal{T}} \lambda \bar{r}_j x_j = y^* \quad (18)$$

$$0 \leq x_j \leq v_j u \quad j \in \mathcal{T}, \quad (19)$$

whose optimal objective value is denoted by  $\zeta(u)$ . For the moment, assume that there exists  $\hat{u}$  satisfying  $\zeta(\hat{u}) = 1 - \hat{u}$ . We establish the existence of such  $\hat{u}$  later in the proof. We let  $\hat{x} = (\hat{x}_0, \dots, \hat{x}_\tau)$  be an optimal solution to the problem above when we solve this problem with  $u = \hat{u}$ . In

this case, we have  $\sum_{j \in \mathcal{T}} \hat{x}_j = \zeta(\hat{u}) = 1 - \hat{u}$ ,  $\sum_{j \in \mathcal{T}} \lambda \bar{r}_j \hat{x}_j = y^*$ ,  $\hat{x}_j/v_j \leq \hat{u}$  for all  $j \in \mathcal{T}$ , which implies that  $(\hat{x}, \hat{u})$  is a feasible solution to problem (12)-(15) providing an objective function value of  $R(y^*)$ . Thus, the solution  $(\hat{x}, \hat{u})$  is also optimal for problem (12)-(15).

Problem (17)-(19) is a knapsack problem where the items are indexed by  $\mathcal{T}$ , the disutility of each item is 1 and the space requirement of item  $j$  is  $\lambda \bar{r}_j$ . We can put at most  $v_j u$  units of item  $j$  into the knapsack. We can solve this knapsack problem by starting from the item with the smallest disutility to space ratio and filling the knapsack with the items in that order. Noting that  $\bar{r}_0 \geq \bar{r}_1 \geq \dots \geq \bar{r}_\tau$ , the disutility to space ratios of the items satisfy  $1/(\lambda \bar{r}_0) \leq 1/(\lambda \bar{r}_1) \leq \dots \leq 1/(\lambda \bar{r}_\tau)$  so that it is optimal to fill the knapsack with the items in the order  $0, 1, \dots, \tau$ . Therefore, if we solve problem (17)-(19) with  $u = \hat{u}$ , then the optimal solution  $\hat{x}$  satisfies  $\hat{x}_0 = v_0 \hat{u}$ ,  $\hat{x}_1 = v_1 \hat{u}, \dots, \hat{x}_{k-1} = v_{k-1} \hat{u}$ ,  $\hat{x}_k \in [0, v_k \hat{u}]$ ,  $x_{k+1} = 0, \dots, x_\tau = 0$  for some  $k \in \mathcal{T}$ . Therefore, the solution  $(\hat{x}, \hat{u})$ , which is optimal to problem (12)-(15), satisfies (16) and we obtain the desired result.

It remains to show that there exists  $\hat{u}$  with  $\zeta(\hat{u}) = 1 - \hat{u}$ . Note that  $\zeta(\cdot)$  is continuous. By the definitions of  $(x^*, u^*)$  and  $y^*$ , if we solve problem (17)-(19) with  $u = u^*$ , then  $x^*$  is a feasible solution providing an objective value of  $\sum_{j \in \mathcal{T}} x_j^* = 1 - u^*$ , where the equality is by (13). Since the solution  $x^*$  may not be optimal to problem (17)-(19), we obtain  $\zeta(u^*) \leq 1 - u^*$ . Also, we clearly have  $\zeta(1) \geq 0$ . Letting  $g(u) = 1 - u$ , we obtain  $\zeta(u^*) \leq g(u^*)$  and  $\zeta(1) \geq g(1)$ . Since  $\zeta(\cdot)$  and  $g(\cdot)$  are continuous, there exists  $\hat{u}$  such that  $\zeta(\hat{u}) = g(\hat{u}) = 1 - \hat{u}$ .  $\square$

We emphasize that the critical assumption in Proposition 2 is that  $\bar{s}_j = \bar{r}_j s_0$  for all  $j \in \mathcal{T}$  and it is possible to modify this Proposition 2 to accommodate the cases where we do not necessarily have the ordering  $\bar{r}_0 \geq \bar{r}_1 \geq \dots \geq \bar{r}_\tau$ . The key observation is that whatever ordering we have among the probabilities  $\{\bar{r}_j : j \in \mathcal{T}\}$ , the disutility to space ratios of the items in the problem (17)-(19) satisfy the reverse ordering as long as  $\bar{s}_j = \bar{r}_j s_0$ . In this case, we can modify the ordering of the decision variables  $(x_0, \dots, x_\tau)$  in the chain of inequalities in (16) in such a way they follow the ordering of  $\{r_j : j \in \mathcal{T}\}$  and the proof of Proposition 2 still goes through.

We can build on Proposition 2 to solve problem (12)-(15) through bisection search. In particular, for fixed values of  $k$  and  $u$ , (16) shows that the decision variables  $x_0, x_1, \dots, x_{k-1}$  can be fixed at  $x_j = u v_j$  for all  $j = 0, 1, \dots, k-1$ , whereas the decision variables  $x_{k+1}, \dots, x_\tau$  can be fixed at 0. So, for a fixed value of  $k$ , to find the best values of  $u$  and  $x_k$ , we can solve the problem

$$\begin{aligned} Z_k = \max \quad & R\left(\sum_{j=0}^{k-1} \lambda \bar{r}_j v_j u + \lambda \bar{r}_k x_k\right) \\ \text{subject to} \quad & \sum_{j=0}^{k-1} v_j u + x_k + u = 1 \\ & \frac{x_k}{v_k} - u \leq 0 \\ & x_k, u \geq 0, \end{aligned}$$

where we set the optimal objective value  $Z_k$  of the problem above to  $-\infty$  whenever the problem is

infeasible. To find the best value of  $k$ , we can simply compute  $Z_k$  for all  $k \in \mathcal{T}$  and pick the value that provides the largest value of  $k$ . Thus,  $\max\{Z_k : k \in \mathcal{T}\}$  corresponds to the optimal objective value of problem (12)-(15). Although the problem above, which provides  $Z_k$  for a fixed value of  $k$ , appears to involve the two decision variables  $x_k$  and  $u$ , we can solve one of the decision variables in terms of the other one by using the first constraint, in which case, the problem above becomes a scalar optimization problem. Furthermore, noting the definition of  $R(\cdot)$  and the discussion in the last two paragraphs of Section 4.1, the objective function of the problem above is concave. Therefore, we can solve the problem above by using bisection search.

The next corollary shows two intuitive properties of the optimal subset availability probabilities. First, the optimal subset offer probabilities from problem (2)-(4) makes only subsets of the form  $\{0, 1, \dots, j\}$  available for some  $j \in \mathcal{T}$ . Therefore, it is optimal to make a certain number of days into the future available without skipping any days in between. Second, the optimal subset offer probabilities randomize at most between two possible subsets and these two subsets differ from each other only in one day. These results indicate that the randomized nature of the static policy is not a huge concern, as we randomize between only two subsets that are not too different from each other. The proof of this corollary can be found in the appendix.

**Corollary 3.** *Assume that  $\{s_j : j \in \mathcal{T}\}$  share a common value. Then, there exists an optimal solution  $h^* = \{h^*(S) : S \subset \mathcal{T}\}$  to problem (2)-(4) with only two of the decision variables satisfying  $h^*(S_1) \geq 0$ ,  $h^*(S_2) \geq 0$  for some  $S_1, S_2 \subset \mathcal{T}$  and all of the other decision variables are equal to 0. Furthermore, the two subsets  $S_1$  and  $S_2$  are either of the form  $S_1 = \emptyset$  and  $S_2 = \{0\}$ , or of the form  $S_1 = \{0, 1, \dots, j\}$  and  $S_2 = \{0, 1, \dots, j + 1\}$  for some  $j \in \mathcal{T}$ .*

Thus, the optimal subset offer probabilities randomize between at most two subsets of days. Each of these subsets of days contains consecutive days in them and they differ from each other by only one day. When the show-up probabilities  $\{s_j : j \in \mathcal{T}\}$  do not share a common value, an optimal static policy may randomize among multiple subsets that contain nonconsecutive days. Even if that is the case however, considering scheduling policies that only offer consecutive days might be of interest in practice given their simplicity. It turns out that it is not difficult to obtain the static policies that are optimal under such restriction. In other words, it is not difficult to find the optimal subset offer probabilities when we focus only on subsets that include consecutive days. To see this, we let  $\mathcal{S}_C$  denote the set of the offer sets that only contain consecutive days in  $\mathcal{T} = \{0, 1, \dots, \tau\}$  plus the empty set. We note that  $|\mathcal{S}_C| = \frac{(\tau+1)(\tau+2)}{2} + 1$ , as there are only  $k$  different offer sets that contain  $\tau + 2 - k$  consecutive days,  $k = 1, 2, \dots, \tau$ . Therefore, to obtain the desired scheduling policy with the requirement above, we can set  $h(S) = 0$  if  $S \notin \mathcal{S}_C$  in the problem (2)-(4). This approach reduces the number of decision variables from exponentially increasing in  $\tau$  to quadratically. Since the objective function is jointly concave in all decision variables and the constraints are linear, solving problem (2)-(4) is not difficult.

### 4.3 Performance Guarantee

The static model in problem (2)-(4) makes each subset of days available with a fixed probability and does not consider the current state of the scheduled appointments when making its decisions. A natural question is what kind of performance we can expect from such a static model. In this section, we develop a performance guarantee for our static model. In particular, we study the policy obtained from the simple deterministic approximation

$$Z_{DET} = \max \sum_{j \in \mathcal{T}} \lambda \bar{s}_j x_j - \theta \left[ \sum_{j \in \mathcal{T}} \lambda \bar{r}_j x_j - C \right]^+ \quad (20)$$

$$\text{subject to } \sum_{j \in \mathcal{T}} x_j + u = 1 \quad (21)$$

$$\frac{x_j}{v_j} - u \leq 0 \quad j \in \mathcal{T} \quad (22)$$

$$x_j, u \geq 0 \quad j \in \mathcal{T}. \quad (23)$$

The objective function above is the deterministic analogue of the one in problem (5)-(8), where the Poisson random variable is replaced by its expectation. Similar to problem (5)-(8), problem (20)-(23) characterizes a static model, but this static model is obtained under the assumption that all random quantities take on their expected values. In this section, we show that the policy obtained from problem (20)-(23) has a reasonable performance guarantee, even though this problem ignores all uncertainty. Since problem (5)-(8) explicitly addresses the uncertainty, the policy obtained from this problem is trivially guaranteed to perform better than the one obtained from problem (20)-(23). It is common in the revenue management literature to develop performance bounds for policies obtained from deterministic approximations; see Gallego and van Ryzin (1994), Gallego and van Ryzin (1997), Levi and Radovanovic (2010), and Topaloglu et al. (2011). Our analysis has a similar flavor but it nevertheless provides useful insights into identifying parameters which have the strongest effect on the performance of the optimal static policies.

In the next lemma, we begin by showing that the optimal objective value of problem (20)-(23) provides an upper bound on the expected profit per day generated by the optimal policy, which can be a policy that depends on the state of the system. The proof of this result follows from a standard argument that uses Jensen's inequality and is given in the appendix. Throughout this section, we use  $V^*$  to denote the expected profit per day generated by the optimal policy.

**Lemma 4.** *We have  $Z_{DET} \geq V^*$ .*

We let  $\Pi(x)$  with  $x = (x_0, \dots, x_\tau)$  be the objective function of problem (5)-(8). If  $(x^*, u^*)$  is an optimal solution to this problem, then the static policy that uses the subset offer probabilities  $h^* = \{h^*(S) : S \subset \mathcal{T}\}$  defined as in (9) generates an expected profit of  $\Pi(x^*)$  per day. Since  $V^*$  is the expected profit per day generated by the optimal policy, we have  $\Pi(x^*)/V^* \leq 1$ . In the next proposition, we give a lower bound on  $\Pi(x^*)/V^*$ , which bounds the optimality gap of the policy that we obtain by solving problem (5)-(8).



**Proposition 5.** *Letting  $(x^*, u^*)$  be an optimal solution to problem (5)-(8), we have*

$$\frac{\Pi(x^*)}{V^*} \geq 1 - \frac{\theta \sqrt{\frac{\bar{r}_0}{2\pi}} \sqrt{\frac{\lambda \bar{r}_0}{C}}}{\bar{s}_0 \min \left\{ \frac{v_0}{1+v_0}, \frac{C}{\lambda \bar{r}_0} \right\} \sqrt{\lambda}}. \quad (24)$$

The proof of Proposition 5 appears in the appendix and is based on the observation that if  $(\hat{x}, \hat{u})$  is an optimal solution to problem (20)-(23), then since this solution is feasible for problem (5)-(8), we have  $\Pi(x^*)/V^* \geq \Pi(\hat{x})/V^*$ . In this case, the proof of Proposition 5 shows that  $\Pi(\hat{x})/V^*$  is lower bounded by the expression on the right side of (24), which yields the desired result. The performance guarantee in (24) has useful practical implications. Since  $\lambda \bar{r}_0$  is an upper bound on the expected number of appointments we retain on a particular day, if  $C$  and  $\lambda$  satisfy  $C \geq \lambda \bar{r}_0$ , then we essentially have a situation where the capacity exceeds the expected demand. In this case, the quantity  $\sqrt{\lambda \bar{r}_0 / C}$  in the numerator on the right side of (24) is upper bounded by 1 and the min operator in the denominator evaluates to  $\frac{v_0}{1+v_0}$ , which implies that  $\Pi(x^*)/V^*$  is lower bounded by  $1 - \theta \sqrt{\bar{r}_0 / (2\pi)} / (\bar{s}_0 \frac{v_0}{1+v_0} \sqrt{\lambda})$ . So, the performance guarantee improves with rate  $1/\sqrt{\lambda}$ , indicating that as long as capacity exceeds the expected demand, the performance of static policies improves with the demand volume. In other words, we expect static policies to perform well in systems handling large demand volumes. Similarly, if the capacity and the expected demand increase with the same rate so that the ratio  $\frac{C}{\bar{r}_0 \lambda}$  stays constant, then the performance of state independent policies still improves with rate  $1/\sqrt{\lambda}$ , even when  $\frac{C}{\lambda \bar{r}_0}$  evaluates to a number less than 1. These observations support using static models for large systems with high demand volumes.

To get a feel for the performance guarantee in (24), we consider a system with  $C = 12$ ,  $\lambda = 16$ ,  $v_0 = 1$ ,  $\bar{r}_0 = 1$ ,  $\bar{s}_0 = 0.9$  and  $\theta = 1.5$ . These parameters correspond to a case where if we offer only the current day for an appointment, then a patient leaves without scheduling any appointments with probability  $N(\{0\}) = 1/(1+v_0) = 1/2$ . The probability that we retain a patient with a same day appointment is 1 and the probability that this patient shows up for her appointment is 0.9. Considering the fact that we generate a nominal revenue of 1 from each served patient, the overtime cost of 1.5 is reasonable. The capacity and the arrival rate are reasonable for a small clinic. For these parameters, the right side of (24) evaluates to about 62%, indicating that there exists a static policy that generates at least 62% of the optimal expected profit per day. For a larger clinic with  $C = 36$  and  $\lambda = 48$ , the performance guarantee comes out to be about 78%.

#### 4.4 Deterministic Scheduling Policies

As we have shown in Section 4.2, at least under certain conditions, the optimal static policy has a simple structure and thus is easy to implement. If one is interested in using even a simpler policy, then one possibility is to restrict attention to the class of deterministic policies only. Such a deterministic policy would be achieved if we impose the constraints  $h(S) \in \{0, 1\}$  for all  $S \subset \mathcal{T}$ , in which case, problem (2)-(4) looks for exactly one set of days to offer to patients. It turns out that

a transformation similar to the one in Section 4.1 still holds even when we have the constraints  $h(S) \in \{0, 1\}$  for all  $S \subset \mathcal{T}$  in problem (2)-(4). In particular, if we have the constraints  $h(S) \in \{0, 1\}$  for all  $S \subset \mathcal{T}$  in problem (2)-(4), then we can add the constraints  $x_j/v_j \in \{0, u\}$  for all  $j \in \mathcal{T}$  in problem (5)-(8). Under this modification to problem (5)-(8), problem (5)-(8) becomes equivalent to problem (2)-(4) with the additional constraints that  $h(S) \in \{0, 1\}$  for all  $S \subset \mathcal{T}$ . We record this result in the next proposition. The proof follows from an argument similar to that of Proposition 1 and we defer it to the appendix.

**Proposition 6.** *Consider problem (2)-(4) with the additional constraints  $h(S) \in \{0, 1\}$  for all  $S \subset \mathcal{T}$  and problem (5)-(8) with the additional constraints  $x_j/v_j \in \{0, u\}$  for all  $j \in \mathcal{T}$ . These two problems are equivalent in the sense that given an optimal solution to one problem, we can generate a feasible solution to the other one providing the same objective value.*

In short, to find one set of days to offer to customers, we can solve problem (5)-(8) with the additional constraints  $x_j/v_j \in \{0, u\}$  for all  $j \in \mathcal{T}$ . When we have the constraints  $x_j/v_j \in \{0, u\}$  for all  $j \in \mathcal{T}$  in problem (5)-(8), the algorithmic strategies that we can use to solve this problem closely mirror those described at the end of Section 4.1. In particular, since the nonlinearity in the objective function of problem (5)-(8) involves the scalar function  $F(\alpha) = \mathbb{E}\{[\text{Pois}(\alpha) - C]^+\}$ , it is straightforward to construct accurate piecewise linear approximations to this scalar function. In this case, problem (5)-(8) with the additional constraints that  $x_j/v_j \in \{0, u\}$  for all  $j \in \mathcal{T}$  can be formulated as a mixed integer linear program. Another option is to use the dynamic program in (10)-(11). In particular, if we impose the constraint  $x_j/v_j \in \{0, u\}$  in each decision epoch, then this dynamic program solves problem (5)-(8) for a fixed value of  $u$  and with the additional constraints  $x_j/v_j \in \{0, u\}$  for all  $j \in \mathcal{T}$ .

It is important to emphasize that the reformulation in Section 4.1 is still instrumental to solving problem (2)-(4) under the additional constraints  $h(S) \in \{0, 1\}$  for all  $S \subset \mathcal{T}$ . Without this reformulation, problem (2)-(4) has  $2^{|\mathcal{T}|}$  decision variables, whereas our reformulation yields a problem with  $|\mathcal{T}|+1$  decision variables, which can either directly be solved as a mixed integer linear program or through dynamic programming.

## 5 Dynamic Model

The solution to the static model in the previous section identifies a fixed set of probabilities with which to offer each subset of days, independently of the currently booked appointments. In other words, the policy from this static model does not take the state of the system into consideration. Clearly, there is potential to improve such a policy if decisions can take into account the current system state information. In this section, we begin by giving a Markov decision process formulation of the appointment scheduling problem. Using this formulation, we develop a dynamic policy which prescribes a set of scheduling decisions as a function of the current state of the system.

## 5.1 Dynamic Program

As mentioned at the end of Section 3, for  $1 \leq i \leq j \leq \tau$ , we can use  $Y_{ij}(t)$  to denote the number of patients that called in  $i$  days ago (on day  $t - i$ ) and scheduled an appointment  $j$  days into the future (for day  $t - i + j$ ) given that we are at the beginning of day  $t$ . We call the vector  $Y(t) = \{Y_{ij}(t) : 1 \leq i \leq j \leq \tau\}$  as the appointment schedule at the beginning of day  $t$ . For a given policy  $\pi$ , let  $h^\pi(Y(t), S)$  denote the probability that the subset  $S$  of days is offered to the patients when the system state is  $Y(t)$  under policy  $\pi$ . In this case, the evolution of the appointment schedule  $Y^\pi(t)$  under policy  $\pi$  is captured by

$$Y_{ij}^\pi(t+1) = \begin{cases} \text{Pois}(\sum_{S \subset \mathcal{T}} \lambda r_{0j} P_j(S) h^\pi(Y^\pi(t), S)) & \text{if } 1 = i \leq j \leq \tau \\ \text{Bin}(Y_{i-1,j}^\pi(t), r_{i-1,j}) & \text{if } 2 \leq i \leq j \leq \tau, \end{cases}$$

where we use  $\text{Bin}(n, p)$  to denote a Binomial random variable with parameters  $n$  and  $p$ . Assuming that we are on day  $t$ , let  $U_i^\pi(t)$  denote the number of patients who called on day  $t - i$  to make an appointment for today and show up for their appointment. As a function of the appointment schedule on day  $t$ , we can characterize  $U_i^\pi(t)$  by

$$U_i^\pi(t) = \begin{cases} \text{Pois}(\sum_{S \subset \mathcal{T}} \lambda s_0 P_0(S) h^\pi(Y^\pi(t), S)) & \text{if } i = 0 \\ \text{Bin}(Y_{ii}^\pi(t), s_i) & \text{if } 1 \leq i \leq \tau. \end{cases}$$

Finally, we let  $Y_{00}^\pi(t)$  be the number of patients who call on day  $t$  and choose to schedule their appointments on the same day, in which case,  $Y_{00}^\pi(t)$  is characterized by

$$Y_{00}^\pi(t) = \text{Pois}\left(\sum_{S \subset \mathcal{T}} \lambda P_0(S) h^\pi(Y^\pi(t), S)\right).$$

Define  $\phi^\pi(y)$  to be the long-run average expected reward under policy  $\pi$  given the initial state  $y = Y^\pi(0)$ , which is to say that

$$\phi^\pi(y) = \lim_{k \rightarrow \infty} \frac{\mathbb{E}\left\{ \sum_{t=0}^k \left( \sum_{i=0}^{\tau} U_i^\pi(t) - \theta \left[ \sum_{i=1}^{\tau} Y_{ii}^\pi(t) + Y_{00}^\pi(t) - C \right]^+ \right) \mid Y^\pi(0) = y \right\}}{k}.$$

A scheduling policy  $\pi^*$  is optimal if it satisfies  $\phi^{\pi^*}(y) = \sup_{\pi} \phi^\pi(y)$  for all  $y$ .

We can try to bound the number of appointments on a given day by a reasonably large number to ensure that there are a finite number of possible values for the appointment schedule  $Y(t)$ . However, even if we bound the number of appointments, the appointment schedule  $Y(t)$  is a high-dimensional vector and the number of possible values for the appointment schedule would get extremely large even for small values for the length of the scheduling horizon  $\tau$ . This precludes us from using conventional dynamic programming algorithms, such as value iteration or policy iteration, to compute the optimal scheduling policy. Therefore, it is of interest to develop computationally efficient approaches to obtain policies that consider the status of the booked appointments. We describe one such approach in the next section.

## 5.2 Policy Improvement Heuristic

In this section, we develop a dynamic policy, which makes its decisions by taking the current appointment schedule into consideration. The idea behind our dynamic policy is to start with a static policy that ignores the current appointment schedule and apply the single step of the policy improvement algorithm. It is important to note that we do not simply call for applying the policy improvement step and computing the improved policy for every possible state of the system. Rather, we carry out policy improvement only for the states that we run into as the system evolves over time, and furthermore as it turns out, computing the decisions made by our policy improvement heuristic for a particular state of the system requires solving a mathematical problem similar to the one solved for our static model.

Our policy improvement heuristic is developed by building on a static policy mainly because we can, in this case, identify closed-form expressions for the value functions in the policy-improvement step, as described below. Ideally, when implementing the policy improvement heuristic, one uses the optimal static policy obtained by solving problem (2)-(4), but if there are reasonable alternative static policies, then one can simply use them instead. Our derivation of the policy improvement heuristic does not depend on whether or not the initially chosen static policy is obtained through problem (2)-(4) and we keep our presentation general.

We let  $h(S)$  for  $S \subseteq \mathcal{T}$  be the probability that the subset  $S$  of days is offered to each appointment request under a static policy, which, as mentioned above, may or may not have been obtained through problem (2)-(4). Given that we are at the beginning of a particular day, the current appointment schedule is given by  $y = \{y_{ij} : 1 \leq i \leq j \leq \tau\}$  before we observe the appointment requests on the current day. Similar to  $Y_{ij}(t)$ , the component  $y_{ij}$  of the appointment schedule  $y$  represents the number of appointments made  $i$  days ago for  $j$  days into the future. Consider a policy  $\pi$  that makes the subset  $S$  of days available with probability  $f(S)$  on the current day and switches to the static probabilities  $h = \{h(S) : S \subset \mathcal{T}\}$  from tomorrow on.

Given that we start with the appointment schedule  $y$ , define  $\Delta(y, f, h)$  to be the difference in the long-run total expected rewards obtained by following policy  $\pi$  rather than the static policy that uses the subset offer probabilities  $h$  all along. To conduct one step of the policy improvement algorithm, we need to maximize  $\Delta(y, f, h)$  with respect to  $f = \{f(S) : S \subset \mathcal{T}\}$ . Since policy  $\pi$  and the static policy that uses  $h$  make identical decisions after the current day and appointments cannot be scheduled beyond  $\tau$  days into the future, the appointment schedules under the two policies are stochastically identical beyond the first  $\tau + 1$  days. Thus, we can write  $\Delta(y, f, h) = Q_\pi(y, f, h) - Q_{SP}(y, h)$ , where  $Q_\pi(y, f, h)$  and  $Q_{SP}(y, h)$  are respectively the total expected rewards accumulated over the next  $\tau + 1$  days under policies  $\pi$  and the static policy.

When determining  $f$  that maximizes  $\Delta(y, f, h)$ , the function  $Q_{SP}(y, h)$  is simply a constant. Thus, our objective is equivalent to maximizing  $Q_\pi(y, f, h)$ . We proceed to deriving an expression for  $Q_\pi(y, f, h)$ . For  $1 \leq i \leq j \leq \tau$ , we let  $V_{ij}(y)$  denote the number of patients

who called for appointments  $i$  days ago and will not cancel by the morning of their appointment, which is  $j - i$  days from today. We let  $W_{ij}(y)$  denote the number of these patients who show up for their appointments. Similarly, for  $0 \leq j \leq \tau$ , we let  $\bar{V}_j(f)$  denote the number of patients who call to make an appointment today, schedule an appointment for day  $j$  in the scheduling horizon and do not cancel by the morning of their appointment. We let  $\bar{W}_j(f)$  denote the number of these patients who show up for their appointments. Finally, for  $1 \leq k \leq j \leq \tau$ , we define  $\hat{V}_{kj}(h)$  to be the number of patients who will call to make an appointment on day  $k$  in the scheduling horizon and will not have canceled their appointment by the morning of their appointment, which is on day  $j$  of the scheduling horizon. We define  $\hat{W}_{kj}(h)$  to be the number of these patients who will show up for their appointments. Under the assumption that the cancellation and no-show behavior of the patients are independent of each other and the appointment schedule, we characterize the six random variables defined above as  $V_{ij}(y) = \text{Bin}(y_{ij}, \tilde{r}_{ij})$ ,  $W_{ij}(y) = \text{Bin}(y_{ij}, \tilde{s}_{ij})$ ,

$$\begin{aligned} \bar{V}_j(f) &= \text{Pois}\left(\sum_{S \subset \mathcal{T}} \lambda \bar{r}_j P_j(S) f(S)\right), & \bar{W}_j(f) &= \text{Pois}\left(\sum_{S \subset \mathcal{T}} \lambda \bar{s}_j P_j(S) f(S)\right), \\ \hat{V}_{kj}(h) &= \text{Pois}\left(\sum_{S \subset \mathcal{T}} \lambda \bar{r}_{j-k} P_{j-k}(S) h(S)\right), & \hat{W}_{kj}(h) &= \text{Pois}\left(\sum_{S \subset \mathcal{T}} \lambda \bar{s}_{j-k} P_{j-k}(S) h(S)\right), \end{aligned}$$

where  $\tilde{r}_{ij} = r_{ij} r_{i+1,j} \dots r_{jj}$  is the probability that a patient who scheduled her appointment  $i$  days ago will not have canceled her appointment by the morning of the day of her appointment, which is  $j - i$  days from today and  $\tilde{s}_{ij} = s_j \tilde{r}_{ij}$  is the probability that a patient who scheduled her appointment  $i$  days ago will show up for her appointment, which is  $j - i$  days from today.

Using the random variables defined above, we can capture the reward obtained by the policy that uses the subset offer probabilities  $f$  today and switches to the static probabilities  $h$  from tomorrow on. In particular, letting  $q_j(y, f, h)$  denote the total expected reward obtained by this policy  $j$  days from today, we have

$$\begin{aligned} q_j(y, f, h) &= \mathbb{E}\left\{ \sum_{k=1}^{\tau-j} W_{k,k+j}(y) + \bar{W}_j(f) + \sum_{k=1}^j \hat{W}_{kj}(h) \right. \\ &\quad \left. - \theta \left[ \sum_{k=1}^{\tau-j} V_{k,k+j}(y) + \bar{V}_j(f) + \sum_{k=1}^j \hat{V}_{kj}(h) - C \right]^+ \right\}, \end{aligned}$$

where we set  $\sum_{k=1}^j \hat{W}_{kj}(h) = 0$  and  $\sum_{k=1}^j \hat{V}_{kj}(h) = 0$  when we have  $j = 0$ . Thus, it follows that  $Q_\pi(y, f, h) = \sum_{j=0}^{\tau} q_j(y, f, h)$  and we can implement the policy improvement heuristic by maximizing  $\sum_{j \in \mathcal{T}} q_j(y, f, h)$  subject to the constraint that  $\sum_{S \subset \mathcal{T}} f(S) = 1$ . The optimal solution  $f^* = \{f^*(S) : S \subset \mathcal{T}\}$  to the last problem yields the probability with which each subset of days should be offered on the current day. We note that since  $q_j(y, f, h)$  depends on the current state  $y$  of the appointment schedule, the offer probabilities  $f$  we obtain in this fashion also depend on the current appointment schedule. Finally, we observe that solving the last optimization problem simplifies when we approximate the binomial random variables  $V_{ij}(y)$  and  $W_{ij}(y)$  for  $1 \leq i \leq j \leq \tau$  with Poisson random variables with their corresponding means. In that case, one can check that

the last optimization problem is structurally the same as problem (2)-(4) and we can use arguments similar to those in Section 4 to obtain the decisions made by the policy improvement heuristic.

## 6 Computational Results

In this section, we begin by giving the findings from our discrete choice experiment conducted to elicit patient preferences and to obtain data for our computational work. We proceed to describe our benchmark policies and experimental setup. We conclude with our computational results.

### 6.1 Discrete Choice Experiment

Our experiment took place in the Farrell Community Health Center, a large urban primary care health center in New York City. This center serves about 26,000 patient visits per year. A self-report survey instrument was created in both English and Spanish to collect data (see a snapshot of the survey in the appendix). Adult patients waiting were approached and informed about this study. After verbal consent was obtained, each patient was given a questionnaire in her native language, which described a hypothetical health condition entailing a need for a medical appointment and listed a randomly generated set of two possible choices of appointment days (for example, the same day and 2 days from now) as well as the option of seeking care elsewhere. The patient was asked to consider her current work/life schedule and mark the appointment choice she preferred. Since the nature of the health problem is likely to influence the preference of the patient, each patient was asked to make a choice for two different problems, one that can be characterized as “ambiguous” and the other as “urgent.” When describing the health conditions, we used the almost exact same wording Cheraghi-Sohi et al. (2008) used. In particular, a patient faces an ambiguous condition when s/he has been feeling tired and irritable and has had difficulty sleeping over the past few months, and the patient has tried several things to remedy this but is not feeling any better; whereas urgent condition is characterized by severe chest pains when taking deep breath in.

Data were collected from December 2011 to February 2012. To reduce sampling bias, we conducted surveys across different times of day and different days of week. Overall, 240 patients were eligible to participate, out of which 161 of them agreed to participate, yielding a response rate of 67%. Each patient was provided appointment slots over the next five days including the current day, corresponding to a scheduling horizon length of  $\tau = 5$ . We separated the data under the ambiguous health condition and the urgent health condition. Excluding the participants who provided missing data, the sample size for the ambiguous condition turned out to be 156 patients and that for the urgent condition turned out to be 158 patients. For each health condition, we let  $\mathcal{K}$  be the set of patients who picked an appointment day within the options offered and  $\bar{\mathcal{K}}$  be the set of patients who picked the option of seeking care elsewhere. For each patient  $k \in \mathcal{K}$ , the data provide the set of appointment days  $S^k$  made available to this patient and the appointment day  $j^k$  chosen by this patient, whereas for each patient  $k \in \bar{\mathcal{K}}$ , the data provide the set of appointment

Choice Set	Choice Prob. under $v^A$			Choice Prob. under $v^U$		
	1st choice	2nd choice	3rd choice	1st choice	2nd choice	3rd choice
{0, 2, elsewhere}	0.51	0.39	0.10	0.51	0.25	0.23
{1, 3, elsewhere}	0.45	0.45	0.10	0.59	0.10	0.30
{2, 4, elsewhere}	0.52	0.35	0.13	0.39	0.25	0.36
{3, 5, elsewhere}	0.62	0.25	0.13	0.22	0.13	0.66

Table 1: Patient choice probabilities under sample scenarios.

days  $S^k$  made available to the patient. As a function of the parameters  $v = (v_0, v_1, \dots, v_\tau)$  of the choice model, the likelihood function is given by

$$L(v) = \left( \prod_{k \in \mathcal{K}} \frac{v_{j^k}}{1 + \sum_{j \in S^k} v_j} \right) \left( \prod_{k \in \bar{\mathcal{K}}} \frac{1}{1 + \sum_{j \in S^k} v_j} \right)$$

We maximized the likelihood function above by using a nonlinear programming package to obtain the estimator for the preference weights  $v$ .

We use  $v^A$  and  $v^U$  to respectively denote the preference weights estimated by using the data obtained under the ambiguous and urgent health conditions. Our results yield the estimates  $v^A = (4.99, 4.66, 3.84, 4.58, 2.54, 1.83)$  and  $v^U = (2.19, 1.95, 1.09, 0.33, 0.71, 0.19)$ . These estimates conform with intuition. The preference weights derived from the patients with an ambiguous health condition reveal that these patients are less sensitive to appointment delays. They appear to favor any appointment within the four day window more or less equally. Conversely, the patients with urgent symptoms in general desire faster access. They prefer appointments on the day during which they call or the day after much more strongly than other days. We can also see that while the utilities for the unambiguous condition have a monotonic structure those for the ambiguous condition do not, which might be due to the fact that in the patient population, there could be two different attitudes towards an ambiguous health condition. While one group of patients take it seriously and demand seeing a doctor as soon as possible, the other group could assume a more relaxed wait-and-see attitude and prefer scheduling an appointment a few days later and go see a doctor then if their condition is still present.

Also, we observe that the magnitudes of the preference weights are smaller when patients face an urgent health condition compared to when patients face an ambiguous health condition. Noting that the preference weight of seeking care elsewhere is normalized to one, this observation suggests that patients facing an urgent health condition have a higher tendency to seek care elsewhere (e.g., emergency rooms). These can also be observed in Table 1, which lists patient choice probabilities for a given set of options and under each one of the two health conditions.

## 6.2 Description of the Policies

We compare the performances of our static and dynamic policies through a series of computational experiments and test them against other benchmark policies. We work with a total of five

policies. Two of them are policies developed based on our formulation and analysis in this paper. The other three are either meant to capture or inspired by what some clinics already do in practice. So, they can serve as benchmark policies for the policies that we propose.

**Static Policy (SP):** This policy is obtained by solving the static optimization problem described in Section 4. Specifically, SP solves problem (5)-(8) and transforms the optimal solution to this problem to an optimal solution to problem (2)-(4) by using (9). Using  $h^*$  to denote an optimal solution to problem (2)-(4) obtained in this fashion, SP offers the subset  $S$  of days with probability  $h^*(S)$ .

**Dynamic Policy (DP):** This policy is obtained as described in Section 5. Letting  $h^*$  be the subset offer probabilities obtained by SP, given that the appointment schedule on the current day is given by  $x$ , DP maximizes  $\sum_{j \in \mathcal{T}} q_j(x, f, h^*)$  subject to the constraint that  $\sum_{S \subset \mathcal{T}} f(S) = 1$ . In this case, using  $f^*$  to denote the optimal solution to this problem, on the current day, we offer the subset  $S$  of days with probability  $f^*(S)$ . We note that DP recomputes the subset offer probabilities at the beginning of each day by using the current appointment schedule, but the value of  $h^*$  stays constant throughout.

**Controlled Open Access Policy (CO):** This is actually a capacity controlled implementation of open access. CO makes only the current day available for appointments, but to balance the demand with the available capacity, it also allows the possibility of offering no appointment days to an arriving patient. To achieve this, we solve problem (2)-(4) only by considering the sets  $S = \{0\}$  and  $S = \emptyset$ , in which case, problem (2)-(4) becomes a convex optimization problem with two decision variables  $h(\{0\})$  and  $h(\emptyset)$ . Letting  $(h^*(\{0\}), h^*(\emptyset))$  be the optimal solution we obtain, CO offers the current day for appointments with probability  $h^*(\{0\})$  and does not offer any appointment days with probability  $h^*(\emptyset)$ .

**All or Nothing Policy (AN):** This policy either makes all days in the scheduling horizon available for appointments or does not offer an appointment day at all. The probabilities of offering these two options is obtained by using a similar approach to CO. In particular, we solve problem (2)-(4) only by considering the sets  $S = \{0, 1, \dots, \tau\}$  and  $S = \emptyset$ .

**Next- $n$  Policy:** This policy is a smarter, dynamic alternative to CO and AN, which both describe static policies that do not take system state into account. For each day in the scheduling horizon, let  $\mathcal{Q}$  be the set of days, ordered chronologically, for which the number of scheduled appointments is below the regular capacity  $C$ . Under the Next- $n$  policy, we open the first  $\min\{n, |\mathcal{Q}|\}$  days for patients to choose from at the beginning of the day. Note that  $\mathcal{Q}$  depends on the current system state and so is the offer set. This policy essentially tries to schedule appointments as early as possible while making sure that the regular capacity is not exceeded. Our interactions with primary care practices suggest that this policy is reasonably close to how some of the clinics which actively work for high capacity utilization schedule appointments.



There is one issue that needs to be clarified regarding the implementation of the Next- $n$  policy. Clearly, one needs to pick a value for  $n$ , the number of days that will be offered to the patients. However, our computational experiments quickly revealed that there is no one fixed value of  $n$  that works uniformly well in all the problem parameter settings. What works well in one particular scenario may perform badly in others. This poses a challenge for clinics since this would require some sort of optimization to determine the “optimal” value of  $n$ . It also means that a fair comparison of this policy with the other policies is somewhat challenging. The reason is that if one were to set  $n$  to a fixed value, say 3, and test this policy across all scenarios, that would clearly put this policy in a disadvantaged position in some of these scenarios. To avoid this problem, for each scenario, we chose  $n$  “optimally” by simulating the appointment system under each possible value of  $n$  and picking the one under which the expected profit is the largest. In the following, we name this policy Next- $n^*$  policy to make it clear that the choice of  $n$  is not arbitrary. However, it should be noted that by making this assumption we grant a significant advantage to this policy as it assumes a sophisticated clinic, which is capable of making this determination in practice. Nevertheless, with this approach, we at least know that the performance of Next- $n^*$  provides an upper bound on what a clinic can achieve with a Next- $n$  type policy.

### 6.3 Experimental Setup

In all of our computational experiments, we set the daily arrival rate  $\lambda = 16$ . We consider three capacity levels  $C = 0.9\lambda, 1.2\lambda$  and  $1.5\lambda$ , representing tight, medium and ample capacity situations, respectively. These specific capacity levels were chosen so that we cover the same range of capacity-demand ratios considered in Green et al. (2008). We assume that the probability of retaining an appointment  $\bar{r}_j = 1 - 0.04j$  if the appointment delay is  $j$  days, and set  $s_j = 1$  so that all patients who have not canceled by the day of their appointments show up.

Recalling that the revenue of serving a patient is normalized to one, we vary the overtime cost  $\theta$  over three different values, 1.25, 1.5 and 1.75. We use two different values for the length of the scheduling horizon  $\tau$ . In one set of test problems, we set  $\tau = 5$ . For these test problems, we work with two different preference weights,  $v^A$  and  $v^U$ , which are estimated under the ambiguous and urgent health conditions in Section 6.1. In another set of test problems, we set  $\tau = 15$ . For these test problems, we use two different sets of preference weights that we generate ourselves. The first one is  $v^I = (1, 1, \dots, 1)$ , corresponding to the case where a patient is indifferent over the days offered to her. The second set of preference weights is  $v^D = \{1.5, 1.4, \dots, 0.1\}$ , capturing the situation where patients prefer later appointments less. Thus, our experimental setup varies the capacity of the clinic  $C$  over  $\{0.9\lambda, 1.2\lambda, 1.5\lambda\}$ , the overtime cost  $\theta$  over  $\{1.25, 1.5, 1.75\}$  and the preference weights  $v$  over  $\{v^A, v^U, v^I, v^D\}$ , which amounts to a total of 36 test problems.

We use a simulation study to test the performances of the five policies described in the previous section. On each day, we sample the number of appointment requests that arrive, and determine the offer set under the policy in consideration. The patient making the appointment request

chooses among the offered days according to the multinomial logit model or leaves without booking an appointment. Once we collect the booked appointments on the current day and update the appointment schedule, we sample cancellations and no-shows. Based on the no-shows, we account for the revenue and the overtime cost on the current day, after which, we move on to the next day. We simulate each one of the five policies for 135 days and for 100 replications. The first 45 days of each replication is used as a warm up period, during which we do not collect any cost information.

## 6.4 Computational Results

Our computational results are given in Tables 2 and 3. Table 2 is for the test problems with  $\tau = 5$ , having the preference weights  $v^A$  and  $v^U$ . Table 3 is for the test problems with  $\tau = 15$ , having the preference weights  $v^I$  and  $v^D$ . The first three columns in each table indicate the values for  $v$ ,  $C$ , and  $\theta$ , characterizing each scenario. The next five columns show the expected net reward per day obtained under the five policies considered. Along with the expected net reward for Next- $n^*$ , we report the best value of  $n$  that we find for this policy. As DP is usually the best performing policy, we also give the percent gap between the expected daily net reward under DP and that under the remaining four policies in the last four columns. DP performances marked with  $\times$  and  $*$  indicate scenarios where DP performance is statistically better than SP and Next- $n^*$ , respectively at 0.05 level of significance. For a majority of test problems, DP is significantly better than AN or CO. So, we do not statistically compare the performance of AN or CO with that of DP.

From the two tables, we observe that the performances of AN and CO are very much scenario dependent. Each has its own set of scenarios where they perform reasonably well, but they can perform quite poorly in some cases. The performance gap with DP can exceed 34% for AN and 15% for CO. AN and CO can be seen as policies at the two ends of the policy spectrum with AN offering all days (or none) and CO offering only the current day (or none). They are also both static policies. Thus, their uneven and sometimes inferior performance can be attributed to the fact that they are not responsive to changes in the appointment schedule and they each focus on only one of the two dimensions of the problem, with AN focusing on being more accommodating to appointment requests by making a large window of days available and thereby avoiding losing patient demand, whereas CO focusing on preventing no-shows and late cancellations and as a result improving operational efficiency.

The performance of SP is either superior or equivalent to both AN and CO across all scenarios and the performance gap with DP never exceeds 6%. Thus, even if dynamic policies are not considered, clinics can gain quite a bit by choosing the offer sets carefully based on the available capacity with respect to demand and patient preferences. However, we also observe that there are still significant benefits to be gained by making decisions dynamically. As shown in the two tables, in 18 out of the 36 scenarios, DP performs statistically better than SP. As indicated above, the improvements can reach up to 6%. More importantly, the performance improvement from DP is largest when the capacity is tight with respect to the overall patient demand, as we can observe in

Scenario Parameters			Expected Net Reward Per Day					% Gap with DP			
$v$	$C$	$\theta$	DP	SP	AN	CO	Next- $n^*$	SP	AN	CO	Next- $n^*$
$v^A$	$0.9\lambda$	1.25	$\times^*$ 12.52	12.06	11.65	11.97	12.39 (5)	3.67	6.95	4.39	1.04
$v^A$	$0.9\lambda$	1.5	$\times^*$ 12.32	11.74	11.41	11.61	12.16 (5)	4.71	7.39	5.76	1.65
$v^A$	$0.9\lambda$	1.75	$\times^*$ 12.11	11.39	11.18	11.28	11.91 (6)	5.95	7.68	6.85	1.65
$v^A$	$1.2\lambda$	1.25	$\times$ 13.67	13.57	12.72	13.34	13.57 (3)	0.73	6.95	2.41	0.73
$v^A$	$1.2\lambda$	1.5	$\times$ 13.61	13.50	12.65	13.24	13.50 (3)	0.81	7.05	2.72	0.81
$v^A$	$1.2\lambda$	1.75	$\times$ 13.57	13.45	12.61	13.22	13.43 (4)	0.88	7.07	2.58	1.03
$v^A$	$1.5\lambda$	1.25	13.77	13.79	12.81	13.47	13.76 (3)	-0.15	6.97	2.18	0.073
$v^A$	$1.5\lambda$	1.5	13.73	13.72	12.77	13.44	13.75 (3)	0.073	7.00	2.11	-0.15
$v^A$	$1.5\lambda$	1.75	13.72	13.70	12.75	13.45	13.73 (3)	0.15	7.07	1.97	-0.073
Average								1.87	7.13	3.44	0.71
$v^U$	$0.9\lambda$	1.25	$\times^*$ 11.80	11.52	10.20	11.50	11.64 (6)	2.37	13.56	2.54	1.36
$v^U$	$0.9\lambda$	1.5	$\times^*$ 11.70	11.32	10.14	11.36	11.47 (5)	3.24	13.33	2.91	1.97
$v^U$	$0.9\lambda$	1.75	$\times^*$ 11.55	11.10	10.05	11.12	11.25 (6)	3.90	12.99	3.72	2.60
$v^U$	$1.2\lambda$	1.25	12.44	12.43	10.49	12.41	12.43 (6)	0.080	15.68	0.24	0.080
$v^U$	$1.2\lambda$	1.5	12.44	12.47	10.52	12.44	12.48 (6)	-0.24	15.43	0	-0.32
$v^U$	$1.2\lambda$	1.75	12.34	12.33	10.44	12.44	12.37 (5)	0.080	15.40	0	0.08
$v^U$	$1.5\lambda$	1.25	12.52	12.52	10.53	12.52	12.51 (5)	0	15.89	0	0.08
$v^U$	$1.5\lambda$	1.5	12.57	12.57	10.56	12.56	12.58 (5)	0	15.99	0.080	-0.08
$v^U$	$1.5\lambda$	1.75	12.45	12.48	10.47	12.41	12.47 (5)	-0.24	15.90	0.32	-0.16
Average								1.02	14.91	1.09	0.5

Table 2: Performances of the four policies on the test problems with  $v = v^A$  or  $v = v^U$ . DP performances marked with  $\times$  and  $*$  indicate scenarios where DP performance is statistically better than SP and Next- $n^*$ , respectively at 0.05 level of significance. In the Next- $n^*$  column, the number in parenthesis indicates the value of  $n$  for which the Next- $n$  policy has the best performance.

scenarios with  $C = 0.9\lambda$ .

The most serious competitor to DP is the Next- $n^*$  policy with a performance that is close to that of DP across all scenarios. In 9 out of the 36 scenarios tested, DP still performs statistically better than Next- $n^*$  at 0.05 level of significance and in the remaining 27 scenarios, the performances of the two policies are statistically indistinguishable. Furthermore, most of the 9 scenarios where DP is statistically superior are the scenarios where the capacity is tight with respect to the patient demand. These are the scenarios where DP beats SP with the largest margins. This means that DP has the potential to make a difference for clinics with heavy loads, which is important since capacity management for clinics experiencing low patient demand is typically less of a challenge. Second, as the two tables indicate, the value of  $n^*$  changes from scenario to scenario, ranging from 2 to 7. Thus, the performance of Next- $n^*$  essentially provides an upper bound on what clinics can possibly achieve with a simple Next- $n$  type policy that does not necessarily use the best value of  $n$ . In addition, our computational experiments indicated that if  $n$  for the Next- $n$  policy is not chosen optimally, then the performance of Next- $n$  can suffer significantly. This means that, perhaps not surprisingly, there is not a fixed value of  $n$  for which Next- $n$  performs universally well. Clinics need to make a numerical search to determine the right choice for  $n$ . On the other hand, the implementation of DP is more complex than that of Next- $n$  as DP requires solving an optimization problem in an online fashion. Nevertheless, this is not a serious issue as we have found out that

Scenario Parameters			Expected Net Reward Per Day					% Gap with DP			
$v$	$C$	$\theta$	DP	SP	AN	CO	Next- $n^*$	SP	AN	CO	Next- $n^*$
$v^I$	$0.9\lambda$	1.25	$\times$ 11.29	11.04	7.62	9.91	11.31 (6)	2.21	32.51	12.22	-0.18
$v^I$	$0.9\lambda$	1.5	$\times$ 11.24	10.96	7.59	9.84	11.22 (7)	2.49	32.47	12.46	0.18
$v^I$	$0.9\lambda$	1.75	$\times$ 11.22	10.8	7.65	9.83	11.17 (7)	3.47	31.82	12.39	0.45
$v^I$	$1.2\lambda$	1.25	11.66	11.68	7.66	10.15	11.72 (6)	-0.17	34.31	12.95	-0.51
$v^I$	$1.2\lambda$	1.5	11.71	11.71	7.65	10.14	11.74 (6)	0	34.67	13.41	-0.26
$v^I$	$1.2\lambda$	1.75	11.72	11.76	7.7	10.23	11.77 (6)	-0.34	34.3	12.71	-0.43
$v^I$	$1.5\lambda$	1.25	11.72	11.75	7.63	10.16	11.75 (6)	-0.26	34.90	13.31	-0.26
$v^I$	$1.5\lambda$	1.5	11.77	11.77	7.68	10.14	11.73 (6)	0	34.75	13.85	0.34
$v^I$	$1.5\lambda$	1.75	11.81	11.79	7.69	10.2	11.83 (6)	0.17	34.89	13.64	-0.17
Average								0.87	33.84	12.99	-0.093
$v^D$	$0.9\lambda$	1.25	$\times$ 12.66	12.29	12.27	11.37	12.59 (5)	2.92	3.08	10.19	0.55
$v^D$	$0.9\lambda$	1.5	$\times$ 12.32	11.83	11.81	11.13	12.30 (5)	3.98	4.14	9.66	0.16
$v^D$	$0.9\lambda$	1.75	$\times^*$ 12.14	11.42	11.15	11.03	11.95 (5)	5.93	8.15	9.14	1.57
$v^D$	$1.2\lambda$	1.25	$\times$ 14.35	14.25	14.12	12.24	14.24 (2)	0.70	1.60	14.70	0.77
$v^D$	$1.2\lambda$	1.5	$\times^*$ 14.25	14.11	13.98	12.13	14.11 (2)	0.98	1.89	14.88	0.98
$v^D$	$1.2\lambda$	1.75	$\times^*$ 14.29	14.10	13.99	12.17	14.10 (2)	1.33	2.10	14.84	1.33
$v^D$	$1.5\lambda$	1.25	14.54	14.55	14.39	12.26	14.56 (2)	-0.069	1.11	15.68	-0.14
$v^D$	$1.5\lambda$	1.5	14.45	14.46	14.29	12.2	14.44 (2)	-0.069	1.11	15.57	0.069
$v^D$	$1.5\lambda$	1.75	14.53	14.54	14.39	12.31	14.54 (2)	-0.069	0.96	15.28	-0.069
Average								1.73	2.67	13.33	0.58

Table 3: Performances of the five policies on the test problems with  $v = v^I$  or  $v = v^D$ . DP performances marked with  $\times$  and  $*$  indicate scenarios where DP performance is statistically better than SP and Next- $n^*$ , respectively at 0.05 level of significance. In the Next- $n^*$  column, the number in parenthesis indicates the value of  $n$  for which the Next- $n$  policy has the best performance.

for the test problems we considered, the decision made by DP can be computed in less than 0.01 seconds on a standard desktop computer.

## 7 Conclusion

This paper develops new methods, which can be used in making appointment scheduling decisions while taking into account patients' no-show and cancellation behavior and their preferences for different appointment days. Even though management of healthcare capacity has always been an important problem, a number of societal and technological developments over the last decade as well as more recent changes in the U.S. healthcare system through the Patient Protection and Affordable Care Act, have made the decision problem considered here particularly timely. First, in the U.S., due to the Affordable Care Act, it is largely expected that physicians will see significant increase in their patient demand. This, combined with the fact that longer appointment delays lead to higher no-show and cancellation rates, implies that the management of the valuable physician time through appointment scheduling will be increasingly important. Second, somewhat paradoxically, physicians may have increasingly more difficulty in keeping their own patients since thanks to web sites like Zocdoc and other online appointment scheduling platforms, patients will find it easier to make alternative arrangements when their own physician is not available. In other words, even

as the patient demand increases, the competition among the providers will intensify. Therefore, providers will need to pay attention to their patients' preferences regarding when they would like to be seen and somehow balance that with their operational concerns. Third, it will also be reasonable to expect that with time more and more appointments will be scheduled through online scheduling platforms, which not only will make it much easier to collect data to estimate various model parameters and learn individual patient characteristics such as their no-show behavior, but will also provide a more natural environment for application of the type of methods we propose in this paper.

Despite the fact that appointment scheduling has been widely studied in operations research literature, with a few exceptions, patient preferences have been assumed away. This paper is directed to fill this gap. Specifically, the models we develop incorporate patient no-shows and cancellations that are dependent on the appointment delays, embed patient preferences among different appointment days, and take the current appointment schedule into consideration when making decisions. Arguably, one of the main reasons patient preferences have been largely ignored is the analytical and computational difficulty (which is partially a consequence of the fact that models that incorporate preferences involve a large number of decision variables) to the extent that solving the problems for realistic instances could be practically not possible. Clearly, the problem becomes even bigger in the dynamic scheduling case. One of the important contributions of this paper is to propose a formulation along with a set of analytical results which help bypass the computational challenges such preference models typically present.

Our computational study not only demonstrates that the policies we propose perform well but also provides a number of useful practical insights. We find that choosing the offer sets in consideration of patient preferences help quite a bit in some cases significantly even when decisions are not made dynamically. But the improvements are even higher when the updated appointment schedule is also taken into account. Thus, clinics are likely to see clear benefits when making decisions based on both. In addition, we find that the next- $n$  policy, a scheduling policy of a simple form that allows patients to choose appointment slots in the next  $n$  days with regular capacity left, may perform quite well if  $n$  is chosen appropriately.

One limitation of our formulation is that it does not explicitly consider the dynamics of appointment slots within a particular day. The simultaneous consideration of appointment day and time of appointment within the day is clearly an important problem but it is also a very difficult problem, which is possibly the reason why not many have studied it even under the assumption that patients do not have any preferences. Despite the fact that this paper also shies away from this complexity, it brings us one step closer to coming up with a proper way of approaching it. In fact, our formulation can easily be extended in a way that would allow patient preferences to depend on time of the day and appointments to be scheduled not on a day only but at a specific time on a day, and the same analysis would go through as long as one is willing to assume that the cost on a given day only depends on the total number of appointments (see our discussion in

the appendix). Admittedly, however, this last assumption can be problematic because simply knowing the total number of appointments on a given day may not be sufficient to come up with a reasonable estimate on the cost in many cases. Nevertheless, this approach appears to have significant potential in “solving” the simultaneous optimization problem and to that end two directions seem worthwhile to explore: are there any “good” within-day scheduling policies under which the total number of appointments is mostly sufficient to estimate the expected daily cost? If the total number of appointments is not sufficient to reliably estimate the expected cost, what minimal set of information is needed to make the estimation? More research on these two questions will help flesh out the feasibility of extending the framework developed in this paper to develop policies that also take into account within-day dynamics of the appointment scheduling problem.

## Acknowledgment

We thank the Editor, the Area Editor, the Associate Editor and two anonymous referees for their constructive feedback, which significantly improved our work. The work of the first and third authors is supported in part by the National Science Foundation grant CMMI 0969113. The work of the fourth author is supported in part by the National Science Foundation grant CMMI 1234212.

## References

- Ayvaz, N. and Huh, W. (2010), ‘Allocation of hospital capacity to multiple types of patients’, *Journal of Revenue & Pricing Management* **9**(5), 386–398.
- Bean, A. G. and Talaga, J. (1995), ‘Predicting appointment breaking’, *Journal of Health Care Marketing* **15**(1), 29–34.
- Berwick, D., Nolan, T. and Whittington, J. (2008), ‘The triple aim: Care, health, and cost’, *Health Affairs* **27**(3), 759–769.
- Bowser, D., Utz, S., Glick, D. and Harmon, R. (2010), ‘A systematic review of the relationship of diabetes mellitus, depression, and missed appointments in a low-income uninsured population’, *Archives of psychiatric nursing* **24**(5), 317–329.
- Bront, J. J. M., Diaz, I. M. and Vulcano, G. (2009), ‘A column generation algorithm for choice-based network revenue management’, *Operations Research* **57**(3), 769–784.
- Cayirli, T. and Veral, E. (2003), ‘Outpatient scheduling in health care: A review of literature’, *Production and Operations Management* **12**(4), 519–549.
- Cheraghi-Sohi, S., Hole, A., Mead, N., McDonald, R., Whalley, D., Bower, P. and Roland, M. (2008), ‘What patients want from primary care consultations: A discrete choice experiment to identify patients priorities’, *The Annals of Family Medicine* **6**(2), 107–115.
- Dreiherr, J., Froimovici, M., Bibi, Y., Vardy, D., Cicurel, A. and Cohen, A. (2008), ‘Nonattendance in obstetrics and gynecology patients’, *Gynecologic and Obstetric Investigation* **66**(1), 40–43.
- Gallego, G., Iyengar, G., Phillips, R. and Dubey, A. (2004), Managing flexible products on a network, Computational Optimization Research Center Technical Report TR-2004-01, Columbia University.

- Gallego, G., Ratliff, R. and Shebalov, S. (2011), A general attraction model and an efficient formulation for the network revenue management problem, Technical report, Columbia University, New York, NY.
- Gallego, G. and van Ryzin, G. (1994), ‘Optimal dynamic pricing of inventories with stochastic demand over finite horizons’, *Management Science* **40**(8), 999–1020.
- Gallego, G. and van Ryzin, G. (1997), ‘A multiproduct dynamic pricing problem and its applications to yield management’, *Operations Research* **45**(1), 24–41.
- Gallucci, G., Swartz, W. and Hackerman, F. (2005), ‘Brief reports: Impact of the wait for an initial appointment on the rate of kept appointments at a mental health center’, *Psychiatric Services* **56**(3), 344–346.
- Gerard, K., Salisbury, C., Street, D., Pope, C. and Baxter, H. (2008), ‘Is fast access to general practice all that should matter? A discrete choice experiment of patients’ preferences’, *Journal of Health Services Research & Policy* **13**(suppl 2), 3–10.
- Gerchak, Y., Gupta, D. and Henig, M. (1996), ‘Reservation planning for elective surgery under uncertain demand for emergency surgery’, *Management Sci.* **42**(3), 321–334.
- Green, J., McDowall, Z. and Potts, H. (2008), ‘Does choose & book fail to deliver the expected choice to patients? a survey of patients’ experience of outpatient appointment booking’, *BMC Medical Informatics and Decision Making* **8**(1), 36.
- Green, L. V. and Savin, S. (2008), ‘Reducing delays for medical appointments: A queueing approach’, *Operations Research* **56**(6), 1526–1538.
- Grunebaum, M., Lubner, P., Callahan, M., Leon, A., Olfson, M. and Portera, L. (1996), ‘Predictors of missed appointments for psychiatric consultations in a primary care clinic’, *Psychiatric Services* **47**(8), 848–852.
- Gupta, D. and Denton, B. (2008), ‘Appointment scheduling in health care: Challenges and opportunities’, *IIE Transactions* **40**(9), 800–819.
- Gupta, D. and Wang, L. (2008), ‘Revenue management for a primary-care clinic in the presence of patient choice’, *Operations Research* **56**(3), 576–592.
- Gupta, D. and Wang, W. (2011), Patient appointments in ambulatory care, in R. W. Hall, ed., ‘Handbook of Healthcare System Scheduling: Delivering Care When and Where It is Needed’, Springer, New York.
- Hole, A. (2008), ‘Modelling heterogeneity in patients’ preferences for the attributes of a general practitioner appointment’, *Journal of Health Economics* **27**(4), 1078–1094.
- Huh, T., Liu, N. and Truong, V. (2013), ‘Multiresource allocation scheduling in dynamic environments’, *Manufacturing & Service Operations Management* **15**(2), 280–291.
- Institute for Healthcare Improvement (2012), ‘The IHI triple aim’. Retrieved May 8, 2012, from <http://www.ihl.org/offerings/Initiatives/TripleAim/Pages/default.aspx>.
- Kök, A., Fisher, M. and Vaidyanathan, R. (2009), ‘Assortment planning: Review of literature and industry practice’, *Retail supply chain management* pp. 1–55.
- Kunnumkal, S. and Topaloglu, H. (2008), ‘A refined deterministic linear program for the network revenue management problem with customer choice behavior’, *Naval Research Logistics Quarterly* **55**(6), 563–580.
- Levi, R. and Radovanovic, A. (2010), ‘Provably near-optimal LP-based policies for revenue management in systems with reusable resources’, *Operations Research* **58**(2), 503–507.

- Liu, N., Ziya, S. and Kulkarni, V. G. (2010), ‘Dynamic scheduling of outpatient appointments under patient no-shows and cancellations’, *Manufacturing & Service Operations Management* **12**(2), 347–364.
- Liu, Q. and van Ryzin, G. (2008), ‘On the choice-based linear programming model for network revenue management’, *Manufacturing & Service Operations Management* **10**(2), 288–310.
- McFadden, D. (1974), Conditional logit analysis of qualitative choice behavior, in P. Zarembka, ed., ‘*Frontiers in Economics*’, Academic Press, pp. 105–142.
- Murray, M. and Tantau, C. (2000), ‘Same-day appointments: Exploding the access paradigm’. *Family Practice Management*. September, 2000.
- Patrick, J., Puterman, M. L. and Queyranne, M. (2008), ‘Dynamic multi-priority patient scheduling for a diagnostic resource’, *Operations Research* **56**(6), 1507–1525.
- Rau, J. (2011), ‘Medicare to begin basing hospital payments on patient-satisfaction scores’, *Kaiser Health News*. Retrieved on Jan 19, 2012, <http://www.kaiserhealthnews.org/Stories/2011/April/28/medicare-hospital-patient-satisfaction.aspx>.
- Rohleder, T. and Klassen, K. (2000), ‘Using client-variance information to improve dynamic appointment scheduling performance’, *Omega* **28**(3), 293–302.
- Rubin, G., Bate, A., George, A., Shackley, P. and Hall, N. (2006), ‘Preferences for access to the GP: A discrete choice experiment’, *The British Journal of General Practice* **56**(531), 743.
- Rusmevichientong, P., Shmoys, D. B. and Topaloglu, H. (2010), Assortment optimization with mixtures of logits, Technical report, Cornell University, School of Operations Research and Information Engineering. Available at <http://people.orie.cornell.edu/huseyin/publications/publications.html>.
- Ruszczynski, A. (2006), *Nonlinear Optimization*, Princeton University Press, Princeton, New Jersey.
- Ryan, M. and Farrar, S. (2000), ‘Using conjoint analysis to elicit preferences for health care’, *Bmj* **320**(7248), 1530.
- Sampson, F., Pickin, M., O’Cathain, A., Goodall, S. and Salisbury, C. (2008), ‘Impact of same-day appointments on patient satisfaction with general practice appointment systems’, *The British Journal of General Practice* **58**(554), 641–643.
- Schectman, J., Schorling, J. and Voss, J. (2008), ‘Appointment adherence and disparities in outcomes among patients with diabetes’, *Journal of general internal medicine* **23**(10), 1685–1687.
- Silvestre, A., Sue, V. M. and Allen, J. Y. (2009), ‘If you build it, will they come? the kaiser permanente model of online health care’, *Electronic Health Records* **28**, 334–344.
- Talluri, K. and van Ryzin, G. (2004), ‘Revenue management under a general discrete choice model of consumer behavior’, *Management Science* **50**(1), 15–33.
- Topaloglu, H. (2013), ‘Joint stocking and product offer decisions under the multinomial logit model’, *Production and Operations Management* **22**(5), 1182–1199.
- Topaloglu, H., Birbil, S., Frenk, J. and Noyan, N. (2011), ‘Tractable open loop policies for joint overbooking and capacity control over a single flight leg with multiple fare classes’, *Transportation Science* (to appear).
- US Department of Health and Human Services (2011), ‘Electronic health records and meaningful use’. Retrieved on Jan 19, 2012, <http://healthit.hhs.gov/>.



- Wang, W. and Gupta, D. (2011), ‘Adaptive appointment systems with patient preferences’, *Manufacturing & Service Operations Management* **13**(3), 373–389.
- Weiner, M., El Hoyek, G., Wang, L., Dexter, P. R., Zerr, A. D., Perkins, A. J., James, F. and Juneja, R. (2009), ‘A web-based generalist-specialist system to improve scheduling of outpatient specialty consultations in an academic center’, *Journal of General Internal Medicine* **6**, 710–715.
- Zhang, D. and Adelman, D. (2009), ‘An approximate dynamic programming approach to network revenue management with customer choice’, *Transportation Science* **42**(3), 381–394.
- Zocdoc (2012), ‘About us’. Retrieved on Jan 19, 2012, <http://www.zocdoc.com/aboutus>.

## A Appendix: Model Extensions

In this section, we describe a number of extensions of our approach that can be used to formulate certain dimensions of the appointment scheduling problem at a more detailed level.

### A.1 Generalization of the Multinomial Logit Choice Model

Multinomial logit model assumes that the preference weight associated with one alternative does not depend on the offer set. However, in the appointment scheduling context, it is possible that patients place a higher value on the choice of seeking care elsewhere when too few appointment choices are offered. To capture this dependence, we can use a general multinomial logit model introduced by Gallego et al. (2011). In this extended choice model, each day in the scheduling horizon has two parameters, denoted by  $v_j$  and  $w_j$ . If we offer the subset  $S$  of days, then a patient chooses day  $j$  in the scheduling horizon with probability  $P_j(S) = v_j / (1 + \sum_{k \notin S} w_k + \sum_{k \in S} v_k)$ , where the quantity  $\sum_{k \notin S} w_k$  captures the increase in the preference weight of not booking an appointment as a function of the days that are not offered. All of our results in the paper continue to hold under this more general form of the multinomial logit model.

### A.2 Capturing Heterogeneity in Patient Choice Behavior

Our approach implicitly assumes that the choice behavior represented by the multinomial logit model captures the overall patient population preferences. This approach is reasonable either when the patient population is homogeneous in its preference profile, or if we do not have access to additional information that can classify patients in terms of their preferences at time of making scheduling decisions. When additional information is available, subset offer decisions may depend on such information. For example, considering our computational study, if patients having urgent and ambiguous conditions arrive simultaneously into the system and we have information about the condition of the patient before offering the available appointment days, then we can define two sets of decision variables  $h^U = \{h^U(S) : S \subset \mathcal{T}\}$  and  $h^A = \{h^A(S) : S \subset \mathcal{T}\}$  to respectively capture the probability that each subset of days is offered to a patient with urgent and ambiguous conditions. If we make the subset offer decisions with information about the patient condition, then we can use an approach similar to the one in this paper and transform a static model with exponentially many subset offer probabilities into an equivalent static model whose decision variables grows only linearly in the length of the scheduling horizon. Furthermore, building on this static model, we can also formulate a dynamic model by using one step of the policy improvement algorithm.

### A.3 Capturing Day-of-the-Week Effect on Preferences

In our static model, we assume that the demand is stationary and the preference of a patient for different appointment days depends on how many days into the future the appointment is made,

rather than the particular day of the week of the appointment. This assumption allowed us to focus on the expected profit per day in problem (2)-(4). In practice, however, the patient demand may be nonstationary, depending on the day of the week. Furthermore, the preference of a patient for different days of the week may be as pronounced as the preference for different appointment delays. It is possible to generalize our model so that both of these effects are captured.

For instance, assume that nonstationarities follow a weekly pattern and the scheduling horizon  $\tau$  is a multiple of a week. In this case, we can work with the extended set of decision variables by letting  $h_t(S)$  be the probability with which we offer the subset  $S$  of days given that we are on day  $t$  of the week. Using these decision variables, we can construct a model analogous to problem (2)-(4), but we account for the total expected profit per week, rather than expected profit per day. As an example, consider a clinic that is open 7 days a week and is using an appointment window of 14 days. Without loss of generality, we consider the week consisting of days 15 to 21, so the earliest day when patients can be schedule into this week is day 1. Adapting the notation used earlier to the non-stationary setting, we let  $\lambda_t$  denote the mean demand for day  $t$ .

Let  $\mathcal{T}_t = \{t, t + 1, \dots, t + 14\}$  represent the set of days in which patients arriving on day  $t$  can be scheduled. The decision variable in this model is  $h_t(S)$ , which represents the probability of offering the subset of days  $S \subset \mathcal{T}_t$  on day  $t$ . Note that the offering set probability distribution for future days are nonstationary as well and depends on the day of the week. To be concise, we write the set  $\{j - 7 : j \in S\}$  as  $S - 7$ , i.e., the set  $S - 7$  is obtained by subtracting every element in  $S$  by 7 days. Then,  $h_t(S) = h_{t-7}(S - 7)$ . Let  $P_t(S)$  be the probability that an arriving patient will choose day  $t$  when offered set  $S$ . Then, for a patient who arrives on day  $t - j$ , the probability that she will choose day  $t$  is given by  $\sum_{S \subset \mathcal{T}_{t-j}} P_t(S) h_{t-j}(S)$ . Thus the number of patients who schedule an appointment to day  $t$  and are retained until the morning of day  $t$  is given by a Poisson random variable with mean  $\sum_{j=0}^{14} \lambda_{t-j} \bar{r}_j \sum_{S \subset \mathcal{T}_{t-j}} P_t(S) h_{t-j}(S)$ . Similarly, the number of patients who actually show up on day  $t$  is also a Poisson random variable with mean  $\sum_{j=0}^{14} \lambda_{t-j} \bar{s}_j \sum_{S \subset \mathcal{T}_{t-j}} P_t(S) h_{t-j}(S)$ . Then, we can formulate a static model as

$$\begin{aligned}
\max \quad & \sum_{t=15}^{21} \sum_{j=0}^{14} \lambda_{t-j} \bar{s}_j \sum_{S \subset \mathcal{T}_{t-j}} P_t(S) h_{t-j}(S) \\
& - \theta \sum_{t=15}^{21} \mathbb{E} \left\{ \left[ \text{Pois} \left( \sum_{j=0}^{14} \lambda_{t-j} \bar{r}_j \sum_{S \subset \mathcal{T}_{t-j}} P_t(S) h_{t-j}(S) \right) - C \right]^+ \right\} \\
\text{subject to} \quad & \sum_{S \subset \mathcal{T}_t} h_t(S) = 1 \quad \forall t \in \{1, 2, \dots, 21\} \\
& h_t(S) = h_{t-7}(S - 7) \quad \forall t \in \{8, 9, \dots, 21\}, \forall S \subset \mathcal{T}_t \\
& h_t(S) \geq 0, \quad \forall t \in \{1, 2, \dots, 21\}, \forall S \subset \mathcal{T}_t.
\end{aligned}$$

The first constraint ensures that the probability we offer a set of days in any given day is one, whereas the second constraint captures the weekly pattern of decision variables. For this extended version of problem (2)-(4), we can still come up with a transformation similar to the one in Section

4.1 that reduces the number of decision variables from exponential in the length of the scheduling horizon to only linear.

Finally, another simplifying assumption our model makes is that while patients have preferences for which day of the week they would like to be seen they do not have any preferences for the specific appointment time of the day. It is important to note that the same way we formulate preferences on different days of the week (as shown above), we can also incorporate preferences for different times of the day, as long as we make the assumption that the expected cost the clinic incurs on a given day only depends on the total number of patients scheduled for that day, but not the specific times of the appointments.

## B Appendix: Omitted Results

In this section, we give the proofs of the results that are omitted in the paper.

### B.1 Proof of Proposition 1

We complete the proof of Proposition 1 in two parts. First, assume that  $h^* = \{h^*(S) : S \subset \mathcal{T}\}$  is an optimal solution to problem (2)-(4). Letting  $x_j^* = \sum_{S \subset \mathcal{T}} P_j(S) h^*(S)$  and  $u^* = \sum_{S \subset \mathcal{T}} N(S) h^*(S)$ , we need to show that  $(x^*, u^*)$  is a feasible solution to problem (5)-(8) providing the same objective value as the solution  $h^*$ . We have

$$\sum_{j \in \mathcal{T}} x_j^* + u^* = \sum_{S \subset \mathcal{T}} h^*(S) \left[ \sum_{j \in \mathcal{T}} P_j(S) + N(S) \right] = \sum_{S \subset \mathcal{T}} h^*(S) = 1,$$

where the second equality follows by the definition of the multinomial logit model and the third equality follows since  $h^*$  is feasible to problem (2)-(4). Thus, the solution  $(x^*, u^*)$  satisfies the first constraint in problem (5)-(8). On the other hand, using  $\mathbf{1}(\cdot)$  to denote the indicator function, we have  $x_j^*/v_j = \sum_{S \subset \mathcal{T}} [\mathbf{1}(j \in S) h^*(S) / (1 + \sum_{k \in S} v_k)]$  by the definition of  $P_j(S)$  in the multinomial logit model. Noting that  $u = \sum_{S \subset \mathcal{T}} [h^*(S) / (1 + \sum_{k \in S} v_k)]$ , it follows that  $x_j^*/v_j \leq u$ , indicating that the solution  $(x^*, u^*)$  satisfies the second set of constraints in problem (5)-(8). Since  $x_j^* = \sum_{S \subset \mathcal{T}} P_j(S) h^*(S)$ , comparing the objective functions of problems (2)-(4) and (5)-(8) shows that the solutions  $h^*$  and  $(x^*, u^*)$  provide the same objective values for their problems.

Second, assume that  $(x^*, u^*)$  is a feasible solution to problem (5)-(8). We construct the solution  $h^*$  as in (9). To see that the solutions  $(x^*, u^*)$  and  $h^*$  provide the same objective values for their respective problems, we observe that

$$\sum_{S \subset \mathcal{T}} P_j(S) h^*(S) = \sum_{i=0}^{\tau} P_j(S_i) h^*(S_i) = \sum_{i=j}^{\tau} P_j(S_i) h^*(S_i) = \sum_{i=j}^{\tau} v_j \left[ \frac{x_i^*}{v_i} - \frac{x_{i+1}^*}{v_{i+1}} \right] = x_j^*,$$

where the first equality is by the fact that  $h^*(S)$  takes positive values only for the sets  $S_0, \dots, S_\tau$  and  $\emptyset$ , the second equality is by the fact that  $j \in S_i$  only when  $j \leq i$  and the third equality follows

by the definition of  $h^*(S_i)$  and noting that  $x_{\tau+1}^* = 0$ . Using the equality above and comparing the objective functions of problems (2)-(4) and (5)-(8) show that the solutions  $h^*$  and  $(x^*, u^*)$  provide the same objective values for their problems. To see that the solution  $h^*$  is feasible to problem (2)-(4), we let  $V_i = 1 + \sum_{k \in S_i} v_k$  for notational brevity and write  $\sum_{S \subset \mathcal{T}} h^*(S)$  as

$$u^* - \frac{x_0^*}{v_0} + \sum_{j=0}^{\tau} V_j \left[ \frac{x_j^*}{v_j} - \frac{x_{j+1}^*}{v_{j+1}} \right] = u^* + (V_0 - 1) \frac{x_0^*}{v_0} + (V_1 - V_0) \frac{x_1^*}{v_1} + \dots + (V_\tau - V_{\tau-1}) \frac{x_\tau^*}{v_\tau} = 1,$$

where the first equality follows by rearranging the terms and using the convention that  $x_{\tau+1}^* = 0$  and the second equality is by noting that  $V_i - V_{i-1} = v_i$  and using the fact that  $(x^*, u^*)$  is feasible to problem (5)-(8) so that  $\sum_{j=0}^{\tau} x_j^* + u^* = 1$ .  $\square$

## B.2 Lemma 7

The following lemma is used in Section 4.

**Lemma 7.** *Letting  $F(\alpha) = \mathbb{E}\{[\text{Pois}(\alpha) - C]^+\}$ ,  $F(\cdot)$  is differentiable and convex.*

*Proof.* The proof uses elementary properties of the Poisson distribution. By using the probability mass function of the Poisson distribution, we have

$$\begin{aligned} F(\alpha) &= \sum_{i=C+1}^{\infty} \frac{e^{-\alpha} \alpha^i}{i!} (i - C) = \sum_{i=C+1}^{\infty} \frac{e^{-\alpha} \alpha^i}{(i-1)!} - \sum_{i=C+1}^{\infty} \frac{e^{-\alpha} \alpha^i}{i!} C \\ &= \alpha \sum_{i=C}^{\infty} \frac{e^{-\alpha} \alpha^i}{i!} - C \sum_{i=C+1}^{\infty} \frac{e^{-\alpha} \alpha^i}{i!} = \alpha \mathbb{P}\{\text{Pois}(\alpha) \geq C\} - C \mathbb{P}\{\text{Pois}(\alpha) \geq C + 1\}. \end{aligned} \quad (25)$$

Thus, the differentiability of  $F(\cdot)$  follows by the differentiability of the cumulative distribution function of the Poisson distribution with respect to its mean. For the convexity of  $F(\cdot)$ , we have

$$\begin{aligned} \frac{d\mathbb{P}\{\text{Pois}(\alpha) \geq C\}}{d\alpha} &= -\frac{d\mathbb{P}\{\text{Pois}(\alpha) \leq C - 1\}}{d\alpha} = -\sum_{i=0}^{C-1} \frac{d\left(\frac{e^{-\alpha} \alpha^i}{i!}\right)}{d\alpha} \\ &= \sum_{i=0}^{C-1} \frac{e^{-\alpha} \alpha^i}{i!} - \sum_{i=1}^{C-1} \frac{e^{-\alpha} \alpha^{i-1}}{(i-1)!} = \mathbb{P}\{\text{Pois}(\alpha) = C - 1\}. \end{aligned}$$

In this case, if we differentiate both sides of (25) with respect to  $\alpha$  and use the last chain of equalities, then we obtain

$$\begin{aligned} \frac{dF(\alpha)}{d\alpha} &= \mathbb{P}\{\text{Pois}(\alpha) \geq C\} + \alpha \mathbb{P}\{\text{Pois}(\alpha) = C - 1\} - C \mathbb{P}\{\text{Pois}(\alpha) = C\} \\ &= \mathbb{P}\{\text{Pois}(\alpha) \geq C\} + \alpha \frac{e^{-\alpha} \alpha^{C-1}}{(C-1)!} - C \frac{e^{-\alpha} \alpha^C}{C!} = \mathbb{P}\{\text{Pois}(\alpha) \geq C\}. \end{aligned}$$

To see that  $F(\cdot)$  is convex, we use the last two chains of equalities to observe that the second derivative of  $F(\alpha)$  with respect to  $\alpha$  is  $\mathbb{P}\{\text{Pois}(\alpha) = C - 1\}$ , which is positive.  $\square$

### B.3 Proof of Corollary 3

By Proposition 2, there exists an optimal solution  $x^*$  to problem (5)-(8) that satisfies (16). We define the subsets  $S_0, S_1, \dots, S_\tau$  as in the proof of Proposition 1 and construct an optimal solution  $h^*$  to problem (2)-(4) by using  $x^*$  as in (9). In this case, since  $x^*$  satisfies (16) for some  $k \in \mathcal{T}$ , only two of the decision variables  $\{h(S) : S \subset \mathcal{T}\}$  can take on nonzero values and these two decision variables are  $h^*(S_{k-1})$  and  $h^*(S_k)$ . Thus, the desired result follows by observing that  $S_k$  is a subset of the form  $\{0, 1, \dots, k\}$ .  $\square$

### B.4 Proof of Lemma 4

We let  $\pi^*(S)$  be the steady state probability with which we offer the subset  $S$  of days under the optimal, possibly state-dependent, policy. So, if we consider a particular day in steady state, then the number of patients that are scheduled for this day  $j$  days in advance is given by a Poisson random variable with mean  $\sum_{S \subset \mathcal{T}} \lambda P_j(S) \pi^*(S)$ . Therefore, if we use the random variable  $A_j^*$  to denote the number of patients that we schedule for a particular day  $j$  days in advance in steady state, then  $A_j^*$  has mean  $\sum_{S \subset \mathcal{T}} \lambda P_j(S) \pi^*(S)$ . We note that  $A_1^*, A_2^*, \dots, A_\tau^*$  are not necessarily independent of each other, since the decisions under the optimal state-dependent policy on different days can be dependent. Similarly, in steady state, we let  $S_j^*$  be the number of patients that we schedule for a particular day  $j$  days in advance and that show up under the optimal state-dependent policy. Finally, we let  $R_j^*$  be the number of patients that we schedule for a particular day  $j$  days in advance and that we retain until the morning of the appointment under the optimal state dependent policy. Noting that the show-up and cancellation decisions of the patients are independent of how many patients we schedule for a particular day, we have  $\mathbb{E}\{S_j^*\} = \bar{s}_j \mathbb{E}\{A_j^*\}$  and  $\mathbb{E}\{R_j^*\} = \bar{r}_j \mathbb{E}\{A_j^*\}$ . In this case, the average profit per day generated by the optimal state-dependent policy satisfies

$$\begin{aligned} V^* &= \mathbb{E}\left\{\sum_{j \in \mathcal{T}} S_j^*\right\} - \theta \mathbb{E}\left\{\left[\sum_{j \in \mathcal{T}} R_j^* - C\right]^+\right\} \leq \sum_{j \in \mathcal{T}} \mathbb{E}\{S_j^*\} - \theta \left[\sum_{j \in \mathcal{T}} \mathbb{E}\{R_j^*\} - C\right]^+ \\ &= \sum_{j \in \mathcal{T}} \sum_{S \subset \mathcal{T}} \lambda \bar{s}_j P_j(S) \pi^*(S) - \theta \left[\sum_{j \in \mathcal{T}} \sum_{S \subset \mathcal{T}} \lambda \bar{r}_j P_j(S) \pi^*(S) - C\right]^+ \leq Z_{DET}. \end{aligned}$$

In the chain of inequalities above, the first inequality is by the Jensen's inequality. The second equality is by  $\mathbb{E}\{S_j^*\} = \bar{s}_j \mathbb{E}\{A_j^*\}$  and  $\mathbb{E}\{R_j^*\} = \bar{r}_j \mathbb{E}\{A_j^*\}$ . To see the second inequality, we note that  $\{\pi^*(S) : S \subset \mathcal{T}\}$  is a feasible but not necessarily an optimal solution to the problem

$$\begin{aligned} &\max \sum_{j \in \mathcal{T}} \sum_{S \subset \mathcal{T}} \lambda \bar{s}_j P_j(S) w(S) - \theta \left[\sum_{j \in \mathcal{T}} \sum_{S \subset \mathcal{T}} \lambda \bar{r}_j P_j(S) w(S) - C\right]^+ \\ &\text{subject to } \sum_{S \subset \mathcal{T}} w(S) = 1 \\ &\quad w(S) \geq 0 \quad S \subset \mathcal{T} \end{aligned}$$

and the optimal objective values of the problem above and problem (20)-(23) are equal to each other, which can be verified by using the argument in the proof of Proposition 1.  $\square$

## B.5 Proof of Proposition 5

Letting  $(\hat{x}, \hat{u})$  be an optimal solution to problem (20)-(23), we have  $\Pi(x^*) \geq \Pi(\hat{x})$ . Since we can always offer the empty set with probability one, the optimal objective value of problem (5)-(8) is nonnegative and we obtain  $\Pi(x^*)/V^* = [\Pi(x^*)]^+/V^* \geq [\Pi(\hat{x})]^+/V^*$ . Using Lemma 4, we continue this chain of inequalities as  $[\Pi(\hat{x})]^+/V^* \geq [\Pi(\hat{x})]^+/Z_{DET} \geq \Pi(\hat{x})/Z_{DET} = 1 - (Z_{DET} - \Pi(\hat{x}))/Z_{DET}$ . So, it is enough to show that the second term on the right side of (24) upper bounds  $(Z_{DET} - \Pi(\hat{x}))/Z_{DET}$ . For a Poisson random variable with mean  $\alpha$ , we claim that

$$\mathbb{E}\{[\text{Pois}(\alpha) - C]^+\} \leq [\alpha - C]^+ + \alpha/\sqrt{2\pi C}. \quad (26)$$

This claim will be proved at the end. With this claim and letting  $\beta = \sum_{j=1}^T \lambda \bar{s}_j \hat{x}_j$  and  $\alpha = \sum_{j=1}^T \lambda \bar{r}_j \hat{x}_j$  for notational brevity, we obtain

$$\frac{Z_{DET} - \Pi(\hat{x})}{Z_{DET}} = \frac{[\beta - \theta[\alpha - C]^+] - [\beta - \theta\mathbb{E}\{[\text{Pois}(\alpha) - C]^+\}]}{Z_{DET}} \leq \frac{\theta\alpha}{\sqrt{2\pi C}} \leq \frac{\theta\lambda\bar{r}_0}{Z_{DET}}, \quad (27)$$

where the second inequality is by noting that  $\sum_{j \in \mathcal{T}} \hat{x}_j \leq 1$ ,  $\bar{r}_0 \geq \bar{r}_1 \geq \dots \geq \bar{r}_T$  so that  $\alpha \leq \lambda\bar{r}_0$ .

We proceed to constructing a lower bound on  $Z_{DET}$ . The solution  $(\tilde{x}, \tilde{u})$  we obtain by setting  $\tilde{x}_0 = \frac{v_0}{1+v_0}$ ,  $\tilde{u} = \frac{1}{1+v_0}$  and all other decision variables to zero is feasible to problem (20)-(23). Thus, if  $\lambda\bar{r}_0 \frac{v_0}{1+v_0} \leq C$ , then we can lower bound  $Z_{DET}$  as  $Z_{DET} \geq \lambda\bar{s}_0 \tilde{x}_0 - \theta[\lambda\bar{r}_0 \tilde{x}_0 - C]^+ = \lambda\bar{s}_0 \frac{v_0}{1+v_0}$ . On the other hand, if  $\lambda\bar{r}_0 \frac{v_0}{1+v_0} > C$ , then the solution  $(\tilde{x}, \tilde{u})$  we obtain by setting  $\tilde{x}_0 = \frac{C}{\lambda\bar{r}_0}$ ,  $\tilde{u} = 1 - \frac{C}{\lambda\bar{r}_0}$  and all other decision variables to zero is feasible to problem (20)-(23). Thus, if  $\lambda\bar{r}_0 \frac{v_0}{1+v_0} > C$ , then we can lower bound  $Z_{DET}$  as  $Z_{DET} \geq \lambda\bar{s}_0 \tilde{x}_0 - \theta[\lambda\bar{r}_0 \tilde{x}_0 - C]^+ = \bar{s}_0 \frac{C}{\bar{r}_0}$ . Collecting the two cases together, we lower bound  $Z_{DET}$  by  $\bar{s}_0 \min\left\{\lambda \frac{v_0}{1+v_0}, \frac{C}{\bar{r}_0}\right\}$ . Continuing the chain of inequalities in (27) by using the lower bound on  $Z_{DET}$ , we obtain

$$\frac{\theta\lambda\bar{r}_0}{\sqrt{2\pi C}} \leq \frac{\theta\lambda\bar{r}_0}{\sqrt{2\pi C}} \frac{1}{\bar{s}_0 \min\left\{\lambda \frac{v_0}{1+v_0}, \frac{C}{\bar{r}_0}\right\}}.$$

Arranging the terms in the last expression above yields the desired result.

Finally, we prove the claim (26) made above. For  $k \geq C+1$ , we observe that  $[k-C]^+ - [\alpha-C]^+ \leq k - \alpha$ . In particular, for  $k \geq \alpha$ , this inequality follows by the Lipschitz continuity of the function  $[\cdot - C]^+$ . For  $k < \alpha$ , we have  $C+1 \leq k < \alpha$  and it follows  $[k-C]^+ - [\alpha-C]^+ = k - \alpha$ , establishing the desired inequality. In this case, the result in the lemma follows by noting that

$$\begin{aligned} \mathbb{E}\{[\text{Pois}(\alpha) - C]^+\} &= \sum_{k=C+1}^{\infty} [k-C]^+ \frac{e^{-\alpha}\alpha^k}{k!} \leq [\alpha-C]^+ + \sum_{k=C+1}^{\infty} [[k-C]^+ - [\alpha-C]^+] \frac{e^{-\alpha}\alpha^k}{k!} \\ &\leq [\alpha-C]^+ + \sum_{k=C+1}^{\infty} (k-\alpha) \frac{e^{-\alpha}\alpha^k}{k!} = [\alpha-C]^+ + \frac{e^{-\alpha}\alpha^C}{C!} \alpha \leq [\alpha-C]^+ + \frac{e^{-C}C^C}{C!} \alpha, \end{aligned}$$

where the first inequality follows by adding and subtracting  $[\alpha - C]^+$  to the expression on the left side of this inequality, the second inequality follows by the inequality derived at the beginning of the proof, the last equality is by arranging the terms in the summation on the left side of this inequality and the third inequality is by noting that the function  $f(\alpha) = e^{-\alpha}\alpha^C$  attains its maximum at  $\alpha = C$ . In this case, the result follows by noting that  $C! \geq \sqrt{2\pi C}(C/e)^C$  by Stirling's approximation and using this bound on the right side of the chain of inequalities above.  $\square$

## B.6 Proof of Proposition 6

The proof follows from an argument similar to the one in the proof of Proposition 1. Assume that  $h^* = \{h^*(S) : S \subset \mathcal{T}\}$  is an optimal solution to problem (2)-(4) with the additional constraints  $h_t(S) \in \{0, 1\}$  for all  $S \subset \mathcal{T}$ . Letting  $x_j^* = \sum_{S \subset \mathcal{T}} P_j(S) h^*(S)$  and  $u^* = \sum_{S \subset \mathcal{T}} N(S) h^*(S)$ , we can follow the same argument in Section B.1 to show that  $(x^*, u^*)$  with  $x^* = (x_0^*, \dots, x_\tau^*)$  is a feasible solution to problem (5)-(8) with the additional constraints  $x_j/v_j \in \{0, u\}$  for all  $j \in \mathcal{T}$ . Furthermore, the objective values provided by the two solutions for their respective problems are identical. On the other hand, assume that  $(x^*, u^*)$  with  $x^* = (x_0^*, \dots, x_\tau^*)$  is an optimal solution to problem (5)-(8) with the additional constraints  $x_j/v_j \in \{0, u\}$  for all  $j \in \mathcal{T}$ . We reorder and reindex the days in the scheduling horizon so that we have  $u^* = x_0^*/v_0 = x_1^*/v_1^* = \dots = x_{j-1}^*/v_{j-1}^* \geq x_j^*/x_j = x_{j+1}^*/v_{j+1} = \dots = x_\tau^*/v_\tau^* = 0$ . We define the subsets  $S_0, S_1, \dots, S_\tau$  as  $S_j = \{0, 1, \dots, j\}$ . For notational convenience, we define  $x_{\tau+1}^* = 0$ . In this case, letting

$$h^*(\emptyset) = u^* - \frac{x_0^*}{v_0} \quad \text{and} \quad h^*(S_j) = \left[1 + \sum_{k \in S_j} v_k\right] \left[\frac{x_j^*}{v_j} - \frac{x_{j+1}^*}{v_{j+1}}\right]$$

for all  $j = 0, 1, \dots, \tau$  and letting  $h^*(S) = 0$  for all other subsets of  $\mathcal{T}$ , we can follow the same argument in Section B.1 to show that  $\{h^*(S) : S \subset \mathcal{T}\}$  is a feasible solution to problem (2)-(4) with the additional constraints  $h_t(S) \in \{0, 1\}$  for all  $S \subset \mathcal{T}$ . Furthermore, we can check that the two solutions provide the same objective value for their respective problems.  $\square$

## B.7 Sample Survey Questions



**Q3.** For the following questions, imagine that you are in your current state of health but over the last few months you have been feeling tired and irritable and have had difficulty sleeping. You have tried several things yourself to remedy this but are not feeling any better. You decide to seek a medical opinion from doctors at the Farrell clinic. Now you are making calls to Farrell to schedule an appointment.

If the clinic can only offer you an appointment today or 2 days from now, *considering your current work/life schedule*, which appointment date would you take or would you seek care elsewhere, e.g., try to find another physician or go to an emergency room? Choose one option below by placing an “x” in the box that best represents your response.

- Today
- 2 Days from Now
- Seek Care Elsewhere

Figure 1: Sample choice question under ambiguous health condition.

**Q7.** For the following questions, imagine that you have a heavy cough and cold today. Over the past 2 days you have started to get some pain in the right side of your chest. It is very sharp and worse if you cough or take a deep breath in. You decide to seek a medical opinion from doctors at the Farrell clinic. Now you are making calls to Farrell to schedule an appointment.

If the clinic can only offer you an appointment today or 2 days from now, *considering your current work/life schedule*, which appointment date would you take or would you seek care elsewhere, e.g., try to find another physician or go to an emergency room? Choose one option below by placing an “x” in the box that best represents your response.

- Today
- 2 Days from Now
- Seek Care Elsewhere

Figure 2: Sample choice question under urgent health condition.