Calendar mechanisms

Toomas Hinnosaar[†]

September 2013

Abstract

I study a repeated mechanism design problem where a revenue-maximizing monopolist sells a fixed number of service slots to randomly arriving buyers with private values and increasing exit rates. In addition to characterizing the fully optimal mechanism, I study the optimal mechanisms in two restricted classes. First, the pure calendar mechanism, where the seller allocates future service dates instead of general promises. The unique optimal pure calendar mechanism is characterized in terms of the opportunity costs of allocating additional service slots. Second, I analyze the waiting list mechanism, where promises of delayed service can depend on future arrivals, but the seller cannot discriminate among buyers who are offered the same position in the waiting list. Both the waiting list and the fully optimal mechanism are implemented by non-standard auctions with a scoring rule where the distance between buyers' bids affects the allocation. A novel property of these auctions is that for buyers it is better to win by a close margin and it is worse to lose by a close margin. Finally, I model partial commitment power as a penalty that the seller has to pay when forfeiting a promise. All the results are given for general partial commitment and therefore include full commitment and no commitment as special cases.

JEL: D82, C73, D44

Keywords: dynamic mechanism design, services, ticket sales, restricted mechanisms

This paper is based on the first chapter of my Ph.D. thesis at Northwestern University. I am grateful to Jeff Ely, Alessandro Pavan, and Asher Wolinsky for many helpful discussions, and Eddie Dekel, Dan Garrett, Marit Hinnosaar, Rakesh Vohra, Dirk Bergemann as well as seminar participants at Northwestern University, Aalto University, Bocconi University, Collegio Carlo Alberto, University of Bonn, University of Exeter, HEC Paris, McGill University, University College London, Vanderbilt University, ESEM 2012 and the Midwest Economics Theory meeting at Washington University, and "Conference on Private Information, Interdependent Preferences and Robustness: Theory and Applications" in Bonn. The financial support from the Center for Economic Theory at Economics Department of Northwestern University is acknowledged.

[†]Collegio Carlo Alberto, `mailto:toomas.hinnosaar@carloalberto.org`.

1 Introduction

I study a dynamic mechanism design problem where a monopolist sells services to randomly arriving privately informed forward-looking customers. Service slots could be seats in a train, which are usually allocated by ticket sales, so that tickets are sold at constant prices and each ticket guarantees a seat in a particular moment of time. There are several ways to improve this. First, we could sell tickets by the optimal mechanism instead of posted prices. Second, we could replace tickets by contracts that guarantee seats only when the demand is not too high. For example airline companies have already moved towards this direction: passengers who buy cheap tickets will face the risk of being bumped when the demand among higher-price ticket classes turns out to be high. Other applications include Internet ad-auctions, ticket sales for events or transportation, restaurant reservations or the assignment of service slots by service providers ranging from hair salons to hospitals. The common features in all these examples are that 1) goods are perishable—once the corresponding date is passed, the good disappears; 2) the situation is repeated; and 3) demand is random, with customers willing to wait for some time but not for very long.

In the model, a monopolistic seller has a fixed number of identical service slots every period. Buyers arrive randomly in time and have independent private values. They prefer instant service, but are willing to wait a little, if the price is right. The seller's goal is to maximize the expected discounted revenue[1] by choosing the optimal mechanism from a potentially restricted class of mechanisms. The goal of this paper is to characterize the fully optimal mechanism as well as optimal mechanisms in situations where the seller has only a limited set of "simpler" allocation rules available.

First, in the pure calendar mechanism, the seller can only give promises of future service that are fulfilled with certainty. In other words, the seller has a calendar according to which the service is supplied. In each period, the seller assigns available dates from the calendar to new buyers. Selling tickets and making appointments are examples of this kind of mechanism. Second, in the waiting list mechanism, the seller can give promises that are conditional on future arrivals but he cannot personalize the promises to buyers' types. This restriction operates as if the seller keeps a list of buyers who were promised delayed service and serves them according to some rule. Customers in different positions in the list may be served differently. However, only the position in the list—and not the buyers type—will determine the allocation she gets. In some sense this is a natural relaxation of the pure calendar mechanism assumption—the seller may want to schedule appointments but with an understanding that when a very valuable buyer arrives in the future, some of the appointments will be canceled. Third, in the fully optimal mechanism, the seller is free to choose any kind of promise of the future service. In contrast to the waiting list the promises will be personalized to each buyer according to their valuations.

Finally, I study the sales of future service, one interesting question is how the seller's ability

[1] As it is often the case in mechanism design literature, maximizing efficiency is not fundamentally different. In particular, all the results presented in this paper can be interpreted as efficient mechanisms when virtual values in the expressions are replaced by true values.

to commit to providing service affects the optimal mechanisms. To study commitment power I assume that the seller is able to commit only to the promises of delayed services that are not too tempting to break in the future. This is modeled using a partial commitment constant that can be interpreted as a penalty—if the seller does not serve a customer who was promised service, she must incur this cost. All the results presented in this paper are for general partial commitment constant; therefore full commitment and no commitment come as special cases.

In this paper I take a different approach from the related literature by assuming that each buyer interacts with the seller only once. This assumption is motivated by applications, where buyers do not interact with the seller for a very long time. This is presumably because it would be too costly for them to participate in the mechanism more than necessary or that they prefer to have the information about the future service as soon as possible (for example to purchase complementary services, such as to reserve hotel rooms at the same time as buying plane tickets). Note that if the class of mechanisms that the seller can use is unrestricted, the one-time-interaction assumption is without loss. For any multiple-interactions mechanisms, there exist a corresponding one-time-interaction mechanism, where the seller gives each buyer a complete history-dependent allocation rule as a promise.

There are two special cases of this model, where the fully optimal mechanism is relatively simple to characterize. First, if the number of objects is large enough to serve all potential customers at their arrival, the optimal mechanism would be a constant posted-price mechanism. This is a version of Stokey's (1979) no pure price-discrimination result. Here is the reason. From the buyers' perspective, being offered a delayed service is equivalent to getting a fractional allocation. Without the need for rationing and without taking the continuation values into account, the optimal solution would use only corner solutions[2] (Myerson, 1981). Offering a delayed service can only decrease the continuation values, so the dynamic feasibility considerations would not change the conclusion. This result does not hold when at least in some periods there are fewer objects than new buyers, so that rationing is needed. Second, when the buyers discount the future service at a constant rate and the seller has the option to give promises of arbitrary length, it is optimal to give the current service to the buyers who have the highest values[3] among all buyers who have arrived but not yet received service. This result follows from Said (2012). The reason is that two buyers who have arrived in different periods look at the future the same way, which makes it easy to compare buyers—only their valuations matter. The result fails, however, if buyers do not discount the future at a constant rate or exit at an increasing rate. In these cases the seller needs to treat buyers from different periods differently. The buyers' types would be multi-dimensional, characterized by both the valuation and the exit process.

This paper concentrates on the remaining case, where rationing is needed and buyers' continuation values are dependent on their arrival time. There are many applications that fall in this category. First, there could be practical or legal reasons why buyers interact with the seller once and the seller cannot give promises about service in the too distant future. Many types of

[2]In particular give a buyer the instant service if her virtual surplus is positive and no service otherwise.
[3]Highest positive virtual surplus if the goal is to maximize revenue.

selling tickets and making appointments fall into this situation. Second, it may be because of buyers' preferences—it is possible that buyers accept service if it is offered in the near future, but they cannot wait for a long time. Getting a haircut, appointments for emergency services, and transportation seem to satisfy this description. Finally, it could be due to competition that we do not model explicitly. For example buyers would wait, if they expected to get the service soon, but if the delay was long, they would look for alternatives.

I focus on the extreme case where the simple characterizations fail. In particular, we assume that buyers interact with the seller only at their arrival, and the seller cannot give promises for more than one period ahead. In this case two buyers who have arrived in two consecutive periods are different, since the buyer who has already waited one period must be served now or will not be served, whereas a new buyer can be offered either current or delayed service. The heterogeneity of customers is now two-dimensional—not only their values but also their exit rates differ. As noted, there are at least three mathematically equivalent interpretations of this assumption. First, the seller simply cannot contract service more than one period ahead. Second, buyers have time preferences such that they get a positive value from current and next-period consumption, but they do not get any value from the service after that. Third, buyers exit the model two periods after their arrival. As we would expect, the optimal mechanism in this framework offers delayed service with strictly positive probability and treats buyers from different periods differently. Alternatively, one could think of it as a mechanism design problem in overlapping generations model.

I show that the fully optimal mechanism is implemented by a non-standard auction with a scoring rule, which favors buyers who arrived earlier and in which the allocation that the new buyers receive depends not only on the order of their values, but also on the differences in their values. In particular I show that some types of buyers strictly prefer to win in the auction by a small margin rather than by a large margin, because this is the only way they receive an instant service. Similarly, some types of buyers prefer to lose by a large margin rather than a small margin, since in the latter case they would be refused service instead of being offered a delayed service.

Characterizing the fully optimal mechanism is one contribution of the paper. In addition to that I derive the optimal mechanism in the two classes of restricted mechanisms described above. The optimal pure calendar mechanism is characterized by a unique vector of opportunity costs of the corresponding service slots. It can be implemented by a simple two-stage mechanism. The qualitative properties of the optimal waiting list mechanism are similar to the fully optimal mechanism, but the delayed service contracts cannot be personalized to the buyers receiving the delayed service. The optimal contracts are therefore designed for the "average" buyer who will be offered delayed service. The optimal waiting list will have non-trivial dynamics even if the arrival process is constant. Since the value of an average buyer to whom the seller expects to assign each contract depends on previous promises, it is possible that the optimal contracts change every period.

Finally, I characterize all three classes of mechanisms under the general partial commitment assumption, which shows how the optimal mechanism changes with commitment power. In particular there is a sufficient and, in most cases, necessary level of commitment that allows the seller to use the optimal mechanism in all three classes. If the commitment power is relatively low, the optimal calendar mechanism is static, whereas the revenue from the optimal waiting list mechanism and the unrestricted mechanism are strictly increasing in commitment power.

As with most of the mechanism design literature, this modeling approach is similar to the seminal papers by Myerson (1981) and Riley and Samuelson (1981). In particular, Myerson's optimal mechanism is the optimal static mechanism in the model analyzed in this paper. Static mechanism belongs to all three classes of mechanisms that we consider, and it is therefore a benchmark against which to compare all the dynamic mechanisms. We show, however, that it will not generally be optimal.

This paper belongs to a growing body of literature on dynamic mechanism design. A related branch of literature studies the sale of durable goods, where the seller has a fixed number of durable goods and a deadline by which the goods must be sold. Gershkov and Moldovanu (2009, 2010) characterize the optimal online mechanism where buyers must be assigned an object at their arrival, and Li (2009); Board and Skrzypacz (2010) characterize the optimal mechanism when the buyers are patient. Pai and Vohra (2008); Mierendorff (2011) consider a model where buyers' unobservable types include their arrival and departure times and give partial characterization of the optimal mechanism. In these papers the most important trade-offs are between extraction of rents and the option value—giving away an object before the deadline means that the object cannot be assigned to potentially higher-valuing buyers who arrive later. The option value decreases in time, however, since it becomes less likely that such buyers will arrive. In contrast to these papers, our model is designed to analyze the situation where the seller has a new set of goods every period, and one set of goods perishes every period. This situation creates a different set of trade-offs.

Bergemann and Välimäki (2010) and Pavan, Segal, and Toikka (2013) give results for general classes of dynamic mechanisms, where both adverse selection and moral hazard issues could be studied. Pavan, Segal, and Toikka give the envelope formula for a large class of dynamic mechanism design questions and show how to compute transfers. One application of their general result is the sale of a durable good to buyers with changing types. Similarly to our results they show that the optimal mechanism is implementable with a non-standard auction with scoring rule, where the differences between bids affect the allocation. However, in their model, winning by a large margin is good and losing by a large margin bad, which is the opposite result from the implications of our model. Bergemann and Välimäki's main result is the characterization of transfers that implement any efficient mechanism.

This paper is closely related to literature on price discrimination with durable goods. After Stokey (1979); Conlisk, Gerstner, and Sobel (1984) showed that it is not optimal to delay sales purely for price discrimination reason there have been documented several reasons why it could be optimal to delay sales. Board (2008) showed that when demand fluctuates, some delay would be necessary, which is similar demand-smoothing argument as in this paper, but with in his model the monopolist had much more limited set of instruments available. Garrett (2013) offered another reason for price discrimination: changes in buyers' valuations. The intersection of this paper and the current paper is Ely, Garrett, and Hinnosaar (2012), where buyers valuations change and the seller has relatively sophisticated contracts available. Similarly to current paper, it is optimal for the seller to bump people with relatively low valuations, but for another reason: to extract more surplus from partially uninformed buyers.

All the papers noted above study the fully optimal mechanism. The pure calendar mechanism and waiting list mechanism are new to this paper. The particular way we model partial commitment is also new, but analogous to renegotiation-proofness models—for example Hart and Tirole (1988). In this literature the mechanism-designer cannot offer a contract if it is known that it will be mutually beneficial to renegotiate this contract in the future. In our model buyers interact with the seller only once, which means that renegotiation is impossible, but the approach we take is the same: the seller with partial commitment power can give only those promises that she will not find it optimal to break in the future. For simplicity, we assume that the cost of commitment is exogenous[4].

Section 2 shows that most of the trade-offs of the problem can be studied in a simple illustrative example, with only one object and two new buyers every period. In particular, Section 2.1 characterizes the optimal pure calendar mechanism, Section 2.3 the optimal waiting list mechanism, and Section 2.2 the fully optimal mechanism. Section 2.4 introduces the partial commitment assumption, and shows how all the previous results generalize for partial commitment. Section 3 generalizes the results for an arbitrary number of service slots and random arrivals. Section 4 concludes and discusses potential extensions.

2 Illustrative example

In this section we discuss an example of the model with just one object and two new buyers each period. The results in this section are special cases of the general results in Section 3, and therefore formal statements and proofs are omitted.

A monopolistic seller has one service slot each period $t \quad \{0, 1, 2, \dots\}$. It could be a time slot for providing service, a seat, or a perishable good produced on the same day. The seller maximizes expected discounted revenue, with discount factor $\quad (0, 1)$.

Each period two new buyers arrive. Each buyer i has unit demand for the service, with independent private value $v_i \quad [0, 1]$. Value v_i is a random variable with cumulative density function F and probability density function $f(v_i) > 0$ for all $v_i \quad [0, 1]$. Function $w(v_i) = v_i - \frac{1-F(v_i)}{f(v_i)}$ denotes the standard static virtual surplus function, which is assumed to be strictly increasing[5]. To simplify the notation, we rename the buyers so that buyer 2 is the one with higher value and buyer 1 with lower value. Denote the CDF[6] of buyer i by F_i and corresponding

[4] There would be several interesting ways to endogenize it, including reputation or direct compensation to buyers.

[5] Monotone hazard rate condition is a sufficient condition.

[6] $F_2(v_2) = F(v_i)^2, F_1(v_1) = F(v_1)[2 - F(v_1)]$.

PDF[7] by f_i. Note that we are not following this convention in the figures, where we allow v_1 and v_2 to be arbitrary.

Buyer i with type v_i who is served s_i periods after her arrival and pays p_i gets payoff $\;\;^{s_i} v_i - p_i$. In particular, if she is served at the same period she arrives, she gets payoff $v_i - p_i$. Discount factor $\quad (0, 1)$ is the same for all buyers as well as the seller. We concentrate on mechanisms where a buyer who is not served one period after arrival is never served. As argued in the introduction, this may be the case for several reasons—restriction on the mechanism, preferences, or exit process. It describes situations where in the short run the dates are imperfect substitutes for the buyers, but in the long run the value decreases faster than a constant rate.

Each buyer interacts with the seller only at arrival. This means that each period the two new buyers announce simultaneously their types, and the seller responds by giving both of them either instant service or promises about future service and requests transfers. A promise of future service could be a refusal of service or delayed service conditional on next period arrivals. A promise of delayed service is a complete description of the allocation that the buyer gets conditional on the future arrivals. In the illustrative example it is characterized by a set $\hat{D} \quad [0, 1]^2$, such that if buyer i receives promise $\hat{D}$ she will be served in the next period if and only if the new buyers who arrive in the next period announce values $v = (v_1, v_2) \quad \hat{D}$. This means that the expected payoff from promise $\hat{D}$ to buyer i with value v_i is $Pr(\hat{D}) \;\; v_i + Pr(\hat{D}^c)0 - p_i = \;\; \hat{d} v_i - p_i$, where $\hat{d} = Pr(\hat{D})$. The set $\hat{D}$ can be interpreted as a contract or the terms of a ticket that the seller can give to a buyer one period before the service in exchange for money.

A state captures all relevant details from the history up to a particular period. In particular, it characterizes the promises of delayed service given to buyers in the previous period. In the illustrative example a state is characterized by set $D \quad [0, 1]^2$, which includes all the combinations of types v for which the instant service is unavailable. Of course, a state is equivalent to a promise given to a buyer in the previous period—the current object is unavailable because it was promised to a buyer in the previous period.[8] Unless explicitly stated otherwise, we denote the state by D and a contract by $\hat{D}$. As an illustration of possible dynamics, consider the following example. In period 0 the seller has not made any promises, so the initial state $D_0 = \quad$. Two new buyers arrive and announce their types. The seller serves one of them instantly (at period 0) and promises the other that she will be served in the next period if and only if all new buyers announce values below $\underline{v}$. This is described by a contract $\hat{D}_0 = [0, \underline{v}]^2$, which is the state in period 1, $D_1 = \hat{D}_0 = [0, \underline{v}]^2$. Now, in period 1 two new buyers arrive and announce their values $v = (v_1, v_2)$. If $v \quad D_1$, the seller has to serve the buyer from the previous period and therefore cannot serve any of the new buyers instantly, whereas if $v \;/\; D_1$, the instant service is available. When the seller does not offer delayed service to either of the new buyers, we return to the initial state $\quad$.

[7] $f_2(v_2) = 2F(v_2)f(v_2), f_1(v_1) = 2[1 - F(v_1)]f(v_1)$.

[8] There is a minor abuse of notation here, since it is possible that two of the buyers from the previous period received promises of delayed service, say in sets D_1 and D_2 respectively. At the optimum this happens with zero probability, but is feasible when $D_1 \quad D_2 = \quad$ and $D_1 \quad D_2 \quad [0, 1]$. In this case we can simply redefine the state as $D = D_1 \quad D_2$.

In the general analysis in Section 3 we extend all the results to m 0 service slots, random number n $\{0, \dots, N\}$ buyers every period, and allow the seller's discount factor differ from the buyers' discount factor (denoted by). In Appendix A we discuss how the results in the example would change if we relax the assumption = that we use here.

The goal of the mechanism designer is to find mechanisms that are Bayesian incentive-compatible and ex-ante individually rational. As it turns out, all mechanisms that we will discuss below are implementable in dominant-strategy incentive-compatible and ex-post individually rational transfers.

One feasible mechanism in this framework is the repetition of the static mechanism, where the good is always sold to one of the buyers arriving in the particular period. In other words, the seller never gives promises of delayed service, so the state is $D =$ in all periods. By Myerson (1981) the optimal static mechanism assigns instant service to the buyer with the higher value if his virtual valuation is positive, and refuses service to both buyers otherwise. This allocation rule is illustrated by Figure 2.1.

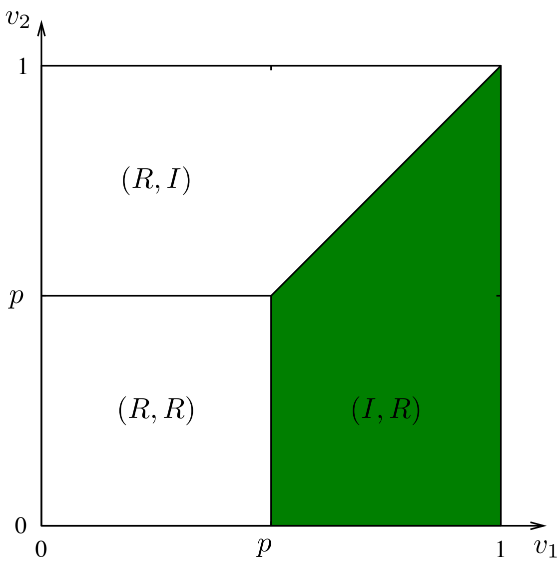


Figure 2.1: Optimal static allocation. Shaded area shows when buyer 1 is served instantly.

The optimal static mechanism can be implemented by the second price auction with reserve price p such that $w(p\ ) = p\ - \frac{1-F(p\ )}{f(p\ )} = 0$. This ensures expected revenue ${}^{s} = {}_{p}^{1}\ w(v_2)dF_2(v_2)$. This mechanism will typically not be optimal in a dynamic setting, but the reserve price and the static revenue ^{s} will be useful in the following analysis.

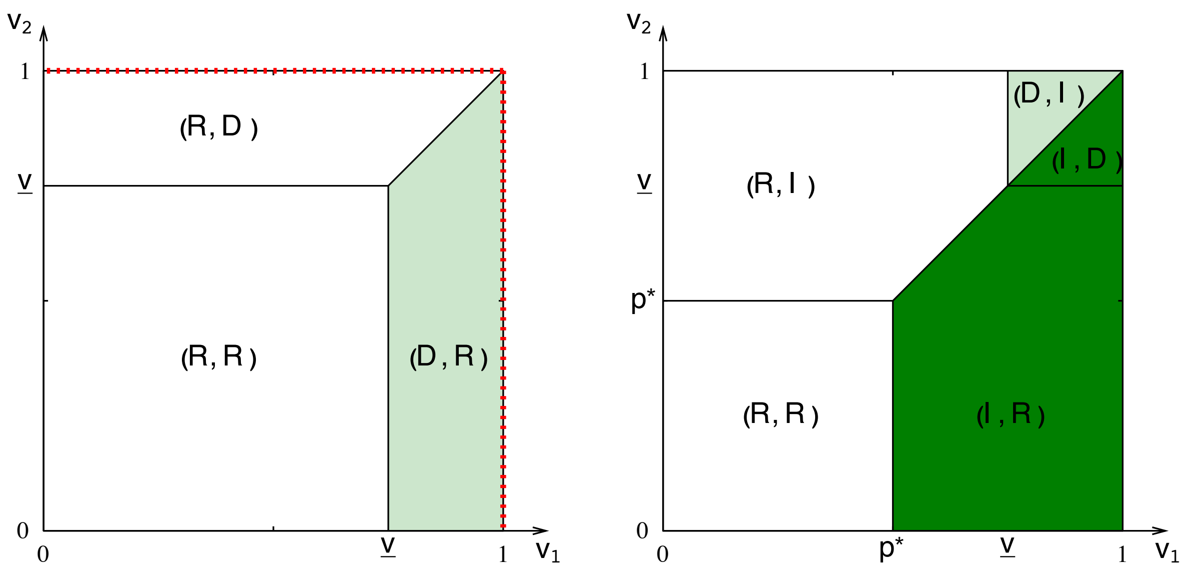
2.1 Pure calendar mechanism

A pure calendar mechanism is a mechanism where the promise of delayed service is unconditional on next period arrivals. It is like marking dates on a calendar and then providing services according to this calendar, or selling tickets that guarantee the service. This means that a promise of delayed service needs to be fulfilled with certainty, and the seller cannot serve anyone instantly when a buyer from the previous period was promised delayed service.

In the notation introduced above, the promise of delayed service can be only $D =$ or $D = [0, 1]^2$, which correspond to probabilities 0 and 1 respectively. There are only two possible states: the instant service is either available, or it is already promised to someone and is therefore unavailable. To shorten the notation, we denote the first state by 0 and the second by 1, both of which correspond to the promises given to buyers in the previous periods. Denote the revenue from the state when the instant service is available by (0) and when it is unavailable by (1). Clearly, the expected revenue from the former is higher than from the latter. Denote this difference by $= (0) - (1) > 0$.

The difference in revenues in the two states is the opportunity cost of a promise one period ahead. In particular, if the seller promises delayed service to a current buyer, the discounted loss is $\frac{}{1-}$. The denominator $1 -$ is needed to compare the difference in continuation values, , to the difference in flow values, which is the revenue extracted from current buyers.

Consider the state where instant service is unavailable. Then the only object that can be allocated is a promise of next-period service. Buyers' values of the object are v_1 and v_2, and allocating this object costs $\frac{}{1-}$ to the seller. The optimal mechanism therefore assigns the object to the buyer with the higher value, if his discounting-adjusted virtual surplus $w(v_2)$ is greater than the opportunity cost $\frac{}{1-}$. It is illustrated by Figure 2.2(a). This allocation rule can be implemented by a standard second price auction for the delayed service with reserve price $\underline{v}$, defined by $w(\underline{v}) = \frac{}{1-}$.



(a) Instant service unavailable

(b) Instant service available

Figure 2.2: Optimal pure calendar mechanism. Intensity of shading color describes the discounted expected quantity for buyer 1, the area below dotted line is the set where previous-period customer is served.

Now, consider the state where instant service is available, so that the seller sells two goods to two buyers. The instant service is more valuable to both buyers and costless to allocate, whereas allocating delayed service changes the state exactly as in the case studied above and has the same opportunity cost $\frac{}{1-}$ as above. This means that the instant service is allocated often, and the delayed service is promised only if both buyers have sufficiently high values.

This is a demand-smoothing argument. When in some period both buyers have high values, the demand is high. Since it is unlikely that in the next period the values would be as high, it is optimal to extract more revenue from the current high-valuing buyers and lose some revenue from tomorrow's buyers (whose values are uncertain).

In particular, the optimal mechanism allocates the instant service to the buyer with higher value when his virtual surplus is positive. It promises delayed service to the buyer with lower type, when his discounting-adjusted virtual surplus exceeds the opportunity cost. This allocation rule is illustrated by Figure 2.2(b).

This allocation rule is implemented by a simple two-stage auction mechanism. At the first stage, the instant service is allocated at a standard second price auction with reserve price p . In the second stage, the loser is offered the next period service at a fixed price $\underline{v}$. When the loser accepts the offer, the winner of the first-round auction gets a discount $(v_l - \underline{v}) > 0$, where v_l is the loser's bid in the auction. Offering the delayed service to the loser of the second price auction creates surplus for the loser. This means that the winner must be compensated, since losing is not as bad as it would be without the second stage. If types of buyers are $v_2 \quad w_1 \quad \underline{v}$, the high type 2 would prefer winning to losing in the first round if compensation p is high enough to satisfy $v_2 - v_1 + p \quad (v_2 - \underline{v})$. In particular, in the critical case when $v_1 = v_2$, we get the compensation given above, whereas if $v_2 > v_1$, this compensation is sufficient and even gives some extra surplus to the buyer.

Note that the threshold $\underline{v}$ that we used to determine whether or not to allocate the delayed service is the same in both states. The opportunity cost of a promise comes from being in a less restrictive state in the next period. This cost is independent of the type who receives the delayed service.

The discussion above took the opportunity cost $\frac{}{1-}$ as given, but the difference between the two states is endogenous and yet to be determined. The difference in revenue in the two states can be decomposed to two parts. First, if the instant service is available, it is assigned to the higher-valuing buyer whenever her virtual valuation is positive. This is exactly the same as static optimum and thus gives the static revenue ^{s}. The second difference is dynamic and comes from the promises of delayed service. When the instant service is unavailable, the delayed service is promised to the higher buyer whenever his discount-adjusted virtual surplus $w(v_2)$ is higher than the opportunity cost, which is equal to $w(\underline{v})$. In this case the seller gets the difference. This gives expected dynamic flow revenue $_2(\underline{v}) = \int_{\underline{v}}^{1} [\ w(v_2) - \ w(\underline{v})] dF_2(v_2)$. In the state where the instant service is available, the delayed service is offered under the same condition, but only to the lower-valued buyer. This ensures expected dynamic flow revenue $_1(\underline{v}) = \int_{\underline{v}}^{1} [\ w(v_1) - \ w(\underline{v})] dF_1(v_1)$, which is smaller than $_2(\underline{v})$.

Using these observations, we can express the opportunity cost $\frac{}{1-}$ in two ways as a function of $\underline{v}$ to get equation (2.1) that uniquely determines $\underline{v}$ and therefore , assuming that the interior solution exists.

$$w(\underline{v}) = \frac{}{1-} = \quad (\ ^{s} + \ [\ _1(\underline{v}) - \ _2(\underline{v})]) \, . \tag{2.1}$$

It is easy to see that (2.1) has an interior solution and it is unique. The left hand side of the equation is strictly increasing by assumption and takes value 0 when $\underline{v} = p$ and when $\underline{v} = 1$. The right hand side on the other hand is strictly decreasing[9] and takes value ^{s} $(0, 1)$ at $\underline{v} = 1$. Since both expressions are continuous, we get existence and uniqueness. The general proof extends the same ideas for any number of slots and buyers using Brouwer's fixed point theorem.

2.2 Fully optimal mechanism

In this section we analyze the fully optimal mechanism, where the seller can choose any allocation rule she might like. Compared to the pure calendar mechanism this means relaxing two restrictions. First, the seller can offer conditional contracts, so that the buyer is served in some situations and is not served in other situations. Second, the seller can give out personalized contracts, i.e. different promise for each type of buyers. In the next section we are a class of mechanisms that relaxes only the first assumption while maintaining the second: waiting list mechanisms, where the seller can offer contingent contracts, but cannot personalize the contracts for each particular type.

The relaxation of these restrictions changes the mechanism design problem. In the pure calendar mechanism, the problem was discrete, since promising future service came with strictly positive cost and therefore the main question was who and when should receive this promise. Now, the the problem becomes continuous and the seller can promise some probability of delayed service for free. Indeed, with some probability all new buyers have negative virtual surpluses, therefore the seller would not want to serve them even without previous commitments. Therefore the main question is not whether to offer the contract or not, but what kind of contract to offer.

We show that the optimal mechanism has interesting properties. It is implemented by non-standard auctions with scoring rule where the services that buyers are offered will depend on how similar their bids are to their opponent's bids. In particular, it is sometimes better to win by a little than by a lot, and it is sometimes better to lose by a lot than by a little. After we have characterized the optimal mechanism we will make these statements precise, but the general intuition is that distortions from serving a buyer from a past period depends on the values of both new buyers. Therefore whether or not the seller uses instant service to serve new buyers or a buyer from the previous period will depend on the values of both new buyers.

There are some straightforward observations that make the analysis simpler. First, the seller

[9] Derivative with respect to $\underline{v}$ is $w\ (\underline{v})[F_1(\underline{v}) - F_2(\underline{v})] < 0$, because $w\ (\underline{v})$ by assumption and $F_1(\underline{v}) > F_2(\underline{v})$ for all $\underline{v} < 1$ by the properties of order statistics.

will promise delayed service to only one buyer[10]. Therefore a state is still simply D, the promise given to one of the previous buyers. Second, we can again separate the optimal assignment decision to probability of delayed service $\hat{d}$, and contract $\hat{D}$ which will be optimal given this probability. Note that initial state may be sub-optimal, but continuation states are always optimal. Even if the initial state is such that there are two buyers waiting, their transfers are already sunk, so we can treat of them as a single buyer.

We proceed by characterizing the fully optimal mechanism in four steps. First, we take the expected revenue functions $(\hat{d})$ as given and characterize the optimal allocation rule. The optimal allocation rule will now include the optimal promise function $\hat{d}(\cdot)$, which at we first characterize a function of continuation values. Second, we discuss how we construct the contract that fulfills the probabilistic promises of delayed service. Third, we characterize the optimal promise function $\hat{d}(v_i)$ explicitly and discuss its properties. Finally, we verify that functions $(\hat{d})$ exist.

Lets start by assuming that the continuation functions $(\hat{D})$ and $(\hat{d})$ functions exist and fix a state D. We denote the optimal contract of delayed service offered to a buyer in the current period by $\hat{D}$. Since the buyers are interested in probability of service, it is also useful to denote the promise in terms of probability of being served, $\hat{d} = Pr(\hat{D})$.

Suppose the new buyers announce types $v \quad D$. Then the seller has to serve the previous buyer and therefore can offer delayed service to only one of the buyers. This is offered to the highest of the two buyers with some probability $\hat{d}$. The optimal choice of $\hat{d}$ maximizes expected discounted revenue and therefore solves the following maximization problem

$$\_(v) = \max_{\hat{d}} \quad (0) + (1 - \quad) \hat{d} w(v_2) - \quad [\quad (0) - \quad (\hat{d})] \quad , \tag{2.2}$$

where the first component is the static continuation revenue that the seller gains by not giving any promises; the second term is the revenue extracted from the high type by giving a promise of service with probability $\hat{d}$ in the next period; and the third term is the loss in continuation value from giving this promise. The revenue extracted from the high type is now his standard static virtual surplus $w(v_2)$ adjusted by discount factor and probability $\hat{d}$. The first order condition from the maximization problem is

$$(1 - \quad) w(v_2) + \quad (\hat{d}) \quad 0. \tag{2.3}$$

We will come back to the function $\hat{d}$ once we know how to compute (d) and differentiate it. From (2.3) we see that $\hat{d}(v_2)$ is only a function of v_2 and it balances the trade-off from extracting revenue from the buyer and having higher opportunity cost because of being in a more restrictive state in the next period. In particular, $\hat{d}(v) = 0$ whenever it is promised to a buyer with negative virtual surplus and $\hat{d} \quad F_2(p \;)$ for all $v_2 > p$, since with $v_2 < p$ the seller would not want to serve new buyers instantly and therefore $(d) = \quad (0)$ in this interval.

[10] Instead of promising D_1 and D_2 two two current buyers, promising $D_1 \quad D_2$ to the higher of them would increase flow revenue while keeping continuation value unchanged.

When buyers announced $v \ / \ D$, the instant service is available, so that the higher of the new buyers is served instantly if and only if he has strictly positive virtual surplus, whereas the lower new buyer is offered delayed service with probability $\hat{d}$, that solves the following optimization problem

$$\bar{\ }(v) = \max_{\hat{d}} \quad (1- \quad) \max\{0, w(v_2)\} + \quad (0) + (1- \quad) \ \hat{d}w(v_1) - \ [\ (0) - \ (\hat{d})] \quad . \tag{2.4}$$

There are two differences with the previous maximization problem. First, the static revenue is increased, since some revenue is extracted from serving the high type instantly. Second, the delayed revenue comes now from the low type rather than the high type, which means lower dynamic revenue in expectation. The first order condition is the same as previously given with (2.3), but for v_1 instead of v_2. Therefore the optimal promise is the same function $\hat{d}(v_i)$ in both cases; the difference is that it is promised to a different buyer in each of the two cases. Graphical illustration of the combined allocation rule is by Figure 2.3.

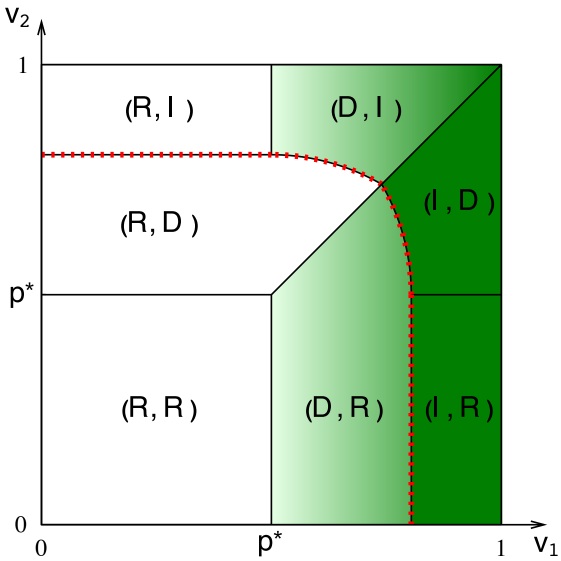


Figure 2.3: Optimal allocation in fully optimal mechanism. Intensity of shading color describes the expected discount factor for buyer 1, the area below dotted line is the set where delayed service is offered.

The next question is how to fulfill probabilistic promise $\hat{d}$ optimally, i.e. which values v to include to the contract $\hat{D}$ and which not, while satisfying constraint that $Pr(\hat{D}) \quad \hat{d}$. Suppose $\hat{D}$ is an optimal contract. From the description above we can calculate the expected revenue for each v in both of the two cases discussed. Denote the expected revenue from v such that the instant service is unavailable by $\_(v)$ and the expected revenue when the instant service is available by $\bar{\ }(v)$. The difference between the two, $\bar{\ }(v) - \_(v)$ can be interpreted as the loss in continuation revenue from adding particular type profile v to the contract $\hat{D}$. The expected revenue from probability $\hat{d}$ can be expressed as

$$(\hat{d}) = E_v\left[^{-}(v)\right] - \hat{d} \min_{\hat{D}:Pr(\hat{D}) \quad \hat{d}} E_v \quad ^{-}(v) - \_(v)|v \quad \hat{D} \quad . \tag{2.5}$$

From (2.5), it is clear that in optimum $\hat{D}$ must include points where the loss in revenue, $^{-}(v)-\_(v)$, is the smallest. It is easy to see that when $v_2 \quad p$, distortion $^{-}(v) - \_(v)$ is strictly increasing in v_2, and therefore the constraint must be binding. That is, the optimal way to fulfill promised probability $\hat{d}$ is to serve the buyer from the previous period if and only if the distortion from doing so is less than the upper bound for the loss in revenue, which we denote by $\bar{b}$. Therefore we can write $\hat{D} = \{v : {}^{-}(v) - \_(v) \quad \bar{b}\}$ for $\bar{b}$ such that $Pr(\hat{D}) = \hat{d}$.

This representation also means that instead of characterizing the optimal contract in terms of probability $\hat{d}$ we can represent it with respect to corresponding maximum loss in revenue, which we denoted by $\bar{b}$. This is more convenient, because it represents exactly the marginal loss of the optimal contract. That is, $(\hat{d}(v_i)) = -\bar{b}(v_i)$. We can now analyze equation (2.3), the first order condition that characterizes optimal promise $\hat{d}(v_i)$, which takes the form

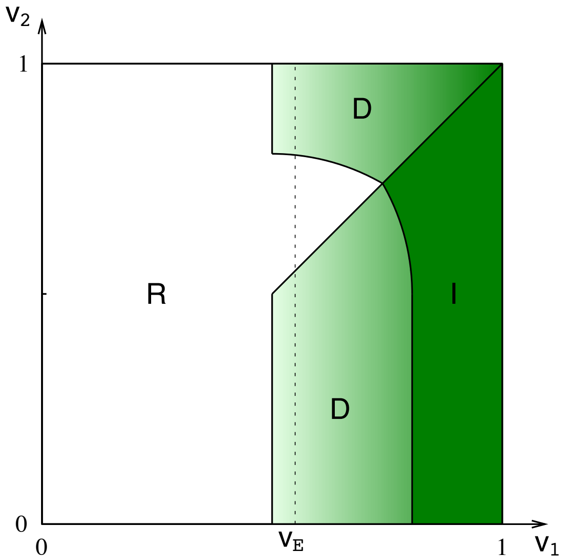
$$(1 - \quad)w(v_i) = \bar{b}(v_i). \tag{2.3'}$$

We see that whenever $w(v_i) > 0$, the maximum distortion $\bar{b}(v_i) > 0$ and so the optimal probability $\hat{d}(v_i) > F_2(p \ )$. Moreover, maximum distortion $\bar{b}(v_i)$ and therefore probability $\hat{d}(v_i)$ are strictly increasing functions of v_i for all interior solutions. Finally, we get the upper corner solution where the probability $\hat{d}(v_i) = 1$, or equivalently maximum distortion $\bar{b}(v_i) = 1 -$ exactly when $w(v_i) = 1$.

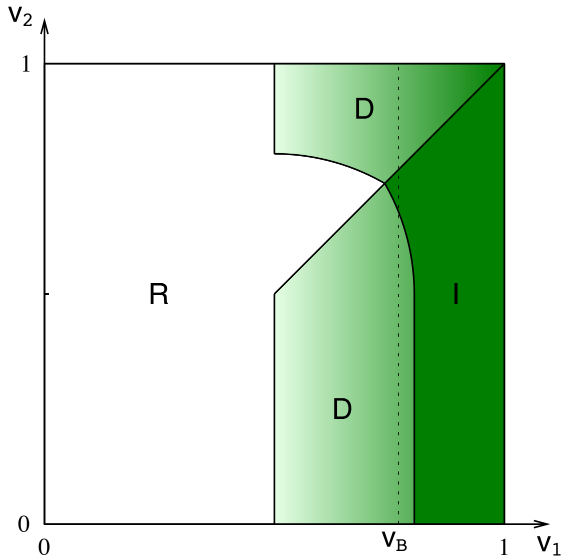
Finally, we have to verify that the continuation values (d) exist and are uniquely defined. The description above shows how to compute maximal revenue for each profile v, taking continuation value function (d) as fixed. Let $T \ (d)$ denote the profit corresponding to the optimal contract with service probability d. We need to show that there is unique fixed point $= T$. We use Blackwell's sufficient conditions to argue that T is a contraction with rate and then apply the Contraction mapping theorem to get the existence and uniqueness. Verifying the conditions is intuitive. First, monotonicity. If $\hat{}(d) \quad (d)$ for all d, then any policy under $\hat{}$ ensures higher revenue than the same policy under and therefore the optimum cannot be lower either. Second, to verify discounting, suppose that $\hat{}(d) = \quad (d) + a$. Since the dynamic part of the revenue depends only on the difference $(0) - \quad (d)$, the optimal policy is unchanged. The only remaining dependence is through static continuation value, $\hat{}(0)$ and therefore $T\hat{}(d) = T \ (d) + \ a$ for all d and $a > 0$.

The optimal mechanism can be implemented by a non-standard auction with scoring rule, where whether or not buyers receive instant or delayed services depends not only on their own types and order of types, but also on the similarity of their types. In particular, there are some types of buyers who receive delayed service unless the other buyer has only slightly smaller value (in this case they get instant service). There are also types of buyers who receive delayed service unless the other buyer has only slightly higher value.

The reason is that the loss in revenue is increasing in both new buyers' values, and therefore the allocation that a buyer gets also depends on the opponent's bid. Consider for example a situation where buyer P from previous period had value v_P has an optimal contract and new buyer 1 has value $v_E < v_P$, as described in Figure 2.4(a). If new buyer 2 has value below v_E, then buyer 1 wins the auction and gets assigned the delayed service. Now, if we increase buyer 2's value then eventually it will be higher than v_E, which means that 2 wins the auction and gets the delayed service instead, so buyer 1 is refused service. However, if we increase buyer 2's value even further then at some point (v_1, v_2) is not in contract set anymore, since the loss in revenue from not serving buyer 2 instantly and offering 1 delayed service will become too high, so that buyer P is not served and buyer 1 ends up getting delayed service again. This is the non-monotonicity: buyer 1 with given value v_E is strictly worse off by losing by a relatively small margin than either winning or losing by a large margin. Similarly Figure 2.4(b) describes a situation where new buyer with value v_B that is slightly higher value than v_P, prefers to win by a small margin to either losing or winning by a large margin, because this is the case where he will be offered instant service.



(a) It is sometimes bad to lose by a small margin



(b) It is sometimes good to win by a small margin

Figure 2.4: Winning margin matters in the fully optimal mechanism. Intensity of shading color describes the discounted expected quantity for buyer 1 at a particular state, R is short for "refused service", D for delayed service, and I for "instant service".

2.3 Waiting list

Since implementing the fully optimal mechanism requires quite flexible tools: the seller has to be able to offer personalized contingent contracts. This may be possible under some circumstances, but difficult or impractical in others. Therefore it is natural to also ask, what would be the optimal mechanism in restricted class of "simpler" mechanisms. We already discussed the pure calendar mechanisms, that corresponds to very simple ticket-selling. In this class, the seller is not able to sell contingent contracts and therefore also not able to personalize the contracts of delayed service. In this subsection we consider intermediate case: we assume that the seller is able to offer contingent contracts, but cannot personalize it for different types of buyers.

The particular assumption we make here is that at any period t, possible promises are such that each buyer gets either instant service, no service, or a promise of delayed service in the next period, when the new arrivals are in some fixed set $\hat{D}_t \quad [0,1]^2$, where $\hat{D}_t$ is chosen optimally. That is, the seller can give contingent promises, but of just one type every period, rather than offering different contracts for each type of buyers. Note that we allow $\hat{D}$ be different for all states D or even for all periods. We have already covered two special cases: if $\hat{D}_t = \quad$ at all periods, we get static mechanism, whereas if $\hat{D}_t = [0,1]^2$ each period, we get the pure calendar mechanism.

We are calling this class of mechanisms waiting list mechanisms, because the assumption can be interpreted as "waiting list" assumption. Of course, there are potentially many ways for how a waiting list could be organized in practice. In our interpretation, the crucial element of waiting list is that for the people who wait are put on a list, and once the person is on this list in a certain position, the type is forgotten and the process becomes the same for all types buyers who wait in the same position. This is exactly what the formal assumption means. In the case when the number of objects is one, the optimal waiting list will include at most one individual. In the general model we will analyze in Section 3 the optimal waiting list includes up to m individuals.

We will show that the optimal waiting list mechanism shares features with both the pure calendar mechanism and the fully optimal mechanism, but also has some new elements. In particular, the allocation rule is analogous with the pure calendar mechanisms and is characterized by the opportunity costs of service slots, whereas the structure of the optimal contract of delayed service is similar to the fully optimal mechanism and has the same non-monotonicity properties.

The new aspect is that contract must be designed for an average person who gets the contract, rather than particular one which has new implications. In particular, the dynamics of waiting list is non-trivial. It is possible that the optimal contracts are different each period. The distribution of buyers who receive the promise of delayed service will depend on the contract given to a buyer in the previous period. It is possible that in each period this contract is different from the previous contract and therefore the new optimal contract is also going to be different. As long as the new arrivals are such that one of them receives delayed service, the mechanism could therefore be in a new state every period.

The characterization of the optimal waiting list follows the same logic as in the previous subsections, so we will omit some steps and emphasize the differences. We start by assuming again that the continuation value functions $\quad(d)$ are well-defined and consider state D. Suppose that the optimal contract in this state is[11] $\hat{D}$ and denote the corresponding probability with $\hat{d}$.

[11]Remember that by waiting-list assumption, this cannot depend on the values of buyers, therefore we can

Now, notice that $[0, p\ ]^2 \quad \hat{D}$ and so $\hat{d} \quad F_2(p\ )$, since with those profiles the seller would never serve new buyers, so they are costless to add to the contract.

Characterization of the allocation rule with a fixed contract $\hat{D}$ follows exactly the same logic as with the pure calendar mechanism, only the opportunity cost of allocating the service is now $\frac{}{1-}[\ (\ ) - \ (\hat{D})]$, which obviously depends on the contract. That is, the allocation rule is characterized by threshold $\underline{v}$ that satisfies (2.6)

$$\hat{d}w(\underline{v}) = \frac{}{1-}[\ (\ ) - \ (\hat{D})]. \tag{2.6}$$

The next question to address is the relation between optimal probability $\hat{d}$ and contract $\hat{D}$. The answer is exactly the same as with the fully optimal mechanism, the optimal mechanism promises service in the situations where the total loss in revenue is low, which means with profiles where the loss in revenue, which we denoted by $^{-}(v) - \_(v)$, is smaller than some maximum value $\bar{b}$.

Now we need to decide, what is the optimal probability of delayed service and this is where we deviate from previous analysis. Before optimizing with respect to the probability of delay $\hat{d}$, we have to find how it affects the expected revenue. An increase in probability $\hat{d}$ has two opposite effects on revenue. First, the seller extracts a higher proportion of discounting-adjusted virtual surplus $\ w(v_i)$ from each buyer who receives the promise. Second, the contract will be more restrictive and therefore it decreases the continuation revenue whenever delayed service is promised. To study both effects we have to know the probability of assigning the contract to someone and the distribution of buyers' types who receive the promise. We already know from previous analysis that inside set D the delayed service is promised to the high type and outside the set D it is promised to the low type. In both cases the condition to assign the delayed service is that the corresponding type must be higher than threshold $\underline{v}$.

Let us define probability density function $g_v(\cdot|D)$ as[12] and the corresponding cumulative density function be $G_v(v_i|D)$.The distribution G_v that we defined here plays a crucial role in the analysis. Remember that the seller cannot discriminate between buyers who receive the delayed service contract. The distribution G_v characterizes the distribution of the type of buyer who receives the delayed service contract. It is a mixture of f_1 and f_2, the distributions of high and low type, where the precise mixture depends on the current state D. In particular, if the seller did not give any promises in the previous period, $D = \quad$, then it is the same as the distribution of low type, whereas if the seller promised delayed service with certainty (as in pure calendar mechanism), it is the distribution of high type.

From the allocation rule above we know that the seller assigns the delayed service contract to

without loss assume that it depends only on state D

[12]In particular,

$$g_v(v_i|D) = \int_0^{v_i} \mathbb{1}[(v_1, v_i) \quad D] f_1(v_1) f_2(v_i) dv_1 + \int_{v_i}^{1} \mathbb{1}[(v_i, v_2) \ / \ D] f_1(v_i) f_2(v_2) dv_2.$$

The function g_v is a PDF, since by construction $g_v(v_i|D) \quad 0$ for each v_i and $\int_0^1 g_v(v_i|D) dv_i = Pr(D) + Pr(D^c) = 1$.

a corresponding buyer if and only if her type $v_i > \underline{v}$. This happens with probability $1 - G_v(\underline{v}|D)$. The expected revenue extracted from the buyer who receives the promise of delayed service is now $\ _{\underline{v}}^{1} \ \hat{d}w(v_i)dG_v(v_i|D)$.

As in the case of fully optimal mechanism, it is convenient to change variables. Instead of working with probability $\hat{d}$, we work with maximum distortion $\bar{b}$ that characterizes the contract $\hat{D}$ and therefore probability $\hat{d}$. In particular, let $G_b(\bar{b}) = Pr(^{-}(v) - \_(v) \quad \bar{b})$. Then $\hat{d} = G_b(\bar{b})$. Moreover, we also know how to compute $\ (\ ) - \ (\hat{D})$. Compared to non-constrained optimization problem at state , state $\hat{D}$ forces the seller to serve the previous buyer at set $\hat{D}$ and therefore lose $^{-}(v) - \_(v)$ at all these realizations. A convenient way to compute this value is to integrate over G_b. We get that $\ (\ ) - \ (\hat{D}) = \ _0^{\bar{b}} bdG_b(b)$.

We can therefore rewrite the contract design problem in terms of $\bar{b}$ to obtain the following maximization problem[13].

$$\max_{\bar{b} \quad [0,1- \ ]} \quad (1 - \ ) \ G_b(\bar{b})E_{G_v}[w(v_i)|v_i \quad \underline{v}] - \ _0^{\bar{b}} bdG_b(b) \quad ,$$

which gives a first order condition[14]

$$(1 - \ )E_{G_v}[w(v_i)|v_i \quad \underline{v}] = \bar{b} \tag{2.7}$$

Similarly to the fully optimal case, (2.7) balances the trade-off from extracting slightly more revenue from the average type who receives the contract $\hat{D}$ and the loss from added distortion in the next period.

We have now characterized the optimal optimal mechanism at state D with two variables, the maximum distortion $\bar{b}$ from contract $\hat{D}$ and threshold $\underline{v}$ that characterizes the opportunity cost of promising the contract $\hat{D}$. The two variables are related by equations (2.6) and (2.7), which can be rewritten as (2.6') and (2.7') respectively.

$$w(\underline{v}) = E_{G_b} \quad \frac{b}{1 -} \quad \frac{b}{1 -} \qquad \frac{\bar{b}}{1 -} \quad , \tag{2.6'}$$

$$\frac{\bar{b}}{1 -} = E_{G_v}\left[w(v_i)|w(v_i) \quad w(\underline{v})\right]. \tag{2.7'}$$

The first condition (2.6') characterizes the trade-off between promising a delayed service to marginal type $\underline{v}$ and some revenue next period. In particular, the opportunity cost of $\hat{D}$ is the average distortion it creates next period. The distortion has conditional distribution function $\frac{g_b(b)}{G_b(\bar{b})}$, which gives us the condition. The second condition (2.7') characterizes the trade-off between increasing the set $\hat{D}$ slightly and therefore extracting slightly more revenue from average

[13] $E_{G_v}[w(v_i)|v_i \quad \underline{v}] = \ _{\underline{v}}^{1} \ w(v_i)\frac{dG_v(v_i|D)}{1-G_v(\underline{v}|D)}$ is the expected surplus of a buyer who receives the delayed service contract

[14] Note that we can ignore the effect though $\underline{v}$ by using the envelope theorem.

type who receives the promise $\hat{D}$ and the additional distortion this increase creates.

Both equations (2.6') and (2.7') describe strictly increasing continuous relationships between $\frac{\bar{b}}{1-}$ and $w(\underline{v})$, and it is straightforward to check that the system has a unique solution. Notice that at $\underline{v} = 1$ (2.7') gives $\frac{\bar{b}}{1-}$, whereas at $w(\underline{v}) = 0$ the value is in $(0, 1)$. Similarly, at $\bar{b} = 0$, (2.6') gives $w(\underline{v}) = 1$ and at $\frac{\bar{b}}{1-} = 1$ the $w(\underline{v})$ $(0, 1)$. Therefore indeed the solution exists, is unique, and is interior. Finally, the proof of existence and uniqueness for (d) functions is analogous to the fully optimal mechanism and uses the Contraction mapping theorem.

The implementation of the mechanism relies on the same ideas as the pure calendar mechanism and the fully optimal mechanism above. In particular, the allocation is characterized by the opportunity costs and contract design is similar to the fully optimal mechanisms, which means that we have to use non-standard auction with scoring rule again and get similar non-monotonicities. The graphic illustration is given by Figure A.2

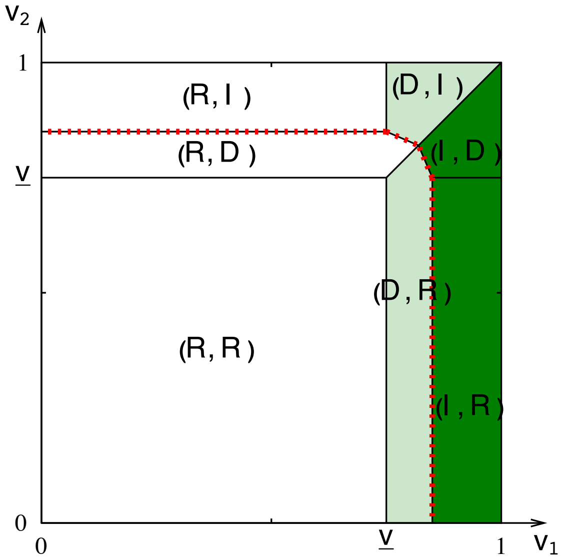


Figure 2.5: Optimal waiting list mechanism. Intensity of shading color describes the discounted expected quantity for buyer 1, the area below dotted line is the set where the previous-period customer is served.

A new aspect of the waiting list mechanism is its potentially non-trivial dynamics in the following sense. Suppose the initial state is such that no promises have been given, $D_0 =$. Then the seller designs some optimal contract D_1 that is offered to some types of buyers, which will always be the lower of the two. Now, at the next period the state is D_1, which is different from D_0. Therefore the types of buyers who now receive a promise of delayed service come from a different distribution, which is a mixture of high and low types of buyers, characterized by a function $G_v(v_i|D_1)$. The optimal contract is now different, say D_2. If this contract is given to a new buyer, in period 2 the state is D_2, which is again different from D_0 and D_1. Continuing the same way, we get a sequence of promises $D_1, D_2, \dots$, which are all different and optimal, given that there has been a sequence of arrivals such that one buyer always got a promise of delayed service.

2.4 Partial commitment

Since this paper focuses on promises of delayed service, commitment power plays a crucial role. Whenever a promise is restrictive, the seller would have incentive to break it in the future. In this section and in the general model in the next section, we are extending the results from the three classes of mechanisms studied thus far to a general partial commitment assumption. The full commitment analysis we did above and no commitment are special cases of the general assumption.

There are many reasons why the promises are kept: legal rules and fines, reputation of the seller, or perhaps intrinsic motivation. In this paper we are taking a reduced-form approach and define partial commitment as an exogenous cost of breaking a promise to one buyer. This could be interpreted as a fine that has to be paid to a government agency. Of course, the model could be extended to make the commitment constant endogenous, for example by modeling reputation explicitly or allowing the seller to pay the compensation directly to the consumer, but this is not the focus of the current paper.

As always in mechanism design literature, we assume that the sets of states and promises are such that the seller is able to fully commit to any promise she has made. To study the lack of commitment, we are restricting the set of possible promises.

We define partial commitment with constant c as a restriction to possible promises and therefore states. A seller cannot give any promise with which at any future realization the benefit from breaking a promise to one buyer would increase revenue by more than c. Then c can be interpreted as a commitment cost or fine for breaking the promise. If $c = 0$ we say that the seller has no commitment power and if c she has full commitment power.

Benchmark: no commitment. As a benchmark, let's consider the case when the seller does not have any commitment power. This will be the limiting solution for both the waiting list and the unrestricted mechanism. Then the commitment constant is $c = 0$, which means that breaking any promise is costless. This is as if the next period seller is a separate player and the current seller does not have any power to enforce promises. In this situation, promising delayed service is limited. Since the transfers from the previous period are already sunk, the seller always gives the priority to new arrivals. Whenever a new buyer has positive virtual surplus, the seller would like to serve her as soon as possible. Feasibility implies that higher type is offered instant service and lower type is offered delayed service with highest feasible probability, whenever their types are above p. This puts an upper bound to the promises that can be given about the future service. The seller can promise service only in set $\hat{D} = [0, p]^2$, where p is defined so that virtual surplus $w(p) = 0$. Therefore if the buyer with lower type has positive virtual surplus, she is offered to be served with probability $F_2(p)$ and will be served when v $\hat{D} = [0, p]^2$. This allocation rule is illustrated by Figure 2.6.

The optimal no-commitment mechanism can be implemented by the following indirect

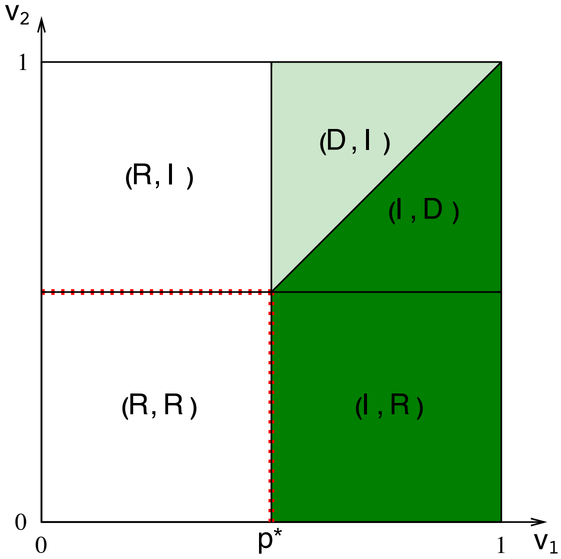


Figure 2.6: Optimal no-commitment allocation. Intensity of shading color describes the discounted expected quantity for buyer 1 , the area below dotted line is the set where the previous-period customer is served.

mechanism. First, both types are asked to pay $F_2(p\ )\ p$ for a chance to be served tomorrow, if the seller has free time. After this, an upgrade to instant service is sold at a second price auction with reserve price $[1 - \ F_2(p\ )]p$.

Pure calendar mechanism with partial commitment. As long as commitment constant c is higher than $1-$, commitment level is sufficient to sustain the optimal pure calendar mechanism described in Section 2.1. However, when $c < 1 -$, the only possible calendar mechanisms are static. The discreteness of the result comes from the pure calendar mechanism assumption, where the seller either does or does not use unconditional promises of delayed service; there are no intermediate cases. Promising delayed service with certainty decreases the revenue in the next period, which means that the pure calendar mechanism requires some commitment power. The question is how much commitment is sufficient.

The incentive to break a promise to a previous buyer would be the highest when both new arrivals have very high type, so the limiting case is when $v = (1, 1)$. In this case when keeping the promise, the seller would get $w(1) =$ from promising delayed service to one of the buyers and would have to refuse service to the other. When breaking the promise, he could instead get $1 +$, with the same continuation value. This means that the maximal gain from breaking the promise in terms of normalized revenue is $(1 - \ )w(1) = 1 -$. This is the necessary commitment level c that ensures that the seller is able to keep any promises.

Waiting list mechanism with partial commitment. Above we computed the optimal no-commitment mechanism, which is a waiting list mechanism, since the delayed service is always with a fixed set $\hat{D} = [0, p\ ]^2$ and therefore does not discriminate between buyers who are offered delayed service. As the opposite benchmark we have also characterized the optimal waiting list with full commitment in Section 2.3.

The optimal waiting list with full commitment at state D gave an optimal promise in the form $\hat{D} = \{v : \bar{\ }(v) - \_(v) \quad \bar{b}\}$, where $b \quad [0, 1 - \ ]$ is the maximum distortion that we know how to compute, and $\bar{\ }(v) - \_(v)$ is the distortion from serving a buyer from the previous period to a current revenue when arriving buyers announce v (the difference between revenues when the instant service is and is not available). Therefore with arrivals v, if the seller breaks the promise to a previous buyer, the additional revenue is exactly $\bar{\ }(v) - \_(v)$. This means that the maximal incentive to break a promise $\hat{D}$ is the maximum distortion $\bar{b}$. If the maximum distortion $\bar{b}$ is lower than commitment constant c, the seller can give promise $\hat{D}$, whereas when $\bar{b} > c$, it is known that the promise will be broken under some realizations and therefore this promise is not possible.

From this argument we see that finding the optimal waiting list mechanism for general partial commitment c is not very different from what we did with full commitment. The only difference is that maximum distortion $\bar{b}$ now has an additional upper bound c.

It is illustrative to consider the two extremes. In the limit when commitment constant c converges to 0 we obviously get the no-commitment solution described above. However, the no-commitment solution is optimal only in the limit. Whenever $c > 0$, the seller would make the contract $\hat{D}$ at least slightly larger than $[0, p\ ]^2$. Increasing the probability of service is valuable for all buyers who receive delayed service and therefore strictly increases revenue. Near the no-commitment solution it is almost costless, since the first types added to $\hat{D}$ have virtual surplus close to zero and therefore there is almost no distortion from not serving them.

The upper bound of $\bar{b}$ in the full commitment analysis was $1 - \ $. Therefore, if $c \quad 1 - \ $, the commitment constraint is never binding. This is the same sufficient commitment level that was needed to commit to the optimal pure calendar mechanism. When $c < 1 - \ $, we will sometimes get a corner solution where maximum distortion $\bar{b} = c$ and therefore the probability of service $\hat{d}$ is always strictly below 1. In this sense a waiting list can be seen as a relaxation of a pure calendar mechanism.

Unrestricted mechanism with partial commitment. The effect of adding partial commitment to the fully optimal mechanism (discussed in Section 2.2) is similar as to the waiting list. Again, it means that any promise of delayed service must be such that the distortion created by this promise to the next period's revenue is not larger than c. The only difference is that in this case we allow the promises to depend on the type of the buyer who receives the promise, so that the upper bound must hold for each type. Formally $\bar{b}(v_i) \quad [0, \min\{1 - \ , c\}]$. Again, the limit when $c \quad 0$ is the no-commitment mechanism (but it is exactly optimal only if $c = 0$) and when $c \quad 1 - \ $, the commitment level is sufficient to sustain any mechanism.

3 General analysis

In this subsection we give formal results for arbitrary fixed number of service slots and random number of buyers. The main trade-offs are the same as with the illustrative example, but obviously the notation and characterizations of optimal mechanisms are somewhat more complex. In contrast to the previous section, we are prove the results with the partial commitment assumption, so that full commitment is a special case.

3.1 Model

Seller has fixed number $m \quad \mathbb{N}$ service slots each period $t \quad \{0, 1, \dots\}$ and maximizes expected discounted revenue with discount factor $\quad (0, 1)$.

Every period random number n new buyers arrive such that the maximum number of buyers arriving with positive probability, $N > m$. Each buyer has unit demand and independent private value v_i which is i.i.d. draw from CDF F and PDF $f(v) > 0$. Let $F_{k:n}$ and $f_{k:n}$ denote the CDF[15] and PDF[16] of k'th smallest value out of n (k'th order statistic). Then for example, $F_{n:n}$ is the CDF of the maximum of the values of all arrived customers.

We maintain two important assumptions. First, each buyer interacts with the seller only at arrival. Buyer i who is promised service s_i periods after her arrival and pays p_i at arrival gets payoff of $\ ^{s_i} v_i - p_i$. Second, we concentrate on mechanisms where the seller never makes promises of delayed service more than one period ahead.

The promise of delayed service to buyer i is a set $D_i \quad \{(n, v) : n \quad \{0, \dots, N\}, v \quad [0, 1]^n\}$ and a state is a vector $D = (D_1, \dots, D_N)$, a collection of promises given to buyers from the previous period. If buyer i receives a promise of delayed service with a contract D_i, she will be served in the next period if and only if the number of buyers n and their announcements v are such that $(n, v) \quad D_i$.

Since buyers' utility functions are linear in value and quasilinear in transfers, the only variables affecting buyer i's payoff are her transfer $p_i(v)$ and discounted expected quantity, denoted by $\ _i(v)$, which is 1 if i receives instant service, 0 if refused service, and $\ Pr(D_i)$ if is promised delayed service in set D_i. These variables are functions of the allocation the seller chooses as a response to vector of types v announced by the buyers in this period. Let $\tilde{p}_i(\hat{v}_i)$ and $\tilde{\ }_i(\hat{v}_i)$ denote the expected values of the respective variables over the other buyers' types assuming they report their types truthfully.

We are looking for (Bayesian) incentive-compatible and individually rational mechanisms. In this notation, the individual rationality constraint (IR) for buyer i with type v_i is $\tilde{\ }_i(v_i)v_i - \tilde{p}_i(v_i) \quad 0$. The incentive-compatibility constraint (IC) is $\tilde{\ }_i(v_i)v_i - \tilde{p}_i(v_i) \quad \tilde{\ }_i(\hat{v}_i)v_i - \tilde{p}_i(\hat{v}_i)$ for all $\hat{v}_i \quad [0, 1]$.

Using the standard methods developed by Myerson (1981) it is straightforward to verify that a mechanism satisfies IC and IR constraints if and only if (1) the discounted expected quantity

---

[15] $F_{k:n}(v_i) = \ {}^{n}_{j=k} \ {}^{n}_{j} \ F(v_i)^j [1 - F(v_i)]^{n-j}$.

[16] $f_{k:n}(v_i) = \frac{n!}{(k-1)!(n-k)!} F(v_i)^{k-1} [1 - F(v_i)]^{n-k} f(v_i)$.

$\tilde{\ }_i$ is weakly increasing in buyer's own type, (2) buyer with type $v_i = 0$ gets payoff at least 0 (an in optimum exactly 0), (3) the transfers are computed from ex-ante envelope condition

$$\tilde{p}_i(v_i) = v_i \tilde{\ }_i(v_i) - {}_0^{v_i} \tilde{\ }_i(\hat{v}_i) d\hat{v}_i + \tilde{p}_i(0). \tag{3.1}$$

Although we are looking for optimal mechanisms that induce truthful behavior when buyers do not know the other buyers' types, it turns out that all the optimal mechanisms are implementable even when they know the other types. Let's define ex-post individual rationality constraint (EPIR) for buyer i when the type vector is $v = (v_1, \ldots, v_n)$ by $\ _i(v)v_i - p_i(v) \quad 0$. The dominant-strategy incentive-compatibility constraint (DSIC) is defined as $\ _i(v)v_i - p_i(v) \qquad _i(\hat{v}_i, v_{-i})v_i - p_i(\hat{v}_i, v_{-i})$ for all i's announcements $\hat{v}_i \quad [0, 1]$ and vectors of opponents' types $v_{-i} \quad [0, 1]^{n-1}$.

Again, it is straightforward to verify using standard methods that a mechanism satisfies DSIC and EPIR constraints if and only if (1) the ex-post discounted expected quantity $\ _i$ is weakly increasing in buyer's own type v_i, (2) buyer with type $v_i = 0$ always gets a payoff at least 0 (and in optimum exactly 0), and (3) the transfers are computed from ex-post envelope condition

$$p_i(v_i) = v_i \ _i(v_i) - {}_0^{v_i} \ _i(\hat{v}_i) d\hat{v}_i + p_i(0). \tag{3.2}$$

We are using these results for two purposes. First, the ex-ante envelope condition (3.1) and the fact that $p_i(0)$ must be 0 are used to express the seller's expected revenue at state z as

$$(z) = \max E_{n,v} \quad (1 - \quad) \sum_{i=1}^{n} \tilde{p}_i(v_i) + \quad (z \ ) \quad = \max E_{n,v} \quad (1 - \quad) \sum_{i=1}^{n} \ _i(v)w(v_i) + \quad (z \ ) \quad ,$$

where the seller maximizes over feasible promises, $w(v_i) = v_i \frac{1-F(v_i)}{v_i}$ is the standard static virtual surplus, and z is the continuation state. This expression is separable in (n, v). That is, the seller can choose optimal promises for each possible combination (n, v) separately and the value of being in a particular state can then be computed as expectation over all realizations.

We will find the optimal allocation rule for different restrictive classes of mechanisms that restrict feasibility. In all the cases, the optimal allocation rule turns out to be monotone with respect to buyers' own types. Therefore, as the second use of the results above, we can apply ex-post envelope condition (3.2) to compute the transfers that implement this mechanism under even stronger conditions (EPIR, DSIC).

Before studying the dynamic mechanisms, there is one special mechanism in this model that serves as a useful benchmark and for which the allocation rule and implementation are well known. We are call a mechanism static, if the seller never gives promises for delayed service. That is, at each period, each buyer either gets instant service or is refused service. This mechanism does use delayed service, so the restrictions for possible promises do not apply and it is therefore a feasible mechanism in all the cases studied below.

Analogously to Myerson (1981), the static optimum in this framework is the allocation rule that assigns objects to m buyers with highest types, given that their virtual valuations are positive. This is implemented under DSIC and EPIR constraints by a standard uniform price auction with reserve price p such that $w(p) = 0$.

3.2 Pure calendar mechanism

The pure calendar mechanism is a mechanism where delayed service is unconditional on new arrivals. Each buyer is either assigned instant service, refused service, or promised service in the next period with certainty. The general set of states $D = (D_1, \ldots, D_N)$ can be simplified to a number z $\{0, \ldots, m\}$. This is true since instead of a promise set we only have to keep track of which buyers received the delayed service and which did not. Moreover, from the seller's point of view, who did and who did not receive the delayed service is not payoff-relevant—the only relevant variable is how many people were promised delayed service. Therefore a state z $\{0, \ldots, m\}$ denotes how many service slots are unavailable, and thus $m - z$ slots are still available.

The main result in this section is Theorem 3.1, which characterizes the optimal pure calendar mechanism (with partial commitment). The result has the same general trade-offs as the illustrative example in Section 2.1. First, delay is optimal only if buyer's discount factor is high enough compared to . In particular, $-$ must be at least $E_{n,v}[\max\{w(v_{n:n}), 0\}]$, which is the average of expected revenue that can be extracted from the highest type buyer in a static mechanism. If is strictly less than times this value, then it would not be optimal to give delayed service even to buyers with highest possible values of 1. The static mechanism would be strictly better. Note that the requirement is satisfied, for example, when .

Second, the seller's commitment power must be high enough. The sufficient commitment power is the same as in the illustrative example, c $1 -$ for the same reason. If this condition is satisfied, even if all new arrivals have types very close to 1, the seller would prefer not to break previous promises. But in general it is necessary only if N $2m$. When $N < 2m$, there can be only $N < 2m$ buyers with high positive surplus next period. This limits the cost of current promises, since the seller is never forced to refuse service to very high customers (he can extract part of revenues by promises of future service).

The main trade-off is again between extracting revenue from current relatively high buyers and the opportunity cost of having one less service slot available next period. Giving away each additional unit of next period service will be associated with opportunity due to being in a more restrictive state next period. The optimal pure calendar mechanism balances extraction from current buyers by promising some of them delayed service and the opportunity costs of each additional slot that is promised away.

In particular, the optimal pure calendar mechanism is characterized by m thresholds, denoted by $\underline{w} = (\underline{w}_1, \ldots, \underline{w}_m)$, where $\underline{w}_k$ can be interpreted as the opportunity cost of the k'th object. The actual choice of promises will be non-trivial. We will describe below how to make this choice and Corollary 3.3 describes a special case where the allocation rule is much simpler.

Theorem 3.1. *Suppose m $\mathbb{N}$ and n . Under sufficient conditions* $E_{n,v}[\max\{w(v_{n:n}), 0\}]$

and $c \quad 1-\quad$, the optimal pure calendar mechanism mechanism with partial commitment uses the following allocation rule at each state $z = \{0, \ldots, m\}$

(i) $m - z$ new buyers with highest strictly positive virtual surplus will get instant service.

(ii) If there are more than $m - z$ buyers with positive virtual surplus, the next k highest of them will be promised service in the next period, where k is defined as

$$k = \underset{k \quad \{0,\ldots,m\}}{\arg\max} \quad \overset{m}{\underset{i=1}{}} [w(v_{n+1-(m-z)-i:n}) - \underline{w}_i]$$

where constants $\underline{w}_1, \ldots, \underline{w}_m$ are uniquely determined by

$$\underline{w}_j = -E_{n,v} \quad \max\{0, w(v_{n-(m-j):n})\}$$

$$+ \quad \max_k \overset{k}{\underset{i=1}{}} w(v_{n+1-(m-j)-i:n}) - \underline{w}_i \quad - \max_k \overset{k}{\underset{i=1}{}} w(v_{n-(m-j)-i:n}) - \underline{w}_i \quad .$$

The result covers the situations where it is optimal to offer delayed service and commitment level is sufficient to do so. The following Corollary 3.2 shows that these sufficient conditions are also almost necessary. In particular, the condition relating and is necessary, whereas the condition on commitment level is necessary when the probability that at least $2m$ new buyers arrive is positive.

Corollary 3.2. *If one of or neither of the sufficient conditions do not hold:*

(i) When $\quad < \quad E_{n,v}[\max\{w(v_{n:n}), 0\}]$, the optimal mechanism is static.

(ii) When $N \quad 2m$ and $c < 1 -\quad$, the optimal mechanism is static.

(iii) When $m < N < 2m$ and $c < 1 -\quad$, depending on details the optimal mechanism could be static or with some delay.

The proofs for both are in Appendix B.1. Here is the overview of the proofs. We start by assuming that and c are high enough, so that there is some optimal delay. We can then make some instant observations: (1) buyers with negative virtual surplus are never served, (2) if there are enough buyers with positive virtual surplus, all current objects will be allocated, and (3) promises are monotone in types (higher types will be served earlier). This gives an almost full description of optimal allocation rule.

The remaining question is for situations where there are more buyers with strictly positive virtual surplus than current service slots. The question is, to how many of them should the seller promise delayed service. To do this, we define constants $\underline{w}_1, \ldots, \underline{w}_m$ where $\underline{w}_k$ can be interpreted as the opportunity cost of giving away kth unit of tomorrow's service. Now, vector $\underline{w}$ is defined in terms of continuation values and is therefore endogenous. To show that $\underline{w}$ exists and is uniquely defined we write the equation system for $\underline{w}$ in the form of $\underline{w} = \quad (\underline{w})$ and show that there exists a fixed point. The solution—if it exists—will be unique because will be strictly monotone.

The final part of the proof is about commitment power. We compute the cost of keeping a promise in all possible situations for each additional buyer. Value $1 -$ will be an upper bound of these costs, therefore if $c \quad 1 -$, commitment power is sufficient. When $N \quad 2m$, this upper bound is the lowest upper bound, since it is possible that $2m$ buyers arrive and all have values close to 1, which gives the bound. When $N < 2m$, this situation is not possible and therefore it is possible to have optimal delay even when c is slightly below $1 -$.

Theorem 3.1 implies that the profit from calendar mechanism at state z can be written as

$$(z) = E_{n,v} \quad (1 - \quad) \sum_{i=1}^{m-z} \max\{0, w(v_{n+1-i:n})\} + \quad (m) + (1 - \quad) \max_k \sum_{i=1}^{k} w(v_{n+1-(m-z)-i:n}) - \underline{w}_i \quad .$$

The first two terms add up to the maximum revenue that can be extracted by static promises—the revenue from highest $m - z$ buyers plus continuation value of having all m object available next period. The last term is positive and captures the extra revenue from delayed service to buyers with lower valuations. In particular, it is the sum of virtual surpluses extracted from the k buyers who are promised delayed service minus the opportunity costs $\underline{w}_i$ that are lost due to the fact that fewer objects are available in the next period.

Suppose (n, v) is such that there are enough buyers with positive surplus to allocate all current and next period objects, that is, $v_{n-2m:n} > p$. In this case, the number of buyers who are served instantly z, which is clearly increasing in z. In the appendix we prove Lemma B.1 which shows that the optimal number of buyers that are offered delayed service is weakly decreasing in $m - z$. This is true since with higher $m - z$ more of the high-value buyers will receive instant service and therefore the remaining buyers have lower values.

In general the optimal mechanism is non-monotone in terms of assigning delayed service. It is possible, for example, that at some state with some vector of values two buyers are offered delayed service, but after weakly increasing all the values only one buyer is offered delayed service. There is one case where the mechanism is relatively easier to interpret and implement—if the thresholds characterized by Theorem 3.1 happen to be ordered. In this case the assignment is monotone and therefore each of the following buyers is served if and only if her value is higher than the corresponding threshold. Corollary 3.3 gives the formal result.

Corollary 3.3. *If the assumptions in Theorem 3.1 hold and the implied vector $\underline{w}$ is such that $\underline{w}_1 \quad \cdots \quad \underline{w}_m$, then the optimal mechanism uses the following allocation rule at each state z*

(i) $m - z$ buyers with highest strictly positive virtual surplus will get instant service.

(ii) If there are more than z buyers with positive virtual surplus, k highest of them will be promised service in the next period, where k is such that $w(v_{n+1-(m-z)-i:n}) > \underline{w}_i$ if and only if $i \quad k$.

The expected revenue at state is

$$(z) = E_n \quad (1 - \quad) \quad {}_{i=1}^{m-z} \quad {}_{n+1-i}(0) + \quad (0) + (1 - \quad) \quad {}_{i=1}^{m} \quad {}_{n+1-(m-z)-i}(\underline{w}_i) \quad .$$

where ${}_k(\hat{w}) = {}_{w^{-1}(\hat{w})}^{1}[w(v_{k:n}) - \hat{w}]dF_{k:n}(v_{k:n})$.

Proof is in Appendix B.1.

3.3 No commitment

As a benchmark, we are considering the limiting case where the seller does not have any commitment power. That is, commitment constant $c = 0$. In this case the seller always gives priority to new arrivals, since the payments from previous buyers is sunk. That is, whenever a new buyer with strictly positive virtual surplus arrives, she will be served before all waiting buyers. This implies that promises can be made only for the cases when there are sufficiently few new buyers with positive virtual surpluses. The following Proposition 3.4 formalizes this.

Proposition 3.4. *Suppose $m \quad \mathbb{N}$ and $n \quad$. The optimal no-commitment mechanism is characterized as follows. Let k be the number of new buyers with strictly positive virtual surpluses.*

(i) If $k \quad m$, then m new buyers with highest strictly positive virtual surpluses are served instantly and no waiting buyers are served. If $k < m$ then all k new buyers with strictly positive virtual surplus as well as $m - k$ waiting buyers with highest positive surpluses are served.

(ii) Buyers with strictly positive surpluses who were not served instantly are promised that they will be served if sufficiently few buyers with strictly positive surpluses arrive next period. In particular, $n + 1 - m - i$th highest buyer will be served next period if and only if his virtual surplus is strictly positive and next period $k \quad m - i$ new buyers with strictly positive virtual surplus arrive.

Proof follows from the discussion above.

3.4 Waiting list

The waiting list mechanism is a mechanism where the promise depends only on the position and not the type of buyer who is in the position. Formally, at each state $D = (D_1, \ldots, D_N)$, the waiting list is characterized by a vector of contracts $\hat{D} = (\hat{D}_1, \ldots, \hat{D}_N)$ that are fixed before the arrival of buyers. Each contract $\hat{D}_i$ is a set $\hat{D}_i \quad \{(n, v) : n \quad \{0, \ldots, N\}, v \quad [0, 1]^n\}$ that describes under what conditions the receiver of the contract will be served.

In addition to contracts $\hat{D}$ being fixed before the number of buyers and their values being observed, the waiting list assumption says that when k new buyers are promised delayed service, they receive contracts $\hat{D}_1, \ldots, \hat{D}_k$. This is the "list" part of the assumption. For example if buyer i was promised set $\hat{D}_k$, it is as if he is added to some list in the position k, after $k - 1$

other buyers. He will be served if and only if n new buyers who arrive next period announce values v such that $(n, v) \quad \hat{D}_k$.

Notice that we do not explicitly assume that the waiting list is a priority ranking in the sense that being higher in the list ensures more likely service than being in a lower position. This is a property of an optimal waiting list. In particular, by Theorem 3.5 the optimal waiting list is such that $\hat{D}_1 \quad \hat{D}_2 \quad \cdots \quad \hat{D}_m$ and $\hat{D}_k = \quad$ for all $k > m$.

Another property of optimal waiting list will be that each contract is determined separately from all other contracts. It will balance the same trade-offs that the illustrative example. On one hand, the assignment of jth contract balances the marginal surplus extracted from a buyer who gets this contract (analogously to the pure calendar mechanism) and on the other hand the precise composition of the contract balances the trade-off between average surplus extracted from a buyer who gets the contact and increased distortion from this contract (which is analogous to the fully optimal case).

Theorem 3.5. *The optimal waiting list mechanism with partial commitment is such that at each state $D = (D_1, \ldots, D_N)$, when n buyers arrive and announce values v, then*

(i) $z = \#\{i : (n, v) \quad D_i\}$ *waiting buyers and $m - z$ new buyers with highest positive virtual surpluses are served instantly.*

(ii) *The next m highest buyers with positive virtual surpluses are offered delayed services with sets $\hat{D} = \quad \hat{D}_1, \hat{D}_2, \ldots, \hat{D}_m \quad$.*

where contracts $\hat{D}$ are characterized as follows.

$$\hat{D}_j = \{(n, v) : \quad ^{j-1}(n, w) - \quad ^{j}(n, w) \quad \bar{b}_j\},$$

$$^{j}(n, w) = (1- \quad) \sum_{i=1}^{m-j} \max\{0, w_{n+1-i}\} + \quad (0, \ldots, 0) + (1- \quad) \max_{k \quad \{0,\ldots,m\}} \sum_{i=1}^{k} \hat{d}_i[w_{n+1-(m-j)-i} - \underline{w}_i],$$

and $\underline{w}_j \quad [0, 1]$, $\bar{b}_j \quad [0, \min\{1 - \quad, c\}]$ *are characterized by*

$$\underline{w}_j = \frac{}{(1- \quad)} \int_0^{\bar{b}_j} b_j \frac{dG_{b_j}(b_j)}{G_{b_j}(\bar{b}_j)} \quad \text{and} \quad \frac{\bar{b}_j}{(1- \quad)} \quad \int_{\underline{w}_j}^{1} \hat{w}_j \frac{dG_{w_j}(\hat{w}_j)}{1 - G_{w_j}(\underline{w}_j)},$$

where $G_{b_j}(b_j) = Pr(\{(n, v) : \quad ^{j-1}(n, w) - \quad ^{j}(n, w) \quad \bar{b}_j\})$ *and* $G_{w_j}(\hat{w}_j) = \int_0^{\hat{w}_j} g_{w_j}(\hat{w}_j) d\hat{w}_j$, *st*

$$g_{w_j}(\hat{w}_j) = E_n \quad 1[D_1^c] f_{n+1-(m-j)}(\hat{w}_j) + \sum_{i=1}^{m-1} 1[D_i \setminus D_{i+1}] f_{n+1-(m-j)+i}(\hat{w}_j) + 1[D_m] f_{n+1-(m-j)+m}(\hat{w}_j) \quad .$$

Proof is in Appendix B.2 and follows the same steps as the illustrative example. Proof structure:

Step 0 We assume that functions $\quad (d)$ are well-defined and make some immediate observations regarding the contracts and the allocation rule.

Step 1 For a fixed vector of contracts $\hat{D}$ we characterize the optimal allocation rule. It will be characterized by a vector of constants $\underline{w} = (\underline{w}_1, \ldots, \underline{w}_m)$. Constant $\underline{w}_j$ can be interpreted as the opportunity cost of allocating the jth contract.

Step 2 For a fixed vector of probabilities of delayed service $\hat{d} = (\hat{d}_1, \ldots, \hat{d}_m)$ we show how to fulfill these promises with these probabilities optimally. In particular, each optimal contract $\hat{D}_j$ minimizes the distortion from not having one extra object j available.

Step 3 We derive the condition for optimal probability of delayed service for jth contact.To do this, we change the decision variables from probabilities of delay $\hat{d}$ to maximum distortions $\bar{b}$ and find the first order condition with respect to this. The optimum balances the revenue extracted from the average person who gets the contract and the marginal increase in the distortion from giving this contract.

Step 4 We argue then that we have characterized the optimal contracts. At this point we have derived two equations and of two variables that each contract j has to satisfy. The two variables are opportunity cost of allocating jth contract, $\underline{w}_j$, and the maximum distortion characterizing the contract j, $\bar{b}_j$. The equations are the corresponding first order conditions. We argue that each contract is well defined, either it is in the interior or in the upper corner. The upper bound for $\bar{b}_j$ is either $1 -$ or commitment constant c.

Step 5 Finally, the existence of (d) functions can be verified using Blackwell's sufficient conditions and the Contraction mapping theorem.

3.5 Fully optimal mechanism

In this section again, a state is a vector $D = (D_1, \ldots, D_N)$, where D_k denotes the realizations in which the seller has to serve a particular buyer corresponding to k from the previous period. Promises of delayed service are again denoted by $\hat{D} = (\hat{D}_1, \ldots, \hat{D}_N)$, but now this vector can depend on the announcements of current buyers.

The optimal mechanism has to fulfill all previous promises and assigns the remaining instant service slots to buyers with highest positive virtual values. If there still remains one or more buyers with positive surplus, the seller must decide what to promise. The result states that the promises they get are ordered in terms of both probabilities and sets. The highest of the remaining buyers is promised a probability $\hat{d}_1 = Pr(\hat{D}_1)$ that depends only on her valuation and is increasing in her type.

Now, the seller would want to assign all the possible delayed promises to the same buyer, since she is the remaining buyer with the highest value, but she cannot use more than one service slot next period. Therefore the seller reserves the first slot for her and uses this first slot to serve only her or new buyers. The next slot is matched with the next highest remaining buyer with the positive valuation—he will receive a promise that is increasing in his type, but since his value is lower than the first buyer's value, this probability is lower. The assignment continues until all m future slots are matched with buyers (some of whom may receive probability zero as a promise). The optimal way to fulfill the promise is such that the difference between maximized expected revenue when this slot is available relative to when it is unavailable is minimized.

Theorem 3.6. *The optimal unrestricted mechanism with partial commitment is such that at each state $D = (D_1, \ldots, D_N)$, when n buyers arrive and announce values v, then*

(i) $z = \#\{i : (n, v) \quad D_i\}$ *waiting buyers and* $m - z$ *new buyers with highest positive virtual surpluses are served instantly.*

(ii) The next m *highest buyers with positive virtual surpluses are offered delayed services with sets* $\hat{D}(v) = \quad \hat{D}_1(v_{n+1-(m-j)-1:n}), \hat{D}_2(v_{n+1-(m-j)-2:n}), \ldots, \hat{D}_m(v_{n+1-(m-j)-m:n}) \;$,

where contracts $\hat{D}(v)$ *are characterized as follows.*

$$\hat{D}_j(v_i) = \{(n, v) : \; ^{j-1}(n, v) - \; ^{j}(n, v) \quad \bar{b}(w_i)\}$$

$$^{j}(n, v) = (1- \;) \sum_{i=1}^{j} \max\{0, w(v_{n+1-i:n})\} + \max_{\hat{d}} \quad (1- \;) \sum_{i=1}^{m} \hat{d}_i w(v_{n+1-(m-j)-i:n}) + \quad (\hat{d}) \quad ,$$

$$(d) = E_{n,v}[\; ^{0}] - \sum_{j=1}^{m} E_{n,v} \; \mathtt{1}[\; ^{j-1}(n, w) - \; ^{j}(n, w) \quad \bar{b}_j][\; ^{j-1}(n, w) - \; ^{j}(n, w)] \;\; ,$$

$$\bar{b}(w_i) = \begin{array}{lll} 0 & & w_i \quad 0, \\ \frac{(1- \;)}{} w_i & & 0 < w_i < \frac{}{(1- \;)} \min\{1 - \;\; , c\}, \\ \min\{1 - \;\; , c\} & & w_i \quad \frac{}{(1- \;)} \min\{1 - \;\; , c\}. \end{array}$$

Proof is in Appendix B.3. The structure is the same as in the illustrative example and parts of it are analogous to the waiting list. Proof structure:

Step 0 We assume (d) is well-defined function and make the same immediate observations regarding the allocation rule and the optimal contracts as in the waiting list case.

Step 1 We derive the optimal allocation rule. In this case optimal delayed services are personalized, so we get first-order conditions relating probabilities of service $\hat{d}_k$ and buyer's virtual valuation who gets the promise.

Step 2 Next we describe how to fulfill promises optimally. The solution minimizes distortions as in the waiting list case.

Step 3 We can now analyze the first-order conditions we got in Step 1. We show that the optimal probability of delayed service depends only of buyer's own virtual surplus and is strictly increasing in the interior.

Step 4 Finally to prove the existence of (d) we use the Contraction mapping theorem.

4 Discussion

We studied the problem of selling service slots. From previous literature we already knew the solution to the problem in two special cases. First, when the capacity is high, the optimal mechanism is a constant posted price mechanism. Second, when the buyers' discount rate and exit rate are constant, the fully optimal allocation rule always assigns the objects to the highest-valuing[17] buyers among all remaining buyers. This paper showed how to solve the optimal mechanism design problem in the remaining case, where rationing is necessary and buyers are heterogeneous not only in their valuations for service, but also in their exit rates, so that the buyers from different periods should be treated differently. This introduces a new set of trade-offs which are balanced optimally by a new type of auction mechanism.

In addition to the fully optimal mechanism, we analyzed two classes of restricted mechanisms where the possible contracts of delayed service are limited: pure calendar mechanisms and waiting list mechanisms. The optimal pure calendar mechanism was a simple mechanism, characterized by a vector of opportunity costs. The qualitative properties of waiting lists were similar to the fully optimal mechanisms. Finally, we showed how the optimal mechanisms change in all three classes when the seller has less than perfect commitment power.

We made several simplifying assumptions that could be relaxed without changing the conclusions. We assumed that the arrivals of buyers are observable both for the other buyers and for the seller. This assumption is without loss in the sense that if buyers had a chance to choose whether to enter the mechanism at their arrival or later—without knowing anything about the other buyers—they would always enter immediately.

It would also be relatively straightforward to introduce more general arrival processes. We assumed that the number of objects m is constant and the number of buyers n is an independently and identically distributed random variable. We could assume instead that m and n are independent draws from some joint distribution, where the distribution can be different each period. This is a generalization that includes, for example, seasonal fluctuations in demand and supply.

In real life applications the most common mechanisms to allocate service are posted price mechanisms. As we argued, whenever there is a need for rationing, the optimal mechanism would involve an auction. The assumption that the seller has to use a posted price mechanism would be another restriction to the allocation rule. Some related papers[18] have shown that if the arrival process is continuous, then the optimal mechanism would be implementable with posted prices, since two buyers never arrive at the same time. The same could be done with the model here, by assuming arrival in continuous time and leaving the production unchanged. Alternatively, we could assume that buyers in one period arrive sequentially and the seller has to assign the allocation to each buyer at the moment of their arrival, before observing the other new buyers' arrivals and their types.

[17] If the goal is to maximize revenue, then the mechanism assigns service only to buyers with positive virtual surplus.

[18] See Gershkov and Moldovanu (2009) and Board and Skrzypacz (2010) for example.

Perhaps the most restrictive technical assumption was concentrating on contracts only one period ahead. A generalization of current models would involve a common discounting process , where payoff from being served eventually at date s_i is $(s_i)v_i - p_i$, such that (s_i) is decreasing in s_i and $\lim_s (s_i)/ \ ^{s_i} \quad 0$. Special cases of this model would be $(s_i) = \ ^{s_i}$ which follows from Said (2012) and $(0) = 1$, $(1) =$, $(s_i) = 0$, which was studied in this paper.

References

Bergemann, D., and J. Välimäki (2010): "The Dynamic Pivot Mechanism," *Econometrica*, 78(2), 771–789.

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

Board, S., and A. Skrzypacz (2010): "Optimal dynamic auctions for durable goods: Posted prices and fire-sales," .

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 (2012): "Overbooking," .

Garrett, D. (2013): "Incoming Demand with Private Uncertainty," .

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.

Hart, O. D., and J. Tirole (1988): "Contract Renegotiation and Coasian Dynamics," *The Review of Economic Studies*, 55(4), 509–540.

Li, A. (2009): "Selling Storable Goods to a Dynamic Population of Buyers: A Mechanism Design Approach," .

Mierendorff, K. (2011): "Optimal dynamic mechanism design with deadlines," .

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

Pai, M., and R. Vohra (2008): "Optimal Dynamic Auctions," Discussion Papers 1461, Northwestern University, Center for Mathematical Studies in Economics and Management Science.

Pavan, A., I. Segal, and J. Toikka (2013): "Dynamic Mechanism Design: A Myersonian Approach," .

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.

Appendices

A Different discount factors

In contrast to the illustrative example in Section 2, the general results in Section 3 are derived under assumption that seller's discount factor is not necessarily the same as buyers' discount factor .[19] In this subsection we will discuss which results change if we relax this assumption in the illustrative example.

A.1 Pure calendar mechanism

The only qualitative change in the pure calendar mechanism analysis is that (2.1) that defines threshold $\underline{v}$ takes the form (2.1') and this equation may not have a solution.

$$w(\underline{v}) = \frac{}{1-} = \quad (\ ^s + \ [\ _1(\underline{v}) - \ _2(\underline{v})]) \,. \tag{2.1'}$$

To see why equation (2.1') may not have a solution in $\underline{v}$ $[0,1]$, consider the case when the buyers' discount factor is very low compared to the seller's discount factor . With sufficiently small , the value of delayed service is negligible to the buyers and therefore it is never optimal to offer delayed service. In this case the optimal calendar mechanism is static. We already know that the optimal static mechanism is the second price auction with reserve price p , which guarantees revenue ^{s}.

Suppose now that is slightly higher, so that the seller is exactly indifferent between offering delayed service to a very small set of potential buyers and using the optimal static mechanism. The incentive to offer delayed service is clearly highest for highest types. In particular, the gain from offering delayed service to a buyer with type $v_i = 1$ would be , but the cost would be lost revenue ^{s} next period, with discounted value ^{s}. This gives ^{s} as a necessary and sufficient condition for optimal delay. If $/ \; > \; ^s$, it is optimal to offer delayed service with

[19] However, we maintain assumption that discount factor is the same for all buyers.

positive probability, whereas when $/ < {}^{s}$ the optimal pure calendar mechanism is always static.

A.2 Fully optimal mechanism

The effect of very low buyers' discount factor is less extreme for the fully optimal mechanism. As we argued above, the seller can costlessly offer a positive probability of service by including only the set $[0, p\ ]^2$ in the contract. Therefore, whenever > 0, the optimal mechanism offers some delayed service. In particular, the first-order condition (2.3') takes the form (2.3''), which has a solution for arbitrarily low values of $-$.

$$(1 - \ )\ w(v_i) = \ \bar{b}(v_i). \tag{2.3''}$$

However, it is possible to get a corner solution for high values of $-$ when v_i is high enough. This is not surprising. In the example above, the optimal contract for the highest type $v_i = 1$ was to offer delayed service with certainty. Of course, when $- > 1$, there is even bigger reason to offer this type the delayed service with very high probability. Moreover, even some types close to 1 will be offered the same contract. In the extreme where $-$ is very high, the seller is very impatient compared to the buyers and therefore it is profitable to sell future service as much as possible, since it gives current revenue and the continuation value does not matter much.

A.3 Waiting list

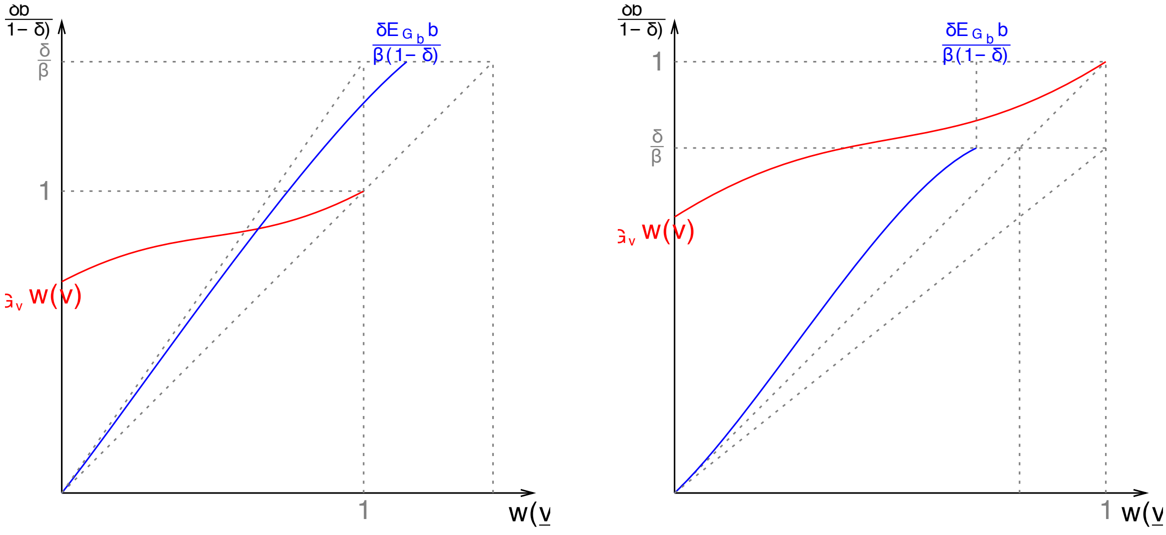
In the waiting list case, we characterized the optimal mechanisms by two Equations (2.6') and (2.7') that gave us unique interior values for $\frac{\bar{b}}{1-}$ and $w(\underline{v})$. With different discount factors, it is convenient to derive the equation system for variables $\frac{\bar{b}}{(1- \ )}$ and $w(\underline{v})$ and it is given by Equations (2.6'') and (2.7'')

$$w(\underline{v}) = E_{G_b} \quad \frac{b}{(1 - \ )} \quad \frac{b}{(1 - \ )} \qquad \frac{\bar{b}}{(1 - \ )} \quad , \tag{2.6''}$$

$$\frac{\bar{b}}{(1 - \ )} = E_{G_v} \left[w(v_i) | w(v_i) \qquad w(\underline{v}) \right]. \tag{2.7''}$$

Both Equations (2.6'') and (2.7'') still describe strictly increasing continuous relationships, so the remaining question whether the system has a solution. Note that by assumptions $w(\underline{v}) \quad [0, 1]$ and $\bar{b} \quad [0, 1 - \ ]$. If the solution to these equations is such that $\bar{b} > 1 -$, then it means that the revenue is strictly increasing for each $\bar{b}$, so that we get the corner solution where $\bar{b} = 1 -$ or equivalently $\hat{d} = 1$. Depending on the ordering of and we get one of two cases illustrated by Figure A.1.

Let's first consider the case when the seller's discount factor is not smaller than the buyers' discount factor . This includes $=$ that we discussed in Section 2 as a special case and all the



(a) Case   : only interior solutions.   (b) Case < : corner solution possibility.

Figure A.1: Plots illustrating the characterization of the optimal waiting list mechanism.

qualitative properties will be the same as in the example. Intuitively, when   is close to 0, we should find that it is optimal to choose very low $\hat{d}$. This case is illustrated by Figure A.1(a). In this case there is always an interior solution $(\bar{b}, \underline{v})$. Indeed, for example when   is very low, the solution has to be such that $\bar{b}$ is very small, which corresponds to $\hat{d}$ close to $F_2(p\ )$ and therefore $\hat{D}$ close to $[0, p\ ]^2$.

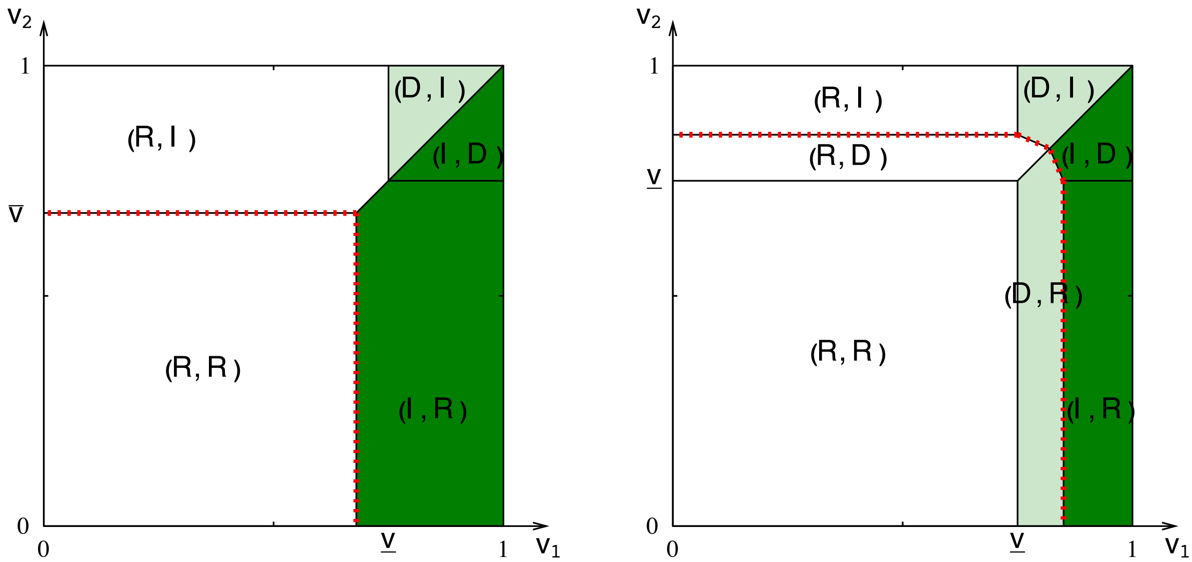
Consider now the case when $<$, which is illustrated by Figure A.1(b). Now there are two possibilities. Either there is an interior solution $(\bar{b}, \underline{v})$ satisfying both equations, or we get the corner solution where the contract ensures service in the next period with certainty. If this is the case both at states $D =$ and $D = [0, 1]^2$ we get the calendar mechanism studied in Section 2.1.

The condition for getting the corner solution is relatively simple. Let $\underline{v}^{cm}$ be the threshold used to offer delayed service in the calendar mechanism case. The average discount-adjusted virtual surplus of a buyer who receives a promise of delayed service with threshold $\underline{v}^{cm}$ is $E_{G_v}[w(v_i)|v_i\ \ \underline{v}^{cm}]$. If this value is higher than , then even at the highest possible distortion $\bar{b}$, the revenue would be increasing in $\bar{b}$, so we would be in the corner solution. Note that by definition, $w(\underline{v}^{cm}) = \frac{}{1-}[\ (\ ) -\ ([0, 1]^2)] = \frac{}{1-} E_{G_b}[b]$, which gives the formal condition for the corner solution in (A.1).

$$E_{G_v}\ \ w(v)\ \ w(v)\ \ \ \frac{E_{G_b}[b]}{(1-\ \ )} = E_{G_v}[w(v)\,|v\ \ \ \underline{v}^{cm}] \qquad \text{(A.1)}$$

When we map the optimal probability of service $\hat{d}$ back to the optimal contract $\hat{D}$, there are two possible types optimal contracts. In Section 2 we already discussed one type of optimal contract, that had similar characteristics to a fully optimal contract. For comparison, the shape of this contract is repeated on Figure A.2(b). But it is also possible, that the optimal contract

$\hat{D}$ is characterized by a simple threshold rule, $\hat{D} = [0, \overline{v}]^2$ as described by Figure A.2(a). This threshold is defined such that $(1- \quad)w(\overline{v}) = \overline{b}$. It is optimal to use this type of contract only if optimal $\overline{b}$ and therefore probability of service is small enough.



(a) Low /

(b) High /

Figure A.2: Optimal waiting list mechanism. Intensity of shading color describes the discounted expected quantity for buyer 1, the area below dotted line is the set where the previous-period customer is served.

Which type of optimal contract we have in a particular situation depends mainly on the difference in discount factors. In the analysis above we got that $\frac{\overline{b}}{(1- \ )} \quad w(\underline{v})$. Using the definition of $\overline{v}$ we get $\quad w(\underline{v}) \qquad w(\overline{v})$. Therefore $\underline{v}$ can be higher than $\overline{v}$ only if buyers' discount factor is sufficiently larger than seller's discount factor . A sufficient condition for being in the "high / " region is when , whereas when is close to 0 we are clearly in the low " / " region.

In particular, when / is relatively low (Figure A.2(a)), the optimal mechanism promises delayed service with relatively low probability and it is possible to fulfill this promise in a region characterized by a simple threshold $\overline{v}$, so that $\hat{D} = [0, \overline{v}]^2$. In this case the implementation of the mechanism is analogous to the pure calendar mechanism, but with a different reserve price. The mechanism is the following: In the first stage the instant service is sold at a second price auction with reserve price $\overline{v}$. If neither of the buyers bid above this value, the waiting buyer is served and new buyers are refused service. However, if both buyers bid above the reserve price, the loser is offered delayed service with price $\quad \hat{d}\underline{v}$. If he accepts the offer, the winner will be given discount $\quad \hat{d}[v_l - \underline{v}]$, where v_l is the loser's bid in the second price auction. The reason for this discount is the same as in pure calendar mechanism case—the second stage creates surplus for the loser of the auction, therefore extra surplus must be given to the winner to avoid bid shading.

B Proofs

To shorten the notation, for a given realization (n, v) we denote the kth lowest virtual surplus by $w_k = w(v_{k:n})$, so that w is always the ordered virtual surplus vector corresponding to (n, v). We denote its distribution function by F_k, so that $F_k(w) = Pr(w_k \quad w) = Pr(v_{k:n} \quad w^{-1}(w)) = F_{k:n}(w^{-1}(w))$.

B.1 Calendar mechanism

Proof of Theorem 3.1 Remember that state $z \quad \{0, \ldots, m\}$ denotes how many service slots are already promised to buyers in the previous period. Then at state z the seller can serve up to $m - z$ new buyers instantly. Let $\ (z)$ denote maximum revenue from state z. It is clear that $\ (z)$ is strictly decreasing in z.

We will start by assuming that c is large enough so that the seller can commit to all promises. Fix state z, number of buyers n and the corresponding ordered virtual surplus vector w. Let $\{I, D, R\}$ be a partition of buyers $\{1, \ldots, n\}$, such that I is the set of buyers who will be served instantly, D receive delayed service, and R are refused service. Feasibility requires that $\#I \quad z$, $\#R \quad m$. We can express the maximum revenue from state z and realization (n, w) by

$$^{z}(n, w) = (1 - \quad) \quad_{i \quad I} w_i + (1 - \quad) \quad_{i \quad D} w_i + \quad (\#D)$$

We can make some instant observations. First, when $w_k \quad 0$, buyer k will not get the object, since refusing him service would be feasible and would increase instant revenue without decreasing continuation payoff. Second, if $z + \#I < m$, so that some of the current objects remain unallocated, it must be because there were exactly $\#I$ buyers with $w_i > 0$. This is true, since by replacing the policy that assigns $w_{n-(\#I+1)}$ current object, the seller would strictly increase flow payoff while still satisfying feasibility and not decreasing continuation payoff. Finally, it must be the case that if $i \quad I, d \quad D$, and $r \quad R$, then $w_i \quad w_d \quad w_r$. That is, we can express I, D, and R as $I = \{n, \ldots, n + 1 - \#I\}$, $D = \{n - \#I, \ldots, n - \#I + 1 - \#D\}$, $R = \{n - \#I - \#D, \ldots, n - \#I - \#D + 1 - \#R\}$). If this is not true, the seller could swap the service dates of the buyers for whom the order does not hold, thus preserving feasibility and continuation payoffs, but increasing flow revenue.

This means that the optimal policy such that up to $m - z$ buyers with highest positive virtual surpluses will receive the current service. If there are less than $m - z$ buyers with positive surpluses, the rest of the buyers are refused service and the continuation state will be m (and $k = 0$).

Consider the situation where there are more than $m - z$ buyers with strictly positive virtual surpluses, so that $w_{n-(m-z)} > 0$. Then $\#I = \{n, \ldots, n + 1 - (m - z)\}$ and the seller chooses optimal number k of buyers with highest positive virtual surpluses among remaining buyers to whom to offer delayed service, so that $D = \{n - (m - z), \ldots, n + 1 - (m - z) - k\}$.

Denote the revenue difference between having all m service slots available and having one slot unavailable by $_1 = (0) - (1) > 0$. Similarly, let $_k$ denote the revenue decrease from having kth slot unavailable, so that $_k = (k-1) - (k)$. Analogously to the illustrative example, the normalized discounted version of this difference is the opportunity cost, which we denote by $\underline{w}_k = \frac{\quad}{1-}\frac{k}{}$. Now, k is optimal if it maximizes

$$^z(n,w) = (1- \ ) \sum_{i=n+1-(m-z)}^{n} w_i + (1- \ ) \sum_{i=1}^{k} w_{n+1-(m-z)-i} + \ (k).$$

Since $(k) = (0) - (0) + (1) - \cdots + (k) = (0) - \ {}_{i=1}^{k} \ {}_i$, we can rewrite the objective as

$$^z(n,w) = (1- \ ) \sum_{i=n+1-(m-z)}^{n} w_i + \ (0) + (1- \ ) \sum_{i=1}^{k} [w_{n+1-(m-z)-i} - \underline{w}_i]. \tag{B.1}$$

The first part, $(1- \ ) \ {}_{i=n+1-(m-z)}^{n} w_i + \ (0)$, is the maximum static revenue that the seller can achieve, whereas the last part is the dynamic revenue, which in optimum is always non-negative.

To compute the expected revenue in state z, we have to take expectation of $^z(n,w)$ with respect to the number of buyers n and valuations v. This gives

$$(z) = E_n \ (1- \ ) \ {}_{i=1}^{m-z} \ {}_0^1 w dF_{n+1-i}(w) + \ (0) + (1- \ ) \ E_v W((w_{n-(m-z)}, \ldots, w_{n+1-(m-z)-m}), \underline{w}) \ ,$$

where $W((w_{n-(m-z)}, \ldots, w_{n+1-(m-z)-m}), \underline{w}) = \max_{k \ \{0,\ldots,m\}} \ {}_{i=1}^{k} \ w_{n+1-(m-z)-i} - \underline{w}_i$. Therefore for each $j \ \{1, \ldots, m\}$,

$$\underline{w}_j = \frac{\quad}{1-} \frac{(j-1) - \ (j)}{} = \frac{\quad}{1-} \frac{E_{n,w}[\ ^{j-1}(n,w) - \ ^j(n,w)]}{} = \ {}_j(\underline{w}),$$

where

$$_j(\underline{w}) = -E_n \ {}_0^1 w dF_{n-(m-j)}(w)$$

$$+ \ E_v \ W((w_{n-(m-j)}, \ldots, w_{n+1-(m-j)-m}), \underline{w}) - W((w_{n-(m-j)-1}, \ldots, w_{n-(m-j)-m}), \underline{w})$$

To show that vector $\underline{w} \ [0,1]^m$ exists and is uniquely defined, we show that is continuous and strictly increasing and $(\underline{w}) \ [0,1]^m$ for all $\underline{w} \ [0,1]^m$. Then $\underline{w}$ exists by Brouwer's fixed point theorem and is unique because of strict monotonicity. We will need some properties of W functions that are proven separately in Lemmas B.1 to B.3.

Continuity comes directly from the fact that expectation operator is continuous and the function W is continuous (Lemma B.1). For boundedness we look at $\underline{w} = \mathtt{1} = (1, \ldots, 1)$ and $\underline{w} = \mathtt{0} = (0, \ldots, 0)$.

At $\underline{w} = \mathtt{0}$, $\max_{k \ \{0,\ldots,m\}} \ {}_{i=1}^{k} w_{n+1-(m-j)-i}$ includes $w_{n+1-(m-j)-i}$ if and only if $w_{n+1-(m-j)-i} >$

0. Therefore[20]

$$E_v[W((w_{n-(m-j)},\dots,w_{n+1-(m-j)-m}))] = \quad {}^{m}_{i=1} \quad {}^{1}_{0}[1-F_{n+1-(m-j)-i}(w)]dw.$$

We get an analogous expression for the other W term. Plugging this into $_i$ and rearranging gives

$$_j(\mathtt{0}) = -E_n \quad 1 - \quad {}^{1}_{0} \quad (1- \quad)F_{n-(m-j)}(w) + \quad F_{n-(m-j)-m}(w) \quad dw \qquad 0.$$

Next, at $\underline{w} = \mathtt{1}$ note that $[w_k - 1] \quad 0$ for any $k > 0$, therefore $W(\cdot, \mathtt{1}) = 0$. Therefore

$$_j(\mathtt{1}) = -E_n \quad {}^{1}_{0}[1-F_{n-(m-j)}(w)]dw \qquad -E_n \quad {}^{1}_{0}[1-F_n(w)]dw \qquad 1,$$

where the first inequality comes from the stochastic dominance of higher order statistics and the second inequality from the assumptions. When $-E_n \quad {}^{1}_{0}[1-F_n(w)]dw \quad > 1$, not only is the existence not guaranteed, but also optimal calendar mechanism with positive probability of delay cannot exist, since at any (n, v, z), the surplus extracted from delay is less than the expected value from having one more object available next period. Therefore in this case the optimal mechanism would be the optimal static mechanism.

Finally, strict monotonicity of $_j(\underline{w})$ with respect to $\underline{w}$ follows from Lemma B.3. Therefore $_j(\underline{w}) \quad [\quad _j(\mathtt{0}), \quad _j(\mathtt{1})] \quad [0,1]$ and so we have continuous $\quad : [0,1]^m \quad [0,1]^m$ which means that by Brouwer's fixed point theorem $\underline{w}$ exists and by strict monotonicity it is unique.

To complete the proof we have to verify that when $c \quad 1-$, the optimal mechanism described above satisfies commitment conditions. Fix any state $z < m$ (at state m there are obviously no commitment issues), realization (n, w). By fulfilling all promises, the seller gets revenue $^z(n, w)$, whereas leaving a promise to one previous customer unfulfilled will give one additional instant service spot, which therefore gives revenue $^{j-1}(n, w)$. Therefore the cost of fulfilling the promise is

$$^{j-1}(n,w) - \quad ^{j}(n,w) = (1- \quad)w_{n-(m-z)}$$

$$+(1- \quad) \quad \max_k \quad {}^{k}_{i=1}[w_{n+1-(m-z)-i} - \underline{w}_i] - \max_k \quad {}^{k}_{i=1}[w_{n-(m-z)-i} - \underline{w}_i] \quad .$$

By Lemma B.1, $\max_k \quad {}^{k}_{i=1}[w_{n+1-(m-z)-i} - \underline{w}_i] \qquad \max_k \quad {}^{k}_{i=1}[w_{n-(m-z)-i} - \underline{w}_i]$, and by assumptions $w_{n-(m-z)} \quad w(1) = 1$, therefore maximum possible value for $^{j-1}(n, w) - \quad ^{j}(n, w)$ is $1-$. This means that $c \quad 1-$ is sufficient for the commitment.

It is also necessary if at each z there exists (n, w) such that $^{j-1}(n, w) - \quad ^{j}(n, w)$ achieves value $1-$. When $N \quad 2m$, we have that $n \quad z + m$ at each $z \quad \{0, \dots, m\}$, and therefore a possible realization is such that $N = 2m$, $w_{n-(m-z)} = \cdots = w_{n-(m-z)-m} = 1$, which gives the

[20] Integration by parts gives $\quad {}^{1}_{0} wdF_k(w) = \quad {}^{1}_{0}[1-F_k(w)]dw$.

equality. When $N \quad \{m+1,\dots,2m-1\}$ the necessary condition (especially for low states) is even weaker and needs to be computed directly from the expression above. □

Lemma B.1. *Let* $W(w,\underline{w}) = \max_{k \quad \{0,\dots,m\}} \quad {}_{i=1}^{k}[w_i - \underline{w}_i]$ *for all* $w,\underline{w} \quad [0,1]^m$ *such that* $w_1 \quad w_2 \quad \cdots \quad w_m$. *Let* $k \ (w,\underline{w}) = \min \operatorname{argmax}_{k \quad \{0,\dots,m\}} \quad {}_{i=1}^{k}[w_i - \underline{w}_i]$. *Then*

(i) W *continuous, non-decreasing in* w*, non-increasing in* $\underline{w}$*, and convex.*

(ii) k *is non-decreasing in* w *and non-increasing in* $\underline{w}$*.*

Remark: k is defined as the smallest maximizer simply to avoid the complications that arise when there are multiple maximizers and the changes do not affect the maximum. In this case the set of maximizers may change (monotonically), but the maximum does not. Since we are not interested in characterizing all possible tie-breaking cases, we will pick just the smallest one.

Proof Since W is the maximum of linear non-decreasing functions of $(w,-\underline{w})$, it is continuous, non-decreasing in each argument (thus non-increasing in $\underline{w}$), and convex.

Suppose k is not non-decreasing in $(w,-\underline{w})$. Then there exist $(w,\underline{w})$ and $(\hat{w},\hat{\underline{w}})$ such that $\hat{w} \quad w$ and $-\hat{\underline{w}} \quad -\underline{w}$, but $k \ (\hat{w},\hat{\underline{w}}) < k \ (w,\underline{w})$. Optimality of $k \ (w,\underline{w})$ and $k \ (\hat{w},\hat{\underline{w}})$ requires that

$$\sum_{i=1}^{k \ (\hat{w},\hat{\underline{w}})}[\hat{w}_i - \hat{\underline{w}}_i], > \sum_{i=1}^{k \ (w,\underline{w})}[\hat{w}_i - \hat{\underline{w}}_i] \quad 0 \quad \sum_{i=k \ (\hat{w},\hat{\underline{w}})}^{k \ (w,\underline{w})}[\hat{w}_i - \hat{\underline{w}}_i]$$

$$\sum_{i=1}^{k \ (w,\underline{w})}[w_i - \underline{w}_i] > \sum_{i=1}^{k \ (\hat{w},\hat{\underline{w}})}[w_i - \underline{w}_i] \quad \sum_{i=k \ (\hat{w},\hat{\underline{w}})}^{k \ (w,\underline{w})}[w_i - \underline{w}_i] > 0.$$

Since $[\hat{w}_i - \hat{\underline{w}}_i] \quad [w_i - \underline{w}_i]$ for all i, this is a contradiction. □

Lemma B.2. *Let* $W(w,\underline{w}) = \max_{k \quad \{0,\dots,m\}} \quad {}_{i=1}^{k}[w_i - \underline{w}_i]$ *for all* $w,\underline{w} \quad [0,1]^m$ *such that* $w_1 \quad w_2 \quad \cdots \quad w_m$. *Fix* $w_0 \quad w_1 \quad \cdots \quad w_m$ *and denote* $w = (w_1,\dots,w_m)$ *and* $\hat{w} = (w_0,w_1,\dots,w_{m-1})$. *Fix* $\underline{w} \quad [0,1]^m$, $l \quad \{1,\dots,m\}$, $\quad > 0$ *and let* $\hat{\underline{w}}$ *be such that* $\hat{\underline{w}}_l = \underline{w}_l +$ *and* $\hat{\underline{w}}_i = \underline{w}_i$ *for all* $i = l$. *Then*

$$W(w,\hat{\underline{w}}) - W(\hat{w},\hat{\underline{w}}) \quad W(w,\underline{w}) - W(\hat{w},\underline{w}).$$

Proof First note that by Lemma B.2 $k \ (w,\underline{w}) \quad k \ (\hat{w},\underline{w})$ and $k \ (w,\hat{\underline{w}}) \quad k \ (\hat{w},\hat{\underline{w}})$. Moreover, by the same result $k \ (w,\hat{\underline{w}}) \quad k \ (w,\underline{w})$ and $k \ (\hat{w},\hat{\underline{w}}) \quad k \ (w,\hat{\underline{w}})$.

Suppose $l > k \ (w,\underline{w})$. Then, since the relevant terms in the summation are not affected by $\ $, we have $W(w,\hat{\underline{w}}) = W(w,\underline{w})$. Formally,

$$W(w,\hat{\underline{w}}) \quad \sum_{i=1}^{k \ (w,\underline{w})}[w_i - \hat{\underline{w}}_i] = \sum_{i=1}^{k \ (w,\underline{w})}[w_i - \underline{w}_i] = W(w,\underline{w}) \quad W(w,\hat{\underline{w}}).$$

Since we have $W(\hat{w}, \underline{\hat{w}}) \quad W(\hat{w}, \underline{w})$, Lemma B.1 implies that $W(w, \underline{\hat{w}}) - W(\hat{w}, \underline{\hat{w}}) \quad W(w, \underline{w}) - W(\hat{w}, \underline{w})$.

Suppose now that $l \quad k\ (w, \underline{w})$. Then $l \quad k\ (\hat{w}, \underline{w})$. Now observe that either $k\ (w, \underline{\hat{w}}) = k\ (w, \underline{w})$ or $k\ (w, \underline{\hat{w}}) < l$, because if this were not the case, it would imply that

$$W(w, \underline{\hat{w}}) = \quad {}^{k\ (w,\underline{\hat{w}})}_{i=1} [w_i - \underline{\hat{w}}_i] \quad {}^{k\ (w,\underline{w})}_{i=1} [w_i - \underline{\hat{w}}_i] \quad {}^{k\ (w,\underline{\hat{w}})}_{i=1} [w_i - \underline{w}_i] - \quad {}^{k\ (w,\underline{w})}_{i=1} [w_i - \underline{w}_i] - \ ,$$

but this contradicts the optimality of $k\ (w, \underline{w})$ at $(w, \underline{w})$. Similarly either $k\ (\hat{w}, \underline{\hat{w}}) = k\ (\hat{w}, \underline{w})$ or $k\ (\hat{w}, \underline{\hat{w}}) < l$. This means that we have four remaining potential cases to consider.

1. When $k\ (w, \underline{\hat{w}}) = k\ (w, \underline{w})$, $k\ (\hat{w}, \underline{\hat{w}}) = k\ (\hat{w}, \underline{w})$, it implies that $W(w, \underline{\hat{w}}) = W(w, \underline{w}) -$ and $W(\hat{w}, \underline{\hat{w}}) = W(\hat{w}, \underline{w}) -$, so $W(w, \underline{\hat{w}}) - W(\hat{w}, \underline{\hat{w}}) = W(w, \underline{w}) - W(\hat{w}, \underline{w})$.

2. When $k\ (w, \underline{\hat{w}}) < l$, $k\ (\hat{w}, \underline{\hat{w}}) = k\ (\hat{w}, \underline{w})$, we know that $W(\hat{w}, \underline{\hat{w}}) = W(\hat{w}, \underline{w}) -$ and

$$W(w, \underline{\hat{w}}) \quad {}^{k\ (w,\underline{w})}_{i=1} [w_i - \underline{\hat{w}}_i] = \quad {}^{k\ (w,\underline{w})}_{i=1} [w_i - \underline{w}_i] - \quad = W(w, \underline{w}) - \ .$$

   Moreover, the inequality has to be strict, since otherwise $k\ (w, \underline{\hat{w}})$ would also be optimal in $(w, \underline{w})$. Therefore $W(w, \underline{\hat{w}}) - W(\hat{w}, \underline{\hat{w}}) > W(w, \underline{w}) - \quad - W(\hat{w}, \underline{w}) + \quad = W(w, \underline{w}) - W(\hat{w}, \underline{w})$.

3. Case $k\ (w, \underline{\hat{w}}) = k(w, \underline{w})$, $k\ (\hat{w}, \underline{\hat{w}}) < l$ is not possible, since it would mean that $k\ (w, \underline{\hat{w}}) = k(w, \underline{w}) \quad l > k\ (\hat{w}, \underline{\hat{w}}) \quad k\ (w, \underline{\hat{w}})$.

4. Finally, when $k\ (w, \underline{\hat{w}}) < l$, $k\ (\hat{w}, \underline{\hat{w}}) < l$, we can write

$$[W(w, \underline{\hat{w}}) - W(\hat{w}, \underline{\hat{w}})] - [W(w, \underline{w}) - W(\hat{w}, \underline{w})]$$

$$= - \quad {}^{k\ (\hat{w},\underline{\hat{w}})}_{i=k\ (w,\underline{\hat{w}})+1} [w_i - \underline{w}_i] + \quad {}^{k\ (w,\underline{w})}_{i=k\ (\hat{w},\underline{\hat{w}})+1} [\hat{w}_i - w_i] + \quad {}^{k\ (\hat{w},\underline{w})}_{i=k\ (w,\underline{w})+1} [\hat{w}_i - \underline{w}_i] > 0,$$

   because (1) ${}^{k\ (\hat{w},\underline{\hat{w}})}_{i=k\ (w,\underline{\hat{w}})+1} [w_i - \underline{w}_i] \quad 0$, since otherwise $k\ (w, \underline{\hat{w}})$ would not be optimal at $(w, \underline{\hat{w}})$, (2) $\hat{w}_i \quad w_i$ for each i, (3) ${}^{k\ (\hat{w},\underline{w})}_{i=k\ (w,\underline{w})+1} [\hat{w}_i - \underline{w}_i] > 0$, since otherwise $k\ (\hat{w}, \underline{w})$ would not be optimal at $(\hat{w}, \underline{w})$.

We have shown that the weak inequality holds for all possible arguments. □

Lemma B.3. *Let* $W(w, \underline{w}) = \max_{k\ \{0,\ldots,m\}} \ {}^{k}_{i=1} [w_i - \underline{w}_i]$ *for all* $w, \underline{w} \quad [0,1]^m$ *such that* $w_1 \quad w_2 \quad \cdots \quad w_m$. *Let* $w_{n-j+1} \quad \cdots \quad w_{n-j+1-m}$ *be generated so that* $Pr(n = \hat{n}) = \ _{\hat{n}}$, $w_k = w(v_{k:n})$ *when* $k > 0$ *and* $w_k = 0$ *otherwise, and* $Pr(v_{k:n} \quad \hat{v}) = F_{k:n}(v)$. *Fix* $\underline{w} \quad (0,1)^m$, $l \quad \{1, \ldots, m\}$, $\quad (0, 1 - \underline{w}_l)$ *and let* $\underline{\hat{w}}$ *be such that* $\underline{\hat{w}}_l = \underline{w}_l +$ *and* $\underline{\hat{w}}_i = \underline{w}_i$ *for all* $i = l$. *Denote* $\underline{w} = (\underline{w}_m, \ldots, \underline{w}_1)$, $w = (w_{n-j}, \ldots, w_{n-j+1-m})$, $\hat{w} = (\hat{w}_{n-(j-1)}, \ldots, \hat{w}_{n-(j-1)-m})$. *Then*

$$E_{n,v}[W(w, \underline{\hat{w}}) - W(\hat{w}, \underline{\hat{w}})] > E_{n,v}[W(w, \underline{w}) - W(\hat{w}, \underline{w})].$$

Proof Lemma B.2 gives weak inequality for any realization of (n, v). Moreover, the proof shows that whenever $k \ (w, \underline{\hat{w}}) < l \quad k \ (w, \underline{w})$, the inequality is strict. Therefore it suffices to construct a positive measure set of w_i's, where $k \ (w, \underline{\hat{w}}) < l \quad k \ (w, \underline{w})$. We construct the set as follows:

(i) $w_{n-j+1} \quad \cdots \quad w_{n-j+1-(l-1)} > \max\{\underline{w}_m, \ldots, \underline{w}_{l+1}\}$

(ii) $w_{n-j+1-l} \quad (\underline{w}_l, \underline{w}_l + \quad)$

(iii) $w_{n-j+1-m} \quad \cdots \quad w_{n-j+1-l-1} < \min\{\underline{w}_1, \ldots, \underline{w}_{l-1}\}$

A combination of w_i's satisfying these restrictions occur with strictly positive probability. By construction,

- $[w_{n-j+1-i} - \underline{w}_{m+1-i}] > 0$ for all $i \quad l$ and $[w_{n-j+1-i} - \underline{w}_{m+1-i}] < 0$ for all $i > l$, therefore $k \ (w, \underline{w}) = l$.
- $[w_{n-j+1-i} - \underline{\hat{w}}_{m+1-i}] > 0$ for all $i < l$ and $[w_{n-j+1-i} - \underline{\hat{w}}_{m+1-i}] < 0$ for all $i \quad l$, therefore $k \ (w, \underline{\hat{w}}) = l - 1 < l$.

Therefore under all these realizations the inequality is indeed strict. □

Proof of Corollary 3.3 We apply Theorem 3.1 to get $\underline{w}$ and the general structure of the optimal mechanism. Suppose $\underline{w}_1 \quad \cdots \quad \underline{w}_m$. The claim is that $(n + 1 - (m - z) - i)$th highest customer with virtual surplus $w_{n+1-(m-z)-i}$ is offered delayed service if and only if $w_{n+1-(m-z)-i} > \underline{w}_i$. To see this, let

$$k \ = \arg\max_{k \quad \{0,\ldots,m\}} \ \sum_{i=1}^{k} [w_{n+1-(m-z)-i} - \underline{w}_i],$$

$$\hat{k} = \max \ k \quad \{0, \ldots, m\} : [w_{n+1-(m-z)-k} - \underline{w}_k] > 0 \ .$$

For each $i \quad \hat{k}$ we have $w_{n+1-(m-z)-i} \quad w_{n+1-(m-z)-k}$ and $\underline{w}_i \quad \underline{w}_k$, so $[w_{n+1-(m-z)-i} - \underline{w}_i] \ [w_{n+1-(m-z)-k} - \underline{w}_k] \quad 0$. For each $i > \hat{k}$ we have by definition that $[w_{n+1-(m-z)-i} - \underline{w}_i] \quad 0$. Therefore we must have $k \ = \hat{k}$. □

B.2 Waiting list

Proof of Theorem 3.5 By assumptions, allocation assigns each buyer i either instant service, refusal of service, or promise of delayed service at some set $\hat{D}_i$, so that this buyer will be served next period if and only if $(n, v) \quad \hat{D}_i$. The waiting list assumption says that vector $\hat{D} = (\hat{D}_1, \ldots, \hat{D}_N)$ must be fixed before observing (n, v). Feasibility requires that the total number of buyers served in the current period must be at most m, and promises of delayed service are such that $\#\{i : (n, v) \quad \hat{D}_i\} \quad m$ for each (n, v). Remember that we denote state by $D = (D_1, \ldots, D_N)$, the vector of contracts by $\hat{D} = (\hat{D}_1, \ldots, \hat{D}_N)$, and the corresponding probabilities $d = (d_1, \ldots, d_N)$, $\hat{d} = (\hat{d}_1, \ldots, \hat{d}_N)$ st $d_i = Pr(D_i)$ and $\hat{d}_i = Pr(\hat{D}_i)$.

Let (D) denote the maximized revenue at state D and with some abuse of notation let (d) be the maximized revenue for given probabilities d. That is, $(d) = \max_{D:Pr(D_i) \quad d_i} \quad (D)$. We assume that the functions (d) are well-defined and will verify it later.

After the standard manipulations we can express the revenue recursively as

$$(d) = \max E_{n,v} \quad (1 - \quad) \sum_{i=1}^{n} \quad {}_i w_i + \quad (\hat{d}) \quad , \tag{B.2}$$

where ${}_i = 1$ if i is assigned instant service, ${}_i = 0$ if refused service, and ${}_i = \quad \hat{d}_i$ when promised delayed service with probability $\hat{d}_i$.

We can make some immediate observations. It is clear that giving instant service or promising delayed service to buyers with negative virtual surpluses is dominated by not serving them. Moreover, it is optimal to serve higher-valued buyers before lower valued buyers. That is, if buyer i is served, buyers j and j are offered delayed service with respective sets $\hat{D}_j$ and $\hat{D}_j$ such that $\hat{d}_j > \hat{d}_j$, and r is refused service, then it must be that $w_i \quad w_j \quad w_j \quad w_r$, since otherwise the seller could strictly increase revenue by swapping promises given to these buyers. Finally, the seller will never assign positive probability of delayed service to more than m buyers. Then we could take the lowest-valued buyer who receives positive probability of service and reassign his probability to higher-valued buyers without changing the feasibility or continuation states, but increasing the flow revenue.

The final immediate observation is that the monotonicity of promises also holds in terms of sets. In particular, contracts are ordered so that[21] $\hat{D}_1 \quad \cdots \quad \hat{D}_m$. This is a priority ranking—the keeper of the first contract will always be served before the keeper of the second contract. If the ordering were not true for some j, the seller could replace promises for $\hat{D}_{j-1}$ and $\hat{D}_j$ by new promises $\hat{D}_{j-1} = \hat{D}_{j-1} \quad \hat{D}_j$ and $\hat{D}_j = \hat{D}j - 1 \quad \hat{D}_j$ while keeping other contracts unchanged. The new contracts do not violate feasibility but extract more revenue from the current buyers.

Step 1: allocation rule. Suppose first that the contracts $\hat{D}$ are already optimally designed. Fix realization (n, v) and the corresponding ordered virtual surplus vector w and denote the number of slots unavailable by $z = \#\{i : (n, v) \quad D_i\}$. By the observations above, it is clear that $m - z$ highest-valuing buyers with strictly positive surplus will receive instant service and buyers with negative virtual surplus are never served. If there are more than $m - z$ buyers with strictly positive virtual surplus, k highest of them will receive contracts $\hat{D}_1, \ldots, \hat{D}_k$. The remaining question is the choice of optimal $k \quad \{0, \ldots, m\}$.

Let ${}_1$ be the difference between revenue when no promises are given compared to revenue when one buyer received promise $\hat{D}_1$, that is, ${}_1 = \quad (0, \ldots, 0) - \quad (\hat{D}_1, 0, \ldots, 0)$. More generally, let the revenue effect of having assigned the contract k by

$${}_k = \quad (\hat{D}_1, \ldots, \hat{D}_{k-1}, 0, \ldots, 0) - \quad (\hat{D}_1, \ldots, \hat{D}_k, 0, \ldots, 0).$$

[21] More precisely this relation must hold with probability one, that is $Pr(\hat{D}_j \setminus \hat{D}_{j-1}) = 0$ for all j.

Again, it will serve as the opportunity cost of allocating kth contract $\hat{D}_k$ in the allocation problem. Define $\underline{w}_k = \frac{}{1-} \frac{k}{\hat{d}_k}$.

Let $\;^j(n,w)$ denote the maximum revenue when j instant service slots are unavailable at realization (n,w). Then we can express it as

$$^j(n,w) = (1- \ ) \sum_{i=1}^{m-j} \max\{0, w_{n+1-i}\} + \ (0,\ldots,0) + (1- \ ) \max_{k \ \{0,\ldots,m\}} \sum_{i=1}^{k} \hat{d}_i[w_{n+1-(m-j)-i} - \underline{w}_i].$$

Again, the sum of the first two terms is the maximal static revenue that the seller can extract without giving any promises of delayed service. The last term is the extra revenue extracted through delay and always non-negative.

Step 2: fulfilling promises optimally. The next question is how to choose particular contracts $\hat{D}$ if for fixed probabilities of service $\hat{d}$. Using the fact that $\hat{D}_1 \ \cdots \ \hat{D}_m$ we can express the expected revenue from a given $\hat{d}$ as

$$(\hat{d}) = \max_{\hat{D}:Pr(\hat{D}_i) \ \hat{d}_i} E_{n,v} \ 1[\hat{D}_1^c] \ ^0(n,w) + 1[\hat{D}_1 \setminus \hat{D}_2] \ ^1(n,w) + \cdots + 1[\hat{D}_m] \ ^m(n,w) \ .$$

With some straightforward manipulations, we can rewrite it as

$$(\hat{d}) = E_{n,v}[\ ^0(n,v)] - \sum_{j=1}^{m} \min_{\hat{D}_j:Pr(\hat{D}_j) \ \hat{d}_j} E_{n,v} \ 1[\hat{D}_j][\ ^{j-1}(n,w) - \ ^j(n,w)] \ . \tag{B.3}$$

Note that the problem in (B.3) is separable in j, since both the objective and the constraint only depend on $\hat{D}_j$ and $\hat{d}_j$. Now, the difference in revenues, $\;^{j-1}(n,w) - \;^j(n,w)$, is equal to 0 whenever there is less than $m-j$ new buyers with strictly positive virtual valuations, and is strictly increasing otherwise. That is, it is strictly positive and strictly increasing if and only if $w_{n-(m-j)} > 0$.

This means that we have two cases to consider. When (n,v) such that $w_{n-(m-j)} \ \ 0$, then $\;^{j-1}(n,w) - \;^j(n,w) = 0$. Therefore if $\hat{d}_j < Pr(w_{n-(m-j)} \ \ 0) = E_n[F_{n-(m-j)}(p \ )]$, then the problem is simple to solve, any subset of $\{(n,v) : w_{n-(m-j)} \ \ 0\}$ works. However, it is not optimal, since by increasing the probability $\hat{d}_j$ to $E_n[F_{n-(m-j)}(p \ )]$ the seller can increase the revenue while not violating the feasibility and not decreasing the continuation value.

Therefore the optimum must be in the second case where $\hat{d}_j \ \ E_n[F_{n-(m-j)}(p \ )]$. In this case the contract set $\hat{D}_j$ has to include some points where $w_{n-(m-j)} > 0$ so that $\;^{j-1}(n,w) - \;^j(n,w)$ is positive. Since then $\;^{j-1}(n,w) - \;^j(n,w)$ is strictly increasing in $w_{n-(m-j)}$, we get that each contract $\hat{D}_j$ includes exactly measure $\hat{d}_j$ of points where $\;^{j-1}(n,w) - \;^j(n,w)$ is the smallest.

This means that we can characterize contract $\hat{D}_j$ by the upper bound to the distortion $\;^{j-1}(n,w) - \;^j(n,w)$. We will denote the maximum distortion level by $\bar{b}_j$. Then

$$\hat{D}_j = \{(n,v) : \ ^{j-1}(n,w) - \ ^j(n,w) \ \ \bar{b}_j\}$$

for uniquely determined maximum distortion level $\bar{b}_j$ such that $Pr(\hat{D}_j) = \hat{d}_j$.

Step 3: optimal probabilities of delayed service. Since $\bar{b}_j$ is strictly increasing with $\hat{d}_j$ in the relevant region, there is a one-to-one mapping between $\hat{d}_j \quad [E_n[F_{n-(m-j)}(p \ )], 1]$ and $\bar{b}_j \quad [0, 1 - \ ]$.

The choice of optimal contracts takes the form

$$\max_{\bar{b}_j} \ E_{n,v} \ \ \mathbb{1}[\hat{D}_1^c] \ \ ^0(n,w) + \mathbb{1}[\hat{D}_1 \setminus \hat{D}_2] \ \ ^1(n,w) + \cdots + \mathbb{1}[\hat{D}_m] \ \ ^m(n,w) \quad ,$$

subject to $0 \quad \bar{b}_j \quad 1 -$ and $\bar{b}_j \quad c$ (commitment constraint). Notice that here is when the commitment level affects the decision problem. So far we analyzed allocation rule for a given contract and choice of contracts for a given probability (which was not necessarily possible under partial commitment). Now the choice of probability of delay or equivalently the maximum distortion $\bar{b}_j$ must be such that when contract j is assigned, then under any realization (n, v) the maximum benefit from not fulfilling contract j is less than c, which is exactly what $\bar{b}_j \quad c$ guarantees.

We will now define two new distribution functions. First, $G_{b_j}(b_j) = Pr(\{(n, v) : \ ^{j-1}(n,w) - \ ^j(n,w) \quad b_j\})$. This allows us to compute the probability of delayed service from its corresponding maximum distortion $\bar{b}_j$ by $\hat{d}_j = G_{b_j}(\bar{b}_j)$ as well as the average distortion from promising the optimal contract corresponding to the probability $\hat{d}_j$ by $\ _j = \ _0^{\bar{b}_j} b_j dG_{b_j}(b_j)$.

The second new distribution will be the distribution of the virtual surplus of an average person who gets the contract $\hat{D}_j$. Let's define

$$g_{w_j}(\hat{w}_j) = E_n \ \ \mathbb{1}[D_1^c] f_{n+1-(m-j)}(\hat{w}_j) + \sum_{i=1}^{m-1} \mathbb{1}[D_i \setminus D_{i+1}] f_{n+1-(m-j)+i}(\hat{w}_j) + \mathbb{1}[D_m] f_{n+1-(m-j)+m}(\hat{w}_j)$$

and $G_{w_j}(\hat{w}_j) = \ _0^{\hat{w}_j} g_{w_j}(\hat{w}_j) d\hat{w}_j$. Remember that $w_k = w(0)$ whenever $k \quad 0$. The distribution G_{w_j} is now a mixture of distributions of virtual surpluses $w(v_{n+1-(m-j):n}), \ldots, w(v_{n+1-(m-j)+m:n})$ with partition $\{D_1^c, D_1 \setminus D_2, \ldots, D_{m-1} \setminus D_m, D_m\}$.

We can now compute the marginal effect of $\bar{b}_j$ to the revenue $\ ^i(n,w)$. For all (n, w) and all i,

$$\frac{\ ^i(n,w)}{\bar{b}_j} = \begin{array}{ll} 0 & w_{n+1-(m-i)-j} \quad \underline{w}_j, \\ (1 - \ ) \ \ g_{b_j}(\bar{b}_j) w_{n+1-(m-i)-j} - \ \ \bar{b}_j g_{b_j}(\bar{b}_j) & w_{n+1-(m-i)-j} > \underline{w}_j. \end{array}$$

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

$$0 = E_{n,v} \ \ \mathbb{1}[D_1^c] \frac{\ ^0}{\bar{b}_j} + \mathbb{1}[D_1 \setminus D_2] \frac{\ ^1}{\bar{b}_j} + \cdots + \mathbb{1}[D_{m-1} \setminus D_m] \frac{\ ^m}{\bar{b}_j} + \mathbb{1}[D_m] \frac{\ ^m}{\bar{b}_j} \quad ,$$

which we can rewrite as

$$\frac{\bar{b}_j}{(1-\quad)} = \,_{\underline{w}_j}^{1} \hat{w}_j \frac{dG_{w_j}(\hat{w}_j)}{1-G_{w_j}(\underline{w}_j)} = E_{G_{w_j}} \quad \hat{w}_j|\hat{w}_j > \underline{w}_j \quad . \tag{B.4}$$

As in the illustrative example, we get that the maximum distortion from jth contract—adjusted with discount factor differences and de-normalized—must be equal to the average virtual surplus of a customer whom the seller expects to allocate this contract.

Step 4: characterization of optimal contracts By definition of $\underline{w}_j$ we can now write

$$\underline{w}_j = \frac{\,_0^{\bar{b}_j} b_j dG_{b_j}(b_j)}{(1-\quad)\quad G_{b_j}(\bar{b}_j)} = \frac{E_{G_{b_j}}[b_j|b_j \quad \bar{b}_j]}{(1-\quad)}. \tag{B.5}$$

For any $j \quad \{1,\dots,m\}$ have now two equations that relate $\underline{w}_j$ and $\bar{b}_j$, equations (B.5) and (B.4). Both describe a strictly increasing continuous relationships between $\bar{b}_j$ and $\underline{w}_j$. Let us first look at (B.4). At the lower bound $\underline{w}_j = 0$ we get that $\frac{\bar{b}_j}{(1-\ )} = E_{G_{w_j}}[\hat{w}_j|\hat{w}_j > 0] \quad (0,1)$ and near the upper bound $\underline{w}_j = 1$ we have $\frac{\bar{b}_j}{(1-\ )} = E_{G_{w_j}}[\hat{w}_j|\hat{w}_j \quad 1] = 1$. Equation (B.5) gives that at the lower bound $\bar{b}_j = 0$ we have $\underline{w}_j = \frac{E_{G_{b_j}}[b_j|b_j \quad 0]}{(1-\ )} = 0$ and at the upper bound $\bar{b}_j = 1 -\quad$ we have $\underline{w}_j = \frac{E_{G_{b_j}}[b_j]}{(1-\ )} \quad 0, - \quad$.

Therefore we have either that there exists interior solution $(\underline{w}_j, \bar{b}_j)$, so that the probability $\hat{d}_j < 1$ or that the first order condition (B.4) always holds as strict equality and therefore the optimal $\bar{b}_j = 1 -\quad$ which means $\hat{d}_j = 1$ or $\bar{b}_j = c$ which means $\hat{d}_j < 1$ because promising delayed service with certainty is impossible due to partial commitment.

Step 5: existence of $\quad(d)$ To complete the analysis, we need to argue that the mechanism description given above gives well-defined functions $\quad(d)$ which we initially assumed to exist. The characterization above is a mapping from a continuation value function $\quad(\hat{d})$ to current expected revenue $\quad(d)$. We denote the mapping by $T\quad$. We will first use Blackwell's sufficient conditions to show that the mapping T is a contraction with speed of convergence $\quad$. Then we can apply the contraction mapping theorem to show that $\quad(d)$ exists.

From the recursive formulation (B.2) it is clear that since $\quad_i \quad_i w_i \quad [0,N]$, mapping T maps bounded functions $\quad(\hat{d})$ to bounded functions. If we have two functions, $\quad, \hat{}$ such that $\hat{}(\hat{d}) \qquad (\hat{d})$ for all $\hat{d}$, then $T\hat{}(\hat{d}) \quad T\hat{}(\hat{d})$ for all $\hat{d}$, since at each state with any promises the continuation revenue is increased and the flow revenue unchanged. Therefore the optimal promises cannot lead to lower revenue. This guarantees monotonicity. To verify discounting, suppose $\hat{}(\hat{d}) = \quad(\hat{d}) + a$. Then each $\quad^j(n,w)$ will simply have an extra $\quad a$ at the end. In the dynamic part of the problem $\quad a$ terms cancel in $\quad^{j-1}(n,w) - \quad^j(n,w)$ and therefore $T\hat{}(\hat{d}) = E_{n,v}[\quad^0(n,w) + \quad a] = T\quad(\hat{d}) + \quad a$. □

B.3 Fully optimal mechanism

Proof of Theorem 3.6 We can make the same immediate observations as in the proof of Theorem 3.5: (1) all buyers with negative virtual surplus are never served; (2) the allocation is monotone in values in the sense that the highest types get instant service and the next highest will get delayed service; (3) only up to m of the new buyers are offered delayed service; (4) the optimal promises are ordered both in probabilities $\hat{d}_1 \quad \cdots \quad \hat{d}_m$ and also in sets[22] $\hat{D}_1 \quad \cdots \quad \hat{D}_m$; and (5) the recursive definition of the revenue function is again

$$(d) = \max E_{n,v} \quad (1 - \quad) \quad {}_{i=1}^{n} \quad {}_i w_i + \quad (\hat{d}) \quad , \tag{B.6}$$

where ${}_i = 1$ if i receives instant service, ${}_i = 0$ if is refused service, and ${}_i = \quad \hat{d}_i$ if is promised delayed service with probability $\hat{d}_i$.

Step 1: allocation rule. Let ${}^j(n, w)$ denote the maximum revenue that the seller can achieve when reserving exactly j current service slots for buyers from the previous period (for example $j = \#\{i : (n, v) \quad D_i\}$). That is, the seller can serve only up to $m - j$ of the new buyers instantly and gives optimal promises for next period service to m buyers who were not served instantly. Because of monotonicity in the order of the promises, the buyers with virtual surpluses $w_n, \ldots, w_{n+1-(m-j)}$ receive instant service (if their virtual surpluses are positive) and buyers with virtual surpluses $w_{n-(m-j)}, \ldots, w_{n+1-(m-j)-m}$ receive the corresponding promises such that $\hat{d}_1 \quad \cdots \quad \hat{d}_m$. Then ${}^j(n, w)$ can be computed as

$${}^j(n, w) = \max_{\hat{d}}(1 - \quad) {}_{i=1}^{m-j} \max\{0, w_{n+1-i}\} + (1 - \quad) {}_{i=1}^{m} \hat{d}_i w_{n+1-(m-j)-i} + \quad (\hat{d})$$

such that $\hat{d}_j \quad [0, 1]$ for each $j \quad \{1, \ldots, m\}$ and the commitment constraint is satisfied. Differentiation with respect to the probability of delay $\hat{d}_i$ gives

$$\frac{d \quad {}^j(n, w)}{d\hat{d}_i} = (1 - \quad) \quad w_{n+1-(m-j)-i} + \quad \frac{(\hat{d})}{\hat{d}_i}. \tag{B.7}$$

The m first order conditions define vector of optimal probabilities $\hat{d}$ as a function of virtual surpluses $w_{n-(m-j)}, \ldots, w_{n+1-(m-j)-m}$, which we will discuss in detail later when we have determined the properties of $\frac{(\hat{d})}{\hat{d}_i}$.

Step 2: fulfilling promises optimally. We have determined the optimal probabilities of service, $\hat{d}$. Next we will find the precise optimal promises in terms of sets $\hat{D}$, where these promises are fulfilled. We argued above that the promises are ordered in the sense that higher types are ordered higher probability of service. Analogously to the waiting list mechanism case we can

[22] With the exception of zero-measure sets. The precise statement is: $Pr(\hat{D}_j \setminus \hat{D}_{j-1}) = 0$ for all $j \quad \{2, \ldots, m\}$.

write

$$(\hat{d}) = E_{n,v}[\ {}^{0}(n,w)] - \sum_{j=1}^{m} \min_{\hat{D}:Pr(\hat{D}_j)\ \ \hat{d}_j} E_{n,v}\ \ \mathbb{1}[\hat{D}_j][\ {}^{j-1}(n,w) - \ {}^{j}(n,w)]\ \ . \tag{B.8}$$

Since both the objective and the constraints are separable in j, we can again optimize point-wise. That is, to choose each contract $\hat{D}_j$ such that $\ {}^{j-1}(n,w) - \ {}^{j}(n,w)$ is the smallest, conditional on $Pr(\hat{d}_j)\ \ \ \hat{d}_j$.

Again, when $w_{n-(m-j)}\ \ \ 0$, then all the buyers starting from $n-(m-j)$ will be refused service, so $\ {}^{j-1}(n,w) - \ {}^{j}(n,w) = 0$, whereas when $w_{n-(m-j)} > 0$, this difference is strictly positive. Therefore the binding promise constraint is always binding in the optimum, $\hat{d}_j = Pr(\hat{D}_j)\ \ \ E_n[F_{n-(m-j)}(p\ \ )]$. Finally, note that $\ {}^{j-1}(n,w) - \ {}^{j}(n,w)$ is still strictly increasing in $w_{n-(m-j)}$.

Now, for each j, the optimal contract is $\hat{D}_j = \{(n,v) :\ \ {}^{j-1}(n,w) - \ {}^{j}(n,w)\ \ \ \ \bar{b}_j\}$ for uniquely defined $\bar{b}_j$ such that $Pr(\hat{D}_j) = \hat{d}_j$.

Step 3: optimal probabilities of delayed service. We can again change the optimization variable, instead of choosing the optimal probability of delayed service $\hat{d}_j$ we can choose the optimal maximal distortion $\bar{b}_j$. Using this we get that $\frac{(\hat{d})}{\hat{d}_j} = -\bar{b}_j$, which is independent of $\hat{d}_i$ for $i = j$.

Moreover, if $\hat{d}_j = 1$ (seller is promised service with certainty), then we know that $\hat{D}_j$ must include all pairs (n,w), therefore $\bar{b}_j = 1 -\ \ $ and so $\frac{(\cdot,1,\cdot)}{\hat{d}_j} = -(1-\ \ )$. If $\hat{d}_j\ \ \ \ E_n[F_{n-(m-j)}(p\ \ )]$, the constraint is not binding and therefore $\frac{(\cdot,\hat{d}_j,\cdot)}{\hat{d}_j} = 0$, whereas if $\hat{d}_j > E_n[F_{n-(m-j)}(p\ \ )]$ we have that $\bar{b}_j > 0$ and so $\frac{}{\hat{d}_j} = -\bar{b}_j < 0$.

Whenever $E_n[F_{n-(m-j)}(p\ \ )] < \hat{d}_j < \hat{d}_j$ we must have that $\hat{D}_j\ \ \ \ \hat{D}_j$ and therefore $\bar{b}_j < \bar{b}_j$. Therefore $\frac{(\cdot,\hat{d}_j,\cdot)}{\hat{d}_j} = -\bar{b}_j > -\bar{b}_j = \frac{(\cdot,\hat{d}_j,\cdot)}{\hat{d}_j}$, therefore $\frac{2}{\hat{d}_j^2} > 0$.

We can now study the first order condition (B.7). Suppose for a moment that commitment level c is high. Let's consider the upper corner solution first. Suppose $\hat{d}_j = 1$. This is optimal only if $(1-\ \ )\ \ w_{n+1-(m-z)-j}\ \ \ \ \ (1-\ \ )$, which is equivalent to $w_{n+1-(m-z)-j}\ \ \ \ -$. Now, let's take the lower corner solution. We know that $\frac{}{\hat{d}_j} = 0$ for all $\hat{d}_j\ \ \ \ E_n[F_{n-(m-j)}(p\ \ )]$, so $\hat{d}_j\ \ \ E_n[F_{n-(m-j)}(p\ \ )]$ implies that $(1-\ \ )\ \ w_{n+1-(m-z)-j}\ \ \ \ 0$, which means that $w_{n+1-(m-z)-j}\ \ \ \ 0$. This is what we already argued above—whenever the buyer to whom the seller is supposed to promise delayed service (according to her position in the ordered values) has negative virtual surplus, it will be optimal to promise service with zero probability. Finally, interior solutions, $\hat{d}_j\ \ \ \ \ E_n[F_{n-(m-j)}(p\ \ )], 1$ have to satisfy $(1-\ \ )\ \ w_{n+1-(m-z)-j} = \frac{}{\hat{d}_j} = \ \bar{b}_j$. Therefore $\hat{d}_j$ is strictly increasing in $w_{n+1-(m-z)-j}$ for the interior solutions. If commitment level is not high, it gives a tighter upper bound, $\bar{b}_j\ \ \ \ c$. The lower bound and the interior points are not affected.

We can summarize the choice of $\bar{b}_j$ for a buyer[23] i who gets the promise by a function

[23]This is to simplify the notation, in particular the contract j goes to ith highest valuing buyer with $i = n+1-(m-z)-j$.

$\overline{b}_j = \overline{b}(w_i)$, where

$$\overline{b}(w_i) = \begin{cases} 0 & w_i \quad 0, \\ \frac{(1-\ )}{}w_i & 0 < w_i < \frac{}{(1-\ )}\min\{1-\ \ ,c\}, \\ \min\{1-\ \ ,c\} & w_i \quad \frac{}{(1-\ )}\min\{1-\ \ ,c\}. \end{cases}$$

Therefore we showed that the optimal probability of delayed service $\hat{d}_j(w_i)$ is equal to 0 when $w_i \quad 0$, equal to the upper bound $G_{b_j}(\min\{1-\ \ ,c\}) \quad 1$ when $w_i \quad \frac{}{(1-\ )}\min\{1-\ \ ,c\}$ and is strictly increasing function $G_{b_j}\ \frac{(1-\ )}{}w_i \quad (0,1)$ in the interior.

Step 4: existence of $\ (d)$**.** The proof that functions $\ (d)$ are well defined uses the Contraction mapping theorem and is analogous to the proof in the waiting list case. □