



Revenue management by sequential screening [☆]

Mustafa Akan ^a, Barış Ata ^b, James D. Dana Jr. ^{c,*}

^a *Carnegie Mellon University, Tepper School of Business, 5000 Forbes Avenue, Pittsburgh, PA 15213, United States*

^b *University of Chicago, Booth School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637, United States*

^c *Northeastern University, Department of Economics and D'Amore–McKim School of Business, 360 Huntington Avenue, Boston, MA 02115, United States*

Received 13 June 2013; final version received 13 July 2015; accepted 22 July 2015

Available online 29 July 2015

Abstract

Using a mechanism design approach, we consider a firm's optimal pricing policy when consumers are heterogeneous and learn their valuations at different times. We show that by offering a menu of advance-purchase contracts that differ in when, and for how much, the product can be returned, a firm can more easily price discriminate between privately-informed consumers. In particular, we show that screening on when the return option can be exercised increases firm profits, relative to screening on the size of the refund alone, only if the expected gains from trade are higher for consumers who learn later. We show that in some settings (mean-preserving spread) the firm can achieve the complete-information profits and analyze the optimal contract in other settings (first-order stochastic dominance) in which the first-best allocation is not always feasible.

© 2015 Elsevier Inc. All rights reserved.

[☆] We would like to thank Johannes Hörner, Hao Li, Larry Samuelson and Kathryn Spier, as well as the editor, Alessandro Pavan, an anonymous associate editor, and two referees, for very helpful comments and suggestions on earlier versions of the paper. Thanks also go to seminar participants at Carnegie Mellon University, Columbia University, Duke University, INSEAD, New York University, Northwestern University, Stanford University, University of Maryland, University of Rochester, University of Southern California, University of Texas at Austin, University of Toronto, Washington University in St. Louis, Wharton, and Yale University, as well as to participants at the Econometric Society Summer Meetings and the INFORMS Revenue Management and Pricing Conference. We also thank the National Science Foundation (CMMI 1351821 and CMMI 1334194), Booth School of Business, University of Chicago, Carnegie Mellon University, the Kellogg School of Management, Harvard Business School, and Northeastern University for their financial support.

* Corresponding author.

E-mail addresses: akan@andrew.cmu.edu (M. Akan), baris.ata@chicagobooth.edu (B. Ata), j.dana@neu.edu (J.D. Dana).

JEL classification: D82; D4; D42; D47

Keywords: Price discrimination; Revenue management; Dynamic pricing; Intertemporal price discrimination; Dynamic mechanism design

This paper analyzes an optimal advance-purchase pricing, or revenue-management, problem when consumers learn their valuations for future consumption over time. We assume that each consumer begins with private information about both the distribution of her valuation and the time when she will learn her valuation. Later, once she has learned her valuation, she also has private information about the realization of her valuation. For ease of analysis, we assume that each consumer's type, or initial private information, is the time that she will learn her valuation, and that the distribution of each consumer's valuation is a function of her type. In this way, consumers begin with private information about two characteristics, the distribution of their valuations and the time that they will learn their valuation, but their initial private information can be represented as a single dimension of information.

We formulate the firm's optimal pricing problem as a dynamic mechanism design problem. The firm maximizes its profit over the set of dynamically incentive-compatible direct-revelation mechanisms when the firm knows only the ex ante distribution of consumers' private information. We characterize the optimal direct-revelation mechanism and show that the firm can implement the optimal mechanism with a menu of contracts with different prices and different refund terms, including both the size of the refund available and the time at which the refund option can be exercised. Such pricing strategies are widely used by airlines, hotels, and railroads, as well as a variety of other firms in the retail, transportation and event industries.

The main contribution of the paper is to demonstrate that time can be a powerful screening device when consumers learn their valuations at different times and their valuations are correlated with when they learn. By explicitly assuming consumers learn at different times, we are adding additional periods (a continuum of periods since we assume a continuum of consumer types) to the sequential screening model analyzed by [Courty and Li \(2000\)](#). While the information consumers have before learning and after learning is the same as in [Courty and Li \(2000\)](#), the existence of these additional reporting periods creates an additional screening device which the firm can potentially use to earn significantly higher profits and even, in some cases, to implement the first best.

While [Courty and Li \(2000\)](#) focus on screening on the size of the refund that consumers receive in a model in which consumers learn their valuations (and make their return decisions) at the same time, we assume consumers learn their valuations at different times and then allow the firm to screen on both the size of the refund and when the refund option can be exercised. Intuitively, screening on when the refund can be exercised allows the firm to costlessly satisfy all of its "downward" local incentive compatibility constraints. No consumer wants to report that she will learn her valuation just before she actually does, because doing so will require her to report her valuation before she learns it and hence to forgo the option value of returning the good if her valuation is low – a positive discrete cost. Because screening on when the refund option can be exercised may relax incentive constraints, it has the potential to increase profits, but only if those relaxed incentive constraints were binding. We show that profits may increase, and may even increase all the way to the complete-information profits, but that profits may also be the unchanged (relative to a model in which consumers learn their valuations simultaneously).

We first consider conditions under which the firm can implement the first best. These are settings in which the expected gains from trade are higher for consumers who learn later, so if the firm offered a menu of the complete information contracts then consumers who learn later would all imitate the consumer who learns the earliest – “downwards” incentive compatibility constraints bind. The firm would like to relax these constraints by screening on when the return is exercised. These are also settings in which the value of the option to return the good is sufficiently high.

When a consumer who learns late imitates a consumer who learns early, she forgoes the option to return the good if she subsequently learns her valuation is low (because the return option expires before she learns her valuation). If this forgone option value is greater than the difference in the two types’ expected gains from trade, then the firm can extract all of the expected gains from trade from consumers. We show that one setting that satisfies these two conditions is when consumers who learn later have a distribution of valuations that is a mean-preserving spread of the distribution of valuations of consumers who learn earlier.

We then consider conditions under which screening on when consumers can exercise the return option has no added value. These are settings in which the expected gains from trade are decreasing. So if the firm offered a menu of the complete-information contracts, then all consumers would imitate the consumer who learns the latest. In these settings, screening on when the return option can be exercised has no value. The firm is unable to exploit the fact that some customers learn their valuations later because these are not the consumers to whom it would like to set a higher price.

Finally, we consider conditions under which screening on when consumers can exercise the return option increases profits but does not implement the first best. These are settings in which the expected gains from trade are higher for consumers who learn later, but the value of the return option is not high enough to deter consumers who learn late from imitating consumers who learn early. We show that in these settings the firm will screen both on the size of the refund and on the time when it can be exercised.

For comparison to [Courty and Li \(2000\)](#) we emphasize two cases that exhibit increasing expected gains from trade, mean preserving spread (MPS) and first-order stochastic dominance (FOSD). However, in contrast to [Courty and Li \(2000\)](#), our consumers are ordered by when they learn, not just by their distributions. So we differentiate between Forward MPS (consumers who learn late have a distribution of valuations that is a mean-preserving spread of the distributions of consumers who learn early) and Reverse MPS (consumers who learn early have a distribution of valuations that is a mean-preserving spread of the distributions of consumers who learn late); and we differentiate between Forward FOSD (consumers who learn late have a distribution of valuations that first-order stochastically dominates the distributions of consumers who learn early) and Reverse FOSD (consumers who learn early have a distribution of valuations that first-order stochastically dominates the distributions of consumers who learn late). Both Forward MPS and Forward FOSD exhibit increasing expected gains from trade, while both Reverse MPS and Reverse FOSD exhibit decreasing expected gains from trade. We find that the firm can extract the entire surplus in the Forward MPS case – the option value a consumer forgoes by purchasing before they know their valuation is always greater than the difference in expected gains from trade (the difference in the complete information prices). We also find that the firm earns higher profits, but may not be able to extract the entire surplus, in the Forward FOSD case. Finally, we find that in the Reverse MPS and the Reverse FOSD cases screening on when the return option is exercised has no effect on profits.

We characterize the optimal contract under Forward FOSD when the complete-information allocations are not feasible and show that the firm still offers a menu of expiring refund contracts, but offers lower-valuation, early-learning consumers contracts with a higher refund price (a downward distortion in allocation) and offers higher-valuation, later-learning consumers contracts with a lower refund price (an upward distortion in allocation). The distortion for consumers who learn early deters imitation by consumers who learn late. This is the same distortion as in [Courty and Li \(2000\)](#), but it in our model only the very highest type consumer is indifferent between reporting truthfully and imitating consumers who learn early, and the only binding “downwards” incentive-compatibility constraints are global constraints. This distortion increases what consumers who learn early are willing to pay because they are getting a more attractive refund option, but the highest type will never exercise the refund option if she imitates another consumer because she won’t yet know her valuation. The lower return price (upward distortion in allocation) for consumers who learn late deters imitation by consumers who learn earlier (but not very early). This is a binding “upward” local incentive-compatibility constraint that does not exist in [Courty and Li \(2000\)](#) and other settings. The distortion is profitable because the downward global incentive-compatibility constraints don’t bind for these consumers, but these consumers are attracted to the rents captured by the very highest type. And the distortion is effective because these consumers are also less willing to pay up front for the increased chance of being allocated the good, so the distortion reduces the rents captured by these consumers.

Advance-purchase tickets with expiring refund options (or exchange options) are commonly sold by airlines, hotels, theaters and railroads. The US passenger service railroad, Amtrak, has a 90% of purchase price cancellation policy for tickets purchased at list price, but also offers lower promotional discount fares which are non-refundable and must be purchased 14 days in advance. Carnival Cruise Line’s list prices for its products are 50% refundable between 30 and 45 days prior to departure, 25% refundable between 15 and 29 days prior to departure, and non-refundable within 14 days of departure, but tickets purchased through its Early Saver Program are up to 20% less expensive but are non-refundable from the date of purchase and are typically only offered more than 90 days before departure. In the US, several hotels, including Marriott, Hilton, Sheraton, and Westin, offer 21-day advance-purchase discounts that are non-refundable, but most other reservations at these hotels can be canceled prior to 1 day prior to arrival for a full refund. And almost all airlines offer both fully refundable and non-refundable fares. For example, Southwest Airlines offers both fully refundable “Business Select” and “Anytime” fares with no advance-purchase requirement and significantly lower-priced, non-refundable, “Wanna Get Away” fares with an advance-purchase requirement (the industry typically uses 7-day, 14-day, and 21-day advance purchase requirements). And finally, Disney charges a cancellation fee of \$200 for cancellations 6 or more days before arrival and offers no refund for cancellations 5 or fewer days before arrival. Disney also offers bundled discount packages with less generous cancellation policies. These examples illustrate that many firms sell their product at a discount to consumers who accept contracts with no return option or with a return option which expires early.¹

The paper is organized as follows. Section 1 gives a numerical example that demonstrates that screening on when the return option can be exercised increases firm profits. Section 2 briefly reviews the related literature. Section 3 presents the general model and introduces the mechanism

¹ Partially-refundable, or equivalently options, contracts are also used by distributors and very large buyers to purchase electric power. These contracts reduce the risk faced by the buyers, but also may enable electricity suppliers with local market power to extract more surplus from buyers.

design problem. Section 4 presents necessary and sufficient conditions for the first-best solution to be feasible, that is, for screening on when the return option can be exercised to earn the complete information profits, and demonstrates that the first-best is achieved when the distributions satisfy Forward MPS. Section 5 presents sufficient conditions under which screening on when the return option can be exercised is not useful, which includes both Reverse MPS and Reverse FOSD. In these cases, the optimal mechanism is the equivalent to the optimal mechanism in a model in which consumers learn simultaneously (i.e., Courty and Li, 2000). Finally, in Section 6, we derive the optimal mechanism in an interesting case in which screening on when the return option can be exercised is a useful screening device, but the first best is not feasible. That is, when the distributions satisfy Forward FOSD but the option value is insufficient. Section 7 discusses potential extensions of the model and offers some concluding remarks.

1. A numerical example

A numerical example illustrates how a firm can use both the size and the timing of its refund offers to increase its profits. Suppose that there are two types of consumers who know the distribution of their valuations, but don't initially know the realization of those distributions. Type 1 consumers learn early, at time t_1 , and type 2 consumers learn late, at time t_2 , where $t_2 > t_1$. Assume that a type 1 consumer's valuation is $v_1 = 120$ with probability $1/4$ and is $v_1 = 0$ with probability $3/4$, and assume that a type 2 consumer's valuation is $v_2 = 180$ with probability $1/2$ and is $v_2 = 0$ with probability $1/2$. Also, suppose that the firm's unit cost is $c = 100$ so that it is efficient for consumers to consume the good whenever their valuations are high.² Finally, suppose the firm sells in advance, before any consumers learn their valuations.

If consumers' types were common knowledge, then the firm could design a different contract for each type. The firm wants to create incentives for efficient ex post consumption while extracting all of the ex ante surplus. One obvious way to do this is to sell the good with an option to return it later (anytime after she consumer learn her valuation) if the realized valuation is low. To insure ex post efficient allocation, the firm could set the return price for both types equal to its unit cost, $r_1 = r_2 = 100$ (this is not the unique value of the return price in this example), and in order to extract all the surplus the firm would charge type 1 consumers an up-front payment of 105 (equal to $E[\max\{v_1, r_1\}]$) and charge type 2 consumers an up-front payment of 140 (equal to $E[\max\{v_2, r_2\}]$). The expected profit is $E[\max\{v_1 - c, 0\}] = 5$ for each type 1 consumer, and $E[\max\{v_2 - c, 0\}] = 40$ for each type 2 consumer.

Now consider what happens when the firm offers this menu of contracts to privately-informed consumers. Clearly the contracts are not incentive compatible. Both consumer types prefer the contract offered to the type 1 consumers, so both pay a price of 105. Type 1 consumers get a surplus of 0 and type 2 consumers get an expected surplus of 35 (equal to $1/2 \times 180 + 1/2 \times 100 - 105$). The firm earns a profit of 5 for each type 1 and 5 for each type 2 consumer. However the firm can do better.

The firm has two instruments with which it can screen the type 2 consumers. First, the firm offers a higher refund to type 1 consumers. This discourages type 2 consumers from purchasing the contract designed for type 1 consumers because the higher refund is associated with a

² In our two-type example, changing refunds need not lead to inefficient ex post allocations, so there may be a variety of refund values which yield the same profits, but in our general model valuations are drawn from a continuous distribution, so refunds that are not equal to cost lead to allocation distortions, and the profit-maximizing contracts are unique.

larger upfront payment, and type 2 consumers value the return option less than type 1 consumers because they use it less frequently.

Specifically, the firm can increase the size of the refund for type 1 consumers to $r_1 = 120$. By increasing the refund for type 1 consumers to 120, it can increase the price charged to type 1 to $p_1 = 120$ (equal to $1/4 \times 120 + 3/4 \times 120$). Even though the price is higher, profits from each type 1 consumer are still 5 because the refund is also higher (and the increase does not cause a distortion in allocation to type 1 as long as the refund is less than 120 because we have assumed a binary distribution for v_1). type 2 consumers get a surplus of 30 from type 1's contract, so the firm can offer a refund of $r_2 = 100$ and a price of $p_2 = 110$ (equal to $1/2 \times 180 + 1/2 \times 100 - 30$) to type 2 consumers (note that a change in the refund offered to the type 2 consumers has no effect on profits). In this way the firm can increase its profit from 5 to 10 for each type 2 consumer.

Screening in this way on the size of the refund, as originally analyzed by [Courty and Li \(2000\)](#), increases the firm's profits but does not implement the complete-information outcome – the first-best profit is 5 for each type 1 consumer and 40 for each type 2 consumer. But if $t_2 > t_1$, the firm can do even better because it now has a second instrument. The firm can also set an expiration time for the refund offers made to type 1 consumers. This discourages type 2 consumers from purchasing the contract designed for type 1 consumers because type 2 consumers won't know their valuations when they must choose whether to accept the refund. In this example, the complete information contracts can be modified to make them incentive compatible. The firm simply needs to specify that the return option offered to type 1 consumers must be exercised before time t_2 , that is, before type 2 consumers learn their valuations.

However, notice that if $t_2 \leq t_1$, then screening on when the return option is exercised has no effect. The best the firm can do is screen on the size of the refund. So screening on when the return can be exercised is profitable when consumers who learn late want to imitate consumers who learn early.

Finally, notice that if type 2 has valuation $v_2 = 180$ with probability $4/5$ instead of $1/2$, then the firm finds it profitable to use both screening instruments. In particular, the complete information contracts are $p_1 = 105$ and $r_1 = 100$ for type 1 and $p_2 = 164$ and $r_2 = 100$ for type 2, but now the type 2 consumer gets a surplus of 59 from type 1's complete information contract ($4/5 \times 180 + 1/5 \times 100 - 105$); the type 2 consumer gets a surplus of 48 from the contract $p_1 = 120$ and $r_1 = 120$ ($4/5 \times 180 + 1/5 \times 120 - 120$); and the type 2 consumer gets a surplus of 24 from the contract $p_1 = 120$ and $r_1 = 120$ if the firm also screens on when the return option can be exercised ($4/5 \times 180 + 1/5 \times 0 - 120$). So the contract offered to the type 2 consumer is $p_2 = 140$ and $r_2 = 100$, for a surplus of 24, and the firm's profit is highest when both instruments are used together.

This numerical example illustrates that there are two potential elements of the firm's optimal screening strategy. First, the firm increases the price of the ticket sold to type 1 consumers by increasing the size of the refund which increases the option value and increases type 1 consumers' willingness to pay. And second, the firm sets an expiration date for the return option offered to type 1 consumers. The expiration date is before type 2 consumers learn their valuations so that the type 2 consumers can't imitate the type 1 consumers without paying for a return option that they never use.

While the numerical example illustrates the value of screening on when the refund option can be exercised, it is restrictive in several important ways. First, the assumption that each consumer's valuation is drawn from a discrete distribution implies that at least some changes in the return price do not change or distort ex post consumption decisions. In the paper, we analyze a model with much more general valuation distributions in which changes in the refund price are always

distortionary. Second, the assumption that there are only two consumer types oversimplifies the firm's problem. Instead, we analyze a model with a continuum of consumers learning in sequence and identify additional issues that aren't evident in a two-type model. In particular, we show that the binding incentive constraints are not just the constraints preventing consumers from imitating other consumers who learn earlier. When consumers who learn late capture information rents to prevent them from imitating consumers who learn early, consumers who learn at intermediate times may find it more attractive to imitate consumers who learn later than those that learn earlier.

Astute readers may also have noticed that instead of offering a menu of refundable tickets *ex ante*, the firm could also implement the first best by offering a sequence of *ex post* contracts and literally screening on when the product is purchased. The firm could instead offer type 1 consumers a non-refundable contract for 120 that they purchase at time t_1 , *after* they have learned their valuations, but before type 2 consumers have learned their valuations, and offer type 2 consumers a non-refundable contract for 180 that they purchase at time t_2 , *after* they have learned their valuations. These prices are incentive compatible. And this implementation fits the pricing practices in our hotel, airline and railroad examples quite well. Type 1 consumers represent tourist or leisure travelers, and these consumers purchase at a lower price early, but after they learn their valuations are high. Type 2 consumers represent business travelers who purchase at a higher price late, and only if their valuations are high. But extracting the entire surplus with these *ex post* spot prices is only feasible because of the binary distribution assumption in our numerical example (see [Ata and Dana, forthcoming](#) for a more general analysis of optimal pricing under this assumption).

Finally, the most astute readers may have noticed that in our numerical example the firm can also implement the first best without screening on when the return can be exercised by waiting until after type 1 has learned her valuation to offer the menu of contracts. This is a consequence of the fact that in the example there are only two types and that type 1 consumers are homogeneous conditional on having a valuation greater than c . Simply increasing the number of types to three would prevent this strategy from implementing the first best without screening on when the option is exercised. And screening on when the option is exercised is essential to getting the full information profits in the general model.

2. Related literature

We begin by discussing three different closely-related literatures: the literature on advance-purchase pricing when consumers learn over time, the operations management literature on pricing and capacity controls in stochastic environments, and the dynamic mechanism design literature on optimal pricing when consumers learn over time.

The literature on advance-purchase pricing has largely focused on two-period models, that is, one period before consumers learn their demands and one period after. With the exception of [Courty and Li \(2000\)](#), discussed below, this literature also assumes that purchases are not refundable. [Courty \(2003\)](#) considers a monopolist that can commit in advance to its prices and that chooses whether to sell to consumers before (advance sales) or after (spot market sales) they have learned their demand. [Courty \(2003\)](#) shows that *ex ante* sales reduce profits because they lead to inefficient allocation (over consumption or under consumption), but can increase profits for a monopolist when consumers are more homogeneous *ex ante* so the dead weight loss of the monopoly pricing distortions is reduced. [DeGraba \(1995\)](#) considers a monopolist that is unable to commit to its future prices and that intentionally creates a capacity shortage and hence a buying frenzy in the spot market that induces buyers to purchase early, before they know

their valuations. Dana (1998) considers a competitive model in which heterogeneous consumers divide their purchases between advanced sales, when they may not yet know their demands, and spot market sales, after their valuations have been realized. Gale and Holmes (1993) considers a related model in which a monopolist that sells to heterogeneous consumers discriminates by selling to some consumers early before they know their departure-time preferences and selling to other consumers late after they learn their preferences.³ Probably the most closely related paper in this literature is Nocke et al. (2011). Like us they emphasize the use of advance sales as a screening mechanism in a model with full commitment and no aggregate demand uncertainty or capacity constraints, but in their model, consumers vary in how much information they have in advance, not in when they learn their valuations.⁴

Che (1996) is the first paper to explicitly consider the optimal use of refunds (or returns) when consumers learn about their valuations after making their purchase decisions. Che (1996) finds firms are more likely to offer refunds when costs are high and when consumers are risk averse, though Che (1996) restricts attention to nonrefundable or fully refundable sales.

A few papers have considered partially refundable ticket sales when consumers learn over time.⁵ The most closely related paper is Courty and Li (2000) (see also Ringbom and Shy, 2004). In Courty and Li (2000), the firm screens consumers on the size of the refund, but not on when consumers can claim the refund. As in our model, consumers make their purchases before they learn their valuations, but retain an option to cancel and claim a partial refund after they learn their valuations. These refund contracts increase firm profits in two ways. First, they allow the firm to extract more total surplus from consumers, even when consumers are homogenous. And second, because the firm can vary the size of the refund, they allow the firm to better discriminate between heterogeneous consumers.⁶

Historically, the operations management literature on advance-purchase sales has considered both optimal pricing and capacity controls when heterogeneous consumers purchase in an exogenously given, sequential order. Littlewood (1972) considers a setting in which the consumers have either high or low valuations, which are known and observable. Consumers with low valuations arrive in the first period, while the consumers with high valuations arrive in the second period. There is aggregate uncertainty about the number of consumers of each type, and a system manager chooses how much capacity to reserve for the consumers with high valuations. Littlewood (1972) characterizes the optimal policy as a booking limit policy. Brumelle and McGill (1993) and Curry (1990) provide extensions of Littlewood's result to n customer classes, characterizing the optimal capacity control policy by nested booking limits. Talluri and van Ryzin (2004) provides an extensive review of these papers and many other extensions of these capacity control models in the operations management literature.

³ See also Möller and Watanabe (2010) and Gale and Holmes (1992).

⁴ Nocke et al. (2011) also make the restrictive assumption that consumers' valuation distributions are binary. Because of this assumption, and the way that they model consumer heterogeneity, as in Ata and Dana (forthcoming), the optimal mechanism can always be implemented without selling refundable tickets.

⁵ In the absence of risk aversion, by setting the refund price equal to its cost, the firm is assuring that the consumer's consumption decision is ex post efficient and allowing itself to extract all of the consumer surplus through an ex ante lump sum payment, so the optimality of partial refunds is a direct implication of the principal-agent literature on ex ante contracting when the agent is ex post privately informed.

⁶ A few papers in the economics literature have looked at consumer learning and refund contracts empirically. In particular, Escobari and Jindapon (2008) show that the difference between the advance-purchase price of a refundable airline ticket and the advance purchase price of a non-refundable airline ticket declines over time, which is consistent with consumers learning about their valuations over time.

Gallego and Şahin (2010) and other more recent papers in the operations management literature, model consumers as forward-looking and strategic (for a review see Shen and Su, 2007). Gallego and Şahin (2010) (see also Su, 2009) consider a model with ex ante homogeneous consumers, but assume that consumers learn in a multiperiod, stochastic environment and that the firm is capacity constrained. They find that the firm maximizes its profits by selling to consumers as early as possible and allowing consumers to exercise the return option as late as possible. Chen (2011) considers a related model of optimal pricing and refund policy with consumer learning, aggregate demand uncertainty, and capacity constraints.

The final closely-related literature is the dynamic mechanism design literature which considers optimal mechanisms when agents learn over time, and hence mechanisms in which information is revealed sequentially.⁷ Two of the earliest papers are Baron and Besanko (1984) and Riordan and Sappington (1987). Both papers consider optimal regulatory policies when regulated firms are learning about their costs over time. Board (2011) considers a related model in which a firm auctions (in advance) the right to consume in the future after they have learned their private value of consumption. Esö and Szentes (2007a) consider the optimal dynamic mechanism in which an informed client can hire an expert to generate additional information that the client can use to refine their decision making. They find the expert's profit is the same as if the expert perfectly observed the client's private information.⁸ Esö and Szentes (2007b) consider a related auction model in which the seller can release additional information to privately-informed agents who update their valuations based on the new information. As in the single agent model, in the optimal mechanism the seller captures all the rents associated with the information they provide. Krämer and Strausz (2011) generalize both Courty and Li (2000) and Esö and Szentes (2007a) by adding endogenous information acquisition with moral hazard.

Board and Skrzypacz (2010) consider a capacity-constrained firm selling to ex ante homogeneous consumers who arrive in an exogenously given stochastic arrival process but who are forward looking and can delay their purchases (i.e., can imitate consumers who arrive later). While their consumers do not learn additional information after they arrive (see also the closely-related model of pricing with strategic consumers in competitive markets by Deneckere and Peck, 2012), they characterize the optimal pricing and allocation and show that it can be implemented with simple deterministic pricing rules that depend only on time and remaining inventory.

An important paper in this literature is Pavan et al. (2014) whose dynamic envelope theorem characterizes the information rents of privately-informed agents. Since our paper characterizes environments in which the first best can be implemented, an important question is why their envelope theorem does not hold in our model. We discuss this in Section 4 of the paper.

Several related papers in dynamic mechanism design literature consider stochastic arrivals of new consumers after the mechanism is announced. Deb and Said (2015) consider a two-period model related to Courty and Li (2000) in which additional consumers arrive in period two who cannot write contracts in period one. They also assume that the firm that cannot commit in advance to the contracts it offers in period 2. They show that the firm can benefit from postponing ex ante contracting with buyers available in the first period, although this is never optimal under full commitment. Gershkov and Moldovanu (2010) consider the optimal mechanism for a patient firm selling a fixed number of heterogeneous objects to a stochastic sequence of impa-

⁷ For recent surveys of this literature see Bergemann and Said (2011) and Vohra (2012).

⁸ Gale and Holmes (1992) use a mechanism design approach to analyze consumer learning, but in their model it is sufficient to utilize a static direct revelation mechanism because the firm's optimal ex post allocation depends only on the ex ante information.

tient, privately-informed buyers. Garrett (2013) considers a firm that faces stochastic arrivals of privately-informed buyers when the firm sells a durable good.

A paper that is more closely related to ours is Ely et al. (2013). They assume that consumers observe a signal of their demand when they arrive and only learn their final demand in the final period. Because arrivals are stochastic, the optimal mechanism features an auction-like mechanism to determine the refund, or buyback, price. However, consumers' valuations are not correlated with their arrival time, so firms do not screen on when consumers make their final purchase decisions.

Our work and Ely et al. (2013) both explicitly model the consumer learning process (see also Ata and Dana, forthcoming). Further research in this area is important and should improve our understanding consumers' decisions about when to enter the purchasing process and firms' responses to that behavior.

Finally, we analyze a model in which screening on when refunds take place is a useful screening instrument, a result which contrasts with the traditional view that intertemporal price discrimination is not profitable, a literature that began with Stokey (1979).⁹ We show intertemporal price discrimination may be profitable when consumers learn their valuations at different times and when the distribution of their valuations is correlated with when they learn.

3. The model

A single, risk-neutral firm sells a single, homogeneous good with unit cost c . Consumers are heterogeneous and their types are continuously distributed on $[0, T]$ with a strictly positive density function $h(t)$ and cumulative distribution function $H(t)$. That is, $h(t)$ represents the relative frequency of type t consumers in the population. Consumers privately learn their type prior to time 0, and all consumption takes place simultaneously after time T .¹⁰ The type t determines the probability distribution of their valuations as well as the time at which they learn their valuations. Without loss of generality, we assume that a type t consumer privately learns her realized valuation at time t . We let $f(v, t)$ denote the density function of the joint distribution of types and valuations, and we assume that $f(v, t)$ is continuously differentiable. The valuation of a type t consumer is distributed according to the probability density function $f(v|t)$, and the cumulative distribution function $F(v|t) = \int_{\underline{v}}^v f(\tilde{v}|t) d\tilde{v}$ on the interval $[\underline{v}, \bar{v}]$. It follows that

$$\int_{v,t} f(v, t) dv dt = 1, \quad \int_v f(v, t) dv = h(t), \quad \text{and} \quad f(v, t) = f(v|t) h(t).$$

Assuming that $f(v|t)$ has a common support for all t significantly simplifies the analysis and would be difficult to relax.

Note that we have assumed perfect correlation between the buyer's distribution and the time when she learn her valuation. This is a strong assumption and one that is potentially difficult to relax. In particular, if the consumer did not know at time 0 precisely when she would learn her valuation, then the optimal mechanism would require that the consumer report a much more

⁹ See also Salant (1989) and Anderson and Dana (2009) for generalizations of Stokey (1979) in which intertemporal price discrimination may be profitable.

¹⁰ We sometimes refer to the contracting time as time zero and the consumption time at time T , but the contracting time is sometime after consumers learn their types, but before any consumers learn their valuations, and the consumption time is sometime after all consumers learn their valuations.

complicated stream of information, dramatically increasing the complexity of the optimal mechanism. This restrictive assumption allows us to focus on mechanisms that can be implemented with a menu of refund schedules, a remarkably simple class of mechanisms. Much more work remains to be done to consider optimal mechanisms with more realistic assumptions about the relationship between consumers' types and when they learn.

In what follows, we consider profit-maximizing incentive-compatible direct-revelation mechanisms for the seller. There is no loss in generality from restricting the seller to implement a sales mechanism in which consumers truthfully report their types, t , at time zero and then report their realized valuations, v , at time t (see Pavan et al., 2014; Myerson, 1986, and Green and Laffont, 1986). Intuitively the seller can do no better than the maximally centralized communication system in which, at every moment in time, each individual confidentially reports all of her private information.

Before describing the mechanism design problem formally, we introduce some notation. For each pair of reports of valuation v and type t , let $y(v, t)$ be the probability that the seller delivers the good, and let $x(v, t)$ denote the net payment to the seller. Consider a consumer whose type is t and whose valuation is v . Her ex post utility (or surplus) is given by

$$u(v', t'; v, t) = vy(v', t') - x(v', t') \quad (1)$$

if she reports her type as t' and her valuation as v' . We will use $u(v, t)$ to denote the consumer's ex post utility when she reports her type and valuation truthfully. That is,

$$u(v, t) = vy(v, t) - x(v, t). \quad (2)$$

The consumer's ex ante expected utility as a function of her actual type, t , and her reported type, t' , is

$$U(t'; t) = \begin{cases} \mathbb{E}_t [\max_{v'} u(v', t'; v, t)] & \text{if } t' \geq t, \\ \max_{v'} \mathbb{E}_t [u(v', t'; v, t)] & \text{otherwise,} \end{cases} \quad (3)$$

where \mathbb{E}_t denotes the expectation over v given the consumer's type, t , or

$$\mathbb{E}_t [\cdot] = \int_{\underline{v}}^{\bar{v}} \cdot f(v|t) dv.$$

We use $U(t)$ to denote the consumer's ex ante utility when she reports her type, and then later her valuation, truthfully. That is,

$$U(t) = \mathbb{E}_t [u(v, t)] = \int_{\underline{v}}^{\bar{v}} u(v, t) f(v|t) dv, \forall t. \quad (4)$$

Finally, using integration by parts, we write the total expected gains from trade associated with sales to a type t consumer as

$$\mathbb{E}_t [\max(v - c, 0)] = \int_c^{\bar{v}} (v - c) f(v|t) dv = \bar{v} - c - \int_c^{\bar{v}} F(v|t) dv, \forall t. \quad (5)$$

The seller's mechanism design problem (P^0) can be stated as follows:

$$\max_{x(v, t), y(v, t)} \int_{v, t} f(v, t) [x(v, t) - cy(v, t)] dv dt \quad (P^0)$$

subject to

$$U(t) \geq 0, \forall t, \quad (\text{IR})$$

$$u(v, t) \geq u(v', t; v, t), \forall v, v', t, \quad (\text{IC}_t)$$

$$U(t) \geq U(t'; t), \forall t, t', \quad (\text{IC}_0)$$

$$0 \leq y(v, t) \leq 1, \forall v, t, \quad (\text{F})$$

and equations (1) through (4) (which define U and u as functions of x and y).

The first set of constraints are the individual rationality, or participation, constraints. These constraints are imposed to guarantee that the firm gives every consumer nonnegative expected surplus. Note that there is no ex-post individual rationality constraint, i.e., the ex-post utility $u(v, t)$ of a type t consumer with a realized valuation v could be negative. For example, a consumer might purchase a ticket to attend a meeting but not be eligible for a full refund if she later learns the meeting has been canceled.

The second set of constraints are the incentive compatibility constraints with respect to the consumers' realized valuations. These are imposed to guarantee that each consumer, conditional on reporting her type at time zero truthfully, finds it optimal to report her realized valuation truthfully at time t .

The third set of constraints are the incentive compatibility constraints with respect to the reports of consumers' types at time zero. These constraints can be divided into two distinct types because $U(t'; t)$ is defined differently for upward deviations and downward deviations. When a type t consumer reports a lower type, i.e., $t' < t$, she will subsequently be asked to report her valuation before she learns her true valuation, while when a type t consumer reports a higher type, i.e., $t' \geq t$, she will subsequently be asked to report her valuation after she learns her true valuation.

The final set of constraints, denoted by (F), require the delivery rule y to be feasible.

The following lemma, which closely follows Courty and Li (2000), characterizes how a consumer reports her valuation if she does not report her type truthfully at time zero. The proof is omitted.

Lemma 1. *If the mechanism satisfies the incentive compatibility constraints, (IC_t) , regarding the report of the consumers' valuations, then*

(i) *if a type t consumer reports her type as t' at time zero, and if she knows her true valuation at time t' (because $t' \geq t$), then it is optimal for her to report her true valuation, that is, $v \in \arg \max_{v'} u(v', t'; v, t)$; and*

(ii) *if a type t consumer reports her type as t' at time zero, and if she does not know her true valuation at time t' (because $t' < t$), then it is optimal for her to report her expected valuation, that is, $\mathbb{E}_t[v] \in \arg \max_{v'} \mathbb{E}_t[u(v', t'; v, t)]$.*

As in Courty and Li (2000), Lemma 1(i) follows immediately from (IC_t) and holds because once the consumer learns v her payoff is independent of her true type and depends only on her announced type. So if (IC_t) holds then consumers will always reveal v truthfully.

Lemma 1(ii) also follows from (IC_t) and holds because a consumer whose valuation was $\mathbb{E}_t[v]$ would report it truthfully even after reporting her type as t' , and the consumer's payoff is linear in the realization of her valuation conditional on her reports, so if reporting $\mathbb{E}_t[v]$ maximizes $u(v', t'; \mathbb{E}_t[v], t')$ then reporting $\mathbb{E}_t[v]$ must also maximize her expected utility.

Using Lemma 1(i) for reports satisfying $t' \geq t$ and Lemma 1(ii) for reports satisfying $t > t'$, and separating (IC₀) into two constraints, the seller's mechanism design problem becomes (P¹):

$$\max_{x(v,t), y(v,t)} \int_{v,t} f(v,t) [x(v,t) - cy(v,t)] dv dt \quad (\text{P}^1)$$

subject to

$$U(t) \geq 0, \forall t, \quad (\text{IR})$$

$$u(v,t) \geq u(v',t; v,t), \forall v, v', t, \quad (\text{IC}_t)$$

$$U(t) \geq \mathbb{E}_t[u(v,t'; v,t)], \forall t, t'; t' \geq t, \quad (\overline{\text{IC}}_0)$$

$$U(t) \geq \mathbb{E}_t[u(\mathbb{E}_t[v], t'; v,t)], \forall t, t'; t' < t, \quad (\underline{\text{IC}}_0)$$

$$0 \leq y(v,t) \leq 1, \forall v, t. \quad (\text{F})$$

In this formulation, we separate the set of constraints (IC₀) into two subsets in order to emphasize the difference between the upward and downward deviations in the consumer's report of her type. The set of constraints ($\overline{\text{IC}}_0$) corresponds to upward deviations (imitating a type that learns later), whereas ($\underline{\text{IC}}_0$) corresponds to downward deviations (imitating a type that learns earlier).

This is the fundamental difference between the firm's problem when consumers learn sequentially, (P¹), and Courty and Li (2000)'s analysis of the firm's problem when consumers learn simultaneously. The ex ante incentive compatibility constraint in their paper is

$$U(t) \geq \mathbb{E}_t[u(v,t'; v,t)], \forall t, t', \quad (6)$$

instead of ($\overline{\text{IC}}_0$) and ($\underline{\text{IC}}_0$), so it is equally easy to imitate any other type, while in our paper it is more costly for a consumer to imitate consumers who learn earlier than it is to imitate consumers who learn later.

If consumers had no private information, that is, if the seller could ignore the incentive compatibility constraints and maximize only subject to the individual rationality and feasibility constraints, then clearly the optimal solution to the above program is to set $y(v,t) = 1$ if $v \geq c$ and $y(v,t) = 0$ otherwise, and to set $x(v,t)$ to extract all of each consumer's ex ante consumer surplus, that is, to set $x(v,t)$ such that $U(t) = 0$. This is the complete-information or first-best solution. The seller is able to extract all of the consumer surplus and the solution allocates the good efficiently.

The following lemma is standard and is useful for further simplifying the seller's problem. It states that under any optimal mechanism, when the consumer's realized valuation is higher, she receives the good with a greater probability and has a greater consumer surplus. The proof of the lemma is standard in the mechanism design literature and therefore is skipped.

Lemma 2. *The incentive compatibility constraint (IC_t) is satisfied if and only if*

- (i) $\partial u(v,t) / \partial v = y(v,t)$ for almost every v ; and
- (ii) $y(v,t)$ is non-decreasing in v .

4. When screening on time achieves the first-best

The following proposition gives necessary and sufficient conditions under which the firm can implement the complete-information allocation and earn the complete-information profits even

though consumers' types and valuations are privately observed. That is, we provide conditions under which the firm can implement the first-best outcome by screening only on when the return option can be exercised. Its proof is given in [Appendix A](#).

Proposition 1. *The seller can implement the complete-information solution if and only if*

$$\mathbb{E}_{t'} [\max (v - c, 0)] \geq \mathbb{E}_t [\max (v - c, 0)], \forall t' > t, \quad (\text{IEGT})$$

and

$$\mathbb{E}_{t'} [\max (v - c, 0)] \geq \mathbb{E}_t [v] - c, \forall t' < t. \quad (\text{SOV})$$

Intuitively, the complete-information solution is implemented by charging every consumer $p(t) = \mathbb{E}_t [\max (v, c)] = \mathbb{E}_t [\max (v - c, 0)] + c$ at time 0, and offering them a refund of $r(t) = c$ at time t , where $\mathbb{E}_t [\max (v - c, 0)]$ is the expected gains from trade for a consumer of type t . So if [Condition IEGT](#), short for *increasing expected gains from trade*, did not hold, then a type t would strictly prefer the lower-priced contract, $\{p(t'), r(t')\}$, which she could costlessly obtain by imitating type $t' > t$. Imitation of type $t' > t$ is costless because the refunds offered are the same and because type t will know her valuation at time $t' > t$.

[Condition SOV](#), short for *sufficient option value*, is necessary because otherwise type t would imitate type $t' < t$. The cost of this deviation is

$$\mathbb{E}_t [\max (v - c, 0)] - [\mathbb{E}_t [v] - c], \quad (7)$$

which is the value of the option to return the good if type t does not deviate (type t will report her valuation exceeds c when required to report early). This is a strictly positive cost that does not depend on which type $t' < t$ that t imitates. So this cost makes small or local downward deviations unprofitable if the firm screens on when the refund can be exercised.

The benefit of this deviation is the reduction in the upfront payment, $p(t) - p(t')$, or

$$\mathbb{E}_t [\max (v - c, 0)] - \mathbb{E}_{t'} [\max (v - c, 0)], \quad (8)$$

which is positive by [Condition SOV](#). Clearly the cost, (7), is bigger than the benefit, (8), if and only if [Condition SOV](#) holds.

Since the expected gains from trade can be written as

$$\mathbb{E}_t [\max (v - c, 0)] = \mathbb{E}_t [v] - c + \mathbb{E}_t [\max (c - v, 0)],$$

and the option value, $\mathbb{E}_t [\max (c - v, 0)]$, is always positive, [Condition SOV](#) is satisfied whenever the consumer's expected valuation, $\mathbb{E}_t [v]$, is decreasing in t . But [Condition SOV](#) may not be satisfied when the consumer's expected valuation is decreasing. In particular, if $\underline{v} \geq c$ then the option value is zero and [Condition SOV](#) is satisfied if and only if the expected valuation is increasing, so [Condition IEGT](#) and [Condition SOV](#) are mutually exclusive. However, when c is sufficiently large, or when \underline{v} is sufficiently small, or more generally when the option value is sufficiently large, then [Condition IEGT](#) and [Condition SOV](#) can both hold.

[Proposition 1](#) shows that the seller is strictly better off when consumers learn their preferences sequentially; indeed, when [Condition IEGT](#) and [Condition SOV](#) hold, the seller can implement the unconstrained first-best. When [Condition IEGT](#) and [Condition SOV](#) are both satisfied, there is a unique mechanism, $(x(v, t), y(v, t))$, that extracts all of the surplus. One way the firm can implement this mechanism is with a menu of expiring refund contracts: the initial upfront price for a ticket is $p(t) = \mathbb{E}_t [\max (v, c)]$ and the refund is $r(t) = c$ if the ticket is returned any time

up to time t . Since the refund is equal to c , only consumers with valuations higher than the c will consume, which implies that the allocation is ex post efficient. Moreover, the expected utility of all the consumers is equal to zero. Because [Condition SOV](#) of [Proposition 1](#) is satisfied, no type t consumer would want to purchase the ticket designed for a higher type $t' > t$. Similarly, when [Condition SOV](#) of [Proposition 1](#) is satisfied, no type t would want to purchase the ticket designed for a lower type $t' < t$ since the refund of the ticket for type t' expires at time t' , when the type t consumer is still uncertain about her valuation for the ticket.

On the other hand, when the consumers learn their valuations at the same time, the seller cannot exploit the differences in learning times to screen consumers. In particular, in the model analyzed by [Courty and Li \(2000\)](#) the only case in which the seller can implement the first-best is the degenerate case in which all consumers have the same expected surplus, i.e., $\mathbb{E}_t [\max(v, c)] - c$ does not depend on t .¹¹

Importantly, [Proposition 1](#) is inconsistent with the envelope theorem of [Pavan et al. \(2014\)](#), which characterizes the information rents that privately-informed agents earn in optimal dynamic mechanisms. Clearly our assumptions differ in important ways from those of [Pavan et al. \(2014\)](#). In some respects the differences in our setting and that of [Pavan et al. \(2014\)](#) are rather dramatic. In particular, the timing of consumers' valuation reports depends on consumers' earlier reports about their types, and we consider a continuous-time model while [Pavan et al. \(2014\)](#) consider a discrete-time model. One intuitive way to think about the inconsistency is that we introduce screening on when consumers can exercise the return option and this instrument is not distortionary in our model. That is, requiring a type t consumer to exercise the return option at time t (instead of later every consumer has learned) is not costly for type t , but is costly for type $t' > t$. But the fundamental reason for the inconsistency is that the impulse response functions are unbounded in our setting and so the envelope theorem does not apply.

Note also that as in the general model of [Pavan et al. \(2014\)](#), the agent's private information in this paper at any moment τ can be summarized with a one-dimensional type, which is his initial type t at all time $\tau < t$ and his realized valuation v for time $\tau > t$. However, the impulse response of this stochastic process to the agent's initial type is infinite at the moment $\tau = t$ (since a small change in t shifts the agent's type from t to the agent's realized value). In contrast, the PST envelope-theorem derivation of the agent's information rents relies on their assumption of bounded impulse responses. The unboundedness of impulse response in our paper may be the technical reason for the principal's ability to extract the agent's information rents under some conditions.

Note however that the assumption that learning is instantaneous and not gradual is not crucial for implementing the first-best allocation. [Ata and Dana \(forthcoming\)](#) explore a generalization of the binary distribution numerical example discussed above and derive similar results even when information about valuations arrives gradually. In that paper, consumers' beliefs evolve in response to new information between time 0 and time t , and we show that it is a cap on how optimistic consumers can become, not on quickly their beliefs evolve that determines whether the first-best is incentive compatibility. Of course, in that paper we are also assuming the type space is multidimensional, which is inconsistent ([Pavan et al., 2014](#)). Clearly, it is important to understand which assumptions in our model are critical for our economic insights and how those assumptions differ from those in [Pavan et al. \(2014\)](#) and other related research.

¹¹ This is because [Courty and Li \(2000\)](#) assume the distribution of valuations has full support. When the distribution is binary, the complete-information profit may be feasible with heterogeneous consumers because differences in the refundability of purchases do not distort ex post consumption.

Clearly a more realistic assumption is that consumers vary in when they expect to learn, but precisely when they learn their valuations is still random. For example, a type t consumer might learn her valuation at a random time in a small interval around t . In such a setting, requiring consumers to report their valuations at time t would be costly, but it would still be less costly for a type t consumer to report her valuation at time t than for a type t' consumer, where $t' > t$. So while it wouldn't be possible to achieve the first-best in this case, some of the same economic insights should hold.

A simple case in which both [Condition IEGT](#) and [Condition SOV](#) of [Proposition 1](#) are satisfied, and the optimal mechanism achieves the first best, is when consumers who learn their valuations later have more dispersed priors about their valuations, in the sense of a mean-preserving spread (MPS). We call this Forward MPS and in [Section 5](#) contrast this with Reverse MPS in which consumers who learn their valuations later have less dispersion in their valuations. We define forward mean-preserving spread following the definition in [Rothschild and Stiglitz \(1970\)](#).

Assumption 1 (Forward MPS). Let $\mu = \mathbb{E}_t[v]$ for all t . For all t, t' such that $t' > t$,

$$\int_{\underline{v}}^v [F(v'|t') - F(v'|t)] dv' \geq 0, \forall v, \quad (9)$$

with strict inequality for some v .

We now show that [Condition IEGT](#) and [Condition SOV](#) of [Proposition 1](#) are satisfied under [Assumption 1](#), and so the complete-information outcome can be implemented.

Proposition 2. Under [Assumption 1](#) (Forward MPS), the first-best solution can always be implemented.

Proof. Using [\(5\)](#),

$$\begin{aligned} \mathbb{E}_{t'}[\max(v - c, 0)] - \mathbb{E}_t[\max(v - c, 0)] &= \int_c^{\bar{v}} [F(v|t) - F(v|t')] dv, \\ &= - \int_{\underline{v}}^c [F(v|t) - F(v|t')] dv \geq 0, \end{aligned}$$

for all $t' > t$, where the inequality follows from [Assumption 1](#), so [Condition IEGT](#) of [Proposition 1](#) is satisfied. Also, [Assumption 1](#) implies $\mathbb{E}_t[v] = \mathbb{E}_{t'}[v]$, so

$$\mathbb{E}_{t'}[\max(v - c, 0)] \geq \mathbb{E}_{t'}[v] - c$$

implies

$$\mathbb{E}_{t'}[\max(v - c, 0)] \geq \mathbb{E}_t[v] - c,$$

for all t, t' , so [Condition SOV](#) of [Proposition 1](#) is satisfied. \square

Note that unlike most standard sequential screening frameworks, the firm is offering every consumer a contract with the same strike price, or refund option, yet these consumers are not pooled; the price in each case is equal to their individual willingness to pay. Instead the firm separates initial types by varying the time that the option to claim the refund expires, an instrument that does not reduce total surplus.

Intuitively, consumers who learn late have the same expected valuation, but their expected valuation conditional on the valuation being above cost is higher, so [Condition IEGT](#) is clearly satisfied. And [Condition SOV](#) is satisfied because at the complete-information prices a consumer who imitates a consumer who learns earlier must get a negative expected consumer surplus – the price she pays and her expected valuation are both the same as the consumer she imitates, but that consumer gets zero expected surplus while deriving a positive benefit from the option to return the good if her valuation is low.

In Section 6 we will see that both [Condition IEGT](#) and [Condition SOV](#) may also hold when the consumers' distributions can be ordered with respect to first-order stochastic dominance.

5. When screening on time is not useful

In Section 4 we showed that under some conditions the seller can achieve the first best by using the timing of refunds as a screening device. This section considers instead the conditions under which screening on the timing of refunds is ineffective, or equivalently, when the optimal mechanism with sequential learning is equivalent to the optimal mechanism with simultaneous learning.

Screening on time is a potential way to prevent a buyer of type t from imitating a buyer of type $t' < t$, but not vice versa. The asymmetry is because once the buyer learns her valuation she always knows it, and so imitating a consumer with a higher type is easy, while imitating a consumer with a lower type is harder because it is costly for her to report her valuation before learning it. So time is not useful as a screening device when low types would like to imitate high types, which happens when the expected gains from trade are decreasing in t .

Formally, the firm's problem when it does not screen on the timing of its refunds, or equivalently, when consumers learn simultaneously, is the problem analyzed by [Courty and Li \(2000\)](#), and it is identical to [\(P¹\)](#) except that [\(IC₀\)](#) is replaced with

$$U(t) \geq \mathbb{E}_t[u(v, t'; v, t)], \forall t, t'; t' < t, \quad (\text{IC}'_0)$$

that is, except that all of the consumers report their valuation v after they know it. We call this problem [\(P_{CL}¹\)](#).

The following lemma, which is also used in Section 6, helps us characterize the expected surplus function, U , for any mechanism. Its proof is in [Appendix A](#).

Lemma 3. *For any feasible mechanism in which $U(t)$ is monotone, [\(IC₀\)](#) implies that*

$$U'(t) \leq - \int_{\underline{v}}^{\bar{v}} y(v, t) \frac{\partial F(v|t)}{\partial t} dv \text{ for almost every } t. \quad (10)$$

Using the definition of U and u , we can write the firm's objective function as

$$\max_{y(v, t)} \int_{\underline{v}, t} f(v, t) (v - c) y(v, t) dv dt - \int_t U(t) h(t) dt. \quad (11)$$

Lemma 3 implies that if $U(t)$ is non-increasing then for any optimal policy the individual rationality constraint must bind at T , or $U(T) = 0$. Also, if we relax the firm's problem by ignoring the global incentive compatibility constraints, then it is also clear that (10), the local incentive compatibility constraint, must bind. Equation (10) and $U(T) = 0$ imply

$$U(t) = - \int_t^T \int_{\underline{v}}^{\bar{v}} y(v, t) \frac{\partial F(v|t)}{\partial t} dv ds.$$

This holds in both our problem and in (P_{CL}^1) , because (10) binds in both. Substituting this expression into (11) we can write the firm's relaxed problem as

$$\max_{y(v,t)} \int_{v,t} f(v, t) (v - c) y(v, t) dv dt + \int_0^T \int_t^T \int_{\underline{v}}^{\bar{v}} h(t) \frac{\partial F(v|s)}{\partial s} y(v, s) dv ds dt,$$

subject to the constraint that $y(v, t)$ non-decreasing in v . And using integration by parts the firm's problem can be rewritten

$$\max_{y(v,t)} \int_{v,t} \left((v - c) f(v, t) + (1 - H(t)) \frac{\partial F(v|t)}{\partial t} \right) y(v, t) dv dt, \quad (P^2)$$

subject to the constraint that $y(v, t)$ non-decreasing in v . Note that (P^2) is a relaxation of both (P^1) and (P_{CL}^1) under the assumption that $U(t)$ is increasing.

By **Lemma 1**, it is optimal for a consumer to report her true type if she knows it, or $v = \arg \max_{v'} u(v', t'; v, t)$, so $\mathbb{E}_t[u(v, t'; v, t)] \geq \mathbb{E}_t[u(\mathbb{E}_t[v], t'; v, t)]$, which implies that if (IC'_0) holds then so does (IC_0) . This establishes the following lemma.

Lemma 4. *If the solution to (P^2) solves (P_{CL}^1) then it also solves (P^1) .*

In other words, if the constraints omitted from the firm's problem when it does screen on time, (P_{CL}^1) , are not binding, then time is not a useful screening device. The omitted constraints are (IR) for $t < T$ and the global incentive compatibility constraints, (\bar{IC}_0) and (IC'_0) .

A common solution strategy for problems like (P^2) is to make assumptions that guarantee that

$$(v - c) f(v, t) + (1 - H(t)) \frac{\partial F(v|t)}{\partial t} \quad (12)$$

is monotonically increasing. These assumptions would guarantee that the solution to (P^2) is a function $y(v, t)$ that is non-decreasing in v and hence that $y(v, t) \in \{0, 1\}$ almost everywhere, or equivalently that the solution is deterministic.

However, it is not necessary to impose these additional assumptions. Because (P^2) is subject to the constraint that $y(v, t)$ is non-decreasing, it must have a deterministic solution even if (12) is not monotonically increasing. This is because the firm is maximizing a linear objective function over a closed convex subset of a linear space of functions, and it is well known that in this case the maximal value must be achieved at an extreme point of the feasible set. An extreme point in the set of nondecreasing functions from $\mathbb{R} \rightarrow [0, 1]$ is a function that jumps from 0 to 1 at one point, and its value at that point is immaterial for the integral (see for example Segal, 2003). So a solution to (P^2) exists in which $y(v, t) \in \{0, 1\}$ almost everywhere.

Using the above insight allows us to generalize the results of Courty and Li (2000), who do not characterize general conditions under which the solution is non-decreasing, but instead focus on special distributional examples in which it is possible to show the solution is non-increasing.

We begin with some definitions. First, we define Reverse MPS as follows: as the case in which the consumers' distributions can be ranked using a mean-preserving spread and the distribution of valuations of consumers who learn earlier are a mean-preserving spread of the distribution of valuations of consumers who learn later, implying that the expected gains from trade are higher for consumers who learn earlier because the expected valuation is unchanged, but conditional on $v < c$, expected valuations are lower, and conditional on $v > c$, expected valuations are higher.

Assumption 2 (Reverse MPS). $\mathbb{E}_t[v] = \mu$ for all t , and

$$\int_{\underline{v}}^v [F(v'|t) - F(v'|t')] dv' \geq 0, \forall v, t', t; t' > t, \quad (13)$$

with strict inequality for some v , or equivalently,

$$\int_{\underline{v}}^v \frac{\partial F(v'|t)}{\partial t} dv' \leq 0, \forall v, t, \quad (14)$$

with strict inequality for some v , and strict equality for $v = \bar{v}$.

In other words, Reverse MPS is satisfied when $F(v|t)$ is a mean preserving spread of $F(v|t')$ for all $t < t'$. Assumption 2 implies that there exists a \hat{v} such that $\partial F(\hat{v}|t)/\partial t = 0$ for all t , or equivalently that the distributions $F(v|t)$ cross at a common point. Equivalently, for all t , $\partial F(\hat{v}|t)/\partial t \leq 0$ for all $v < \hat{v}$, and $\partial F(\hat{v}|t)/\partial t \geq 0$ for all $v > \hat{v}$.

We define Reverse FOSD as the case in which the consumers' distributions can be ranked by first-order stochastic dominance and the distribution of valuations of consumers who learn earlier first-order stochastically dominates the distribution of valuations of consumers who learn later.

Assumption 3 (Reverse FOSD). $F(v|t') \geq F(v|t)$ for all v, t, t' such that $t' > t$, with strict inequality for some v .

The following Lemma shows that U is decreasing, so we can apply Lemma 4.

Lemma 5. Under Assumption 2 or 3, the optimal expected utility function, $U(t)$, is non-increasing.

Two additional assumptions are required to demonstrate that the solution to (P²) is the solution to the more general problem, (P¹).

Assumption 4. $\left| \frac{(1-H(t))}{h(t)} \frac{\partial F(v|t)/\partial t}{f(v|t)} \right|$ is non-increasing in t for all v, t .

Assumption 5. $\frac{\partial \left[\frac{\partial F(v|t)/\partial t}{f(v|t)} \right]}{\partial v} > -\frac{h(t)}{1-H(t)}$ for all v, t .

Note that $\frac{\partial F(v|t)/\partial t}{f(v|t)} > 0$ by [Assumption 3](#), so sufficient conditions for [Assumption 4](#) are that the monotone hazard rate assumption holds and that $\left| \frac{\partial F(v|t)/\partial t}{f(v|t)} \right|$ is non-increasing in t .

The following is an application of [Lemmas 4 and 5](#) and of [Lemmas 3.3 and 3.4](#) in [Courty and Li \(2000\)](#), and note again that the result generalizes [Courty and Li \(2000\)](#) who impose additional distributional assumptions to prove that the solution is non-increasing.

Proposition 3. *Under [Assumptions 2 \(Reverse MPS\) and 4–5](#), if $c > \hat{v}$ (alternatively, $c < \hat{v}$), the solution to (P^2) is a non-decreasing (alternatively, non-increasing) refund schedule, $r(t)$, with $r(t) > c$ (alternatively, with $r(t) < c$). This solution is also the optimal policy in (P^1) and the optimal policy in a model in which consumers learn simultaneously, (P_{CL}^1) .*

Similarly, under [Assumptions 3 \(Reverse FOSD\) and 4–5](#), the solution to (P^2) , is a non-decreasing refund schedule, $r(t)$. This solution is also the optimal policy in (P^1) and the optimal policy in a model in which consumers learn simultaneously, (P_{CL}^1) .

Proof. First, consider Reverse MPS. Clearly $r(0) = c$ (no distortion at $t = 0$), so if $c < \hat{v}$, then $F(r(0)|0)/\partial t \leq 0$, and under [Assumption 4](#), $\phi_t(r(t), t) \geq 0$ for all t , and under [Assumption 5](#), $\phi_v(r(t), t) > 0$ for all t , so $r'(t) \leq 0$ and $r(t) \leq c$ for all t , and by [Lemma 4](#) the solution to the relaxed problem is the solution to the original problem.

On the other hand, if $c > \hat{v}$, then $F(r(0)|0)/\partial t \geq 0$, and under [Assumption 4](#), $\phi_t(r(t), t) \leq 0$ for all t , and under [Assumption 5](#), $\phi_v(r(t), t) > 0$ for all t , so $r'(t) \geq 0$ and $r(t) \geq c$ for all t , and by [Lemma 4](#) the solution to the relaxed problem is the solution to the original problem.

Optimality when consumers learn simultaneously follows from the observation that a deterministic solution exists and from [Lemma 3.4](#) in [Courty and Li \(2000\)](#), which shows that the omitted global incentive compatibility constraints are satisfied provided that the solution satisfies $r'(t) < 0$ when $r(t) < c$ and $r'(t) > 0$ when $r(t) > c$. And optimality when consumers learn sequentially follows from [Lemma 4](#).

Now consider Reverse FOSD. Using $f(v|t) = f(v, t)h(t)$, the first order condition for (P^2) is $f(v, t)\phi(r(t), t) = 0$, where

$$\phi(v, t) = \left((v - c) + \frac{(1 - H(t))}{h(t)} \frac{\partial F(v|t)/\partial t}{f(v|t)} \right).$$

Using the implicit function theorem,

$$r'(t) = -\frac{\phi_t(r(t), t)}{\phi_v(r(t), t)}, \quad (15)$$

where [Assumption 4](#) is sufficient to guarantee that $\phi_t \leq 0$, and [Assumption 5](#) is sufficient to guarantee $\phi_v > 0$, so $r'(t)$ is non-negative.

Optimality when consumers learn simultaneously follows from the observation that a deterministic solution exists and from [Lemma 3.3](#) in [Courty and Li \(2000\)](#), which shows that the omitted global incentive compatibility constraints are satisfied provided that the solution satisfies $r'(t) \geq 0$. And optimality when consumers learn sequentially follows from [Lemma 4](#). \square

Reverse MPS implies that consumers who learn earlier have more dispersed valuations, a case we contrast with Forward MPS analyzed in the previous section. Reverse FOSD implies that consumers who learn earlier have greater valuations in the sense of first-order stochastic dominance, a case we contrast with forward first-order stochastic dominance in the next section. Intuitively,

both Reverse MPS and Reverse FOSD imply that the expected gains from trade are decreasing in t , so a type T buyer is associated with the smallest gains from trade, and the optimal mechanism satisfies $U(T) = 0$ and $U'(t) \leq 0$. The lower types get positive rents because they can always imitate type T and guarantee themselves a positive surplus.

6. When screening on the time is somewhat useful

So far we have considered two extreme outcomes. First, we considered conditions under which screening on when the refund option can be exercised implements the first best, and second, we considered conditions under which screening on when the refund option can be exercised is not useful, and the firm not do any better than to treat all consumers as if they learned simultaneously. We now analyze a case in which screening on when the refund option can be exercised is useful, but the first best cannot be achieved.

Suppose consumers' conditional valuation distributions, $F(v|t)$, can be ordered on $[0, T]$ with respect to a first-order stochastic dominance (FOSD) and that consumers who learn later have higher valuations. This is a natural assumption in many settings. For example, in an airline context, business travelers often have higher valuations than the leisure travelers, and they typically learn their travel needs much closer to the departure time.

Formally, we define Forward FOSD as follows:

Assumption 6 (Forward FOSD). $F(v|t') \leq F(v|t)$ for all v, t, t' such that $t' > t$, with strict inequality for some v .

Assumption 6 and equation (5) imply that $\mathbb{E}_t[\max(v, c)]$, and as a consequence $\mathbb{E}_t[\max(v - c, 0)]$, is weakly increasing in t , so **Condition IEGT** of **Proposition 1** is satisfied.

When **Condition SOV** of **Proposition 1** is also satisfied, then the firm allocates the good efficiently and extracts all the surplus. However under **Assumption 6**, **Condition SOV** need not be satisfied. For example, if $\underline{v} \geq c$, then **Condition SOV** cannot hold. Indeed, under **Assumption 6**, **Condition SOV** of **Proposition 1** is violated, that is $\mathbb{E}_t[v] > \mathbb{E}_{t'}[\max(v, c)]$, for some $t > t'$, if and only if

$$\mathbb{E}_T[v] > \mathbb{E}_0[\max(v, c)], \quad (16)$$

because $\mathbb{E}_T[v] \geq \mathbb{E}_t[v]$ for all t and $\mathbb{E}_t[\max(v, c)] \geq \mathbb{E}_0[\max(v, c)]$ for all t , both of which follow from **Assumption 6**.

In this section we assume that equation (16) holds:

Assumption 7. **Condition SOV** of **Proposition 1** is violated.

Again, when **Condition SOV** is violated, the firm cannot extract all of the surplus – if the firm offered the complete information contracts, then even with restrictions on when the return option would be exercised, type T would still imitate type 0.

Under **Assumptions 6 and 7**, the optimal mechanism has several interesting features. First, we show that both local and global incentive compatibility constraints bind. The highest type, type T , is offered a price that makes her indifferent between reporting her type as type T or type 0, or more precisely, indifferent between reporting T and reporting any t in an interval $[0, \sigma]$. So downward global incentive compatibility constraints bind for type T . The rents given to type T make it attractive for consumers in an interval $[\tau, T)$ (where $\tau > \sigma$) to imitate type T , so the

optimal mechanism offers these consumers a lower price as well, so the upward local incentive compatibility constraint binds for these consumers. For these consumers, the downward global incentive compatibility constraints do not bind because the cost of imitating type 0 is higher for them (they must give up even more option value by foregoing the option to return) so they prefer instead to imitate type T (they also pay less than type T when they report truthfully).

Second, as is evident from the above discussion, incentive compatibility constraints bind in both directions, that is, both upward (local) and downward (global) incentive compatibility constraints bind.

Third, we find that the optimal mechanism includes both upward distortions in the allocation for some consumer types and simultaneously downward distortions in the allocations for other consumer types. Type 0 is offered a higher refund and a smaller allocation relative to the complete information contract (a downward allocation distortion), increasing the ex ante price paid by type 0 and reducing the incentive for type T to imitate type 0 (because she pays the ex ante price but does not ever enjoy the refund). And in an interval $(\tau, T]$ consumers are offered a lower refund and a greater allocation (an upward allocation distortion) relative to the complete information contract, reducing the incentive for consumers to imitate higher type consumers. While other papers have shown the distortions can go in more than one direction, even in the same mechanism (see for example Pavan et al., 2014), we think that the simultaneous presence of upward and downward distortions in our paper is particularly natural and intuitive.

Finally, the upward distortion increases with t , so we find there is a distortion “at the top” for the buyer with the largest ex ante highest willingness to pay. This counterintuitive result arises because the global incentive constraint implies that type T gets enough additional rents that the direction of the binding local incentives constraints is reversed. Distorting the allocation “at the top” for type T reduces the information rents captured by consumers in the interval $(\tau, T]$.

In what follows, we first solve a relaxed version of the seller’s problem and then (in Proposition 7) show that its solution satisfies all the constraints of the original problem. As intermediate steps, we first prove that the expected surplus $U(t)$ of type t is non-decreasing under FOSD and then show that we can restrict attention to deterministic allocations.

Lemma 6. *The optimal expected utility function, $U(t)$, is non-decreasing.*

Lemmas 3 and 6 imply

$$0 \leq U'(t) \leq - \int_{\underline{v}}^{\bar{v}} y(v, t) \frac{\partial F(v|t)}{\partial t} dv \text{ for almost every } t, \quad (17)$$

and together with Lemma 2, we use equation (17) to further relax the seller’s problem, which we call (P³).

$$\max_{U(t), u(v, t), y(v, t)} \int_{\underline{v}, t} (v - c) y(v, t) f(v, t) dv dt - \int_t U(t) h(t) dt, \quad (\text{P}^3)$$

subject to

$$U(t) \geq 0 \text{ for all } t, \quad (\text{IR})$$

$$y(v, t) = \frac{\partial u(v, t)}{\partial v} \text{ and } y(v, t) \text{ is non-decreasing in } v, \quad (\tilde{\text{IC}}_t)$$

$$0 \leq U'(t) \leq - \int_{\underline{v}}^{\bar{v}} y(v, t) \frac{\partial F(v|t)}{\partial t} dv \text{ for all } t, \quad (\tilde{\text{IC}}_0)$$

$$U(T) \geq U(t; T) \text{ for all } t, \quad (\text{IC}_0^T)$$

$$0 \leq y(v, t) \leq 1 \text{ for all } v, t. \quad (\text{F})$$

The $U(t)$ in the argument of the firm's objective function is redundant since the firm still chooses x and y , or equivalently $u(v, t)$ and $y(v, t)$, which determine $U(t)$ and $U(t; t')$ from equations (3) and (4).

The following proposition partially characterizes the optimal solution to the relaxed problem, and its proof is given in [Appendix B](#).

Proposition 4. *Under [Assumption 6](#) there exists a deterministic solution, $y(v, t) \in \{0, 1\}$, for all v, t , to the relaxed problem (P³). In particular, for each t , there exists a cut off point $r(t)$ such that $y(v, t) = 1$ if and only if $v \geq r(t)$. Moreover, there exists τ such that*

$$U(t) = \begin{cases} 0 & \text{if } t \leq \tau, \\ - \int_{\tau}^t \int_{r(s)}^{\bar{v}} \frac{\partial F(v|s)}{\partial s} dv ds & \text{if } t \geq \tau. \end{cases}$$

[Proposition 4](#) establishes the existence of an optimal solution. The first part of the proof establishes the solution is deterministic, $y(v, t) \in \{0, 1\}$. We use the existence of a non-deterministic solution to define a new choice problem for the firm, and then we demonstrate that the new choice problem has a deterministic solution that does at least as well as any non-deterministic solution and satisfies the constraints. The proof uses the fact that the unconstrained problem has a deterministic solution (see the discussion in [Section 5](#)).

The second part of the proof establishes the existence of a threshold τ below which $(\tilde{\text{IC}}_0)$ binds at its lower bound and above which $(\tilde{\text{IC}}_0)$ binds at its upper bound. Note that if [Assumption 7](#) is violated, then the proposition still holds, but the first best is achieved, and $\tau = T$.

The next proposition holds when the mechanism is deterministic and is instrumental in characterizing the optimal contract. The proof is in [Appendix B](#).

Proposition 5. *Under [Assumption 6](#), we have that for all t*

$$U(t; T) = (\mathbb{E}_T[v] - r(t)) \mathbf{1}_{\{\mathbb{E}_T[v] - r(t) \geq 0\}} + U(t) - \int_{r(t)}^{\bar{v}} (v - r(t)) f(v|t) dv, \quad (18)$$

where the indicator function $\mathbf{1}_{\{x \geq 0\}}$ is 1 if $x \geq 0$, and zero otherwise.

Note that by [Proposition 4](#), the second term, $U(t)$, is equal to zero for $t \leq \tau$. And more importantly, if $\mathbb{E}_T[v] - r(t) < 0$ then Equation (18) implies $U(t; T) < U(t)$. Since [Proposition 4](#) guarantees that any optimal mechanism satisfies $U(T) > U(t)$, this means that whenever $\mathbb{E}_T[v] - r(t) \leq 0$, the incentive constraint $U(T) \geq U(t; T)$ can be ignored. So if type T deviates and reports type t , it must be that type T will not return the good (or $\mathbb{E}_T[v] - r(t) > 0$). We use this below when we further relax the firm's optimization problem.

We can now further simplify the firm's relaxed problem, (P³). First, using [Proposition 4](#) and [Proposition 5](#), we replace $U(t)$ and $U(t; T)$ in the objective function and in the constraints.

Note also that these functions clearly satisfy (IR) and (\tilde{IC}_0) so these constraints can be dropped. Also, Proposition 4 implies that the optimal $y(v, t)$ can be represented by a function $r(t)$ where $y(v, t) = 1$ if $v \geq r(t)$ and 0 otherwise. It follows that we can write the firm's problem as choosing τ and $r(t)$ on $[0, T]$. That is, without loss of generality we can restrict our attention to contracts in which a type t consumer makes an ex ante payment for the good, which we will denote by $p(t)$, and then receives a refund, which we will denote by $r(t)$, if she returns the good at time t after learning their valuations (in which case she will return the good if and only if $v < r(t)$). This menu of expiring refund contracts clearly satisfy (F), so this constraint is dropped. Finally, we ignore (\tilde{IC}_t) and then show in Proposition 7 that the solution to the relaxed problem satisfies the ignored constraints.

After integrating the objective function by parts and dropping (\tilde{IC}_t) , we can write the firm's relaxed problem as (P^4) :

$$\begin{aligned} \max_{\{\tau, r(t)\}} & \int_0^{\tau} \int_{r(t)}^{\bar{v}} f(v, t)(v - c)dvdt \\ & + \int_{\tau}^T \int_{r(t)}^{\bar{v}} f(v, t) \left(v - c + \frac{1 - H(t)}{h(t)} \frac{\partial F(v|t)/\partial t}{f(v|t)} \right) dvdt \end{aligned} \quad (P^4)$$

subject to $U(T) \geq U(t'; T)$, which, using the relaxed definition of $U(t; T)$ from Proposition 5, becomes

$$\begin{aligned} - \int_{\tau}^T \int_{r(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dvdt & \geq \mathbb{E}_T[v] - r(t') - U(t) - \int_{r(t')}^{\bar{v}} (v - r(t'))f(v|t')dv, \forall t' \\ & \leq T. \end{aligned} \quad (\tilde{IC}_0^T)$$

Proposition 6 characterizes the optimal solution to the problem (P^4) and shows that the optimal refund schedule to be decreasing in t . Assumptions 8 and 5 are sufficient conditions.

Assumption 8. $\frac{\partial F(v|t)/\partial t}{f(v, t)}$ is non-decreasing in t for all t and all $v \in [\underline{v}, c]$.

Assumption 9. $\frac{\partial F(v|t)/\partial t}{f(v, t)}$ is non-increasing in v for all t and all $v \in [\underline{v}, c]$.

Intuitively, Assumption 8 requires that the informativeness of a consumer's type about her valuation weakly increases with her type (for all valuations less than c).

Proposition 6. Under Assumptions 6–9 the solution to (P^4) is characterized by two thresholds σ and τ , such that $0 < \sigma \leq \tau < T$ and (\tilde{IC}_0^T) binds for $t \leq \sigma$ and not otherwise. The optimal $r(t)$ is continuous and satisfies:

1. for $t \leq \sigma$, $r(t) \geq c$, $r(t)$ is non-increasing, and $r(t)$ is the unique solution to (\tilde{IC}_0^T) ;
2. for $t \in [\sigma, \tau]$, $r(t) = c$; and
3. for $t \geq \tau$, $r(t)$ is strictly decreasing.

Note that this solution exhibits both downward distortions in allocation (a return price distorted higher than the complete information price) at and near $t = 0$ and upward distortions in allocation (a return price distorted lower than the complete information price) at and near $t = T$.

In the proof we further relax the firm's problem by considering only the downwards incentive constraints are for type T , who is indifferent between reporting type T and type $t \in [0, \sigma]$ at the optimum, and considering only the upwards local incentive constraints for type $t \in [\tau, T]$. We then check that the omitted local and global incentive constraints are satisfied. Type T is most tempted to report her type as type 0, and setting $r(t) \geq c$ reduces the rents paid to type T . However the need for the distortion diminishes as t increases. At σ the global incentive constraint for type T no longer binds and no distortion is needed (of course σ goes to 0 as the option value gets larger so Assumption 4 no longer holds). Again, the constraint does not bind at τ , and becomes increasingly binding as t approaches T , so the distortion increases, and $r(t)$ falls, as t increases.

Although Assumptions 8 and 9 are sufficient to ensure that $r'(t) \leq 0$, they are not necessary. Indeed, the condition $r'(t) \leq 0$ is itself sufficient but not necessary for proving that (IC_0^T) is slack on $[\tau, T]$ (see Lemma 7 in Appendix B).

Given the cutoff points $\{r(t) : 0 \leq t \leq T\}$ characterized in Proposition 6, the payments can be written as follows:

$$x(v, t) = \begin{cases} p(t) - r(t) & \text{if } v < r(t), \\ p(t) & \text{if } v \geq r(t), \end{cases}$$

where again we interpret $p(t)$ as an upfront payment and $r(t)$ as a refund which is optional and must be exercised at time t . Note that the expected surplus of type t consumer can be written as

$$U(t) = -(p(t) - r(t)) + \int_{r(t)}^{\bar{v}} (1 - F(v|t)) dv. \quad (19)$$

Then, since $U(t) = 0$ for $t \in [0, \tau]$, we write

$$p(t) - r(t) = \int_{r(t)}^{\bar{v}} (1 - F(v|t)) dv \text{ for } t \leq \tau.$$

For $t \geq \tau$, we write the following by taking the derivatives of both sides of (19),

$$p'(t) - r'(t) = -U'(t) - r'(t)(1 - F(r(t)|t)) - \int_{r(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv. \quad (20)$$

Therefore, we can calculate $p(t) - r(t)$ for $t \geq \tau$ from (20) and the boundary condition that

$$p(\tau) - r(\tau) = \int_{r(\tau)}^{\bar{v}} (1 - F(v|\tau)) dv.$$

We interpret $\{(p(t), r(t)) : 0 \leq t \leq T\}$ as a menu of expiring refund contracts where a type t consumer is charged the initial price $p(t)$ and is offered a refund of $r(t)$ if he chooses not to consume the good at (or before) time t . In other words, the refund $r(t)$ expires at time t .

The following proposition establishes that in the case of forward first-order stochastic dominance, this refund contract, $\{(p(t), r(t)) : 0 \leq t \leq T\}$, is an optimal contract in the seller's original mechanism design problem, (P^0) , that is, that the contract satisfies the omitted constraints. The proof is standard and follows the proof in Courty and Li (2000); it is included in Appendix B.

Proposition 7. *Under Assumptions 6–9 there exists a solution $\{x(v, t), y(v, t)\}$, to the firm's mechanism design problem which can be implemented as the menu of expiring refund contracts $\{(p(t), r(t)) : 0 \leq t \leq T\}$ characterized above.*

The following corollary investigates how the optimal initial price $\{p(t) : 0 \leq t \leq T\}$ changes over time and is proved in Appendix B.

Corollary 1. *The optimal menu of refund contracts $\{(p(t), r(t)) : 0 \leq t \leq T\}$ has the following properties: The optimal initial price $p(t)$ is constant for $t < \sigma$ and is strictly increasing with rate $p'(t) = -\int_c^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv$ for $t \in [\sigma, \tau)$. Also, the price is strictly decreasing with rate $p'(t) = r'(t) F(r(t)|t)$ for $t \geq \tau$, while $p(T) > p(t)$ for all $t \leq \sigma$. Finally, the effective price $p(t) - r(t) F(r(t)|t)$, defined as the expected transfer from the consumer to the seller, is increasing in t .*

The interval $[0, \sigma)$ is the range in which, like Courty and Li (2000), allocations are distorted downwards in order to extract more profit from (or give less surplus to) type T . The interval $(\tau, T]$ is the range where allocations are distorted upwards, because consumers in this range are receiving positive surplus since they would otherwise imitate type T . Distorting their allocations upwards reduces the incentive of these consumers to overreport their valuations.

7. Concluding remarks

This paper is the first to examine optimal pricing when consumers vary in when they learn their valuations over time, and when the time that consumers learn their valuations is correlated with the ex ante distribution of their valuations. Sequential learning gives the firm an additional instrument with which to screen consumers. In some cases this instrument enables the firm to implement the first-best allocation and extract all of the expected surplus from the consumers using ex ante contracts. In particular, the seller can implement the first best when consumers who learn later have more dispersed valuations, or when consumers who learn later have larger valuations and the option value of waiting to commit is large because the ex post gains from trade can be negative. In other cases, screening on the expiration of the refund option is not profitable, and the profits are the same as when buyers learn simultaneously.

Two aspects of this pricing problem are worth highlighting. First, a natural asymmetry exists because it is easier to charge a higher price to consumers who learn late than to consumers who learn early. This is because consumers who learn early can costlessly imitate consumers who learn late, but not vice versa. And second, unlike many screening problems, imposing restrictions on when they exercise their return option, need not be distortionary.

Much work remains to be done in characterizing optimal pricing when consumers learn over time. First, not all distributions of valuations and learning fit into the cases we analyzed. Second, and more importantly, when learning takes place and the ex ante distribution of valuations are unlikely to be perfectly correlated. In more realistic settings consumers are likely to have multiple

dimensions of heterogeneity, which could significantly increase the complexity of the optimal mechanism. Third, we assumed that consumers learn instantaneously while it is more realistic to suppose that learning takes place gradually.

Allowing for gradual learning is difficult because the amount of information reporting increases significantly, but notice that it should not significantly alter the consumers' incentives. That is, holding the mechanism fixed, we can allow the consumer to have a little more information just prior to time t without violating any binding incentive constraints because if consumers are going to deviate and pretend to be a lower type, they do not deviate locally. Since local downward constraints are slack, giving consumer more information shortly before they report their valuation does violate the constraints (see [Ata and Dana, forthcoming](#) for additional discussion of gradual learning).

Finally, we have focused on a mechanism design analysis which emphasizes the role of ex ante contracts and return options as mechanisms for extracting surplus. However, many firms that face consumers who learn over time may not be able to use ex ante contracts. This could be because of transaction costs, competition with rivals, or because consumers arrive to late for initial contracting (though late arrival may itself require justification using transaction costs). Further analysis of optimal pricing in such environments is clearly needed.

Appendix A. Proofs of the results in Sections 3 through 5

Proof of Proposition 1. The complete-information allocation is $y(v, t) = 1$ if $v \geq c$ and $y(v, t) = 0$ otherwise. The complete-information payments, $x(v, t)$, are not unique, but must extract all of the ex-ante consumer surplus, so

$$\mathbb{E}_t[x(v, t)] = \mathbb{E}_t[v y(v, t)] = \int_c^{\bar{v}} v f(v|t) dv, \forall t. \quad (\text{A.1})$$

To prove the “if” part of [Proposition 1](#), first note that [Lemma 2](#) requires that $\partial u(v, t)/\partial v = y(v, t)$ almost everywhere, so given $y(v, t)$, the payments $x(v, t)$ satisfies [\(IC_t\)](#) if and only if for some $p(t)$

$$x(v, t) = \begin{cases} p(t) & \text{if } v \geq c, \\ p(t) - c & \text{otherwise.} \end{cases} \quad (\text{A.2})$$

for all v, t . So the unique payments that satisfy [\(A.1\)](#) and [\(IC_t\)](#) are

$$x(v, t) = \begin{cases} \mathbb{E}_t[\max(v - c, 0)] + c & \text{if } v \geq c \\ \mathbb{E}_t[\max(v - c, 0)] & \text{otherwise.} \end{cases} \quad (\text{A.3})$$

Clearly the above mechanism, $(x(v, t), y(v, t))$, is individually rational since expected consumer surplus is 0. And clearly [\(IC_t\)](#) holds since conditions (i) and (ii) of [Lemma 2](#) are satisfied. Next, $\bar{\text{IC}}_0$ is satisfied if [Condition IEGT](#) holds because $x(v, t)$ is increasing in t and $y(v, t)$ does not depend on the reported t , so $U(t'; t) \leq U(t)$ for all $t' > t$. Finally, [IC₀](#) is satisfied under [Condition SOV](#) because for all $t' < t$,

$$U(t'; t) = \mathbb{E}_t[v] - \mathbb{E}_{t'}[\max(v - c, 0)] - c \quad (\text{A.4})$$

and $U(t) = 0$, so $U(t'; t) \leq U(t)$ if and only if [Condition SOV](#) holds.

To prove the “only if” part of [Proposition 1](#), we argue by contradiction. Suppose that the seller can implement the first-best solution but [Condition IEGT](#) is not satisfied. If [Condition IEGT](#) is violated, there exists a type $t'' > t'$ such that

$$\mathbb{E}_{t'} [\max (v - c, 0)] > \mathbb{E}_{t''} [\max (v - c, 0)] \quad (\text{A.5})$$

If type t' misrepresents her type as t'' at time zero, the expected utility, $U(t''; t')$, she gets is equal to

$$\begin{aligned} \int_{\underline{v}}^{\bar{v}} [vy(v, t'') - x(v, t'')] f(v|t') dv &= \int_c^{\bar{v}} vf(v|t') dv - \mathbb{E}_{t'} [x(v, t'')], \\ &= \int_c^{\bar{v}} vf(v|t') dv - p(t'') + cF(c|t'), \\ &= \mathbb{E}_{t'} [\max(v, c)] - p(t''), \\ &> \mathbb{E}_{t''} [\max(v, c)] - p(t''), \\ &= \int_c^{\bar{v}} vf(v|t'') dv - p(t'') + cF(c|t''), \\ &= \int_c^{\bar{v}} vf(v|t'') dv - \mathbb{E}_{t''} [x(v, t'')], \\ &= 0, \end{aligned}$$

where the strict inequality follows from (A.5) (and the fact that $\mathbb{E}_t [\max(v, c)]$ is equal to $\mathbb{E}_t [\max(v - c, 0)] + c$), and the last line follows from (A.1). Thus, type t' has an incentive to misreport her type as t'' at time zero and [\(IC₀\)](#) is violated, which contradicts the assumption that the seller can implement the first-best solution. Hence, [Condition IEGT](#) is necessary for implementing the first-best solution.

Next, suppose that the seller can implement the first-best solution but [Condition SOV](#) is not satisfied. Then there exist types $t'' > t'$ such that

$$\mathbb{E}_{t''} [v] > \mathbb{E}_{t'} [\max(v, c)]. \quad (\text{A.6})$$

Clearly this implies $\mathbb{E}_{t''} [v] \geq c$. If type t'' misreports her type as t' , her expected utility $U(t'; t'')$ is

$$\mathbb{E}_{t''} [vy(\mathbb{E}_{t''} [v], t') - x(\mathbb{E}_{t''} [v], t')] = \mathbb{E}_{t''} [v] - x(\mathbb{E}_{t''} [v], t'), \quad (\text{A.7})$$

since from [Lemma 1](#), type t'' will report her valuation as $\mathbb{E}_{t''} [v]$ at time t' and we have $y(\mathbb{E}_{t''} [v], t') = 1$ because $\mathbb{E}_{t''} [v] \geq c$. Then,

$$\begin{aligned} U(t'; t'') &= \mathbb{E}_{t''} [v] - x(\mathbb{E}_{t''} [v], t'), \\ &= \mathbb{E}_{t''} [v] - p(t'), \\ &> \mathbb{E}_{t'} [\max(v, c)] - p(t') \\ &= \int_c^{\bar{v}} vf(v|t') dv - p(t') + cF(c|t'), \end{aligned}$$

$$\begin{aligned}
&= \int_c^{\bar{v}} v f(v|t') dv - \mathbb{E}_{t'}[x(v, t')]. \\
&= 0,
\end{aligned}$$

where the strict inequality follows from (A.6) and the last line follows from (A.1). Thus, Condition SOV is also necessary for implementing the first-best solution. \square

Proof of Lemma 5. The following assumption on the distribution of v and t , is clearly implied by Assumptions 2 and 3:

Assumption 10. For all α and for all t ,

$$\int_{\alpha}^{\bar{v}} \frac{\partial F(v'|t)}{\partial t} dv' \geq 0. \quad (\text{Condition SDEGT})$$

Condition SDEGT is short for *strong decreasing expected gains from trade*, and Assumptions 2 and 3 each clearly imply Assumption 10. In contrast to the definition of Condition IEGT, the inequality reversed (hence decreasing as opposed to increasing). In addition Condition SDEGT is stated just for all α , while Condition IEGT is stated just for $\alpha = c$. So Condition SDEGT is stronger than just decreasing expected gains from trade. Also, note that Assumption 10 is the same as second-order stochastic dominance if $\mathbb{E}_t[v]$ is independent of t , but more generally is different from second-order stochastic dominance because the range of integration is $[\alpha, \bar{v}]$, not $[\underline{v}, \alpha]$.

Assumption 10 clearly implies that

$$\int_{\alpha}^{\bar{v}} [F(v'|t) - F(v'|t')] dv' \geq 0, \forall v, t, t'; t > t', \quad (\text{A.8})$$

or $\mathbb{E}_t[\max(v - \alpha, 0)] \leq \mathbb{E}_{t'}[\max(v - \alpha, 0)]$.

Note that (IC₀) implies that for $t > t'$,

$$U(t) - U(t') \leq U(t) - U(t; t'), \quad (\text{A.9})$$

and Lemma 2 and integration by parts implies

$$\begin{aligned}
U(t) - U(t; t') &= \int_{\underline{v}}^{\bar{v}} u(v, t) [f(v|t) - f(v|t')] dv \\
&= - \int_{\underline{v}}^{\bar{v}} y(v, t) [F(v|t) - F(v|t')] dv,
\end{aligned} \quad (\text{A.10})$$

where the integral exists (for Reiman–Stieltjes integration) because $F(v|t) - F(v|t')$ is continuous, and y is of bounded variation (y is bounded and weakly increasing in v); see Theorem 30.2 and Corollary 30.3 on page 229 of Bartle (1975). Using integration by parts again (see Theorem 29.7 on page 218 of Bartle, 1975),

$$\begin{aligned}
-\int_{\underline{v}}^{\bar{v}} y(v, t) [F(v|t) - F(v|t')] dv &= -(y(\bar{v}, t) - y(\underline{v}, t)) \int_{\underline{v}}^{\bar{v}} [F(v'|t) - F(v'|t')] dv' \\
&\quad + \int_{\underline{v}}^{\bar{v}} dy(v, t) \int_{\underline{v}}^v [F(v'|t) - F(v'|t')] dv' dv \\
&\leq -(y(\bar{v}, t) - y(\underline{v}, t)) \int_{\underline{v}}^{\bar{v}} [F(v'|t) - F(v'|t')] dv' \\
&\quad + \int_{\underline{v}}^{\bar{v}} dy(v, t) \int_{\underline{v}}^v [F(v'|t) - F(v'|t')] dv' = 0,
\end{aligned} \tag{A.11}$$

where the inequality in the expression above follows from [Assumption 10](#), or more precisely [\(A.8\)](#), which implies that

$$\int_{\underline{v}}^v [F(v'|t) - F(v'|t')] dv' \leq \int_{\underline{v}}^{\bar{v}} [F(v'|t) - F(v'|t')] dv', \tag{A.12}$$

for all v , and the final expression is zero because

$$y(\bar{v}, t) - y(\underline{v}, t) = \int_{\underline{v}}^{\bar{v}} dy(v, t). \tag{A.13}$$

Equations [\(A.9\)](#), [\(A.10\)](#), [\(A.11\)](#) and [\(A.12\)](#) show that under [Assumption 10](#), (\bar{IC}_0) implies that $U(t)$ is monotone decreasing. \square

Proof of Lemma 3. The monotonicity assumption implies that U is differentiable at almost every t . Restricting attention to those points where $U(\cdot)$ is differentiable, note by (\bar{IC}_0) that, for any type t and $h > 0$,

$$\begin{aligned}
U(t) - U(t-h) &\leq U(t) - U(t; t-h) \\
&= \int_{\underline{v}}^{\bar{v}} [vy(v, t) - x(v, t)] (f(v|t) - f(v|t-h)) dv,
\end{aligned}$$

from which it follows that

$$\begin{aligned}
\lim_{h \downarrow 0} \frac{U(t) - U(t-h)}{h} &\leq \lim_{h \downarrow 0} \frac{1}{h} \int_{\underline{v}}^{\bar{v}} [vy(v, t) - x(v, t)] (f(v|t) - f(v|t-h)) dv \\
&= \int_{\underline{v}}^{\bar{v}} u(v, t) \frac{\partial f(v|t)}{\partial t} dv \\
&= - \int_{\underline{v}}^{\bar{v}} y(v, t) \frac{\partial F(v|t)}{\partial t} dv
\end{aligned}$$

where the last equality follows from [Lemma 2](#) and integration by parts. \square

Appendix B. Proofs of the results in Section 6

Proof of Lemma 6. To prove that $U'(t) \geq 0$ for almost every t , we argue by contradiction. Suppose U is an expected utility function resulting from an optimal mechanism and there exists some interval (τ_1, τ_2) such that $U'(t) < 0$ for $t \in (\tau_1, \tau_2)$. We show that types $t \in (\tau_1, \tau_2)$ strictly prefer their own contract to those of any other type. That is, the constraint (IC_0) does not bind for those types. We first prove that for $t \in (\tau_1, \tau_2)$,

$$U(t) > U(t'; t) = \mathbb{E}_t[u(\mathbb{E}_t[v], t'; v, t)] \text{ for all } t' < t.$$

Suppose there exist $t \in (\tau_1, \tau_2)$ and $t' < t$ such that $U(t) = \mathbb{E}_t[u(\mathbb{E}_t[v], t'; v, t)]$. Then, for $\varepsilon > 0$ small enough, we obtain

$$\begin{aligned} U(t + \varepsilon) &< U(t) = \mathbb{E}_t[u(\mathbb{E}_t[v], t'; v, t)] \leq \mathbb{E}_{t+\varepsilon}[u(\mathbb{E}_{t+\varepsilon}[v], t'; v, t + \varepsilon)] \\ &= U(t'; t + \varepsilon), \end{aligned} \quad (B.1)$$

implying that (IC_0) constraint is violated for type $t + \varepsilon$, which contradicts the supposition that U is an expected utility function resulting from an optimal mechanism. Note that the strict inequality in (B.1) is true since $U'(t) < 0$ for $t \in (\tau_1, \tau_2)$. The weak inequality follows from Assumption 6 and the fact that $u(\cdot, t'; v, t)$ is non-decreasing from Lemma 2. Hence, for $t \in (\tau_1, \tau_2)$, $U(t) > U(t'; t)$ for all $t' < t$.

Similarly, we prove that $U(t) > U(t'; t)$ for all $t \in (\tau_1, \tau_2)$ and $t' > t$. Suppose there exist $t \in (\tau_1, \tau_2)$ and $t' > t$ such that $U(t) = \mathbb{E}_t[u(v, t'; v, t)]$. Then, for $\varepsilon > 0$ small enough, we obtain

$$U(t + \varepsilon) < U(t) = \mathbb{E}_t[u(v, t'; v, t)] \leq \mathbb{E}_{t+\varepsilon}[u(v, t'; v, t + \varepsilon)] = U(t'; t + \varepsilon),$$

since $U'(t) < 0$ for $t \in (\tau_1, \tau_2)$, Assumption 6 is satisfied and $u(\cdot, t'; v, t)$ is non-decreasing, contradicting U being an expected utility function resulting from an optimal mechanism. Hence, for $t \in (\tau_1, \tau_2)$, $U(t) > U(t'; t)$ for all $t' > t$.

Then we can decrease $U(t)$ slightly over the interval (τ_1, τ_2) by increasing the payments and leaving the allocation unchanged such that (IC_0) and (\bar{IC}_0) constraints are still satisfied for types (τ_1, τ_2) . This modification also discourages types $[0, T] \setminus (\tau_1, \tau_2)$ from imitating types $t \in (\tau_1, \tau_2)$ since the payments made by types (τ_1, τ_2) are inflated. Hence, this modification of the expected utility function is not only feasible (i.e. satisfies all IC and IR constraints) but also strictly improves the objective. Contradiction to U being a utility function resulting from an optimal mechanism. Hence, $U'(t) \geq 0$ for almost every t . This combined with Lemma 3 gives (17). \square

Preparation for Proof of Proposition 4. Let G_0 , G_1 , and G_2 be linear functionals of y , and let \mathcal{N} be a convex set of functions, that characterize the following primal optimization problem:

$$b^* = \min_{y \in \mathcal{N}} G_0(y) \text{ subject to } G_i(y) \leq 0, \text{ for } i = 1, 2. \quad (P_P)$$

Define the Lagrangian $L(y, \lambda)$ for $y \in \mathcal{N}$ and $\lambda \in \mathbb{R}^2$ as:

$$L(y, \lambda) = G_0(y) + \lambda_1 G_1(y) + \lambda_2 G_2(y), \quad (B.2)$$

and let

$$g(\lambda) = \inf_{y \in \mathcal{N}} L(y, \lambda). \quad (B.3)$$

The dual optimization problem of (P_P) is defined as follows:

$$d^* = \sup_{\lambda} g(\lambda) \text{ subject to } \lambda_i \geq 0, \text{ for } i = 1, 2. \quad (P_D)$$

Assumption 11 (Slater's condition). *There exists $\tilde{y} \in \mathcal{N}$ such that $G_i(\tilde{y}) < 0$ for $i = 1, 2$.*

The following proposition is used in the proof of Proposition 4. Its proof is in Appendix C.

Proposition 8. *Under Assumption 11 (Slater's condition), there exist multipliers $\lambda^* \geq 0$ such that y^* is optimal for (P_P) if and only if y^* minimizes $L(y, \lambda^*)$, and $\sum_{i=1}^2 \lambda_i^* G_i(y) = 0$.*

Proof of Proposition 4. Let U_* , y_* (and the associated u_*) be an optimal solution to (P^3) with the corresponding ex-post utility function denoted by u_* . We will use this solution, which need not be deterministic (i.e., $y \in \{0, 1\}$) to demonstrate the existence of an alternative deterministic allocation, y , for which U_* , y is a solution to (P^3) . In particular, the new solution will keep the expected utility function U_* unchanged, while modifying only the allocation probabilities. This will demonstrate that for each t , there exists a cut-off point $r(t)$ such that the modified allocation probabilities satisfy $y(v, t) = 1$ if $v \geq r(t)$ and $y(v, t) = 0$ if $v < r(t)$.

Without loss of generality, we assume y_* is not almost everywhere equal to zero and y_* is not almost everywhere equal to one (otherwise y_* is deterministic except at one end point, since y_* is non-decreasing, which implies that y_* is already a deterministic solution). This implies that $y_*(v, t) \in (0, 1)$ on a subset of $[\underline{v}, \bar{v}]$ that has positive measure.

Define an alternative allocation $y(v)$ as the solution to the following problem, denoted by (P_A) , and note that for notational brevity, we suppress the dependence of y and u on the type, t :

Choose the control $y(v)$ and u_0 , and hence the ex-post utility $u(v)$ to solve

$$\max_{y \in \mathcal{N}, u_0} \int_{\underline{v}}^{\bar{v}} (v - c) f(v|t) y(v) dv \quad (P_A)$$

subject to

$$u(\underline{v}) = u_0,$$

$$\dot{u}(v) = y(v), \forall v,$$

$$\int_{\underline{v}}^{\bar{v}} f(v|t) u(v) dv \geq U_*(t),$$

$$\int_{\underline{v}}^{\bar{v}} \left[\frac{-\partial F(v|t)}{\partial t} \right] y(v) dv \geq \int_{\underline{v}}^{\bar{v}} \left[\frac{-\partial F(v|t)}{\partial t} \right] y_*(v) dv,$$

$$u(\bar{v}) \leq u_*(\bar{v}),$$

where $D = \{y : y \text{ is nondecreasing, and } 0 \leq y(v) \leq 1, \forall v\}$, and where $U_*(t)$ and $u_*(\bar{v}, t)$ are taken as constants. Note that the initial condition u_0 is a decision variable, and hence, it is a “free” variable. It follows from integration by parts that

$$\int_{\underline{v}}^{\bar{v}} f(v|t) u(v) dv = u(\bar{v}) - \int_{\underline{v}}^{\bar{v}} F(v|t) y(v) dv,$$

so the third constraint can be rewritten as

$$-\int_{\underline{v}}^{\bar{v}} F(v|t) y(v) dv \geq U_*(t) - u(\bar{v}). \quad (\text{B.4})$$

It is crucial to observe that because u_0 is a free variable, we can combine the second, third, and fifth constraints in (P_A) and replace them with the following constraint

$$-\int_{\underline{v}}^{\bar{v}} F(v|t) y(v) dv \geq U_*(t) - u_*(\bar{v}), \quad (\text{B.5})$$

(P_A) can be written as

$$\min_{y \in \mathcal{N}} G_0(y) \text{ subject to } G_i(y) \leq 0, \text{ for } i = 1, 2 \quad (\tilde{\text{P}}_A)$$

where

$$G_0(y) = -\int_{\underline{v}}^{\bar{v}} (v - c) f(v|t) y(v) dv,$$

$$G_1(y) = \int_{\underline{v}}^{\bar{v}} F(v|t) y(v) dv + U_*(t) - u_*(\bar{v}),$$

and

$$G_2(y) = \int_{\underline{v}}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} y(v) dv - \int_{\underline{v}}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} y_*(v) dv.$$

Note that G_0 , G_1 , and G_2 are linear functionals, and that \mathcal{N} is convex. Also note that [Assumption 11](#) (Slater's condition) is satisfied for $y = 0$. This is because y_* is not identically zero, so the second term in $G_2(y)$ is strictly negative for all y , and for $y = 0$ the first term is equal to zero, so $G_2(y) < 0$. And because y_* is not identically equal to zero and y_* is not identically equal to one, $U_* < u_*$, so $G_1(y) < 0$ for $y = 0$.

So for every t , either the solution is identically zero, or identically one, or the solution solves $(\tilde{\text{P}}_A)$. But by [Proposition 8](#), any solution to $(\tilde{\text{P}}_A)$ must also minimize $G_0(y) + \lambda_1^* G_1(y) + \lambda_2^* G_2(y)$ and satisfy $\sum_{i=1}^2 \lambda_i^* G_i(y) = 0$ for some λ^* .

So if y is not identically zero or identically one, then there exist multipliers $\lambda_1, \lambda_2 \geq 0$ such that any solution $y(v)$ solves

$$\max_{y \in \mathcal{N}} \int_{\underline{v}}^{\bar{v}} \left((v - c) f(v|t) - \lambda_1 F(v|t) - \lambda_2 \frac{\partial F(v|t)}{\partial t} \right) y(v) dv, \quad (\text{B.6})$$

but this clearly has a deterministic solution. As observed in Section 5, the firm is maximizing a linear objective over a closed convex subset of a linear space of functions, and the maximal value must be achieved at an extreme point of the feasible set. Here the extreme point in the set of nondecreasing functions from $\mathbb{R} \rightarrow [0, 1]$ is a function that jumps from 0 to 1 at one point, that is a deterministic solution. So for each t , there exists a cutoff point $r(t)$ such that $y(v, t) = 1$ if $v \geq r(t)$ and $y(v, t) = 0$ otherwise (i.e., $v < r(t)$).

As argued above, these allocation probabilities, y , also constitute an optimal solution to (P_A) . Also note that for each solution to (P_A) , since the initial value, u_0 , is free, we can decrease u_0 and make the constraint (B.4) bind. Hence, without loss of generality we will consider only solutions in which (B.4) binds.

For type t , the modified solution will have $U_*(t)$ as the expected utility and $y(v, t)$ as the allocation. (The modified ex-post utility function $u(v, t)$ is also derived from the above optimal control problem.) This modified solution is clearly of the desired form and weakly improves the objective of (P^3) . To establish that it is indeed an optimal solution to (P^3) , we only need to check the constraint (IC_0^T) . To check this, note that $u(\bar{v}, t) \leq u_*(\bar{v}, t)$ (for all t) by the last constraint of (P_A) and that $\partial u(v, t) / \partial v = 1$ for v such that $u(v, t) \geq 0$, where the latter assertion follows since $u(\underline{v}, t) \leq 0$ for all t . To see why $u(\underline{v}, t) \leq 0$ for all t , notice that if $u(\underline{v}) > 0$ and the constraint that

$$\int_{\underline{v}}^{\bar{v}} f(v|t) u(v) dv \geq U_*(t)$$

does not bind in problem (P_A) , we can decrease $u(\underline{v})$ and increase the objective of the original mechanism design problem. If $u(\underline{v}) > 0$ and the constraint that

$$\int_{\underline{v}}^{\bar{v}} f(v|t) u(v) dv \geq U_*(t)$$

binds, then it should be that $U_*(t) > 0$. Then, we should have $U_*(t') > 0$ for all $t' < t$ as any type $t' < t$ could get a strictly positive surplus by pretending to be type t . As U_* is increasing, this would imply that $U_*(t) > 0$ for all t , in which case decreasing $u(\underline{v}, t)$ uniformly for all types would increase the profits, which contradicts $u(\underline{v}, t) > 0$.

Since $u(\underline{v}, t) \leq 0$ for all t , it must be that $u(v, t) \leq u_*(v, t)$ for all v such that $u(v, t) \geq 0$. The constraint (IC_0^T) can be rewritten in a more transparent format as follows:

$$U_*(T) \geq \max_t \{u(\mathbb{E}_T[v], t)\}, \quad (\text{B.7})$$

where we obtained (B.7) by rewriting the constraint (IC_0^T) and using the fact that type T reports his valuation as $\mathbb{E}_T[v]$ and gets a utility of $u(\mathbb{E}_T[v], t)$ if he pretends to be type t . Then (IC_0^T) holds trivially if $u(\mathbb{E}_T[v], t) < 0$ for all t since $U_*(T) \geq 0$. Suppose that there exists $t \in [0, T]$ such that $u(\mathbb{E}_T[v], t) \geq 0$. The constraint (IC_0^T) still holds since

$$U_*(T) \geq \max_t \{u_*(\mathbb{E}_T[v], t)\}, \quad (\text{B.8})$$

$$\geq \max_{t \in \{\tau: u_*(\mathbb{E}_T[v], \tau) \geq 0\}} \{u_*(\mathbb{E}_T[v], t)\}, \quad (\text{B.9})$$

$$\geq \max_{t \in \{\tau: u_*(\mathbb{E}_T[v], \tau) \geq 0\}} \{u(\mathbb{E}_T[v], t)\}, \quad (\text{B.10})$$

$$\geq \max_{t \in \{\tau: u(\mathbb{E}_T[v], \tau) \geq 0\}} \{u(\mathbb{E}_T[v], t)\}, \quad (\text{B.11})$$

$$= \max_t \{u(\mathbb{E}_T[v], t)\}, \quad (\text{B.12})$$

where the inequality in the second line is true since maximization is carried out on a smaller set than the first line. The inequalities in (B.10) and (B.11) follow from the fact that $u(v, t) \leq u_*(v, t)$ for all v such that $u(v, t) \geq 0$. Finally, (B.12) is true since type T would not find it profitable to deviate to any type t such that $u(\mathbb{E}_T[v], t) < 0$. This proves that the modified solution satisfies $(\underline{\text{IC}}_0^T)$ and completes the first part of the proof.

For the remainder of the proof, we will use U, y to denote the optimal solution of (P^3) where $y(v, t) \in \{0, 1\}$ for almost every v, t . To establish the second part of the proof first note that $U(0) = 0$; otherwise we can decrease U uniformly over $[0, T]$ and the objective improves. Next, we prove that we must have

$$U'(t) \in \left\{ 0, - \int_{r(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv \right\} \text{ for almost every } t. \quad (\text{B.13})$$

Lemma 3 and Lemma 6 and the first half of the proof imply that

$$0 \leq U'(t) \leq - \int_{r(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv \text{ for almost every } t.$$

Note that using integration by parts and recalling that $U(0) = 0$, we can rewrite the objective function as follows:

$$\begin{aligned} & \int_{v,t} f(v, t) [x(v, t) - cy(v, t)] dv dt \\ &= \int_{v,t} f(v, t) (v - c) y(v, t) dv dt - \int_0^{\bar{t}} (1 - H(t)) dU(t) \end{aligned} \quad (\text{B.14})$$

To prove (B.13), we argue by contradiction. Suppose that instead there exists some interval $[\tau_1, \tau_2]$ such that

$$0 < U'(t) < - \int_{r(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv \text{ for } t \in [\tau_1, \tau_2].$$

Then the objective function can be improved by replacing U with \hat{U} which is the same except that $\hat{U}'(t) = 0$ for $t \in [\tau_1, \tau_1 + \varepsilon]$ for $\varepsilon > 0$ sufficiently small and appropriately increasing $\hat{U}'(t)$ (which can be done so that $\hat{U}(\tau_2) = U(\tau_2)$ since $U'(t) < - \int_{r(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv$ for $t \in [\tau_1, \tau_2]$). So $\int_{\tau_1}^{\tau_2} d\hat{U}(t) = \int_{\tau_1}^{\tau_2} dU(t)$. This modification improves the objective in (B.14) $\int_{\tau_1}^{\tau_2} (1 - H(t)) d\hat{U}(t) = (1 - H(\tau_2))(U(\tau_2) - U(\tau_1)) < \int_{\tau_1}^{\tau_2} (1 - H(t)) dU(t)$, which is because $1 - H(t)$ is decreasing. And it can be achieved by changing $u(v, t)$ on the interval $[\tau_1, \tau_2]$ appropriately so that $y(v, t)$ remains the same. Then, the constraints of (P^3) including $(\underline{\text{IC}}_0^T)$ still hold since $U(T)$ remains unchanged and $U(t)$ strictly decreases for $t \in (\tau_1, \tau_2)$, while $U(t)$ does not

change elsewhere, and hence type T finds the deviation to types (τ_1, τ_2) strictly less profitable. Thus, (B.13) follows.

Finally, we show that if $U'(t) > 0$ for almost every $t \in [t', t' + \varepsilon_1]$ for some t' and $\varepsilon_1 > 0$, then $U'(t) > 0$ for almost every $t \geq t'$. Suppose not, i.e., there exists an interval $[\tau_1, \tau_2] \subset [t', T]$ over which $U'(t) = 0$. Then, following the reasoning in the preceding paragraph, we can modify U such that $U'(t) = 0$ for $t \in [t', t' + \varepsilon_2]$ for some $0 < \varepsilon_2 < \varepsilon_1$ while keeping $U(\tau_2)$ the same as before, which improves the objective function and hence, leads to a contradiction.

Define $\tau \in [0, T]$ as the essential infimum of types for which $U'(t) > 0$. Formally,

$$\tau = \inf \{t' \in [0, T] : U'(t) > 0 \text{ for almost every } t \in [t', t' + \varepsilon) \text{ for some } \varepsilon > 0\},$$

where $\tau = T$ if $U'(t) = 0$ for almost every t and $\tau = 0$ if $U'(t) > 0$ for almost every t . Then, we have $U'(t) = 0$ for almost every $t \leq \tau$, and

$$U'(t) = - \int_{r(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv \text{ for almost every } t > \tau.$$

The result follows from this since U is Lipschitz continuous (and hence absolutely continuous). \square

Proof of Proposition 5. It follows from Lemma 2 and Proposition 4 that

$$y(v, t) = \begin{cases} 1 & \text{if } v \geq r(t), \\ 0 & \text{if } v < r(t), \end{cases} \text{ and } \frac{\partial u(v, t)}{\partial t} = \begin{cases} 1 & \text{if } v \geq r(t), \\ 0 & \text{if } v < r(t), \end{cases} \quad (\text{B.15})$$

from which it follows that

$$x(v, t) = \begin{cases} x(\bar{v}, t) & \text{if } v \geq r(t), \\ x(\underline{v}, t) & \text{if } v < r(t), \end{cases} \quad (\text{B.16})$$

where $x(\underline{v}, t) = -u(\underline{v}, t)$ and $x(\bar{v}, t) = \bar{v} - u(\bar{v}, t)$. Moreover, it follows from (B.15) that $u(\bar{v}, t) - u(\underline{v}, t) = \bar{v} - r(t)$. Thus, $x(\bar{v}, t) = x(\underline{v}, t) + r(t)$. In other words, defining $p(t) = x(\bar{v}, t)$, we have

$$x(v, t) = \begin{cases} p(t) & \text{if } v \geq r(t), \\ p(t) - r(t) & \text{if } v < r(t). \end{cases} \quad (\text{B.17})$$

Then, for all t ,

$$U(t) = \int_{\underline{v}}^{\bar{v}} u(v, t) f(v|t) dv = \int_{r(t)}^{\bar{v}} v f(v|t) dv - p(t) + r(t) F(r(t)|t).$$

Solving for $p(t)$, we obtain

$$p(t) = \mathbb{E}_t[\max\{v, r(t)\}] - U(t). \quad (\text{B.18})$$

Recall that $U(t; T) = \mathbb{E}_t[u(\mathbb{E}_T[v], t; v, T)]$. That is, $U(t; T) = \mathbb{E}_T[v] y(\mathbb{E}_T[v], t) - x(\mathbb{E}_T[v], t)$. It follows from equations (B.17) and (B.18) that

$$x(\mathbb{E}_T[v], t) = \mathbb{E}_t[\max\{v, r(t)\}] - U(t) - r(t) \mathbf{1}_{\{\mathbb{E}_T[v] - r(t) < 0\}},$$

and clearly (B.15) implies $y(\mathbb{E}_T[v], t) = \mathbf{1}_{\{\mathbb{E}_T[v] - r(t) \geq 0\}}$, so it follows that

$$U(t; T) = (\mathbb{E}_T[v] - r(t)) \mathbf{1}_{\{\mathbb{E}_T[v] - r(t) \geq 0\}} + U(t) - \int_{r(t)}^{\bar{v}} (v - r(t)) f(v|t) dv. \quad \square$$

Preparation for Proof of Proposition 6. The following lemma facilitates the proof of Proposition 6.

Lemma 7. If $r(t)$ is decreasing over $[\tau, T]$, then (\mathcal{IC}_0^T) is slack for all $t' \in [\tau, T)$.

Proof of Lemma 7. To see this, recall that using Proposition 4 to define $U(T)$ and $U(t')$, (\mathcal{IC}_0^T) can be written as

$$\begin{aligned} & - \int_{\tau}^T \int_{r(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv dt - \mathbb{E}_T[v] + r(t') + \int_{\tau}^{\hat{t}} \int_{r(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv dt \\ & + \int_{r(t')}^{\bar{v}} (v - r(t')) f(v|t') dv \geq 0, \end{aligned} \quad (\text{B.19})$$

for all $\hat{t} \in [\tau, T)$. The left hand side can be rewritten as

$$- \int_{\hat{t}}^T \int_{r(\hat{t})}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv dt - \int_{\underline{v}}^{\bar{v}} (v - r(\hat{t})) f(v|T) dv + \int_{r(\hat{t})}^{\bar{v}} (v - r(\hat{t})) f(v|\hat{t}) dv$$

or

$$- \int_{\hat{t}}^T \int_{r(\hat{t})}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv dt - \int_{r(\hat{t})}^{\bar{v}} (v - r(\hat{t})) (f(v|T) - f(v|\hat{t})) dv - \int_{\underline{v}}^{r(\hat{t})} (v - r(\hat{t})) f(v|T) dv.$$

Integrating the second term by parts, this becomes

$$- \int_{\hat{t}}^T \int_{r(\hat{t})}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv dt + \int_{r(\hat{t})}^{\bar{v}} [F(v|T) - F(v|\hat{t})] dv - \int_{\underline{v}}^{r(\hat{t})} (v - r(\hat{t})) f(v|T) dv$$

or

$$- \int_{\hat{t}}^T \int_{r(\hat{t})}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv dt + \int_{\hat{t}}^T \int_{r(\hat{t})}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv dt - \int_{\underline{v}}^{r(\hat{t})} (v - r(\hat{t})) f(v|T) dv > 0,$$

where the first two terms on the right are positive since $r(t) \leq r(\hat{t})$ for all $t \in (\hat{t}, T]$, and the third term is strictly positive since the integrand is non-positive, so the constraint, equation (B.19), is slack for all $\hat{t} \in [\tau, T)$. \square

Proof of Proposition 6. We further relax the firm's problem, (P^4) , by assuming that the constraint does not bind for all $t \in [\tau, T]$, so $U(t) = 0$ by Proposition 4. We solve the relaxed problem and show that the constraint binds for all $t \in [0, \sigma] \subset [0, \tau]$ for some $\sigma \leq \tau$. We then characterize the optimal function $r(t)$ on the three intervals, $[0, \sigma]$, $[\sigma, \tau]$, and $[\tau, T]$, as well as the optimal τ , and prove that $0 < \tau < T$. Finally, we verify that relaxing the constraint was without loss of generality, that is, we verify that $r(t)$ is decreasing on $[\tau, T]$, which by Lemma 7 implies the constraint does not bind.

It is useful to divide the support $[0, T]$ into two intervals and give $r(t)$ a different name, r_1 and r_2 , on $[0, \tau]$ and on $[\tau, T]$ respectively.

So using Proposition 4, $U(t) = 0$ for $t \in [0, \tau]$ and the firm's problem can be written as

$$\max_{\{\tau \in [0, T], r_1(t), r_2(t)\}} \int_0^\tau \int_{r_1(t)}^{\bar{v}} f(v, t)(v - c)dvdt + \int_\tau^T \int_{r_2(t)}^{\bar{v}} f(v, t)\phi(v, t)dvdt \quad (B.20)$$

subject to

$$-\int_\tau^T \int_{r_2(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dvdt \geq \mathbb{E}_T[v] - r_1(t') - \int_{r_1(t')}^{\bar{v}} (v - r_1(t'))f(v|t')dv, \forall t' \in [0, \tau],$$

where

$$\phi(v, t) = v - c + \frac{1 - H(t)}{h(t)} \frac{\partial F(v|t)/\partial t}{f(v|t)}.$$

Note that relaxing the constraint is without loss of generality by Lemma 7 as long as the unconstrained solution satisfies the condition that $r_2^*(t)$ is decreasing.

Given any τ and r_2 , unconstrained maximization implies that $r_1(t) = c$. We now show that there exists a cutoff $\sigma \in [0, \tau]$ such that the constraint binds if and only if $t \in [0, \sigma]$. This cutoff, $\sigma^*(\tau, r_2)$, is defined by

$$-\int_\tau^T \int_{r_2(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dvdt = \mathbb{E}_T[v] - \int_c^{\bar{v}} (v - c)f(v|\sigma^*(\tau, r_2))dv - c,$$

when a solution exists in the interval $[0, \tau]$, and defined by $\sigma^*(\tau, r_2) = 0$ when

$$-\int_\tau^T \int_{r_2(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dvdt > \mathbb{E}_T[v] - \int_c^{\bar{v}} (v - c)f(v|0)dv - c,$$

and $\sigma^*(\tau, r_2) = \tau$ when

$$-\int_\tau^T \int_{r_2(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dvdt < \mathbb{E}_T[v] - \int_c^{\bar{v}} (v - c)f(v|\tau)dv - c.$$

Clearly $\sigma^*(\tau, r_2)$ is unique, given τ and r_2 , since Assumption 6 (Forward FOSD) implies that $\int_c^{\bar{v}} (v - c)f(v|t)dv = \bar{v} - c - \int_c^{\bar{v}} F(v|t)dv$ is strictly increasing in t . It follows that for all $t \in (\sigma^*(\tau, r_2), \tau)$, (IC_0^T) does not bind and that $r_1(t) = r_1^*(t; \sigma^*(\tau, r_2)) = c$. Moreover, for all

$t \leq \sigma^*(\tau, r_2)$, $(\underline{c}, \underline{0})$ binds and $r_1(t) = r_1^*(t; \sigma^*(\tau, r_2))$ is uniquely defined by

$$\begin{aligned} \mathbb{E}_T[v] - \int_{r_1^*(t; \sigma^*(\tau, r_2))}^{\bar{v}} (v - r_1^*(t; \sigma^*(\tau, r_2))) f(v|t) dv - r_1^*(t; \sigma^*(\tau, r_2)) \\ = \mathbb{E}_T[v] - \int_c^{\bar{v}} (v - c) f(v|\sigma^*(\tau, r_2)) dv - c, \end{aligned} \quad (\text{B.21})$$

since $\int_x^{\bar{v}} (v - x) f(v|t) + x$ is strictly increasing in x .

Using the implicit function theorem and the definition of $\sigma^*(\tau, r_2)$, we have

$$\begin{aligned} \frac{\partial \sigma^*(\tau, r_2 + \epsilon z)}{\partial \epsilon} \Big|_{\epsilon=0} \\ = - \left[\int_c^{\bar{v}} (v - c) \frac{\partial f(v|\sigma^*(\tau, r_2))}{\partial t} dv \right]^{-1} \int_{\tau}^T \frac{\partial F(r_2(t)|t)}{\partial t} z(t) dt, \end{aligned} \quad (\text{B.22})$$

and

$$\frac{\partial \sigma^*(\tau, r_2)}{\partial \tau} = - \left[\int_c^{\bar{v}} (v - c) \frac{\partial f(v|\sigma^*(\tau, r_2))}{\partial t} dv \right]^{-1} \int_{r_2(\tau)}^T \frac{\partial F(v|\tau)}{\partial t} dv > 0, \quad (\text{B.23})$$

for all τ and r_2 such that $\sigma^*(\tau, r_2) \in (0, \tau)$. When $\sigma^*(\tau, r_2) = 0$, then both derivatives are zero (i.e., $\partial \sigma^*(\tau, r_2)/\partial \tau = 0$ and $\partial \sigma^*(\tau, r_2 + \epsilon z)/\partial \epsilon|_{\epsilon=0} = 0$). And when $\sigma^*(\tau, r_2) = \tau$, then $\partial \sigma^*(\tau, r_2)/\partial \tau = 1$ and $\partial \sigma^*(\tau, r_2 + \epsilon z)/\partial \epsilon|_{\epsilon=0} = 0$.

Using the implicit function theorem and the definition of $r_1^*(t; \sigma^*(\tau, r_2))$ in (B.21) we have

$$\frac{\partial r_1^*(t; \sigma)}{\partial \sigma} = \frac{\int_c^{\bar{v}} (v - c) \frac{\partial f(v|\sigma)}{\partial t} dv}{F(r_1^*(t; \sigma)|t)}$$

which is clearly positive, and

$$\frac{\partial r_1^*(t; \sigma)}{\partial t} = - \frac{\int_{r_1^*(t; \sigma)}^{\bar{v}} (v - r_1^*(t; \sigma)) \frac{\partial f(v|t)}{\partial t} dv}{F(r_1^*(t; \sigma)|t)}$$

which is clearly negative, so $r_1^*(t)$ is strictly decreasing on $[0, \sigma]$, and $r_1^*(t) = c$ on $[\sigma, \tau]$, so $r_1^*(t)$ is weakly decreasing on $[0, \tau]$.

Substituting $r_1^*(t; \sigma^*(\tau, r_2))$ into the objective function, the firm's problem can be written as the following unconstrained calculus of variations problem:

$$\begin{aligned} \max_{\{\tau, r_2(t)\}} \int_0^{\sigma^*(\tau, r_2)} \int_{r_1^*(t; \sigma^*(\tau, r_2))}^{\bar{v}} f(v, t)(v - c) dv dt + \int_{\sigma^*(\tau, r_2)}^{\tau} \int_c^{\bar{v}} f(v, t)(v - c) dv dt \\ + \int_{\tau}^T \int_{r_2(t)}^{\bar{v}} f(v, t) \phi(v, t) dv dt. \end{aligned} \quad (\text{B.24})$$

Using the fact that $r_1^*(\sigma^*(\tau, r_2); \sigma^*(\tau, r_2)) = c$, the first-order conditions are

$$\begin{aligned} & - \int_0^{\sigma^*(\tau, r_2)} f(r_1^*(t'; \sigma^*(\tau, r_2)), t')(r_1^*(t'; \sigma^*(\tau, r_2)) - c) \\ & \quad \times \frac{\partial r_1^*(t'; \sigma^*(\tau, r_2))}{\partial \sigma} \frac{\partial \sigma^*(\tau, r_2 + \epsilon z)}{\partial \epsilon} \Big|_{\epsilon=0} dt' \\ & - \int_{\tau}^T f(r_2(t), t) \phi(r_2(t), t) z(t) dt = 0, \forall z(t), \end{aligned}$$

and

$$\begin{aligned} & - \int_0^{\sigma^*(\tau, r_2)} f(r_1^*(t'; \sigma^*(\tau, r_2)), t')(r_1^*(t'; \sigma^*(\tau, r_2)) - c) \frac{\partial r_1^*(t'; \sigma^*(\tau, r_2))}{\partial \sigma} \frac{\partial \sigma^*(\tau, r_2)}{\partial \tau} dt' \\ & + \int_c^{\bar{v}} (v - c) f(v, \tau) dv - \int_{r_2(\tau)}^{\bar{v}} \phi(v, \tau) f(v, \tau) dv = 0. \end{aligned}$$

Note that the first first-order condition implies that if $\sigma^*(\tau, r_2) = 0$, the $\tau = 0$, so $r(t) = c$ and there is no distortion, and the firm extracts all of the surplus, which contradicts [Assumption 7](#). So $\sigma^*(\tau, r_2) > 0$ by [Assumption 7](#).

The first-order conditions imply that $0 < \tau^* < T$. To see this, first notice that $\tau^* = 0$ implies that $\sigma^*(\tau, r_2) = 0$, which we just showed implies $\tau = T$ which is a contradiction. Second, notice that $\tau^* = T$ implies $r_2(\tau) = c$ and $1 - H(\tau) = 0$, so the 2nd and 3rd terms in second condition are zero and hence the second first-order condition is strictly negative, unless $\sigma^*(\tau, r_2) = 0$, which again is impossible under [Assumption 7](#), so clearly $0 < \tau^* < T$.

So using [\(B.22\)](#) and [\(B.23\)](#) the first-order conditions can be written as:

$$\Psi(\tau, r_2) \int_{\tau}^T \frac{\partial F(r_2(t)|t)}{\partial t} z(t) dt - \int_{\tau}^T f(r_2(t), t) \phi(r_2(t), t) z(t) dt = 0, \forall z(t),$$

and

$$\Psi(\tau, r_2) \int_{r_2(\tau)}^{\bar{v}} \frac{\partial F(v|\tau)}{\partial t} dv + \int_c^{\bar{v}} (v - c) f(v, \tau) dv - \int_{r_2(\tau)}^{\bar{v}} \phi(v, \tau) f(v, \tau) dv = 0,$$

where

$$\Psi(\tau, r_2) = \int_0^{\sigma^*(\tau, r_2)} \frac{f(r_1^*(t'; \sigma^*(\tau, r_2)), t')(r_1^*(t'; \sigma^*(\tau, r_2)) - c)}{F(r_1^*(t'; \sigma^*(\tau, r_2))|t')} dt' > 0.$$

Substituting for $\phi(v, t)$, we can rewrite the first first-order-condition as

$$- \int_{\tau}^T f(r_2(t), t) \left[r_2(t) - c + \frac{1 - H(t) - \Psi(\tau, r_2) \frac{\partial F(r_2(t)|t)}{\partial t}}{h(t)} \frac{\partial F(r_2(t)|t)}{f(r_2(t)|t)} \right] z(t) dt = 0, \forall z(t),$$

which, using $f(v, t) = f(v|t)h(t)$, becomes simply $\Phi(r_2(t), t; \tau, r_2) = 0$ for all $t \in [\tau, T]$ where

$$\Phi(v, t; \tau, r_2) = v - c + (1 - H(t) - \Psi(\tau, r_2)) \frac{\partial F(v|t)/\partial t}{f(v, t)}. \quad (\text{B.25})$$

Substituting for $\phi(v, \tau)$, and using $f(v, t) = f(v|t)h(t)$, we can rewrite the second first-order-condition as

$$\Psi(\tau, r_2) \int_{r_2(\tau)}^{\bar{v}} \frac{\partial F(v|\tau)}{\partial t} dv = \int_{r_2(\tau)}^{\bar{v}} (1 - H(\tau)) \frac{\partial F(v|\tau)}{\partial t} dv + \int_{r_2(\tau)}^c (v - c) f(v, \tau) dv,$$

or

$$1 - H(\tau) - \Psi(\tau, r_2) = - \left[\int_{r_2(\tau)}^{\bar{v}} \frac{\partial F(v|\tau)}{\partial t} dv \right]^{-1} \int_{r_2(\tau)}^c (v - c) f(v, \tau) dv,$$

and since $\int_{r_2(\tau)}^c (v - c) f(v, \tau) dv \leq 0$ and $\partial F(v|\tau)/\partial t < 0$, it follows that $1 - H(\tau) - \Psi(\tau, r_2) \leq 0$. Also, note that $1 - H(\tau) - \Psi(\tau, r_2) \leq 0$ implies $1 - H(t) - \Psi(\tau, r_2) < 0$ for all $t > \tau$. Using (B.25), clearly, $1 - H(\tau) - \Psi(\tau, r_2) \leq 0$ combined with $\Phi(r_2(\tau), \tau; \tau, r_2) = 0$ implies that $r_2^*(\tau) \leq c$.

We now prove that $dr_2^*(t)/dt < 0$. The equation defining $r_2^*(t)$ is $\Phi(r_2(t), t; \tau, r_2) = 0$, so using the implicit function theorem,

$$\frac{dr_2(t)}{dt} = - \frac{\Phi_t(r_2(t), t; \tau, r_2)}{\Phi_v(r_2(t), t; \tau, r_2)} \quad (\text{B.26})$$

and using $f(v, t) = f(v|t)h(t)$ to cancel the $h(t)$ in the second term,

$$\Phi_t(r_2(t), t; \tau, r_2) = (1 - H(t) - \Psi(\tau, r_2)) \frac{\partial \left[\frac{\partial F(v|t)/\partial t}{f(v, t)} \right]}{\partial t} \bigg|_{v=r_2(t)} - \frac{\partial F(r_2(t)|t)/\partial t}{f(r_2(t)|t)},$$

and

$$\Phi_v(r_2(t), t; \tau, r_2) = 1 + (1 - H(t) - \Psi(\tau, r_2)) \frac{\partial \left[\frac{\partial F(v|t)/\partial t}{f(v, t)} \right]}{\partial v} \bigg|_{v=r_2(t)}.$$

Under [Assumption 9](#), the denominator in (B.26) is positive. Under [Assumption 8](#), the first term in the numerator of (B.26) is positive, and under [Assumption 6](#) (FOSD) the second term is positive, so the numerator of (B.26) is positive, and it follows that $dr_2^*(t)/dt < 0$, and more generally that $r(t)$ is decreasing on $[0, T]$, so from [Lemma 7](#), it follows that relaxing the incentive constraint was without loss of generality.

Finally, it must be true that $r_2(\tau) = c$. We prove this by contradiction. Suppose instead that $r_2(\tau) < c$. Then there exists an ϵ , and a new mechanism $\hat{p}(t)$, $\hat{r}_1(t)$ and $\hat{r}_2(t)$, which satisfies $\hat{r}_1(t) = r_1(t)$, $\forall t \in [0, \tau]$, $\hat{r}_2(t) = r_2(t)$, $\forall t \in [\tau + \epsilon, T]$, and $\hat{r}_2(t) = c$, $\forall t \in [\tau, \tau + \epsilon]$, which earns strictly higher profits. The price of the new mechanism, $\hat{p}(t)$, is chosen so that \hat{U} is continuous, and so that $\hat{U}(\tau + \epsilon) = U(\tau + \epsilon)$, but also so that \hat{U} increases more slowly and over a longer interval than U . Specifically, there exists a $v < \tau$, such that we define $\hat{U}(v) = 0$ and

$$\hat{U}'(t) = - \int_c^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv \quad (\text{B.27})$$

for all $t \in [\nu, \tau + \epsilon]$, while $U(t) = 0$ for $t \in [\nu, \tau]$ and

$$U'(t) = - \int_{r_1(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv, \quad (\text{B.28})$$

so $\hat{U}'(t) < U'(t)$ for $t \in [\tau, \tau + \epsilon]$. This implies that incentive compatibility is unchanged. Consumer surplus increases by

$$\int_{\nu}^{\tau+\epsilon} (\hat{U}(t) - U(t)) h(t) dt, \quad (\text{B.29})$$

but total surplus increases by

$$\int_{\tau}^{\tau+\epsilon} \int_{r_2(t)}^c (c - v) f(v, t) dv dt. \quad (\text{B.30})$$

Since $\nu \rightarrow \tau$, and $\hat{U}(t) - U(t) \rightarrow 0$, as $\epsilon \rightarrow 0$, for ϵ sufficiently small, producer surplus, or profit, must be increasing. \square

Proof of Proposition 7. Since the cutoff points $\{r(t) : 0 \leq t \leq T\}$ are optimal for (P^4) , which in turn is more relaxed than the original screening problem, it suffices to check that $\{(p(t), r(t)) : 0 \leq t \leq T\}$ satisfy the constraints of the original problem.

Since $U(t) \geq 0$ for all t (see Proposition 6) the individual rationality constraints (IR) are satisfied. The feasibility constraints (F) are readily satisfied. Moreover, from Lemma 2, the incentive compatibility constraint (\tilde{IC}_t) of (P^4) and the incentive compatibility constraint (IC_t) of the original problem are equivalent and hence the constraint (IC_t) of the original problem is satisfied by the payment and allocation scheme given by

$$x(v, t) = \begin{cases} p(t) - r(t) & \text{if } v < r(t), \\ p(t) & \text{if } v \geq r(t), \end{cases} \quad \text{and } y(v, t) = \begin{cases} 0 & \text{if } v < r(t), \\ 1 & \text{if } v \geq r(t). \end{cases}$$

The incentive compatibility constraints (IC_0) regarding downward deviations were ignored for all types but the highest type T in (P^4) . We next show that the menu of contracts $\{(p(t), r(t)) : 0 \leq t \leq T\}$ satisfy (IC_0) of the original screening problem. For $t > t'$, $U(t'; t) \geq 0$ only if $\mathbb{E}_t[v] \geq r(t)$, in which case

$$U(t'; t) = \mathbb{E}_t[v] + U(t') - \int_{r(t')}^{\bar{v}} (v - r(t')) f(v|t') dv - r(t'),$$

and we have

$$\frac{\partial U(t'; t)}{\partial t} = \frac{d\mathbb{E}_t[v]}{dt} = \int_{\underline{v}}^{\bar{v}} v \frac{\partial f(v|t)}{\partial t} dv = - \int_{\underline{v}}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv > - \int_{r(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv \geq U'(t).$$

If $U(t'; T) \leq U(T)$, then

$$U(t'; T) - \int_t^T \frac{\partial U(t'; s)}{\partial s} ds \leq U(T) - \int_t^T U'(s) ds,$$

and $U(t'; t) \leq U(t)$ for all $t > t'$, so (\underline{IC}_0) is satisfied.

Finally, we show that the incentive compatibility constraints (\overline{IC}_0) regarding upward deviations, which were ignored in problems (P^3) and (P^4) , are satisfied. For $t' > t$,

$$U(t'; t) = \int_{r(t')}^{\bar{v}} (v - r(t')) f(v|t) dv - p(t') + r(t').$$

We can rewrite this as

$$\begin{aligned} U(t'; t) &= \int_{r(t')}^{\bar{v}} (v - r(t')) f(v|t) dv + U(t') - \int_{r(t')}^{\bar{v}} (v - r(t')) f(v|t') dv, \\ &= \int_{r(t')}^{\bar{v}} (v - r(t')) [f(v|t) - f(v|t')] dv + U(t'). \end{aligned}$$

So

$$\begin{aligned} \frac{\partial U(t'; t)}{\partial t'} &= U'(t') - r'(t') \int_{r(t')}^{\bar{v}} [f(v|t) - f(v|t')] dv + \int_{r(t')}^{\bar{v}} (v - r(t')) \frac{\partial f(v|t')}{\partial t'} dv \\ &= U'(t') - r'(t') [-F(r(t')|t) + F(r(t')|t')] + \int_{r(t')}^{\bar{v}} \frac{\partial F(v|t')}{\partial t'} dv \\ &\leq 0, \end{aligned}$$

since $r'(t') \leq 0$ and $U'(t') \leq -\int_{r(t')}^{\bar{v}} \frac{\partial F(v|t')}{\partial t'} dv$, and since Assumption 6 holds. Integrating, we get $U(t'; t) \leq U(t)$ for all $t' > t$ since $U(t; t) = U(t)$, which shows that (\overline{IC}_0) is satisfied.

Hence, $\{(p(t), r(t)) : 0 \leq t \leq T\}$ is optimal for (P^1) . \square

Proof of Corollary 1. For all t ,

$$U(t) = \int_{r(t)}^{\bar{v}} v f(v|t) dv - p(t)(1 - F(r(t)|t)) + (p(t) - r(t))F(r(t)|t). \quad (B.31)$$

Solving for $p(t)$ in (B.31) we get for all t ,

$$p(t) = \mathbb{E}_t[v; v \geq r(t)] + r(t)F(r(t)|t) - U(t) = \mathbb{E}_t[\max\{v, r(t)\}] - U(t). \quad (B.32)$$

Since type T is indifferent between her contract and the contract offer all types $t \in [0, \sigma]$ and $U(t; T) = \mathbb{E}_T[v] - \mathbb{E}_t[\max\{v, c\}]$ for $t \leq \sigma$ by Proposition 5, $\mathbb{E}_t[\max\{v, r(t)\}]$ is constant as

a function of t over the interval $[0, \sigma]$. Moreover, $U(t) = 0$ for $t \leq \sigma$ and hence, $p'(t) = 0$ for $t \leq \sigma$. Since $r(t) = c$ with $U(t) = 0$ for $[\sigma, \tau]$, from [Assumption 6](#) (FOSD) we obtain

$$p'(t) = \frac{d}{dt} \mathbb{E}_t [\max \{v, c\}] = - \int_c^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv > 0 \text{ for } t \in [\sigma, \tau].$$

For $t \geq \tau$, differentiating [\(B.32\)](#) and using [Proposition 4](#) we get

$$\begin{aligned} p'(t) &= \frac{d}{dt} \mathbb{E}_t [\max \{v, r(t)\}] - U'(t), \\ &= \int_{r(t)}^{\bar{v}} v \frac{\partial f(v|t)}{\partial t} dv + r'(t) F(r(t)|t) + \int_{r(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv + r(t) \frac{\partial F(r(t)|t)}{\partial t}, \\ &= \int_{r(t)}^{\bar{v}} v \frac{\partial f(v|t)}{\partial t} dv + r'(t) F(r(t)|t) + r(t) \frac{\partial F(r(t)|t)}{\partial t} \\ &\quad - \frac{\partial F(r(t)|t)}{\partial t} r(t) - \int_{r(t)}^{\bar{v}} \frac{\partial f(v|t)}{\partial t} v dv, \\ &= r'(t) F(r(t)|t), \end{aligned}$$

and hence $p'(t) \leq 0$ since $r'(t) \leq 0$ for $t \geq \tau$.

Notice that the effective price $p(t) - r(t) F(r(t)|t)$ is increasing for all t since $r'(t) \leq 0$ for almost every t and

$$\begin{aligned} \frac{d}{dt} [p(t) - r(t) F(r(t)|t)] &= p'(t) - r'(t) F(r(t)|t) - r(t) r'(t) f(r(t)|t) \\ &\quad - r(t) \frac{\partial F(r(t)|t)}{\partial t} \\ &\geq -r(t) r'(t) f(r(t)|t) - r(t) \frac{\partial F(r(t)|t)}{\partial t}. \end{aligned}$$

Next, observe that $p(T) \geq p(\sigma)$. To see this, recall that highest type is indifferent between her contract and that of type σ , thus, $U(T) = \mathbb{E}_T[v] - p(\sigma)$. From [\(B.32\)](#), $U(T) = \mathbb{E}_T[\max \{v, r(T)\}] - p(T)$ and hence,

$$p(T) - p(\sigma) = \mathbb{E}_T[\max \{v, r(T)\}] - \mathbb{E}_T[v] \geq 0,$$

where the inequality is strict if $r(T) > \underline{v}$. Finally, if $r(\tau) < c$, then there is a downward jump in prices at τ , i.e. $p(\tau) < p(\tau-)$. To see this note that, $U(t) = 0$ for $t \leq \tau$ and $r(\tau) < c$ for $\sigma \leq t < \tau$ whereas $r(\tau) < c$, and hence $p(\tau) < p(\tau-)$ since by assumption $\partial f(v|t)/\partial t$ is continuous in v and t . \square

Appendix C. Proof of [Proposition 8](#)

The proof consists of a series of claims.

Claim 1 (Weak duality). *The minimized value of (P_D) is less than or equal to the minimized value of (P_P) , or $d^* \leq b^*$.*

Proof. Consider any \hat{y} that is feasible for (P_P) (that is, satisfies $\hat{y} \in \mathcal{N}$ and $G_i(\hat{y}) \leq 0$ for $i = 1, 2$) and any $\lambda \geq 0$. Then clearly

$$L(\hat{y}, \lambda) \leq G_0(\hat{y}),$$

from which it follows that

$$g(\lambda) = \inf_{y \in \mathcal{N}} L(y, \lambda) \leq L(\hat{y}, \lambda) \leq G_0(\hat{y}).$$

Taking the supremum of $g(\lambda)$ over $\lambda \geq 0$, and infimum of $G_0(\hat{y})$ over $\hat{y} \in \mathcal{N}$ gives the result. \square

Our second claim uses [Assumption 11](#) and the following theorem:

Theorem 1. (See [Rockafellar, 1970](#), Theorem 11.3.) *Let C_1 and C_2 be two convex sets. In order that there exists a hyperplane separating C_1 and C_2 properly, it is necessary and sufficient that the relative interior of C_1 and the relative interior of C_2 have no point in common.*

Note that we apply this theorem using just the interior of C_1 and C_2 and not the relative interior (since the sets have the same dimensionality as the space in which they are contained).

Claim 2 (Strong duality). *The problem (P_P) and its dual (P_D) attain the same optimum, or $b^* = d^*$.*

Proof. Define

$$S = \left\{ (G_1(y), G_2(y), G_0(y)) \in \mathbb{R}^3 \mid y \in \mathcal{N} \right\},$$

and define the closely related set $A \subset \mathbb{R}^3$:

$$A = \left\{ (u_1, u_2, t) \in \mathbb{R}^3 \mid \exists y \in \mathcal{N} \text{ such that } G_i(y) \leq u_i, \text{ for } i = 1, 2, \text{ and } G_0 \leq t \right\}.$$

The set A can be viewed as the epigraph form of S , since A includes all points in S and points that are worse, i.e., those with higher objective values or higher inequality constraint function values.

Note that $b^* = \inf \{t \mid (0, 0, t) \in A\}$. Also note that A is a convex set because $G_i(y)$ for $i = 0, 1, 2$ are linear functionals, so if (u_1, u_2, t) and $(\tilde{u}_1, \tilde{u}_2, \tilde{t})$ are both in A , and if y and \tilde{y} are the associated values from the definition of A , then for all $\alpha \in (0, 1)$, $(\check{u}_1, \check{u}_2, \check{t}) = \alpha(u_1, u_2, t) + (1 - \alpha)(\tilde{u}_1, \tilde{u}_2, \tilde{t})$ is in A , because $G_i(\alpha y + (1 - \alpha)\tilde{y}) \leq \check{u}_i$ for $i = 1, 2$, and $G_0(\alpha y + (1 - \alpha)\tilde{y}) \leq \check{t}$.

Define another set $B = \{(0, 0, s) \in \mathbb{R}^3 \mid s < b^*\}$, and note that B is convex and that $A \cap B = \emptyset$, so that we can apply the separating hyperplane theorem (Theorem 11.3 in [Rockafellar, 1970](#)). This theorem implies that there exists $(\lambda_1, \lambda_2, \mu) \neq 0$ and β such that

$$(u_1, u_2, t) \in A \Rightarrow \tilde{\lambda}_1 u_1 + \tilde{\lambda}_2 u_2 + \mu t \geq \beta, \quad (\text{C.1})$$

and

$$(u_1, u_2, t) \in B \Rightarrow \tilde{\lambda}_1 u_1 + \tilde{\lambda}_2 u_2 + \mu t \leq \beta. \quad (\text{C.2})$$

Equation (C.1) implies $\tilde{\lambda}_1 \geq 0$, $\tilde{\lambda}_2 \geq 0$, and $\mu \geq 0$ since otherwise the values of $\tilde{\lambda}_1 u_1 + \tilde{\lambda}_2 u_2 + \mu t$ are unbounded below, so no β exists satisfying (C.1). Equation (C.2) implies simply that $\mu t \leq \beta$ for all $t \leq b^*$, hence $\mu b^* \leq \alpha$. So equations (C.1) and (C.2) together imply that

$$\tilde{\lambda}_1 G_1(y) + \tilde{\lambda}_2 G_2(y) + \mu G_0(y) \geq \beta \geq \mu b^*. \quad (\text{C.3})$$

Next we rule out $\mu = 0$ by contradiction. Using (C.3), $\mu = 0$ implies that for any $y \in \mathcal{N}$, $\tilde{\lambda}_1 G_1(y) + \tilde{\lambda}_2 G_2(y) \geq 0$, but this contradicts Assumption 11, which states that there exists \tilde{y} such that $G_i(\tilde{y}) < 0$ for $i = 1, 2$.

Therefore $\mu > 0$, and dividing (C.3) by μ gives

$$L(y, \tilde{\lambda}/\mu) \geq b^*, \forall y \in \mathcal{N}, \quad (\text{C.4})$$

from which it follows that

$$g(\tilde{\lambda}/\mu) = \inf_{y \in \mathcal{N}} L(y, \tilde{\lambda}/\mu) \geq b^*.$$

or $g(\lambda^*) \geq b^*$ where $\lambda^* = \tilde{\lambda}/\mu$. And since we have already shown in Claim 1 that $g(\lambda^*) \leq b^*$, it follows that $g(\lambda^*) = b^*$, and strong duality holds. \square

Claim 3. Claim 2 (Strong duality) implies that there exists $\lambda^* \in \mathbb{R}^2$ such that

$$d^* = \inf_{y \in \mathcal{N}} \left\{ G_0(y) + \sum_{i=1}^2 \lambda_i^* G_i(y) \right\} = \sup_{\lambda} g(\lambda) \text{ subject to } \lambda_i \geq 0, \text{ for } i = 1, 2. \quad (\text{C.5})$$

Claim 4. Claim 2 implies that there exists $\lambda^* \geq 0$ such that y^* is optimal for (P_P) if and only if y^* minimizes $L(y, \lambda^*)$ and $\sum_{i=1}^2 \lambda^* u_i G_i(y) = 0$.

Proof. If y^* is optimal for (P_P), then from Claims 2 and 3

$$b^* = d^* = L(\lambda^*, y^*) = G_0(y^*) + \sum_{i=1}^2 \lambda^* G_i(y^*) \leq G_0(y^*) = d^*$$

which implies that y^* minimizes $L(y, \lambda^*)$ and $\sum_{i=1}^2 \lambda_i G_i(y) = 0$.

Conversely, if y^* minimizes $L(y, \lambda^*)$ and $\sum_{i=1}^2 \lambda_i G_i(y) = 0$, then Claims 2 and 3 imply

$$b^* = d^* = L(y^*, \lambda^*) = G_0(y^*)$$

which implies that y^* is optimal for (P_P). So both Claim 4 and the proposition are proved. \square

References

- Anderson, E., Dana, J., 2009. When is price discrimination profitable? *Manag. Sci.* 55, 980–989.
- Ata, B., Dana, J., forthcoming. Price discrimination on booking time. *Int. J. Ind. Organ.* <http://dx.doi.org/10.1016/j.ijindorg.2015.06.002>.
- Baron, D., Besanko, D., 1984. Regulation and information in a continuing relationship. *Inf. Econ. Policy* 1, 267–302.
- Bartle, R., 1975. *The Elements of Real Analysis*, 2nd edition. John Wiley and Sons.
- Bergemann, D., Said, M., 2011. Dynamic auctions: a survey. In: Cochrane, J. (Ed.), *Wiley Encyclopedia of Operations Research and Management Science*, vol. 2. John Wiley & Sons, pp. 1511–1522.
- Board, S., 2011. Selling options. *J. Econ. Theory* 136, 324–340.

- Board, S., Skrzypacz, A., 2010. Revenue management with forward-looking buyers. Working Paper.
- Brumelle, S., McGill, J., 1993. Airline seat allocation with multiple nested fare classes. *Oper. Res.* 41, 127–137.
- Che, Y.-K., 1996. Consumer return policies for experience goods. *J. Ind. Econ.* 44 (1), 17–24.
- Chen, Y.-J., 2011. Optimal selling scheme for heterogeneous consumers with uncertain valuations. *Math. Oper. Res.* 36, 695–720.
- Courty, P., 2003. Ticket pricing under demand uncertainty. *J. Law Econ.* 46 (2), 627–652.
- Courty, P., Li, H., 2000. Sequential screening. *Rev. Econ. Stud.* 67 (4), 697–717.
- Curry, R., 1990. Optimal airline seat allocation with fare classes nested by origin and destination. *Transp. Sci.* 24, 193–204.
- Dana, J., 1998. Advance-purchase discounts and price discrimination in competitive markets. *J. Polit. Econ.* 106 (2), 395–422.
- Deb, R., Said, M., 2015. Dynamic screening with limited commitment. *J. Econ. Theory* 159 (Part B), 891–928.
- DeGraba, P., 1995. Buying frenzies and seller induced excess demand. *Rand J. Econ.* 26 (2), 331–342.
- Deneckere, R., Peck, J., 2012. Dynamic competition with random demand and costless search: a theory of price posting. *Econometrica* 80, 1185–1247.
- Ely, J., Garrett, D., Hinnosaar, T., 2013. Overbooking. Working Paper.
- Escobari, D., Jindapon, P., 2008. Price discrimination through refund contracts in airlines. San Francisco State University Working Paper.
- Esö, P., Szentes, B., 2007a. The price of advice. *Rand J. Econ.* 38, 863–880.
- Esö, P., Szentes, B., 2007b. Optimal information disclosure in auctions and the handicap auction. *Rev. Econ. Stud.* 74, 705–731.
- Gale, I., Holmes, T., 1992. The efficiency of advance purchase discounts in the presence of aggregate demand uncertainty. *Int. J. Ind. Organ.* 10 (3), 413–437.
- Gale, I., Holmes, T., 1993. Advance purchase discounts and monopoly allocation of capacity. *Am. Econ. Rev.* 83, 135–146.
- Gallego, G., Şahin, O., 2010. Revenue management with partially refundable fares. *Oper. Res.* 58 (4-part-1).
- Garrett, D., 2013. Incoming demand with private information. Working Paper.
- Gershkov, A., Moldovanu, B., 2010. Efficient sequential assignment with incomplete information. *Games Econ. Behav.* 68, 144–154.
- Green, J., Laffont, J.-J., 1986. Partially verifiable information and mechanism design. *Rev. Econ. Stud.* 53, 447–456.
- Krähmer, D., Strausz, R., 2011. Optimal procurement contracts with pre-project planning. *Rev. Econ. Stud.* 78, 1015–1041.
- Littlewood, K., 1972. Forecasting and control of passenger bookings. In: *Proceedings of the Twelfth Annual AGIFORS Symposium*. Nathanya, Israel.
- Möller, M., Watanabe, M., 2010. Advance purchase discounts versus clearance sales. *Econ. J.* 120, 1125–1148.
- Myerson, R., 1986. Multistage games with communication. *Econometrica* 54, 323–358.
- Nocke, V., Peitz, M., Rosar, F., 2011. Advance-purchase discounts as a price discrimination device. *J. Econ. Theory* 146 (1), 141–162.
- Pavan, A., Segal, I., Toikka, J., 2014. Dynamic mechanism design: a Myersonian approach. *Econometrica* 82 (1), 601–653.
- Ringbom, S., Shy, O., 2004. Advance booking, cancellations, and partial refunds. *Econ. Bull.* 13, 1–7.
- Riordan, M., Sappington, D., 1987. Awarding monopoly franchises. *Am. Econ. Rev.* 77, 375–387.
- Rockafellar, R., 1970. *Convex Analysis*. Princeton University Press.
- Rothschild, M., Stiglitz, J., 1970. Increasing risk I: a definition. *J. Econ. Theory* 2, 225–243.
- Salant, S., 1989. When is inducing self-selection suboptimal for a monopolist? *Q. J. Econ.* 104, 980–989.
- Segal, I., 2003. Optimal pricing mechanisms with unknown demand. *Am. Econ. Rev.* 93 (3), 509–529.
- Shen, Z., Su, X., 2007. Customer behavior modeling in revenue management and auctions: a review and new research opportunities. *Prod. Oper. Manag.* 16 (6), 713–728.
- Stokey, N., 1979. Intertemporal price discrimination. *Q. J. Econ.* 93 (3), 355–371.
- Su, X., 2009. Consumer returns policies and supply chain performance. *Manuf. Serv. Oper. Manag.* 11, 595–612.
- Talluri, T., van Ryzin, G., 2004. *The Theory and Practice of Revenue Management*. Kluwer Academic Publishers.
- Vohra, R., 2012. Dynamic mechanism design. *Surv. Oper. Res. Manag. Sci.* 17 (1), 60–68.