Calendar mechanisms[*]

Toomas Hinnosaar[†]

February 2017

Abstract

I study the dynamic mechanism design problem of a monopolist selling a fixed number of service slots to randomly arriving, short-lived buyers with heterogeneous values. The fully optimal mechanism is a non-standard auction in which bidders' payoffs are non-monotone in their opponents' bids. Because its complexity may make the fully optimal mechanism too costly to implement, I also study the optimal mechanisms in restricted classes. The most restrictive are *pure calendar mechanisms*, which allocate service dates instead of contingent contracts. The optimal pure calendar mechanism is characterized by the opportunity costs of service slots and is implementable with a simple mechanism.

JEL: D82, C73, D44
Keywords: dynamic mechanism design, services, ticket sales, restricted mechanisms

1 Introduction

In this paper I use the mechanism design approach to study the problem of selling services. For example, seats on a train are usually allocated by ticket sales: a ticket is sold at a fixed price and guarantees a seat for a particular day and time. There are alternative and perhaps more profitable ways to sell these seats. Tickets could be auctioned rather than sold at posted prices. Moreover, tickets could be replaced by contracts that guarantee seats only when the demand is not too high. Airline companies have already moved in both directions: prices depend on the demand and passengers with cheap tickets face a higher risk of being bumped when demand for expensive tickets is high. The problem of selling services has three distinct features that I

explore in this paper. First, the goods are perishable; once the date of service passes, the goods disappear. Second, the situation repeats; there is a new set of goods on offer every period. Third, demand is random and customers may be willing to wait but not for too long.

In my model, a monopolist has a fixed number of identical service slots available each period. In each period a random number of buyers arrives. Each buyer lives for two periods and has a unit demand and a random valuation for the service. The valuation stays constant over time, but buyers discount the future. Each buyer interacts with the seller only once, on arrival, and is either instantly served, refused service, or given a contract specifying the circumstances under which he will receive the service in the next period. The seller's goal is to maximize expected discounted revenue by choosing an optimal mechanism. The main part of the paper assumes that the seller observes buyers' arrival times and valuations. In a later section I argue that standard methods can be used to extend the analysis to private information.

The paper provides two main results. The first result characterizes the optimal mechanism in a class of simple mechanisms, where the seller can only allocate future services unconditionally. A buyer with a ticket for future service will receive the service irrespective of future demand. This corresponds to the sale of standard tickets—the selling mechanism can be arbitrary (e.g., an auction), but the ticket conditions are simple: the ticket guarantees that the buyer will receive a service on a certain date. I call this class of mechanisms *pure calendar mechanisms*. The optimal pure calendar mechanism is characterized by the opportunity costs of fulfilling the service slots. The optimal pure calendar mechanism is a strict improvement over static mechanisms since it allows the seller to smooth the demand—to sell more services in periods when demand is high.

The second result characterizes the fully optimal mechanism in which future services are sold through personalized contracts. In this case, the seller is not only able to choose an arbitrary selling mechanism but can also choose the situations in which ticket holders will be seated. With optimal contracts, buyers are served in situations where the total cost of serving them is lowest. The fully optimal mechanism is a strict improvement over the optimal pure calendar mechanism as it allows more flexible allocation of future services. The buyers with higher valuations are promised future service with higher probability, and these promises are fulfilled in the situations in which it is least costly to do so.

In addition to the two main results, I study *waiting-list mechanisms*, a restrictive class of mechanisms which are not as restrictive as pure calendar mechanisms. I focus on a particular property of waiting lists: the fact that a list only keeps track of buyers' positions on the list and not their valuations. In other dimensions the seller may freely optimize. For example, the seller may use an auction to assign newly arrived buyers to positions on the waiting list and may use an arbitrary allocation rule for serving buyers on the waiting list. The assumption requires only that once a buyer is on the waiting list; the likelihood that he will be served is independent of his valuation.[1] The optimal waiting-list mechanism combines ideas from both the optimal pure calendar mechanism and the fully optimal mechanism. The contract design

---

[1]For example, most airlines overbook flights and ration seats, at least partially, according to passengers' boarding priorities, but not according to the price they paid for their tickets.

is similar to the fully optimal mechanism, offering future service in situations where it is least costly to do so, but since the contracts are designed for an average buyer expecting to receive a contract rather than for a specific buyer type, assigning a contract comes with strictly positive opportunity cost. Therefore, the allocation rule of the contracts is similar to the optimal pure calendar mechanism. The optimal waiting-list mechanism is a strict improvement over the optimal pure calendar mechanism, and in some situations it may be simpler to implement than the fully optimal mechanism because the waiting-list mechanism requires designing only a small number of contract types, whereas the fully optimal mechanism requires a continuum of different contract types.

There are two alternative assumptions under which the fully optimal mechanism has been studied in previous literature. First, if there are enough service slots to serve all potential customers immediately upon their arrival, the optimal mechanism would be a posted-price mechanism. This is a version of Stokey's (1979) no-pure-price-discrimination result because, from a buyer's perspective, receiving a delayed service is like receiving a fractional allocation. Similar to Myerson (1981), a fractional allocation never maximizes static profit. Offering delayed service would only decrease future profits unless rationing was necessary, at least, under some realizations of uncertainty. Second, when buyers can remain in the model forever, it is always optimal to serve the remaining buyers with the highest values.[2] This result follows from Said (2012). Buyers who arrive in different periods view the future in the same way, so their arrival times are irrelevant. This result does not hold if buyers do not discount the future at a constant rate or if they exit at an increasing rate. In these cases, the seller must treat buyers arriving in different periods differently. Consequently, buyer types would be two-dimensional and characterized by valuation and arrival time.

This paper concentrates on the remaining case where rationing is needed and arrival times matter. Many applications fit this case. First, buyers may not value the service if they have to wait too long for it. Transportation is an obvious example—a passenger who arrives at the station would prefer to be seated on the next bus or train and may be willing to wait an hour or two, but not more than that. Second, in cases where buyers interact with the seller in advance, there could be legal or practical barriers for sellers promising services in the distant future. Many types of ticket sales fall into this category. Finally, buyers may have alternative ways of fulfilling their needs and therefore might be willing to wait up until a certain point, but would look elsewhere when the service was delayed excessively.

This paper focuses on an extreme case where arrival times matter: buyers exit the model one period after they arrive. This behavior makes the model an *overlapping-generations model* in which there are always two kinds of buyers: those who arrived in the previous period (and will soon leave) and those who have just arrived (and will stay for one more period). The optimal mechanism in this model offers delayed service and differentiates buyers according to their arrival times.

_____

[2]If the goal is to maximize revenue, then it is optimal to serve the remaining buyers with the highest non-negative virtual values.

The modeling approach in this paper is similar to the modeling approach in the seminal papers of Myerson (1981) and Riley and Samuelson (1981). In fact, Myerson's optimal mechanism is the optimal static mechanism in the model analyzed in this paper. Myerson's optimal mechanism is a special case of all three classes of mechanisms that this paper characterizes, and thus provides a benchmark against which to compare all dynamic mechanisms.

This paper belongs to a growing body of literature on dynamic mechanism design. In particular, it is methodologically related to the literature on the sale of durable goods in which the seller has a fixed quantity of goods and a deadline for selling them. Gershkov and Moldovanu (2009, 2010) characterize the optimal online mechanism where buyers must be assigned an object upon their arrival, and Board and Skrzypacz (2016) characterize the optimal mechanism when buyers are patient. Pai and Vohra (2013) and Mierendorff (2016) consider a model in which buyers' unobservable types include their arrival and departure times and partially characterize an optimal mechanism. In these papers, the most important trade-offs are between extraction of rents and option value; giving away an object before the deadline means that the object cannot be assigned to potentially higher-valued buyers who arrive later. The option value decreases over time, however, because the arrival of such buyers becomes increasingly less likely. In contrast to these papers, the model presented here focuses on situations where the seller has a new set of perishable goods in every period. These situations have a different set of trade-offs. For example, the current service does not have an opportunity cost because it is not possible to sell it to the buyers who arrive in the future.

Bergemann and Välimäki (2010) and Pavan, Segal, and Toikka (2014) give results for general classes of dynamic mechanisms. Pavan, Segal, and Toikka provide the envelope formula for a large class of dynamic mechanism design questions and show how to compute transfers. Bergemann and Välimäki's main result is the characterization of transfers that implement any efficient mechanism. In an earlier work, Parkes and Singh (2003) studied sequential allocation problems and constructed a dynamic generalization of the Vickrey-Clarke-Groves mechanism that implements an efficient mechanism in this setting. In contrast to these papers, my focus is on revenue maximization in the setting in which buyers have constant values but arrive stochastically and leave soon after arriving. Therefore, the results of these papers are not applicable, but as I will show, the envelope formula in my model is simple and, therefore, transfers are straightforward to characterize.

This paper is closely related to the literature on price discrimination with durable goods. Stokey (1979) and Conlisk, Gerstner, and Sobel (1984) showed that it is not optimal to delay sales purely for the purpose of price discrimination, but subsequent authors have documented several reasons for delaying sales. Board (2008) showed that when demand fluctuates, some delay is necessary. This result is similar to the demand-smoothing argument in this paper, but in his model, the monopolist has a much more limited set of available instruments. Garrett (2016) offered another reason for price discrimination—changes in buyers' valuations. The intersection of his paper and mine is Ely, Garrett, and Hinnosaar (2016), where buyers' valuations change and the seller has relatively sophisticated contracts available. Similar to my paper, it is optimal

for the seller to refuse service to buyers with low values, but the main force—extracting more surplus from partially uninformed buyers—is different.

Two papers have studied mechanisms in overlapping-generations models. Zeithammer (2007) showed that sequential second-price auctions in an overlapping-generations model facilitate learning both from the side of the buyer (in which case they shade their bids) and from the side of the seller (who sells only if the expected demand is high enough). I study optimal mechanisms (with potentially restricted classes) in a similar environment[3] and show that second-price auctions are never optimal. In the model by Bloch and Houy (2012), the seller has a single object[4] that is used by buyers for either one or two periods and then returned. Similar to my model, the older buyers receive preferential treatment, but in their model, assigning the object to an old buyer has lower opportunity cost because the seller expects to get it back soon.

This paper is organized as follows: Section 2 discusses the main trade-offs with a simple three-period example. Section 3 introduces the model and describes some immediate implications. Section 4 characterizes the optimal pure calendar mechanism, section 5 the fully optimal mechanism, and section 6 the optimal waiting-list mechanism. Section 7 extends the analysis to private information. Finally, section 8 concludes and discusses potential extensions. All proofs are presented in appendix A.

2   Illustrative example

This section introduces a simple three-period model that clarifies some of the assumptions made in this paper and illustrates how optimal rationing leads to the type of contracts discussed in subsequent sections. Note that this example is not a special case of the general model[5] but nonetheless has some of its main characteristics.

In this example, there are three periods, three buyers, and two service slots. Buyer 0 arrives in period 0 and has a value v_0 for the service. Buyers 1 and 2 arrive in period 1 and have values v_1 and v_2, respectively. Buyer 0 exits the model prior to the start of period 2. All three values are independent random variables and are distributed uniformly in $[0, 1]$. The seller observes arrivals and buyers' values and has two service slots available, one in period 1 and one in period 2. The seller's goal is to maximize expected revenue. All of the buyers and the seller discount the future at rate $\delta = \frac{1}{2}$. To shorten the notation, I denote $v_H = \max\{v_1, v_2\}$ and $v_L = \min\{v_1, v_2\}$, and I refer to a buyer having value v_H as a high-value buyer and a buyer having value v_L as a low-value buyer.

The example's simple structure leads to a few immediate conclusions. First, because buyers' values are observable and their values are always positive, both service slots are always filled.

---

[3]In contrast to my model, Zeithammer assumed that buyer types are binary; the seller could produce up to one unit of the good every period, and there were exactly two new buyers every period.

[4]Additionally, Bloch and Houy (2012) assumed that buyers arrive deterministically: one buyer with independent, private value arrives every period so that every period there is one old buyer and one new buyer. In most of the paper, they focus on mechanisms without transfers.

[5]The general model has an infinite number of periods, m service slots every period, and a random number of buyers every period.

Second, the high-value buyer is always served earlier than the low-value buyer. Third, if buyer 0 is never served or buyer 0 is served with certainty, then a revenue $\pi^s = \frac{5}{6}$ (in terms of period 1's value) is assured. The value π^s provides a benchmark upon which the optimal allocation strictly improves.

The only remaining question is whether buyer 0 is served in period 1 or both slots are left for the buyers arriving in period 1. The optimal allocation assigns the first slot to buyer 0 if and only if $v_0 + \frac{1}{2}v_H \geq v_H + \frac{1}{2}v_L$ or equivalently $v_0 \geq v_H - \frac{1}{2}[v_H - v_L]$. The key observation is that both values v_1 and v_2 raise the opportunity costs of assigning the slot to buyer 0. Therefore, both buyers—the buyer directly replaced and the buyer indirectly replaced—matter. Figure 1 illustrates this allocation rule for a given v_0.
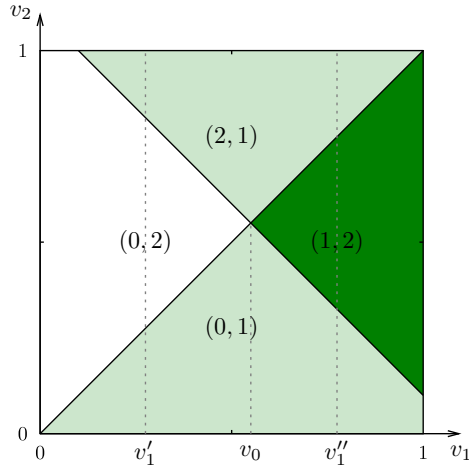

Figure 1: Allocation in the simple example described in section 2. The shading color intensity indicates the discounted expected quantity for buyer 1: darker for instant service, lighter for delayed service, and white for no service. The numbers in parenthesis are the identities of the buyers served in periods 1 and 2, respectively. The value of buyer 0 is $v_0 = 0.55$.

The optimal allocation rule has two main properties. First, it favors buyer 0 because he exits the model earlier. Buyer 0 is assigned the best service slot not only when he has the highest value but also when he has the second highest value and the average of the other two values is lower than v_0. It is one of the key characteristics of the paper's main model that buyers who exit the model earlier receive preferential treatment.

Second, for the newly arriving buyers, the optimal allocation rule is non-monotone in the opponent's value. Consider buyer 1 with value $v_1 > v_0$. (See v_1'' in fig. 1 for illustration.) If v_2 is less than $2v_0 - v_1$, then buyer 0 gets the period 1 slot, and buyer 1, as the highest valued among the other two, receives the period 2 slot. But when v_2 exceeds $2v_0 - v_1$, the opportunity cost of serving buyer 0 becomes too high, so both buyers 1 and 2 are served. As long as $v_2 < v_1$, buyer 1 is assigned the period 1 service. When v_2 increases even further, buyer 1 is again offered service in period 2. Therefore, there is an intermediate range $(2v_0 - v_1, v_1)$ of values for buyer 2 in which buyer 1 gets instant service, whereas for all other values, buyer 1's service is delayed. Similar non-monotonicity holds when $v_1 < v_0$, but with opposite direction. (See v_1' in fig. 1.)

A buyer 1 having value below v_0 can never obtain instant service but can get delayed service if either $v_2 < v_1$ (because buyer 1 is high-value among the new buyers) or $v_2 > 2v_0 - v_1$ (because the opportunity cost is high enough that buyer 2 receives a period 1 slot and buyer 1 is assigned to a period 2 slot). This type of non-monotonicity will be a general property of the main model's optimal mechanisms and implies that the buyers in these mechanisms prefer winning by small margins (rather than large ones) and losing by large margins (rather than small ones).

One of this paper's main contributions is showing how to design optimal mechanisms in the classes of restricted (simpler) mechanisms. To model these restrictions, I take an alternative approach to characterizing mechanisms—I assume that the buyers interact with the seller only on arrival. Without restrictions on mechanisms, this assumption is without loss of generality. In the example, all relevant information concerning buyer 0 is present in period 0 but is used only in period 1. The discussion so far has implicitly assumed that buyer 0 interacts with the seller until the end of period 1 because the seller knows all the relevant values only at that time. However, we can reinterpret the assumption by assuming that buyer 0 interacts with the seller only in period 0, during which buyer 0 pays the seller and receives a *contract* for future service. This contract specifies exact circumstances under which the buyer will be served in period 1. In other words, a contract is a set of realizations of (v_1, v_2) with which buyer 0 will be served. For example, according to the optimal contract, buyer 0 will be served if and only if $v_H - \frac{1}{2}[v_H - v_L] < v_0$. In the general model, it is assumed that buyers interact with the seller only when they arrive, at which time the seller can assign them contingent contracts for future service. As in the illustrative example, if the seller can offer arbitrary contracts, this assumption causes no loss of generality and allows simpler notation.

Because the optimal contract depends on future arrivals and is a function of the buyer's own value, it may be costly to implement in practice. Therefore, we may be interested in mechanisms in which contracts are simpler. The simplest and most restrictive of these mechanisms is the pure calendar mechanism, where the contract for future service is assumed to be unconditional of future events. In this example, the assumption would mean that buyer 0 receives service in period 1 if and only if buyer 0 receives this non-contingent contract. Assigning buyer 0 the future service is optimal when $v_0 + \mathbb{E}[\frac{1}{2}v_H] \geq \mathbb{E}[v_H + \frac{1}{2}v_L]$ or, equivalently, when $v_0 \geq \underline{v} = \mathbb{E}[v_H - \frac{1}{2}[v_H - v_L]] = \frac{1}{2}$. Threshold $\underline{v}$ can be interpreted as the opportunity cost of allocating the service to buyer 0. In contrast to the fully optimal mechanism, this threshold depends on the expected values of the new buyers rather than on their true values. These non-contingent contracts are easy to describe and are sold at a posted price of $\underline{v} = \frac{1}{2}$. Whether the simpler implementation is preferable depends on the particular application and on the costs of implementation. Numeric values are perhaps instructive; in the example in this section, the fully optimal mechanism achieves a 17.5% improvement over the benchmark $\pi^s = \frac{5}{6}$, and the pure calendar mechanism delivers a 15% improvement.

The model that I introduce in the next section has an infinite number of periods, a random number of buyers, and a large number of service slots. However, some of the main features of the example remain. In particular, in every period some of the buyers are about to leave, whereas

other buyers will leave one period later, so there are similar rationing trade-offs. In the fully optimal mechanism, it can be assumed without loss of generality that buyers interact with the seller just once (and receive contingent contracts). I also investigate the optimal mechanism in restricted classes where restrictions are defined by taking the one-time-interaction assumption as a given.

3 Model

The seller has a fixed number $m \in \mathbb{N}$ of service slots in each period $t \in \{0, 1, \dots\}$ and maximizes expected discounted revenue with discount factor $\delta \in (0, 1)$. In every period, an independent random number $n \in \{0, \dots, N\}$ of new buyers arrive such that $Pr(n = k) = \gamma_k$ and $N > m$. Each buyer has unit demand and a value $v_i \in [0, 1]$, which is drawn from CDF F with PDF $f(v_i) > 0$. The profile of values is denoted by $\mathbf{v} = (v_1, \dots, v_n)$.[6] Note that $(n, \mathbf{v})$ is independent across time and dimensions. Let $F_{k:n}$ and $f_{k:n}$ denote the CDF and PDF[7] of the kth smallest value out of n (the kth order statistic). For example, $F_{n:n}$ is the CDF of the highest value of all n newly arrived buyers.

Each buyer interacts with the seller only upon arrival, and buyers' arrival times, as well as their valuations, are observable. A buyer who is served in the arrival period gets payoff $v_i - p_i$, whereas the discounted payoff for serving this buyer one period later is $\delta v_i - p_i$.[8] Buyers who obtain service more than one period after arrival are assumed to derive no value from the service; therefore, service is never offered. If the seller can offer arbitrary contracts, then the one-time interaction is without loss of generality. Consider a mechanism where buyers sometimes interact with the seller later and receive service with some profiles of values for new buyers. The set of all of these profiles can be called a contract. An equivalent mechanism can be constructed in which the seller assigns a buyer a contract at arrival and asks for the total expected transfer as the payment for the contract.

A contract of delayed service given to buyer i is a complete description of all situations in which the buyer will be served. Formally, it is a set $\hat{D}_i \subset \{(n, \mathbf{v}) : n \in \{0, \dots, N\}, \mathbf{v} \in [0, 1]^n\}$. That is, if buyer i receives contract $\hat{D}_i$, buyer i will be served in the next period if and only if the number of buyers n and their values $\mathbf{v}$ are such that $(n, \mathbf{v}) \in \hat{D}_i$.

Contracts from the previous period will restrict the seller's ability to provide instant service to the newly arriving buyers and therefore act as constraints in the maximization problem. To incorporate these constraints into the model, I call the profile of contracts assigned in the previous period the *state*, which is a vector $\mathbf{D} = (D_1, \dots, D_N)$ that includes all contracts assigned

---

[6]Bold font denotes vectors.

[7]$F_{k:n}(v_i) = \sum_{j=k}^{n} \binom{n}{j} F(v_i)^j [1 - F(v_i)]^{n-j}$, and $f_{k:n}(v_i) = \frac{n!}{(k-1)!(n-k)!} F(v_i)^{k-1} [1 - F(v_i)]^{n-k} f(v_i)$.

[8]The term p_i is not discounted because the payment is made at the period of arrival, whereas the service may be provided later. Relaxing the assumption that the buyers' discount factor equals the seller's discount factor would not have much effect on the main results. It would, however, reduce tractability. For example, when the buyers' discount factor is $\beta \gg \delta$, there is a positive mass of values that are optimal to serve regardless of future events; that is, a corner solution is optimal. When $\beta = \delta$, only the extreme value $v_i = 1$ receives the service with certainty with the optimal contract. On the other hand, if $\beta \ll \delta$, the optimal pure calendar mechanism would be static.

in the previous period. If buyer i did not receive a contract, then $D_i = \varnothing$. The initial state is $\mathbf{D} = (\varnothing, \ldots, \varnothing)$.

Because buyers' utility functions are linear in value and quasilinear in transfer, the only variables affecting buyer i's payoff are the transfer $p_i(\mathbf{v})$ and the *discounted expected quantity*, denoted by $q_i(\mathbf{v})$, which is 1 if buyer i receives instant service, 0 if service is refused, and $\delta Pr(\hat{D}_i)$ if delayed service is promised with contract $\hat{D}_i$. Let $\tilde{p}_i(\hat{v}_i)$ and $\tilde{q}_i(\hat{v}_i)$ denote the respective interim values (i.e., the expectation of p_i and q_i over the other buyers' values).

The seller's goal is to find the revenue-maximizing mechanism that satisfies feasibility and individual rationality constraints. In the following sections, we assume the seller observes $(n, \mathbf{v})$ so that there is no need to incentivize truthful reporting. We will later relax this assumption in section 7. In the notation introduced here, the individual rationality constraint (IR) for buyer i with value v_i is $\tilde{q}_i(v_i)v_i - \tilde{p}_i(v_i) \geq 0$. Revenue maximization requires that with any allocation rule, the transfers must be maximal; therefore, the transfer rule must satisfy $\tilde{p}_i(v_i) = \tilde{q}_i(v_i)v_i$.

Although the goal is to find the optimal mechanism when buyers do not know the other buyers' values, all optimal mechanisms will be implementable[9] even when they know the values of other buyers from the same period. I define the periodic ex-post individual rationality (EPIR) constraint for buyer i when values are $\mathbf{v}$ as $q_i(\mathbf{v})v_i - p_i(\mathbf{v}) \geq 0$. The transfer rule that follows from this is $p_i(\mathbf{v}) = q_i(\mathbf{v})v_i$.

Using this transfer rule, we can rewrite the revenue maximization problem as[10]

$$\pi(\mathbf{D}) = \max \mathbb{E}_{n,\mathbf{v}} \left[(1 - \delta) \sum_{i=1}^{n} q_i(\mathbf{v})v_i + \delta \pi(\mathbf{D}') \right], \tag{1}$$

where the seller maximizes over feasible allocation rules, and $\mathbf{D}'$ denotes the continuation state. This expression is separable in $(n, \mathbf{v})$; that is, the seller can choose optimal contracts for each possible combination $(n, \mathbf{v})$ separately, and the value of being in a particular state can then be computed as the expectation over all realizations.

Before considering the dynamic mechanisms, consider a special mechanism that will serve as a useful benchmark. I refer to a mechanism as *static* if the seller never offers delayed service. Thus, in each period, every buyer is either offered instant service or is refused service. This mechanism does not allocate contracts for delayed service, so it is unaffected by the restrictions on contracts, and it is therefore a feasible mechanism in all cases analyzed in this paper.

The static optimum in this framework is the allocation rule that assigns objects to the m buyers with the highest values. This allocation rule is implementable using a standard uniform-price auction without reserve price.

---

[9]The statement is true both with observable and unobservable values. That all optimal mechanisms will be implementable follows from Mookherjee and Reichelstein (1992).

[10]Note that with observable types, the seller's problem is equivalent to maximizing total welfare, so that in the revenue-maximization problem is equivalent to the social welfare maximization problem in terms of allocation rule.

4 Pure calendar mechanism

As the first step, we will discuss the optimal mechanisms in a restricted class, where the set of possible contracts for future service is limited to simple tickets. I define pure calendar mechanisms as mechanisms in which the contracts of delayed service are not conditional on future events. Each buyer is either assigned instant service, refused service, or promised service with certainty in the next period. The advantage of these restricted mechanisms is that they are simple, and the optimal pure calendar mechanism is a strict improvement over static mechanisms.

In the notation introduced above, each contract $\hat{D}_i$ (and therefore each element of the state D_i) is either an empty set or a set of all realizations of $(n, \mathbf{v})$. The notation of state $\mathbf{D} = (D_1, \ldots, D_N)$ can be simplified to a number $z \in \{0, \ldots, m\}$ because it is only necessary to keep track of which buyers received the delayed service and which buyers did not. Moreover, from the seller's point of view, the identities of buyers who were offered delayed service is not payoff-relevant; the only relevant variable is how many buyers were promised delayed service. Therefore, a state z denotes how many service slots are unavailable; thus, $m - z$ slots are still available.

Consider a particular state z. The seller has now $m - z$ instant service slots to allocate and it is clearly optimal to allocate them. These slots are valuable to buyers and costless to allocate because they don't exist in the next period. If there are at most $m - z$ new buyers, then there is nothing left to decide. However, when the number of new buyers is higher, then there are two remaining questions. First, which of the new buyers should get the instant service? Second, when should the seller allocate the future service slots?

The answer to the first question turns out to be simple. It is always optimal to serve buyers with higher valuations earlier. A buyer can receive one of three possible offers: he can receive service instantly, receive delayed service, or never be served. Suppose that a seller offers buyer 1 delayed service and buyer 2 instant service, but the value from service is higher to buyer 1. In this case, just by swapping the offers, the total benefits and therefore the total revenues are increased, whereas total costs are unchanged. The same conclusion is true when we compare offering either instant or delayed service to refusing service. Therefore, the $m - z$ new buyers with the highest valuations should be served instantly and if some buyers receive delayed service, they must be the ones with the highest valuations among the remaining buyers.

The remaining question is how many new buyers should receive delayed service? In this decision, the seller needs to weigh the costs and benefits. Offering delayed service is costly. Assigning each additional unit of next-period service means that it will not be available in the next period and therefore comes with positive opportunity cost. In particular, the optimal pure calendar mechanism is characterized by m thresholds, denoted by $\underline{\mathbf{v}} = (\underline{v}_1, \ldots, \underline{v}_m)$, where $\underline{v}_k$ is the opportunity cost of the kth seat.

To characterize these thresholds, we need to compute the lost revenue from not having the seat available. For this, it is first useful to denote $\pi_k(x) = \int_x^1 [v_{k:n} - x] dF_{k:n}(v_{k:n})$. This function captures the expected revenue from allocating the current service to the buyer with

the kth lowest valuation when the opportunity cost of allocating the slot is x, provided that the buyer's value is at least x. Using this notation, the *direct* opportunity cost of not having the jth slot available is $\pi_{n-(m-j)}(0)$ because the slot would go to the buyer with the value ranked $n - (m - j)$. However, the actual opportunity cost is smaller. If a particular slot becomes unavailable, this means that the first delayed-service slot can go to the buyer with the value ranked $n - (m-j)$ instead of $n - (m-j) - 1$, the second delayed-service slot can go to the buyer with the value ranked $n - (m - j) - 1$ instead of $n - (m - j) - 2$, and so on. Each such swap assigns a delayed service to a buyer with a higher valuation and gives expected revenue increase of $\pi_{n+1-(m-j)-i}(\underline{v}_i) - \pi_{n-(m-j)-i}(\underline{v}_i)$. Equalizing the total expected opportunity cost to the direct revenue that can be extracted from the marginal value $\underline{v}_j$ gives the following equation:

$$\underline{v}_j = E_n \left[\pi_{n-(m-j)}(0) - \delta \sum_{i=1}^m \left(\pi_{n+1-(m-j)-i}(\underline{v}_i) - \pi_{n-(m-j)-i}(\underline{v}_i) \right) \right]. \tag{2}$$

As the proposition 1 shows, the equation system of m equations characterizes a vector of opportunity costs $(\underline{v}_1, \ldots, \underline{v}_m)$, which then fully characterize the optimal pure calendar mechanism.

Proposition 1. *The optimal pure calendar mechanism is such that at each state $z = \{0, \ldots, m\}$, when n buyers arrive and have values $\mathbf{v} = (v_1, \ldots, v_n)$,*

  (i) *the $m - z$ new buyers with the highest valuations v_i will receive instant service.*

  (ii) *if there are more than $m - z$ new buyers, up to the m next highest of them receive delayed service. In particular, the next kth highest of them receives delayed service if and only if $v_{n+1-(m-z)-k:n} > \underline{v}_k$, where the constants $\underline{v}_1 \leq \cdots \leq \underline{v}_m$ are determined by eq. (2).*

Example As an example, consider the special case where there are two new buyers, just one seat is available every period, values are distributed uniformly in $[0, 1]$, and the discount factor is close to one; i.e., $n = 2$, $m = 1$, $v_i \sim \text{unif}[0, 1]$, and $\delta \to 1$. In this case, there are only two states, and the mechanism is fully characterized by a scalar $\underline{v}_1$. In the state $z = 1$, in which instant service is unavailable, the problem is simply selling one product—delayed service—to two buyers with private values (v_1, v_2). The optimal mechanism allocates the service to the high-value buyer whenever that buyer's discounted valuation is higher than the opportunity cost, which gives the condition $v_{2:2} \geq \underline{v}_1$. In the state where instant service is available, allocating it is costless, so it should always be given to the high-value buyer. Moreover, the low-value buyer should receive delayed service whenever that buyer's value is higher than the opportunity cost, which gives the same condition $v_{1:2} \geq \underline{v}_1$. In particular, eq. (2) simplifies in this case to

$$\underline{v}_1 = \pi_2(0) - [\pi_2(\underline{v}_1) - \pi_1(\underline{v}_1)], \tag{2'}$$

where the first term of the right-hand-side is the direct opportunity cost of not being able to assign the seat to the higher-valued buyer and the second term is a correction arising from the fact that the buyer who potentially gets the delayed service would then become the higher-value

11

rather than the lower-value buyer. The solution to eq. (2') is $\underline{v}_1 = \frac{1}{2}$. Figure 2 illustrates these conditions and outcomes.

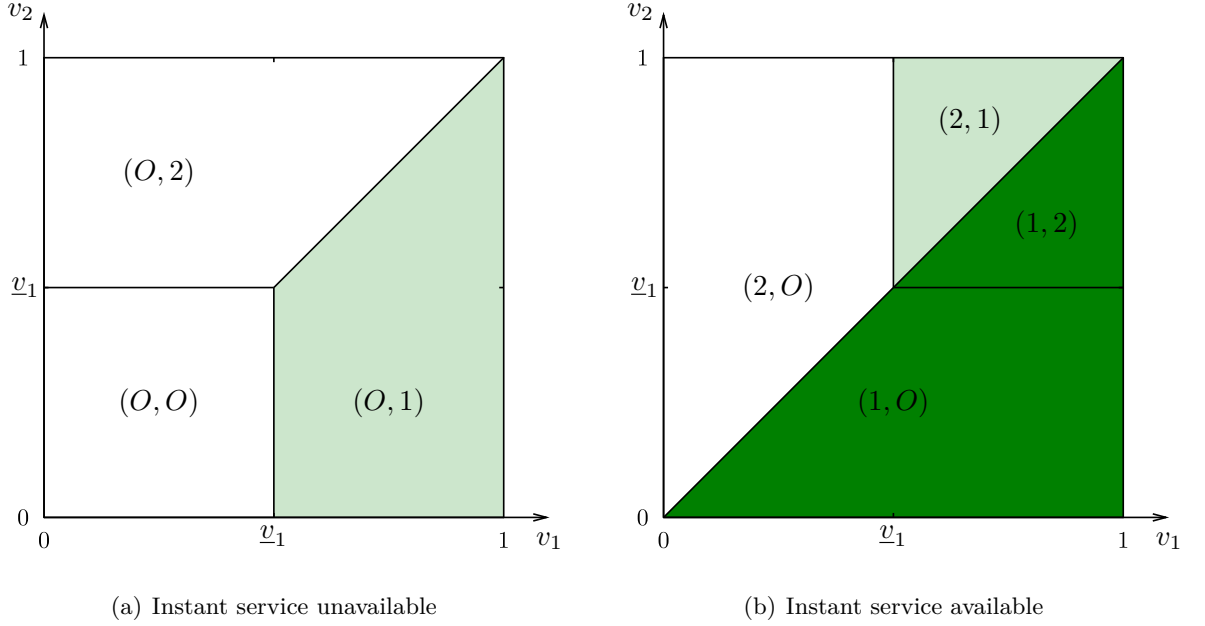(a) Instant service unavailable   (b) Instant service available

Figure 2: The optimal pure calendar mechanism when $m = 1$ and $n = 2$. The shading color intensity indicates the discounted expected quantity for buyer 1. Numbers in parentheses are the identities of the buyers who receive instant and delayed service, respectively, and O denotes other buyers or no buyers.

5   Fully optimal mechanism

In the previous section, the question was how to sell simple tickets optimally. In this section, I will go one step further and study what changes when the seller is able to replace simple tickets with arbitrary contracts of future service. As argued earlier, this is equivalent to a standard dynamic mechanism design problem, where each period the seller can choose which of the remaining buyers will receive the current period service slots, regardless of their arrival times.

I will discuss two questions separately: first, how to best allocate the instant service slots and contracts of delayed service, and second, how to design the contracts optimally. The answer to the first question is straightforward. Namely, among the buyers arriving in the same period, it is always optimal to give higher priority to buyers with the highest valuations. Again, the reason is that the costs of offering instant or delayed service are independent of the valuation of the buyer receiving the service, whereas the benefits are increasing in value. Therefore, when there are enough instant service slots to serve all new buyers, it is optimal to do so. If there are more new buyers, then the buyers with the highest valuations should receive instant service and up to m of the remaining buyers with the highest valuations should receive contracts of delayed

service.

The key question is how to design contracts of next-period service optimally. Before giving the formal result, let me discuss three properties of the optimal contracts. First, buyers from earlier periods are treated preferentially in the sense that they receive service even in situations where some new buyers have higher values and the new buyers get delayed service instead. Consider two buyers with equal valuations, such that the first arrived in the previous period and the second arrived in the current period, and suppose that there is just one current service slot left. By allocating a current service slot to the second buyer, the value to the first buyer is lost. If instead the current slot is allocated to the first buyer, the seller still gets some value from offering the second buyer a contract of future service.

Second, a contract of future service satisfies a threshold rule. That is, if a buyer is promised service for a particular $(n, \mathbf{v})$, then he is also served for all $(n', \mathbf{v}')$ such that $n' \leq n$ and $\mathbf{v}' \leq \mathbf{v}$. Similarly, if he is not served with $(n, \mathbf{v})$, then he is also not served with any $(n', \mathbf{v}')$ such that $n' \geq n$ and $\mathbf{v}' \geq \mathbf{v}$. The buyer's payoff depends only on the probability with which he will be served and not on the particular situations in which the service is offered; therefore, it is always better to serve buyers in situations where the opportunity cost is low. This is true when the demand turns out to be low. When the valuations of the new buyers increase, then the total opportunity cost of not having a slot available increases.

Third, the optimal contracts are such that buyers' payoffs depend non-monotonically on their opponents' values. This also arises from the trade-offs that govern the optimal contracts. The direct cost of not having the slot available is the valuation of a buyer who would take the slot. But the total cost of not having the slot available is typically lower than the value of a buyer who would take the shot because this buyer is then taking the slot of a buyer with lower valuation, and so on. This argument leads to non-monotonicity in other buyer's valuations: new buyers are sometimes better off when they arrive together with other new buyers whose valuations are high, but not higher than their own valuations. Similarly, new buyers are sometimes worse off when they arrive together with other new buyers whose valuations are slightly higher than their own valuations. We will come back to this argument after the formal discussion of the results.

In the notation introduced above, a contract is a set $\hat{D}_i$, which describes under which realizations of $(n, \mathbf{v})$ the buyer will be served next period. The number of available current service slots is therefore limited by the contracts assigned in the previous period, defined as the state $\mathbf{D} = (D_1, \ldots, D_N)$, and choosing the optimal contracts for next-period service is part of the maximization problem.

I denote the contract assigned to the jth highest valued buyers who receive the delayed service by $\hat{D}_j(\cdot)$, and I denote the probability with which the buyer is served in the next period by $\hat{d}_j(\cdot)$; that is, $\hat{d}_j(\cdot) = Pr((n, \mathbf{v}) \in \hat{D}_j(\cdot))$. It turns out that both $\hat{D}_j$ and $\hat{d}_j$ depend only on the value of the buyer who receives contract j, which is the buyer with value ranked $n + 1 - (m - z) - j$ in descending order, where z is the number of seats allocated to buyers from the previous period. To shorten the notation, I denote the profile of contracts by $\hat{\mathbf{D}} = \left(\hat{D}_1, \ldots, \hat{D}_N \right)$.

In particular, the optimal contracts will be such that the total loss from not having a particu-

lar slot available is as small as possible. More formally, I denote the realized maximal profit when j current slots are unavailable and n buyers arrive and have values $\mathbf{v}$ by $\pi^j(n, \mathbf{v})$. In the case of the optimal mechanism, the realized maximal profit expression consists of three components: (1) revenue extracted from the $m - j$ highest-valued buyers who are served instantly, (2) revenue extracted from selling contracts of delayed service to the m buyers having the next-highest values (where $\hat{d}_i$ can be 0), and (3) the continuation profits. This can be written as

$$\pi^j(n, \mathbf{v}) = (1 - \delta) \sum_{i=1}^{m-j} v_{n+1-i:n} + \delta \max_{\hat{\mathbf{D}}} \left\{ (1 - \delta) \sum_{i=1}^{m} Pr(\hat{D}_i) v_{n+1-(m-j)-i:n} + \pi(\hat{\mathbf{D}}) \right\}. \quad (3)$$

The realized loss from not having slot j available is $\pi^{j-1}(n, \mathbf{v}) - \pi^j(n, \mathbf{v})$. On the other hand, the most the seller can extract by assigning the buyer with value $\hat{v}_i$, a slightly higher probability of service, is a slightly larger fraction of the buyer's valuation. Therefore, the optimal contract $\hat{D}_j$ offered to a buyer with value $\hat{v}_i$ is characterized as

$$\hat{D}_j(\hat{v}_i) = \left\{ (n, \mathbf{v}) : \pi^{j-1}(n, \mathbf{v}) - \pi^j(n, \mathbf{v}) \leq (1 - \delta)\hat{v}_i \right\}. \quad (4)$$

Finally, the continuation profit functions $\pi(\mathbf{D})$ above are expectations over realized profits $\pi^j(n, \mathbf{v})$ over $(n, \mathbf{v})$ taking into account that the number of available slots j depends on whether $(n, \mathbf{v})$ is in D_j. In the initial state, the expected revenue is $\pi(\varnothing, \ldots, \varnothing) = \mathbb{E}[\pi^0(n, \mathbf{v})]$. With more restrictive states, expected revenue decreases by the corresponding differences in revenues, so expected revenue is

$$\pi(\mathbf{D}) = \mathbb{E}_{n,\mathbf{v}} \left[\pi^0(n, \mathbf{v}) - \sum_{j=1}^{m} \mathbf{1}_{(n,\mathbf{v}) \in D_j} \left(\pi^{j-1}(n, \mathbf{v}) - \pi^j(n, \mathbf{v}) \right) \right]. \quad (5)$$

Proposition 2. *The optimal (unrestricted) mechanism is such that at state* $\mathbf{D} = (D_1, \ldots, D_N)$, *when n buyers arrive and have values* $\mathbf{v} = (v_1, \ldots, v_n)$,

(i) $z = \#\{i : (n, \mathbf{v}) \in D_i\}$ waiting buyers and $m - z$ new buyers are served instantly.

(ii) The next m highest buyers are offered delayed services with sets

$$\hat{\mathbf{D}} = \left(\hat{D}_1(v_{n+1-(m-z)-1:n}), \hat{D}_2(v_{n+1-(m-z)-2:n}), \ldots, \hat{D}_m(v_{n+1-(m-z)-m:n}) \right)$$

characterized by eqs. (3) to (5).

Example Consider again the example with two new buyers and one seat every period, uniformly distributed values, and discount factor close to one, i.e., $n = 2$, $m = 1$, $v_i \sim \text{unif}[0, 1]$, and $\delta \to 1$. In this case, the optimal contract and allocation rules are illustrated in fig. 3. In particular, suppose that the person receiving the contract of delayed service in the previous period had valuation $\underline{v}$ and the new buyers have valuations v_1 and v_2, respectively. The optimal contract offered to the buyer from the previous period would have to assign him service not only

whenever both new buyers have valuations lower than $\underline{v}$, but also in some profiles (v_1, v_2) where one of the values is slightly higher than $\underline{v}$ if the other valuation is sufficiently low (illustrated by the thick line in fig. 3).
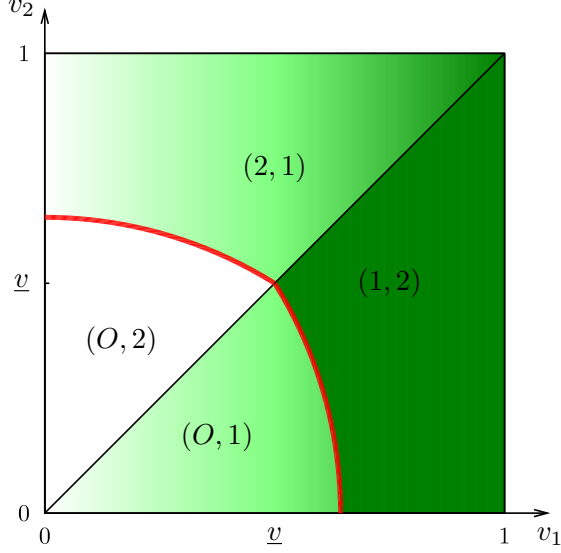

Figure 3: Allocation rule for mechanism when $m = 1$ and $n = 2$. The thick red line indicates the contract offered to a buyer in the previous period (with valuation $\underline{v}$). The shading color intensity indicates the discounted expected quantity for buyer 1. Numbers in parenthesis are the identities of the buyers who receive instant and delayed service, respectively, and O denotes other buyers or no buyers.

The reason for this preferential treatment of the buyer from the previous period is that the optimal contract minimizes the total loss associated with not serving this buyer. Focus first on the 45-degree line where $v_1 = v_2 = \underline{v}$. By offering a contract of delayed service with slightly higher probability, the seller gains a slightly larger fraction of the value $\delta\underline{v}$ as a transfer. However, he loses $\delta\underline{v}$ with the same probability because he can then only offer service to one of the new buyers while the other will be refused service. When we move away from the 45-degree line, the opportunity cost is not as high because the seller can still offer delayed service to the new buyer with a higher valuation.

Figure 3 also illustrates the source of the non-monotonicity of the other buyers' valuations. Suppose that the valuation of new buyer 1 is slightly higher than of the buyer from the previous period who holds an optimal contract, i.e., $v_1 > \underline{v}$. Then if buyer 2 has a low valuation, the contract requires that the buyer from the previous period will be served. Therefore, buyer 1 gets a contract of delayed service. When buyer 2 has a valuation from the intermediate range (i.e., slightly lower than v_1), the total loss from not serving buyer 1 instantly and refusing service to buyer 2 becomes too high; therefore, the contract prescribes that the buyer from the previous period is not served. This means that buyer 1 gets instant service and is better off. However, if buyer 2 has an even higher valuation, then he is the one receiving instant service while buyer 1

receives delayed service again. Similar non-monotonicity in the opposite direction occurs when buyer 1's valuation is below $\underline{v}$. In this case, he receives a contract of delayed service if buyer 2 has either lower value or sufficiently high valuation. But in the intermediate range buyer 1 is not served, which means he is worse off compared to other regions.

6 Waiting-list mechanisms

In this section, I study another restricted class of mechanisms: waiting-list mechanisms. Waiting lists are organized are organized in many ways in real life. In this paper, I focus on a particular characteristic: I assume that for the buyers already on the waiting list, the allocation rule is independent of their value. The seller may choose an optimal way to sell slots on the waiting list , as well as choosing the optimal way to serve buyers from the list. The assumption only requires that once a buyer is on the list, whether he will be served is independent of his valuation. Perhaps, the simplest real life example of this mechanism is the "stand-by list" at the airport. If an airline cannot serve all customers on the current flight, it can assign some customers to the waiting list for the next flight, with the understanding that they will be served only if there is space remaining.

The waiting list mechanism is an extension of the two main results and it serves two purposes. First, it is interesting in its own right as it corresponds to realistic situations where the seller uses a restricted tool to keep track of the promises to buyers. In the language introduced above, this means constructing m different contracts every period[11] instead of the continuum of contracts that is required for the fully optimal mechanism. It will be simpler than the fully optimal mechanism and a strict improvement to the optimal pure calendar mechanism. Second, it illustrates how the ideas from the two main results can be combined to characterize optimal mechanisms in the intermediate cases. The optimal waiting-list mechanism has both the characteristics of the pure calendar mechanisms as well as the fully optimal mechanism.

Formally, at each state, $\mathbf{D} = (D_1, \ldots, D_N)$, a waiting list is characterized by a vector of contracts, $\hat{\mathbf{D}} = (\hat{D}_1, \ldots, \hat{D}_N)$, that are fixed before the realization of $(n, \mathbf{v})$ that is, before the arrival of buyers in this period. Then, upon arrival, some new buyers might be served instantly and $k \in \{0, \ldots, m\}$ new buyers are promised delayed service, and they receive contracts $\hat{D}_1, \ldots, \hat{D}_k$. This is where the waiting-list mechanism is restrictive. The fully optimal mechanism requires offering contracts that are personalized to each buyer, i.e., each contract would be a function of the valuation of the buyer who receives it. Here, however, we assume that the contracts have to be fixed before the realization of $(n, \mathbf{v})$. For example, if buyer i received contract $\hat{D}_k$, it is as if buyer i is assigned a position k on the list after $k-1$ other buyers. Buyer i will be served if and only if the n' new buyers who arrive in the next period have values $\mathbf{v}'$

---

[11]In this section, I allow these m contracts to be different every period and in optimum they will depend on the state. In the states in which many instant service slots are given to buyers from previous periods, the waiting list will be filled with buyers with relatively higher valuations and therefore it is optimal to serve them with relatively higher probability. Of course, the analysis can be extended to the mechanisms in which the contracts of delayed service are the same every period.

such that $(n', \mathbf{v}') \in \hat{D}_k$. This rule is independent of the valuation of buyer i.

To illustrate the assumptions, it is useful to discuss what it means to be in position i (i.e., to receive contract $\hat{D}_i$) on the list. There are two moments in time where the list matters. First, when assigning the buyers to the waiting list, I assume that the seller chooses some buyers to include in the list and assigns them the first contracts $(\hat{D}_1, \hat{D}_2, \dots)$. This implies that it is optimal to make the contracts with lower values more valuable to the buyers, because these contracts are assigned more often. Whenever there are at least i players on the list, there are also $1, \dots, i-1$ players on the list, but not vice versa. Second, the seller must decide in which situations to serve buyers from the list, i.e., design optimal contracts.

Notice that I do not assume that a waiting list is a priority ranking in the sense that being higher on the list must always give the buyer a better chance of getting service. However, this is a property of the optimal waiting list. In particular, by proposition 3, the optimal waiting list is such that $\hat{D}_1 \supset \hat{D}_2 \supset \cdots \supset \hat{D}_m$ and $\hat{D}_k = \varnothing$ for all $k > m$.

The construction of the optimal waiting list combines ideas from the two main results of the paper. On the one hand, each of the contracts is designed analogously to the fully optimal mechanism; the contract includes realizations with the lowest total loss in revenue from not having the slot available. The difference is that in deciding how large the probability should be or, equivalently, how large the largest loss in revenue (denoted by $\bar{b}_j$) should be, the seller designs the contract for the average buyer expected to receive the contract. This means that instead of knowing that the contract will be assigned to a buyer with particular value v_j, the seller knows that v_j will be a random variable according to some distribution that we will denote by G_{v_j}.

Analogous to the fully optimal case (eqs. (3) and (4)), the contracts can be characterized as the set of realizations with the lowest expected loss in revenue from not having the slot available. That is,

$$\hat{D}_j = \left\{ (n, \mathbf{v}) : \pi^{j-1}(n, \mathbf{v}) - \pi^j(n, \mathbf{v}) \leq \bar{b}_j \right\}, \text{where} \tag{6}$$

$$\pi^j(n, \mathbf{v}) = (1 - \delta) \left[\sum_{i=1}^{m-j} v_{n+1-i:n} + \delta \max_k \sum_{i=1}^{k} [\hat{d}_i v_{n+1-(m-j)-i:n} - \underline{v}_i] \right] + \delta \pi(\mathbf{0}). \tag{7}$$

The allocation of contracts follows a similar procedure to the pure calendar mechanism. Because the same contract is designed for buyers with different valuations, allocating it to a buyer comes with a strictly positive opportunity cost. The thresholds $\underline{v}_j$ play the role of these opportunity costs in the allocation problem. Of course, the costs take into account that the contract includes only a subset of realizations, so the opportunity costs are smaller than in the case of the pure calendar mechanism. The following equation is the modified version of eq. (2) for the waiting list case:

$$\underline{v}_j = \frac{\mathbb{E}_{G_{b_j}} \left[b_j \mid b_j \leq \bar{b}_j \right]}{1 - \delta}, \tag{8}$$

where b_j is the (realized) loss in revenue when assigning contract j, and the expectation reflects

the fact that this loss depends on the buyers who arrive in the next period.

The final component of the characterization comes from the local optimality of the contracts. The expected cost of marginally increasing the probability of receiving future service, which is exactly $\bar{b}_j$, must be equal to or greater than the marginal revenue, which is the expected valuation of the buyer who receives the contract. As a result,

$$\frac{\bar{b}_j}{1-\delta} \geq \mathbb{E}_{G_{v_j}} \left[v_j | v_j \geq \underline{v}_j \right]. \tag{9}$$

Proposition 3. *The optimal waiting-list mechanism is such that at state* $\mathbf{D} = (D_1, \ldots, D_N)$, *when* n *buyers arrive and have values* $\mathbf{v}$,

 (i) $z = \#\{i : (n, \mathbf{v}) \in D_i\}$ *waiting buyers and* $m - z$ *new buyers with the highest valuations are served instantly.*

 (ii) *The next* m *buyers with highest values may receive delayed services with sets* $\hat{\mathbf{D}} = \left(\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_m \right)$. *In particular, the next* kth *highest buyer receives a delayed service contract* $\hat{D}_k$ *if and only if* $v_{n+1-(m-z)-k:n} > \underline{v}_k$, *where the thresholds* $\underline{\mathbf{v}}$ *and the contracts* $\hat{\mathbf{D}}$ *are characterized by eqs.* (6) *to* (9).

Example Consider again the example where, in every period, two new buyers arrive and one seat is available, with uniformly distributed valuations and discount factor close to one. In this case, the optimal waiting list has only one position, i.e., there is one contract $\hat{D}_1$. The qualitative properties of the contract are similar to the fully optimal mechanism, i.e,. the contract offers service in situations where the total loss from not having the seat available is relatively low. This means that, again, the payoffs of buyers are non-monotone in their opponents' valuations, for the same reasons as in the case of the fully optimal mechanism. Assigning this contract comes with a strictly positive opportunity cost $\underline{v}_1$, and therefore the contract is assigned only to buyers with valuations above the opportunity cost.

Note that the optimal contract (or the meaning of the waiting list) may change over time: when there are no contracts assigned, the optimal contract is some $\hat{D}_1^0$. Now, if this contract is assigned to a buyer, then in the next period under the circumstances described by $\hat{D}_1^0$, the contract-holder is served and therefore the new buyer with the highest valuation cannot be served. This means the distribution of the valuations of the buyer who is expected to receive a contract of delayed service is now higher and therefore the optimal contract is generally different.

7 Private information

The discussion so far assumes that the seller observes the arrivals and valuations of all buyers. In this section, I extend the analysis to private information and show that the qualitative results will remain unchanged. I also discuss the implementation of the optimal mechanisms in each of the three classes described above. If the buyers are privately informed of their valuations, the

mechanism must guarantee them some information rents, which means that the seller will find it profitable to refuse service to more buyers and sometimes leave slots empty even if there are buyers available. However, the types of mechanisms that are optimal will be the same as those that are optimal with observable valuations. I will also show that buyers do not have incentives to hide their arrivals.

7.1 Unobservable valuations

Let us first consider only private information regarding the valuations, i.e., suppose that each buyer i observes his own valuation v_i, but neither the seller nor the other buyers know this valuation. Moreover, for tractability let us assume that the distribution of valuations, F, satisfies monotone hazard rate condition.[12]

The seller's goal is to find the optimal (Bayesian) incentive-compatible and individually rational mechanism. In the notation introduced above, the incentive compatibility (IC) constraint is $\tilde{q}_i(v_i)v_i - \tilde{p}_i(v_i) \geq \tilde{q}_i(\hat{v}_i)v_i - \tilde{p}_i(\hat{v}_i)$ for all $v_i, \hat{v}_i \in [0,1]$.

Using the standard methods developed by Myerson (1981), it is straightforward to verify that a mechanism satisfies the IR and IC constraints if and only if (1) the discounted expected quantity $\tilde{q}_i$ is weakly increasing in the buyer's own valuation, (2) a buyer with value $v_i = 0$ gets a payoff of at least 0 (exactly 0 in an optimal mechanism), and (3) the transfers are computed from the following ex-ante envelope condition:

$$\tilde{p}_i(v_i) = v_i \tilde{q}_i(v_i) - \int_0^{v_i} \tilde{q}_i(\hat{v}_i)d\hat{v}_i + \tilde{p}_i(0). \tag{10}$$

Again, it turns out that the mechanisms that are optimal under Bayesian incentive-compatibility are implementable even when buyers know the values of other buyers from the same period. The periodic ex-post incentive compatibility (EPIC) constraint is $q_i(\mathbf{v})v_i - p_i(\mathbf{v}) \geq q_i(\hat{v}_i, \mathbf{v}_{-i})v_i - p_i(\hat{v}_i, \mathbf{v}_{-i})$ for all $v_i, \hat{v}_i \in [0,1]$ and all $\mathbf{v}_{-i} \in [0,1]^{n-1}$.

Using standard techniques, it is straightforward to verify that a mechanism satisfies the EPIR and EPIC constraints if and only if (1) q_i is weakly increasing in the buyer's own type v_i, (2) the buyer with valuation $v_i = 0$ gets a payoff of at least 0 (exactly 0 in an optimal mechanism), and (3) the transfers are computed from the following ex-post envelope condition:

$$p_i(v_i, \mathbf{v}_{-i}) = v_i q_i(v_i, \mathbf{v}_{-i}) - \int_0^{v_i} q_i(\hat{v}_i, \mathbf{v}_{-i})d\hat{v}_i + p_i(0, \mathbf{v}_{-i}). \tag{11}$$

Using the ex-ante envelope condition in eq. (10) and the fact that $p_i(0, \mathbf{v}_{-i})$ must be 0 in

---

[12]In particular, $w(v_i) = v_i - \frac{1 - F(v_i)}{f(v_i)}$ is strictly increasing for all v_i.

optimum allows the seller's expected (normalized) revenue at state $\mathbf{D}$ to be expressed as

$$\pi(\mathbf{D}) = \max \mathbb{E}_{n,\mathbf{v}} \left[(1-\delta) \sum_{i=1}^{n} \tilde{p}_i(v_i) + \delta \pi(\mathbf{D}') \right]$$

$$= \max \mathbb{E}_{n,\mathbf{v}} \left[(1-\delta) \sum_{i=1}^{n} q_i(\mathbf{v}) w(v_i) + \delta \pi(\mathbf{D}') \right], \tag{12}$$

where the seller maximizes over feasible promises, $\mathbf{D}'$ is the continuation state, and $w(v_i) = v_i - \frac{1-F(v_i)}{f(v_i)}$ is the standard static virtual value, which is strictly increasing by the monotone hazard rate assumption.

Note that eq. (12) is almost identical to eq. (1) with only one difference—the valuations v_i are replaced with virtual valuations $w(v_i)$, which reflects the fact that buyers get information rents. Therefore, the maximization problem in each of the cases we considered is analogous with the one we solved, with transformed valuations.

Therefore, the results of proposition 1, proposition 2, and proposition 3 still hold, but with the valuations v_i replaced in all cases by virtual valuation $w(v_i)$ and the addition of a qualifier allowing only buyers with positive virtual value to be served.

7.2 Implementation

The characterization results we have discussed here describe the optimal allocation rules for each of the three classes of mechanisms. Equation (11) gives the transfer rules for EPIR and EPIC implementation with a direct mechanism. Of course, if we are only interested in Bayesian incentive-compatible and ex-ante individually rational implementation, there are many other transfer rules that implement the optimal allocation rules. In this subsection, I discuss natural indirect mechanisms that implement the optimal allocation rules.

Let us first consider the optimal pure calendar mechanism. The optimal allocation rule with private values requires that at state z, the $m - z$ instant service slots are allocated to the buyers with the highest positive virtual value $w(v_i)$, which is equivalent to allocating these slots to the buyers with the highest valuations v_i under the condition $v_i \geq r^*$, such that $w(r^*) = 0$. The delayed-service slots should be allocated to the buyers with the highest remaining valuations, provided that the valuations exceed the thresholds characterized above. In particular, let the thresholds that characterize the optimal mechanism with private types be $\underline{\mathbf{w}} = (\underline{w}_1, \ldots, \underline{w}_m)$, where $\underline{w}_j$ is the opportunity cost of the j'th service slot measured in virtual valuation. Then we can find a corresponding vector $\underline{\mathbf{v}} = (\underline{v}_1, \ldots, \underline{v}_m)$, such that $\underline{v}_j = w^{-1}(\underline{v}_j)$ is measured in the true valuation. This allocation rule is implementable using a simple two-stage auction mechanism.

Example For the purpose of illustration, let us consider again the example with $n = 2$ and $m = 1$. In the state where instant service is unavailable, the optimal mechanism allocates the object to the buyer with higher value whenever his valuation is above $\underline{v}_1$. This is implementable through a second-price auction with reserve price $\underline{v}_1$. In the state where the instant service

is available, the mechanism should allocate the instant service to the buyer with higher value whenever the value is above r^* and offer delayed service to the other buyer, but only if his value is above $\underline{v}_1$. This can be implemented as follows. First, the instant service is sold through a standard second-price auction with reserve price r^*. Next, the loser is offered the delayed service at price $\underline{v}_1$. When the loser accepts the offer, the winner of the first-round auction receives a discount $v_{1:2} - \underline{v}_1 > 0$. The reason for the discount is that offering the delayed service creates a surplus for the loser. Without the extra discount, this would make winning in the first round less appealing. Corollary 1 formalizes this idea for general m and n.

Corollary 1. *The optimal pure calendar mechanism is implementable with the following two-stage auction at each state $z \in \{0, \ldots, m\}$:*

(i) *Whenever some instant-service slots are available (i.e., $z < m$), all $m - z$ instant-service slots are sold at a uniform $(m - z + 1)$th-price auction with constant reserve price r^* such that $w(r^*) = 0$.*

(ii) *m delayed-service slots are sold at a uniform $(m - z + 1 + m)$th-price auction with reserve prices $(\delta \underline{v}_1, \ldots, \delta \underline{v}_m)$.*

Whenever $k > 0$ delayed-service slots were sold, all buyers who received the instant service will receive a discount $\delta[v_{n-(m-z):n} - p]$, where p is the price paid in the second auction[13].

The thresholds $\underline{\mathbf{v}} = (\underline{v}_1, \ldots, \underline{v}_m)$ are characterized above.

The indirect implementation of the fully optimal mechanism (as well as the optimal waiting list mechanism) requires running a non-standard auction with a scoring rule, where the rule not only ranks the high bids higher, but also takes into account the extent of the difference between these and other bids.

7.3   Unobservable arrivals

Let us consider the case where the arrivals are unobservable. In particular, let us suppose that buyers may report their arrival one period later (as if they were arriving then) and that they do not observe the state. By announcing their arrival later and claiming value $\hat{v}_i$, a buyer would get an expected discounted-payoff $\delta[\hat{\tilde{q}}_i(\hat{v}_i)v_i - \tilde{p}_i(\hat{v}_i)]$, where $\hat{\tilde{q}}_i(\hat{v}_i)$ is the interim probability that the buyer with value $\hat{v}_i$ is assigned instant service. This term is clearly lower than the interim expected discount factor because it does not include delayed service. Therefore, the deviation is not profitable because

$$\delta[\hat{\tilde{q}}_i(\hat{v}_i)v_i - \tilde{p}_i(\hat{v}_i)] < \tilde{q}_i(\hat{v}_i)v_i - \tilde{p}_i(\hat{v}_i) \leq \max_{\hat{v}_i} \tilde{q}_i(\hat{v}_i)v_i - \tilde{p}_i(\hat{v}_i) = \tilde{U}_i(v_i).$$

Hence, mechanisms that were optimal with observable arrivals will still remain optimal. There is one caveat, however. Not observing the arrivals would require new definitions for the restrictions because late arrivals could make the contracts implicitly contingent on future arrivals. For

---

[13]I.e., $\max \left\{ \underline{v}_k, v_{n-(m-z+m):n} \right\}$.

example, if the seller is constrained to use non-contingent contracts (e.g., pure calendar mechanisms), then the seller might find it profitable to partially relax this restriction by designing the mechanisms so that buyers with some valuations prefer to announce arrivals with delay.

8   Discussion

There are several reasons to use restricted mechanisms in practice. First, the cost of implementing simpler mechanisms could be lower. Second, consumers may prefer to know with certainty whether or not they will receive a service so that they can make complementary investments. The techniques introduced in this paper offer a way to compare the differences in revenues between optimal mechanisms under different restrictions. Although I did not provide comparison results, the revenue from more restrictive mechanisms is strictly lower because the restrictions are binding in each case. The size of the differences depends on the parametric assumptions. For example, when the probability that the need for rationing is low or the discount factor is low, the improvement of any of these dynamic mechanisms compared to the static mechanism will also be small.

I have made several simplifying assumptions that could be relaxed without causing significant changes. First, introducing more general arrival processes would be relatively straightforward. The number of objects m was assumed to be constant, and the number of buyers n was an independently and identically distributed random variable. Alternatively, one could assume that m and n are independent draws from some joint distribution, where the distribution differs each period. This generalization could include, for example, seasonal fluctuations in demand and supply.

Second, in real-life applications, the most common mechanisms for allocating services are posted-price mechanisms. In my model, whenever rationing is needed, the optimal mechanism would involve an auction. The assumption that the seller must use a posted-price mechanism would be another restriction that could be applied to the allocation rule. Alternatively, if we extend the model and assume that the service is provided at discrete points in time, but the buyers arrive in continuous time, then the optimal mechanism would be implementable with posted prices because two buyers never arrive at the same time. This is analogous to Gershkov and Moldovanu (2009) and Board and Skrzypacz (2016).

Third, because this paper examines sales of future services, we might want to know what would happen if the seller was only partially able to commit to delivering the service. This consideration could be added to the model through the introduction of a cost for not fulfilling a promise. If this cost was very high, the seller would have full commitment power; if the cost was very low, there would be no commitment. In this case, the impact to the pure calendar mechanism would be extreme: if the cost is high enough, then the optimal mechanism that we characterized would remain optimal, whereas if the cost is lower, the only feasible pure calendar mechanism would be static, such that no future service would be sold. In both the fully optimal mechanism and in the waiting-list mechanism, the cost would have an impact on the feasible

contracts. The optimal mechanisms in both cases involve offering contracts, which promise service in situations where future demand is not too high, so that the total loss from not using a particular slot falls below a certain threshold. Here the commitment cost would be an additional constraint to the seller that says the total loss from not using the slot must also be lower than the cost of not fulfilling the contract. The qualitative properties of contracts, as well as the allocation rules, would remain otherwise unchanged.

Finally, this paper studied contracts extending only one period into the future and assumed a common discount factor. The model could easily be extended to the case where the buyers' discount factors differ from the seller's discount factor. Studying the structure of longer-term contracts would be important, but is beyond the scope of the current paper.

References

BERGEMANN, D., AND J. VÄLIMÄKI (2010): "The Dynamic Pivot Mechanism," *Econometrica*, 78(2), 771–789.

BLOCH, F., AND N. HOUY (2012): "Optimal assignment of durable objects to successive agents," *Economic Theory*, 51(1), 13–33.

BOARD, S. (2008): "Durable-Goods Monopoly with Varying Demand," *Review of Economic Studies*, 75(2), 391–413.

BOARD, S., AND A. SKRZYPACZ (2016): "Revenue Management with Forward-Looking Buyers," *Journal of Political Economy*, 124(4), 1046–1087.

CONLISK, J., E. GERSTNER, AND J. SOBEL (1984): "Cyclic Pricing by a Durable Goods Monopolist," *The Quarterly Journal of Economics*, 99(3), 489–505.

ELY, J. C., D. GARRETT, AND T. HINNOSAAR (2016): "Overbooking," *Journal of the European Economic Association, forthcoming.*

GARRETT, D. (2016): "Intertemporal price discrimination: dynamic arrivals and changing values," *American Economic Review, forthcoming.*

GERSHKOV, A., AND B. MOLDOVANU (2009): "Dynamic Revenue Maximization with Heterogeneous Objects: A Mechanism Design Approach," *American Economic Journal: Microeconomics*, 1(2), 168–198.

——— (2010): "Efficient sequential assignment with incomplete information," *Games and Economic Behavior*, 68(1), 144–154.

MIERENDORFF, K. (2016): "Optimal dynamic mechanism design with deadlines," *Journal of Economic Theory*, 161, 190–222.

MOOKHERJEE, D., AND S. REICHELSTEIN (1992): "Dominant strategy implementation of Bayesian incentive compatible allocation rules," *Journal of Economic Theory*, 56(2), 378–399.

MYERSON, R. B. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6(1), 58–73.

PAI, M. M., AND R. VOHRA (2013): "Optimal Dynamic Auctions and Simple Index Rules," *Mathematics of Operations Research*, 38(4), 682–697.

PARKES, D. C., AND S. P. SINGH (2003): "An MDP-Based Approach to Online Mechanism Design.," in *NIPS*.

PAVAN, A., I. SEGAL, AND J. TOIKKA (2014): "Dynamic Mechanism Design: A Myersonian Approach," *Econometrica*, 82(2), 601–653.

RILEY, J. G., AND W. F. SAMUELSON (1981): "Optimal Auctions," *The American Economic Review*, 71(3), 381–392.

SAID, M. (2012): "Auctions with dynamic populations: Efficiency and revenue maximization," *Journal of Economic Theory*, 147(6), 2419–2438.

STOKEY, N. L. (1979): "Intertemporal Price Discrimination," *The Quarterly Journal of Economics*, 93(3), 355–371.

ZEITHAMMER, R. (2007): "Research Note-Strategic Bid-Shading and Sequential Auctioning with Learning from Past Prices," *Management Science*, 53(9), 1510–1519.

Appendices

A Proofs

A.1 Pure calendar mechanism

Proof of Proposition 1 As argued in the text, states can be redefined simply as $z \in \{0, \ldots, m\}$, where z denotes the number of current-period service slots promised to buyers in the previous period. If $\pi(z)$ denotes maximum revenue from state z, then $\pi(z)$ is strictly decreasing in z.

Fix state z, the number of buyers n, and the vector of valuations $\mathbf{v}$. Let $\{I, D, R\}$ be a partition of buyers $\{1, \ldots, n\}$ such that I is the set of buyers who are served instantly, D is the set of buyers who receive delayed service, and R is the set of buyers who are refused service. Feasibility requires that $\#I \leq z$, $\#D \leq m$. The following equation expresses the maximum revenue from state z and realization $(n, \mathbf{v})$:

$$\pi^z(n, \mathbf{v}) = (1 - \delta) \sum_{i \in I} v_i + (1 - \delta) \sum_{i \in D} \delta v_i + \delta \pi(\#D).$$

Several immediate observations follow. First, if $z + \#I < m$, so that some of the current slots remain unallocated, this must have occurred because there were enough instant service slots to serve all new buyers, $\#I = n$. Otherwise, the seller could strictly increase revenue by assigning instant service to one more buyer while still satisfying feasibility and not decreasing the continuation payoff. Second, it must be the case that if $i \in I, d \in D$, and $r \in R$, then $v_i \geq v_d \geq v_r$. That is, all values in I are higher than all values in D, which are in turn higher than all values in R. If these inequalities were not true, the seller could swap the service dates of the buyers for whom the order does not hold. This would increase flow revenue while still satisfying feasibility and leaving the continuation revenues unchanged.

Thus, the optimal policy is such that up to $m - z$ buyers having the highest valuations will receive instant service. If there are at most $m - z$ new buyers, they all get the instant service, and the continuation state will be 0. Consider the situation where there are more than $m - z$ new buyers so that the seller offers delayed service to an optimal number $k \in \{0, \ldots, m\}$ of buyers having the highest valuations among the remaining buyers.

Let $\underline{v}_1$ be the opportunity cost of not having the first service slot available. This is the (de-normalized) revenue difference between having all m service slots available and having one slot unavailable, i.e., $\underline{v}_1 = \frac{\pi(0) - \pi(1)}{1 - \delta} > 0$. Similarly, let $\underline{v}_k$ denote the same difference for the kth slot so that $\underline{v}_k = \frac{\pi(k-1) - \pi(k)}{1 - \delta} > 0$. At state z with $(n, \mathbf{v})$, the optimal number of customers who receive delayed service k maximizes the following expression

$$\pi^z(n, \mathbf{v}) = (1 - \delta) \sum_{i=1}^{m-z} v_{n+1-i:n} + (1 - \delta)\delta \sum_{i=1}^{k} v_{n+1-(m-z)-i:n} + \delta\pi(k)$$

$$= (1 - \delta) \sum_{i=1}^{m-z} v_{n+1-i:n} + \delta\pi(0) + (1 - \delta)\delta \sum_{i=1}^{k} \left[v_{n+1-(m-z)-i:n} - \underline{v}_i \right]. \qquad (13)$$

The expected revenue in state z is the expectation of $\pi^z(n, \mathbf{v})$ with respect to the number of buyers n and valuations $\mathbf{v}$ such that

$$\pi(z) = \mathbb{E}_n \Bigg[(1 - \delta) \sum_{i=1}^{m-z} \int_0^1 v \, dF_{n+1-i:n}(v) + \delta\pi(0)$$

$$+ (1 - \delta)\delta \mathbb{E}_{\mathbf{v}} W((v_{n-(m-z):n}, \ldots, v_{n+1-(m-z)-m:n}), \underline{\mathbf{v}}) \Bigg], \qquad (14)$$

where $W((v_{j:n}, \ldots, v_{j+1-m:n}), \underline{\mathbf{v}}) = \max_{k \in \{0, \ldots, m\}} \sum_{i=1}^{k} [v_{j+1-i:n} - \underline{v}_i]$. Let $\hat{\mathbf{\Psi}}(\underline{\mathbf{v}})$ be defined as

$$\hat{\Psi}_j(\underline{\mathbf{v}}) = \frac{\pi(j-1) - \pi(j)}{1 - \delta} = \frac{\mathbb{E}_{n,\mathbf{v}}[\pi^{j-1}(n, \mathbf{v}) - \pi^j(n, \mathbf{v})]}{1 - \delta}. \qquad (15)$$

By definition $\underline{v}_j = \hat{\Psi}_j(\underline{\mathbf{v}})$, so $\underline{\mathbf{v}}$ is a fixed point of $\hat{\mathbf{\Psi}}$; that is, it is defined as $\underline{\mathbf{v}} = \hat{\mathbf{\Psi}}(\underline{\mathbf{v}})$. Instead of showing that $\hat{\mathbf{\Psi}}$ has a fixed point, I am proving a stronger statement: $\hat{\mathbf{\Psi}}$ has a fixed point in which the coordinates are monotone; that is, $0 \leq \underline{v}_1 \leq \cdots \leq \underline{v}_m \leq 1$. Let us denote the set of all such vectors by $S = \{\mathbf{v} : 0 \leq \underline{v}_1 \leq \cdots \leq \underline{v}_m \leq 1\}$, and take $\underline{\mathbf{v}} \in S$. Then $W((v_j, \ldots, v_{j+1-m}), \underline{\mathbf{v}})$

simplifies to $\sum_{i=1}^{m} \max\{0, v_{j+1-i} - \underline{v}_i\}$ so that the expectation can be taken separately. Defining $\pi_k(\underline{v}_i) = \mathbb{E}_{v_{k:n}}[v_{k:n} - x] = \int_x^1 [v_{k:n} - x] dF_{k:n}(v_{k:n})$ gives the expression

$$\pi(z) = \mathbb{E}_n\left[(1 - \delta) \sum_{i=1}^{m-z} \pi_{n+1-i}(0) + \delta\pi(0) + (1-\delta)\delta \sum_{i=1}^{m} \pi_{n+1-(m-z)-i}(\underline{v}_i)\right].$$

Therefore, if $\underline{\mathbf{v}} \in S$, then $\hat{\boldsymbol{\Psi}}$ simplifies to $\boldsymbol{\Psi}$ such that

$$\Psi_j(\underline{\mathbf{v}}) = \mathbb{E}_n\left[\pi_{n-(m-j)}(0) - \delta \sum_{i=1}^{m} \left(\pi_{n+1-(m-j)-i}(\underline{v}_i) - \pi_{n-(m-j)-i}(\underline{v}_i)\right)\right]. \tag{16}$$

To complete the proof, it remains to show that $\boldsymbol{\Psi}$ has a fixed point in S. Notice that Ψ_j is a continuous function, and S is convex and compact. Therefore, it is sufficient to verify that $\boldsymbol{\Psi}(\underline{\mathbf{v}}) \in S$ for all $\underline{\mathbf{v}} \in S$ and then apply Brouwer's fixed point theorem.

First, note that

$$\frac{\partial \Psi_j(\underline{\mathbf{v}})}{\partial \underline{v}_i} = \mathbb{E}_n\left[-\delta\left(F_{n+1-(m-j)-i}(\underline{v}_i) - F_{n-(m-j)-i}(\underline{v}_i)\right)\right] > 0.$$

Therefore,

$$\Psi_1(\underline{\mathbf{v}}) \geq \Psi_1(\mathbf{0}) = \mathbb{E}_n\left[(1-\delta)\pi_{n-(m-j)}(0) + \delta\pi_{n-(m-j)-m}(0)\right] \geq 0, \tag{17}$$

$$\Psi_m(\underline{\mathbf{v}}) \leq \Psi_m(\mathbf{1}) = \mathbb{E}_n\left[\pi_{n-(m-j)}(0)\right] \leq 1. \tag{18}$$

Let us define $\Delta_j(x) = \pi_{j+1}(x) - \pi_j(x) = \int_{\underline{v}}^1 [F_{j:n}(v) - F_{j+1:n}(v)]dv$. Then $\Delta_j(x) > 0$ because of the stochastic dominance of higher order statistics. Therefore, $\frac{\partial \pi_j}{\partial \underline{v}} = -[1 - F_j(\underline{v})] < 0$, so $\frac{\partial \Delta_j(x)}{\partial x} = F_{j+1:n}(x) - F_{j:n}(x) < 0$. Compactly denoting $k = n - (m - j)$ and $\underline{v}_0 = 0$, eq. (16) can be rewritten as

$$\Psi_{j+1}(\underline{\mathbf{v}}) - \Psi_j(\underline{\mathbf{v}}) = \mathbb{E}_n\left[\Delta_k(0) - \delta \sum_{i=1}^{m} \left(\Delta_{k+1-i}(\underline{v}_i) - \Delta_{k-i}(\underline{v}_i)\right)\right]$$

$$= \mathbb{E}_n\left[(1-\delta)\Delta_k(0) + \delta \sum_{i=1}^{m} \left(\Delta_{k+1-i}(\underline{v}_{i-1}) - \Delta_{k+1-i}(\underline{v}_i)\right) + \delta\Delta_{k-m}(\underline{v}_m)\right] > 0$$

because each Δ_j is positive and decreasing and each $\underline{v}_{i-1} \leq \underline{v}_i$. Therefore, $0 \leq \Psi_1(\underline{\mathbf{v}}) \leq \Psi_2(\underline{\mathbf{v}}) \leq \cdots \leq \Psi_m(\underline{\mathbf{v}}) \leq 1$; thus, $\boldsymbol{\Psi}(\underline{\mathbf{v}}) \in S$. $\qquad\square$

A.2 Fully optimal mechanism

Proof of Proposition 2 Let $\pi(\mathbf{D})$ denote the maximized revenue at state $\mathbf{D}$, and, with some abuse of notation, let $\pi(\mathbf{d})$ denote the maximized revenue for a given vector of probabilities $\mathbf{d}$.

That is, $\pi(\mathbf{d}) = \max_{\mathbf{D}:Pr(D_i)\geq d_i} \pi(\mathbf{D})$. I assume that these functions are well defined and $\pi(\hat{\mathbf{d}})$ is differentiable, and I later verify that they are. After standard manipulations, revenue can be expressed recursively as

$$\pi(\mathbf{d}) = \max \mathbb{E}_{n,\mathbf{v}} \left[(1-\delta) \sum_{i=1}^{n} q_i(\mathbf{v})v_i + \delta\pi(\hat{\mathbf{d}}) \right], \tag{19}$$

where $q_i(\mathbf{v}) = 1$ if buyer i is assigned instant service, $q_i(\mathbf{v}) = 0$ if buyer i is refused service, and $q_i(\mathbf{v}) = \delta\hat{d}_i$ if buyer i is assigned a delayed-service contract guaranteeing service with probability $\hat{d}_i$. A few immediate observations follow: First, it is better to serve high-value buyers before serving low-value buyers. Second, only up to m new buyers are offered delayed service. Third, contracts must be ordered both in probabilities $\hat{d}_1 \geq \cdots \geq \hat{d}_m$ and in sets $\hat{D}_1 \supset \cdots \supset \hat{D}_m$.

Let $\pi^j(n,\mathbf{v})$ denote the maximum revenue that the seller can achieve when j current service slots are unavailable so that up to $m-j$ of the new buyers can be served instantly. Because of monotonicity in the order of the promises, buyers with the highest values $v_{n:n},\ldots,v_{n+1-(m-j):n}$ receive instant service and buyers with next highest values $v_{n-(m-j):n},\ldots,v_{n+1-(m-j)-m:n}$ receive the corresponding contracts such that $\hat{d}_1 \geq \cdots \geq \hat{d}_m$. Therefore,

$$\pi^j(n,\mathbf{v}) = \max_{\hat{\mathbf{d}}} (1-\delta) \sum_{i=1}^{m-j} v_{n+1-i:n} + (1-\delta)\delta \sum_{i=1}^{m} \hat{d}_i v_{n+1-(m-j)-i:n} + \delta\pi(\hat{\mathbf{d}}). \tag{20}$$

Differentiation with respect to the probability of receiving the delayed service, $\hat{d}_i$, gives

$$(1-\delta)\delta v_{n+1-(m-j)-i} + \delta\frac{\partial\pi(\hat{\mathbf{d}})}{\partial\hat{d}_i}. \tag{21}$$

So far, the optimal probabilities $\hat{\mathbf{d}}$ have been determined. Next, I characterize the contracts $\hat{\mathbf{D}}$. I argued above that the contracts are ordered. This allows us to express the expected revenue in state $\hat{\mathbf{D}}$ as

$$\pi(\hat{\mathbf{D}}) = \mathbb{E}_{n,\mathbf{v}} \left[\mathbf{1}[\hat{D}_1^c]\pi^0(n,\mathbf{v}) + \sum_{j=1}^{m-1} \mathbf{1}[\hat{D}_j \setminus \hat{D}_{j+1}]\pi^j(n,\mathbf{v}) + \mathbf{1}[\hat{D}_m]\pi^m(n,\mathbf{v}) \right]$$

By adding and subtracting terms $\mathbf{1}[\hat{D}_1]\pi^0(n,\mathbf{v})$ and $\mathbf{1}[\hat{D}_j][\pi^{j-1}(n,\mathbf{v}) - \pi^j(n,\mathbf{v})]$ for each $j = 2,\ldots,m$ and taking into account that $\pi(\hat{\mathbf{d}}) = \max_{\hat{\mathbf{D}}:Pr(\hat{D}_j)\geq\hat{d}_j} \pi(\hat{\mathbf{D}})$ we get

$$\pi(\hat{\mathbf{d}}) = \mathbb{E}_{n,\mathbf{v}} \left[\pi^0(n,\mathbf{v}) \right] - \sum_{j=1}^{m} \min_{\hat{\mathbf{D}}:Pr(\hat{D}_j)\geq\hat{d}_j} \mathbb{E}_{n,\mathbf{v}} \left[\mathbf{1}[\hat{D}_j][\pi^{j-1}(n,\mathbf{v}) - \pi^j(n,\mathbf{v})] \right]. \tag{22}$$

Because both the objective and the constraints are separable in j, each contract $\hat{D}_j$ can be separately chosen to include realizations with the smallest value for $\pi^{j-1}(n,\mathbf{v}) - \pi^j(n,\mathbf{v})$ to satisfy $Pr(\hat{D}_j) \geq \hat{d}_j$. Thus, the optimal contracts can be expressed as $\hat{D}_j = \{(n,\mathbf{v}) : \pi^{j-1}(n,\mathbf{v}) - \pi^j(n,\mathbf{v}) \leq \bar{b}_j\}$ for unique $\bar{b}_j$ such that $Pr(\hat{D}_j) = \hat{d}_j$. I call $\bar{b}_j$ the maximum loss in

revenue (from not having the j'th slot available).

Because there is a one-to-one mapping between probability $\hat{d}_j$ and $\bar{b}_j$, the optimization variable can be changed to $\bar{b}_j$. Therefore, $\frac{\partial \pi(\hat{d})}{\partial \hat{d}_j} = -\bar{b}_j$, which is independent of $\hat{d}_i$ for $i \neq j$. Moreover, if $\hat{d}_j = 1$ (i.e., the buyer receives service with certainty), then $\hat{D}_j$ must include all realizations of $(n, \mathbf{v})$; consequently, $\bar{b}_j = 1 - \delta$, so $\frac{\partial \pi(\cdot, 1, \cdot)}{\partial \hat{d}_j} = -(1 - \delta)$. If $\hat{d}_j \leq Pr(n < m - j) = \sum_{k=0}^{m-j-1} \gamma_k$, the constraint is not binding as with this probability there are not enough new buyers and the $m - j$'th slot would be unused anyway. In this case $\frac{\partial \pi(\cdot, \hat{d}_j, \cdot)}{\partial \hat{d}_j} = 0$ and without loss of generality $\hat{d}_j = 0$. Otherwise, if $\hat{d}_j > Pr(n < m - j)$, then $\bar{b}_j > 0$ and $\frac{\partial \pi}{\partial \hat{d}_j} = -\bar{b}_j < 0$.

I have shown that the continuation profit function $\pi(\hat{\mathbf{d}})$ is decreasing in each argument and strictly decreasing in $\hat{d}_j$ whenever $\hat{d}_j > Pr(n < m - j)$. Moreover, it must be concave over this range. To see this, consider $\hat{d}_j' > \hat{d}_j > Pr(n < m - j)$. Thus, it must be true that $\hat{D}_j \subsetneq \hat{D}_{j'}$, so $\bar{b}_j < \bar{b}_{j'}$. Therefore, $\frac{\partial \pi(\cdot, \hat{d}_j, \cdot)}{\partial \hat{d}_j} = -\bar{b}_j > -\bar{b}_{j'} = \frac{\partial \pi(\cdot, \hat{d}_{j'}, \cdot)}{\partial \hat{d}_j}$, and $\frac{\partial^2 \pi}{\partial \hat{d}_j^2} > 0$.

Next, consider the first-order condition, eq. (21). Consider the upper-corner solution $\hat{d}_j = 1$ first. The optimum is at this corner only if $(1 - \delta)\delta v_{n+1-(m-z)-j:n} \geq \delta(1 - \delta)$, which implies $v_{n+1-(m-z)-j} = 1$. Next, consider the lower-corner solution such that $\frac{\partial \pi}{\partial \hat{d}_j} = 0$. For analogous reasons, this is optimal only if $v_{n+1-(m-z)-j} = 0$. Finally, consider interior solutions, $\hat{d}_j \in (Pr(n < m - j), 1)$, which must satisfy $(1 - \delta)\delta v_{n+1-(m-z)-j} = -\delta\frac{\partial \pi}{\partial \hat{d}_j} = \delta\bar{b}_j$. Therefore, $\hat{d}_j$ is strictly increasing in $v_{n+1-(m-z)-j}$ whenever $v_{n+1-(m-z)-j} \in (0, 1)$.

To complete the proof, I need to show that functions $\pi(\mathbf{d})$ are well defined and differentiable. The characterization above is a mapping from a continuation value function $\pi(\hat{\mathbf{d}})$ to the current expected-revenue $\pi(\mathbf{d})$. I denote this mapping by $T\pi$. First, I use Blackwell's sufficient conditions to show that the mapping T is a contraction with a speed of convergence δ. Then, the contraction mapping theorem implies that $\pi(\mathbf{d})$ exists.

From the recursive formulation in eq. (19), mapping T maps bounded functions $\pi(\hat{\mathbf{d}})$ to bounded functions because in optimum $\sum_i q_i(\mathbf{v})v_i \in [0, N]$. If there are two functions, π and $\hat{\pi}$ such that $\hat{\pi}(\hat{\mathbf{d}}) \geq \pi(\hat{\mathbf{d}})$ for all $\hat{\mathbf{d}}$, then $T\hat{\pi}(\hat{\mathbf{d}}) \geq T\hat{\pi}(\hat{\mathbf{d}})$ for all $\hat{\mathbf{d}}$ because at each state with any set of newly assigned contracts the continuation revenue is increased and the flow revenue remains unchanged. Therefore, the optimal contracts cannot lead to lower revenue, and monotonicity is satisfied. To verify discounting, suppose $\hat{\pi}(\hat{\mathbf{d}}) = \pi(\hat{\mathbf{d}}) + a$. Each $\pi^j(n, \mathbf{v})$ will remain the same except for the addition of δa. In the dynamic part of the problem, δa terms cancel in $\pi^{j-1}(n, \mathbf{v}) - \pi^j(n, \mathbf{v})$, so $T\hat{\pi}(\hat{\mathbf{d}}) = \mathbb{E}_{n,\mathbf{v}}[\pi^0(n, \mathbf{v}) + \delta a] = T\pi(\hat{\mathbf{d}}) + \delta a$. Therefore, T satisfies Blackwell's sufficient conditions, and $\pi = T\pi$ has a unique fixed point by the contraction mapping theorem.

Finally, it suffices to verify that the contraction mapping maps differentiable functions back to differentiable functions. This is true as $\pi(\mathbf{d})$ is an expectation of $\pi^j(n, \mathbf{v})$ over the sets $(D_1, \ldots, D_m)$. It is straightforward to see that each π^j is continuous, and since each set D_j includes the points with lowest $\pi^j(n, \mathbf{v}) - \pi^{j-1}(n, \mathbf{v})$, the expectation is differentiable in their measures $d_j = Pr(D_j)$. $\qquad\square$

A.3 Waiting list

Proof of Proposition 3 Let $\pi(\mathbf{D})$ and $\pi(\mathbf{d}) = \max_{\mathbf{D}:Pr(D_i)\geq d_i} \pi(\mathbf{D})$ be well-defined functions and $\pi(\mathbf{d})$ is differentiable. Again, the revenue can be written recursively as

$$\pi(\mathbf{d}) = \max \mathbb{E}_{n,\mathbf{v}} \left[(1-\delta) \sum_{i=1}^{n} q_i(\mathbf{v})v_i + \delta\pi(\hat{\mathbf{d}}) \right], \tag{23}$$

where $q_i(\mathbf{v}) = 1$ if buyer i is assigned instant service, $q_i(\mathbf{v}) = 0$ if he is refused service, and $q_i(\mathbf{v}) = \delta\hat{d}_i$ if he is offered a contract for delayed service guaranteeing service with probability $\hat{d}_i$.

Several immediate observations follow: First, the higher types are again served earlier in the following sense. Suppose that buyer i is served instantly, buyers d and d' are offered delayed service with respective contracts $\hat{D}_d$ and $\hat{D}_{d'}$ such that $\hat{d}_d > \hat{d}_{d'}$, and buyer r is refused service; it must then be true that $v_i \geq v_d \geq v_{d'} \geq v_r$. Otherwise, the seller could strictly increase revenue by swapping the allocation assigned to these buyers. Second, the seller will never assign a positive probability of delayed service to more than m buyers. Otherwise, it would be possible to take the lowest-valued buyer who receives a positive probability of service and reassign that buyer's probability to higher-valued buyers without affecting feasibility or continuation states while also increasing flow revenue. Finally, contracts are ordered so that $\hat{D}_1 \supset \cdots \supset \hat{D}_m$. This is a priority ranking; the receiver of the first contract will always be served before the receiver of the second contract. If the ordering were not true for some j, the seller could replace promises for $\hat{D}_{j-1}$ and $\hat{D}_j$ with new promises $\hat{D}'_{j-1} = \hat{D}_{j-1} \cup \hat{D}_j$ and $\hat{D}'_j = \hat{D}{j-1} \cap \hat{D}_j$ while keeping other contracts unchanged. The new contracts would not violate feasibility but would extract more revenue from current buyers.

Suppose first that the contracts $\hat{\mathbf{D}}$ are already optimally designed for state $\mathbf{D}$. Fix realization $(n,\mathbf{v})$, and denote the number of unavailable slots by $z = \#\{i : (n,\mathbf{v}) \in D_i\}$. By the observations above, the $m-z$ highest-valued buyers having strictly positive value will receive instant service, and buyers with negative virtual value will never be served. If there are more than $m-z$ buyers with strictly positive virtual value, the k highest of them will receive contracts $\hat{D}_1, \ldots, \hat{D}_k$. The question then becomes how to choose optimal k.

Let $\underline{v}_1$ be the opportunity cost of assigning the first contract $\hat{D}_1$ (i.e., the first slot in the waiting list). This is the (de-normalized) difference between revenue when no assigned contracts are given as compared to revenue when one buyer has received contract $\hat{D}_1$; that is, $\underline{v}_1 = \frac{\pi(0,\ldots,0)-\pi(\hat{D}_1,0,\ldots,0)}{1-\delta}$. More generally, let the opportunity contract of assigning contract $\hat{D}_i$ be $\underline{v}_i$, which can be expressed as

$$\underline{v}_i = \frac{\pi(\hat{D}_1,\ldots,\hat{D}_{i-1},0,\ldots,0) - \pi(\hat{D}_1,\ldots,\hat{D}_i,0,\ldots,0)}{1-\delta}. \tag{24}$$

Let $\pi^j(n,\mathbf{v})$ denote the maximum revenue when j instant-service slots are unavailable at

realization $(n, \mathbf{v})$. Maximum revenue can be expressed as

$$\pi^j(n, \mathbf{v}) = (1 - \delta) \sum_{i=1}^{m-j} v_{n+1-i:n} + \delta\pi(\mathbf{0}) + (1 - \delta)\delta \max_{k \in \{0,\dots,m\}} \sum_{i=1}^{k} [\hat{d}_i v_{n+1-(m-j)-i:n} - \underline{v}_i].$$

The next question is how to choose particular contracts $\hat{\mathbf{D}}$ for a given vector $\hat{\mathbf{d}}$ of probabilities of receiving the service. Using the fact that $\hat{D}_1 \supset \cdots \supset \hat{D}_m$, the expected revenue from a given $\hat{\mathbf{d}}$ can be expressed as

$$\pi(\hat{\mathbf{d}}) = \max_{\hat{\mathbf{D}}: Pr(\hat{D}_i) \geq \hat{d}_i} \left\{ \mathbb{E}_{n,\mathbf{v}} \left[\mathbf{1}[\hat{D}_1^c]\pi^0(n, \mathbf{v}) + \mathbf{1}[\hat{D}_1 \setminus \hat{D}_2]\pi^1(n, \mathbf{v}) + \cdots + \mathbf{1}[\hat{D}_m]\pi^m(n, \mathbf{v}) \right] \right\}$$

$$= \mathbb{E}_{n,\mathbf{v}}[\pi^0(n, \mathbf{v})] - \sum_{j=1}^{m} \min_{\hat{D}_j: Pr(\hat{D}_j) \geq \hat{d}_j} \mathbb{E}_{n,\mathbf{v}} \left[\mathbf{1}[\hat{D}_j][\pi^{j-1}(n, \mathbf{v}) - \pi^j(n, \mathbf{v})] \right]. \tag{25}$$

Note that the minimization problem in eq. (25) is separable in j because both the objective and the constraint depend only on $\hat{D}_j$ and $\hat{d}_j$. The difference in revenues, $\pi^{j-1}(n, \mathbf{v}) - \pi^j(n, \mathbf{v})$, is equal to 0 whenever there are fewer than $m - j$ new buyers and is strictly increasing otherwise. That is, the difference in revenues is strictly positive and strictly increasing whenever $n \geq m - j$.

There are two cases to be considered. When $(n, \mathbf{v})$ such that $n < m - j$, we have that $\pi^{j-1}(n, \mathbf{v}) - \pi^j(n, \mathbf{v}) = 0$. Therefore, if $\hat{d}_j < Pr(n < m - j) = \sum_{k=0}^{m-j-1} \gamma_k$, by increasing the probability $\hat{d}_j$ to $Pr(n < m - j)$, the seller can increase revenue without violating feasibility or decreasing the continuation value. Therefore, the optimum must lie in the second case, where $\hat{d}_j \geq Pr(n < m - j)$. In this case, the contract set $\hat{D}_j$ must include some points where $v_{n-(m-j):n} \geq 0$ so that $\pi^{j-1}(n, \mathbf{v}) - \pi^j(n, \mathbf{v})$ is positive. Because $\pi^{j-1}(n, \mathbf{v}) - \pi^j(n, \mathbf{v})$ is therefore strictly increasing in $v_{n-(m-j):n}$, the constraint must be binding; that is, each contract $\hat{D}_j$ must be such that the probability is $\hat{d}_j$ and the realizations included in the set are the ones with the smallest values of $\pi^{j-1}(n, \mathbf{v}) - \pi^j(n, \mathbf{v})$. Thus, the contract $\hat{D}_j$ can be characterized by the upper bound to the loss in revenue, $\pi^{j-1}(n, \mathbf{v}) - \pi^j(n, \mathbf{v})$, denoted by $\bar{b}_j$. Then the optimal contract must satisfy $\hat{D}_j = \{(n, \mathbf{v}) : \pi^{j-1}(n, \mathbf{v}) - \pi^j(n, \mathbf{v}) \leq \bar{b}_j\}$. Because $\bar{b}_j$ is strictly increasing with $\hat{d}_j$ in the relevant region, there is a one-to-one mapping between $\hat{d}_j \in [Pr(n < m - j), 1]$ and $\bar{b}_j \in [0, 1 - \delta]$.

Next, I define two new distribution functions. The first, G_{b_j}, is the probability of the realized cost of not having the slot available, defined as $G_{b_j}(x) = Pr(\{(n, \mathbf{v}) : \pi^{j-1}(n, \mathbf{v}) - \pi^j(n, \mathbf{v}) \leq x\})$. The density is defined correspondingly as $g_{b_i}(x) = G'_{b_i}(x)$. Using this notation, we can express the mapping between the probability of delayed service and the corresponding maximum loss in revenue $\bar{b}_j$ by $\hat{d}_j = G_{b_j}(\bar{b}_j)$. The opportunity cost of not having slot j available must be equal to the average loss in revenue, which can now be computed as $\underline{v}_j = \frac{\mathbb{E}_{G_{b_j}}[b_j | b_j \leq \bar{b}_j]}{1 - \delta}$.

The second distribution function is for the valuation of the average buyer expected to be impacted by the contract D_j, denoted by[14] G_{v_j}. It is a mixture of valuations $v_{n+1-(m-j):n}, \cdots,$

---

[14]To shorten the notation, D_0 denotes the full set and $D_{m+1} = \varnothing$.

$v_{n+1-(m-j)+m:n}$ with partition $\{D_1^c, D_1 \setminus D_2, \ldots, D_{m-1} \setminus D_m, D_m\}$, defined as

$$g_{v_j}(x) = \mathbb{E}_n \left[\sum_{i=0}^{m} \mathbf{1}[D_i \setminus D_{i+1}] f_{n+1-(m-i)-j:n}(x) \right], \quad G_{v_j}(y) = \int_0^y g_{v_j}(x) dx. \tag{26}$$

The marginal effect of $\bar{b}_j$ to the revenue $\pi^i(n, \mathbf{v})$ can now be computed as

$$\begin{cases} 0 & \forall v_{n+1-(m-i)-j:n} \leq \underline{v}_j, \\ (1-\delta)\delta g_{b_j}(\bar{b}_j) v_{n+1-(m-i)-j:n} - \delta \bar{b}_j g_{b_j}(\bar{b}_j) & \forall v_{n+1-(m-i)-j:n} > \underline{v}_j. \end{cases}$$

Therefore, the optimality condition from the expected revenue at state $\mathbf{D}$ with respect to $\bar{b}_j$ is

$$0 = \mathbb{E}_{n,\mathbf{v}} \left[\mathbf{1}[D_1^c] \frac{\partial \pi^0}{\partial \bar{b}_j} + \mathbf{1}[D_1 \setminus D_2] \frac{\partial \pi^1}{\partial \bar{b}_j} + \cdots + \mathbf{1}[D_{m-1} \setminus D_m] \frac{\partial \pi^m}{\partial \bar{b}_j} + \mathbf{1}[D_m] \frac{\partial \pi^m}{\partial \bar{b}_j} \right],$$

which can be rewritten as

$$\frac{\bar{b}_j}{1-\delta} = \int_{\underline{v}_j}^1 \hat{v}_j \frac{dG_{v_j}(\hat{v}_j)}{1 - G_{v_j}(\underline{v}_j)} = \mathbb{E}_{G_{v_j}} \left[\hat{v}_j | \hat{v}_j > \underline{v}_j \right]. \tag{27}$$

That is, the maximum loss in revenue from the jth contract must be equal to the average value of the customer to whom the seller expects to allocate this contract.

Using the definition of the opportunity cost $\underline{v}_j$,

$$\underline{v}_j = \frac{\int_0^{\bar{b}_j} b_j dG_{b_j}(b_j)}{(1-\delta) G_{b_j}(\bar{b}_j)} = \frac{\mathbb{E}_{G_{b_j}}[b_j | b_j \leq \bar{b}_j]}{1-\delta}. \tag{28}$$

For any $j \in \{1, \ldots, m\}$, eqs. (27) and (28) relate $\underline{v}_j$ and $\bar{b}_j$. Both equations describe a strictly increasing continuous relationship between $\bar{b}_j$ and $\underline{v}_j$. First, consider eq. (27). At the lower bound, $\underline{v}_j = 0$, and $\frac{\bar{b}_j}{1-\delta} = \mathbb{E}_{G_{v_j}}[\hat{v}_j | \hat{v}_j > 0] \in (0, 1)$. Near the upper bound, $\underline{v}_j = 1$, and $\frac{\bar{b}_j}{1-\delta} = \mathbb{E}_{G_{v_j}}[\hat{v}_j | \hat{v}_j \geq 1] = 1$. Next, consider eq. (28). At the lower bound, $\bar{b}_j = 0$, and $\underline{v}_j = \frac{\mathbb{E}_{G_{b_j}}[b_j | b_j \leq 0]}{1-\delta} = 0$. At the upper bound, $\bar{b}_j = 1 - \delta$, and $\underline{v}_j = \frac{\mathbb{E}_{G_{b_j}}[b_j]}{1-\delta} \in (0, 1)$. Therefore, either there exists an interior solution $(\underline{v}_j, \bar{b}_j)$ such that the probability $\hat{d}_j < 1$ or that the first-order condition in eq. (27) always holds as a strict equality. When eq. (27) always holds as a strict equality, the optimal $\bar{b}_j = 1 - \delta$, and consequently $\hat{d}_j = 1$. In both cases, the two equations uniquely characterize the pair $(\underline{v}_j, \bar{b}_j)$.

To complete the analysis, we need to verify that the functions $\pi(\mathbf{d})$ indeed exist. Analogous to the fully optimal mechanism, the characterization is a mapping T from a continuation value function $\pi(\hat{\mathbf{d}})$ to the current expected revenue function $\pi(\mathbf{d})$. As in the proof for the fully optimal case, Blackwell's sufficient conditions hold, so the contraction mapping theorem applies and proves the existence of a fixed point $\pi = T\pi$. Differentiability of $\pi(\mathbf{d})$ can also be verified in the same way as with the fully optimal mechanism. $\qquad \square$