Paper 1: "Price Discrimination on Booking Time" – Baris Ata & James D. Dana, Jr.

Paper 2: "Revenue Management by Sequential Screening" - Mustafa Akan, Barıs Ata, and James D. Dana, Jr.

Price Discrimination on Booking Time*

Barış Ata
University of Chicago

James Dana
Northeastern University

September 28, 2014

Abstract

We consider advance-purchase ticket sales to consumers that differ in when they learn their valuations. We find simple sufficient conditions under which a sequence of increasing prices perfectly price discriminates and fully extracts the consumer surplus even though consumers are ex ante privately informed. The model is closely related to our work on optimal screening with returns contracts, but here we assume a simple binary distribution for consumers' valuations. We also extend the basic model by relaxing the instantaneous learning assumption and show that the perfect discrimination still may be feasible in a more realistic setting in which learning is gradual.

*Thanks to Florian Ederer, Martin Peitz, David Myatt, Kathryn Spier and participants at the EARIE 2014 41st annual conference for helpful comments. Harvard Business School and Northeastern University provided financial support.

1 Introduction

Optimal advance purchase pricing has been studied by many papers in the economics and operations management literatures. Some of this work has emphasized that offering different prices, a lower price before consumers learn and a higher price after consumers learn, can increase firm profits. This paper extends this insight further by analyzing a stylized model in which a continuum of consumers learn their valuations at different times and and by characterizing conditions under which the firm can perfectly price discriminate using a sequence of increasing prices. Unlike our earlier work in which the firm discriminates using a menu of partially refundable tickets at time zero, this paper considers a simpler, binary distribution of valuations for which the optimal pricing policy can be implemented with nonrefundable tickets.[1] This paper also relaxes the strong assumption we make in our earlier work that consumers learn their valuations instantaneously.

We demonstrate that booking time is a powerful screening device when consumers learn their valuations at different times and their valuations are correlated with when they learn. When consumers' valuations are higher when they learn their valuations later, conditional on having positive valuations, then selling nonrefundable tickets at a price that increase over time is profitable and may extract all of the consumer surplus. Consumers who learn their valuations late are deterred from purchasing earlier because doing so requires them to commit to purchase the product before they learn their valuations for the good. And consumers who learn their valuations early are deterred from purchasing later because doing so results in a higher price.

We also find sufficient conditions under which the firm can extract all of the surplus from consumers under the more realistic assumption that consumers learn their valuations gradually over time. This insight is important because it is more realistic and because it generalizes the analysis in our earlier work.

This work, and Akan et al. (2012), builds on a large literature on advance-purchase pricing. Courty (2003) considers the decision to sell to consumers either before or after they have learned their valuations. He shows the tradeoff between allocative efficiency (which favors selling after) and reducing consumer heterogeneity (which typically favors selling before). Dana (1998) analyzes a competitive market with uncertain demand and fixed capacity and shows that in equilibrium some firms

[1]When the distribution of consumers' valuations is more general (see Akan, Ata & Dana (2012)) the refund must be equal to marginal cost to induce efficient return decisions. This cannot be implemented with sequential sales. Assuming a binary distribution of valuations eliminates the distortionary aspect of full price returns options, so fully refundable tickets with expiring refund dates (which are equivalent to sequential spot prices) implement the optimal mechanism.

set lower prices and sell to consumers who purchase in advance before they learn while other firms sell to consumers who wait until after they learn. Gale & Holmes (1993) analyze a related monopoly model in which the firm price discriminates by selling to some consumers before they learn their valuations and other afterwards. Nocke, Peitz & Rosar (2011) use a mechanism design framework to analyze optimal advance purchase pricing when consumers vary in how much information they have ex ante. Courty & Li's (2000) consider optimal screening when the firm can sell returns contracts to all consumers ex ante and vary the size of the refund in order to screen consumers. They show it is optimal to distort the size of the refund to some consumers in order to extract more surplus from others.

The operations management literature has also focused much research on advance-purchase pricing, though that literature frequently considers a stochastic demand environment and capacity constraints. Gallego & Şahin (2010), and other recent work in operations management, models consumers as forward-looking and strategic. An excellent survey of this literature is Talluri & van Ryzin (2004) (see also Shen & Su 2007).

Our work is also related to the literature on dynamic mechanism design, which began in a regulation setting with Baron & Besanko (1984) and Riordan & Sappington (1987).[2] Board (2011) considers a firm that auctions the right to consume in the future after they have learned their private value of consumption. Esö & Szentes (2007*b*) consider dynamic mechanism design when an informed client can hire an expert to generated additional information that the client can use to refine their decision making.[3] Esö & Szentes (2007*a*) allow the seller to release additional information to privately informed agents who update their valuations based on the new information. Krähmer & Strausz (2011) generalize both Courty & Li (2000) and Esö & Szentes (2007*b*) by adding information acquisition. Board & Skrzypacz (2010) consider a dynamic mechanism design model which considers a capacity constrained firm selling to ex ante homogeneous consumers who learn dynamically. Pavan, Segal & Toikka (forthcoming) give a general characterization the optimal mechanism and the information rents of privately informed agents in dynamic mechanism design models.

A few recent dynamic mechanism design papers allow for stochastic arrivals of new consumers, that is, consumers with which firms cannot contract before they arrive. Deb & Said (2013) consider a two-period model like Courty & Li (2000)'s, with additional consumers arriving in period two who cannot write contracts in

[2]For recent surveys of this literature see Vohra (2012) and Bergemann & Said (2011).

[3]Gale & Holmes (1992) also analyze consumer learning, but they can analyze a static direct revelation mechanism because the ex post allocation depends only on the ex ante information.

period one, and with a firm that cannot commit in advance to the contracts it offers in period two. Gershkov & Moldovanu (2010) consider the optimal mechanism for a patient firm selling a fixed number of heterogeneous objects to a stochastic sequence of impatient, privately informed buyers. Garrett (2013) considers a firm facing stochastic arrivals of privately informed buyers when the firm sells a durable good. Ely, Garrett & Hinnosaar (2013) assume that consumers observe only a signal of their demand when they arrive and then all consumers learn their valuations in the final period. None of these papers consider the case in which consumers' valuations are not correlated with their arrival time, so firms do not screen on when consumers arrive. An obvious interpretation of consumer arrivals is that consumers learn something about their valuations, so our work is closely related.

2 The Model

A single, risk-neutral firm sells a homogenous good with unit cost c to heterogeneous, risk-neutral consumers. Suppose that all consumption takes place in the future (think of the firm as selling tickets), and consumers are privately informed about when they will learn their valuations and the distribution of their valuations. The firm starts selling tickets after consumers know their types, but before they begin learning the realized valuations.

Let t denote a consumer's type which represents the time, $t \in [0, T]$, at which the consumer learns her valuation. Suppose this valuation has a simple binary distribution: type t's valuation is high and equal to $V(t) > 0$ with probability $\pi(t)$ and equal to 0 with probability $1 - \pi(t)$. We also assume for simplicity that $V(t) > c$, for all t, so trade is always efficient if consumers learn they have a high valuation. Finally, note that each consumer's type is a pair, (t, v), which represents her ex ante information and the realization of her valuation, but with this simple structure it is straightforward to characterize consumers' behavior as a function of the valuation v, so we often refer to the consumer's type as just t.

Our assumptions imply that consumers' expected valuations, $\pi(t)V(t)$, are correlated with the time t at which they learn their valuation, which makes it possible for the firm to price discriminate using booking time as a screening device.

Under complete information, the firm can extract all of the surplus by setting an ex post price $p(t) = V(t)$. That is, when the firm can observe each consumer's type and charge each consumer a different price, then the firm captures all of the consumer surplus.

However, under incomplete information – when types are unobservable or the firm cannot prevent consumers from buying a ticket designed to be sold to a different consumer – charging different consumers different prices at the same time is impossible. But when consumers learn their valuations at different times, it may still be possible for the firm to extract all the surplus by screening on when consumers make their purchase decisions.

When offered sequentially, the spot prices $p(t) = V(t)$ are incentive compatible under two conditions. First, it must be true that no consumer prefers to wait and purchase later, that is, a type t will not imitate type t'' for any $t'' > t$, which is true as long as $V(t)$ is nondecreasing. And second, it must be true that no consumer prefers to purchase earlier, that is, a type t consumer will not imitate a type t' consumer for any $t' < t$, which is true as long as $V(t') \geq \pi(t)V(t)$, for all t, t' such that $t' < t$. However, since the first condition requires that $V(t)$ is nondecreasing, we can rewrite the two incentive compatibility conditions more succinctly as follows:

Proposition 1 *The firm can implement the complete-information, or first-best, outcome with a sequence of increasing, non-refundable prices $p(t) = V(t)$ if and only if (i) $V(t)$ is nondecreasing, and (ii) $V(0) \geq \pi(t)V(t)$, for all t.*

Proof. To show that (i) and (ii) are sufficient, we need only show that the prices are incentive compatible and individually rational. Let $u(t';t)$ denoted the expected utility of a type t consumer who imitates a type t' consumer. So we need to show that $u(t';t) \geq u(t;t)$ for all $t' \neq t$, and that $u(t;t) \geq 0$. It is natural to assume a type t consumer purchases at price $p(t) = V(t)$ at time t if her valuation is equal to $V(t)$ and does not purchase if her valuation is 0. Formally this follows from incentive compatibility at time t for a consumer of type t, v, but in this binary model it is easy to be informal with respect to this second dimension of a consumer's type. Given this purchasing rule, a type t consumer's utility is clearly 0 whether or not her valuation is high, so $u(t;t) \geq 0$.

Incentive compatibility has two parts. First, each consumer must purchase if only if her valuation weakly exceeds the price (as we just assumed), and second, a type t consumer must prefer to purchase at time t to purchasing before or after. Again, we assume that the consumer always purchases if her valuation weakly exceeds the price, and not otherwise, in order to write the expected utility and explore the second condition.

If a type t consumer imitates a type $t'' > t$ consumer and purchases at time t'', then her consumer surplus is $V(t) - p(t'')$ is weakly negative even if her valuation is high since $V(t)$ is nondecreasing. So $u(t'';t) \geq u(t;t)$ for all $t'' > t$ and imitating type t'' is never profitable.

If a type t consumer imitates a type $t' < t$ consumer and purchases at time t', then her expected utility is $u(t';t) = \pi(t)V(t) - p(t')$, which is negative since $p(t') = V(t')$, and $V(t') \geq V(0)$ by condition *(i)*, and $V(0) \geq \pi(t)V(t)$ by condition *(ii)*.

Condition (i) is necessary since a otherwise there exists a type t and a type t' consumer such that $t' > t$ and $V(t') < V(t)$ and so the type t consumer could earn strictly positive surplus by imitating the type t'' consumer. And condition (ii) is clearly necessary since otherwise there exists a type t consumer who could earn strictly positive surplus by imitating type 0. ■

Proposition 1 establishes that the first best is feasible, but the result that it is feasible with sequential spot prices depends critically on our binary distribution assumption. In particular, it relies on the assumption that the support of v_t is a single point once we condition on the realization of v_t being greater than c. This is crucial to the firm being able to extract all the surplus from a type t consumer. If type t consumers were heterogeneous ex post (after learning their valuations were larger than c) then the firm could not implement the first best with spot prices, though it might be able to do so if it contracted with consumers before they learned their valuations (see Akan et al. 2012).

The following result also follows immediately.

Proposition 2 *If $V(t)$ is nonincreasing, the firm can no better than sell all its tickets at a single price.*

Formally this follows from the analogous result in Akan et al. (2012), and intuitively this follows because in this environment the firm can do no better than design a spot price for each consumer, but when $V(t)$ is nonincreasing, the incentive constraint prevents the firm from charging consumers who learn early a higher price as they can costlessly imitate consumers who learn late.

3 Gradual Learning

A strong assumption in the above model, and also in Akan et al. (2012), is that consumers learn instantaneously at time t, and a natural question is whether the perfect discrimination result is robust to relaxations of this assumption. In this section we extend our analysis by allowing consumers to learn their valuations gradually over time as opposed to instantaneously.

First, it is important to define gradual learning. In this binary environment a consumer's beliefs can be represented by a single real number, the probability of having a high valuation, which will change over time as information arrives and consumers update their valuations using Bayes' rule. A type t consumer in our model without gradual learning begins at time 0 with the prior that her probability of having valuation $V(t)$ is $\pi(t)$ and her probability of having valuation zero is $1-\pi(t)$. This belief doesn't change until time t when the consumer learns her valuation. After time t a fraction $1-\pi(t)$ of the type t consumers believe their valuation is 0 with probability 1, and a fraction $\pi(t)$ of type t consumers believe their valuation is $V(t)$ with probability 1.

We can represent the set of possible beliefs of a type t consumer at time s as a distribution function over the set of binary distributions, or more simply, as a distribution function over probabilities. That is, the distribution of beliefs can be represented by a cumulative distribution function $F(\rho|s,t)$ on $\rho \in [0,1]$, where ρ is the probability that $v = V(t)$ for type t at time s. In our model without gradual learning $F(\rho|s,t)$ has all of the mass at $\pi(t)$ when $s < t$, so $F(\rho|s,t) = 0$ for $\rho < \pi(t)$ and $s < t$ and $F(\rho|s,t) = 1$ for $\rho \geq \pi(t)$ and $s < t$, but when $s \geq t$, the mass is divided between 0 and 1, so $F(\rho|s,t) = \pi(t)$ for $\rho < 1$ and $s \geq t$ and $F(\rho|s,t) = 1$ for $\rho = 1$ and $s \geq t$. Clearly learning is not gradual in this characterization because at $s = t$, the CDF changes discretely from a CDF with all of the mass at $\pi(t)$ to a CDF with some mass at 0 and some mass at 1. That is, a large portion of consumers' beliefs are changing at time t and these changes are large changes.

Instead, we think gradual learning implies that significant changes in beliefs should occur only for an arbitrarily small portion of consumers, and other consumers' beliefs should be changing only in small ways. We suggest that a reasonable necessary condition for learning to be gradual learning that captures this intuition is that F is smooth in s. Formally, we make this our definition of gradual learning, which is as follows:

Definition 1 *Learning is gradual if F is almost everywhere left continuous, or* $\lim_{r\uparrow s} F(\rho|r,t) = F(\rho|s,t)$ *for all $s \in [0,t]$ and for almost every ρ.*

Note that this definition does not restrict how much an individual consumer's beliefs change, but does restrict the total change in beliefs in any small time interval to be small. Hence, some might argue that additional conditions should also be met.

Also, note that if a type t consumer learns that her valuation is high with probability one at any time before t, then she will buy immediately (if she hasn't already done so). Hence perfect discrimination is no longer feasible. But perfect discrimination may be feasible for some gradual learning environments as long as there is a cap on how fast the expected valuations grow. We now focus on a simple model that exhibits such a cap.

Suppose that a type t consumer starts at time 0 with the prior that her valuation is $V(t)$ with probability $\pi(t)$ and then may receive a signal between time 0 and time

t that reveals that her true valuation is 0. The probability density of receiving the signal is uniformly distributed on the time interval $[0,t]$, where $f(s) = (1-\pi(t))/t$ for $s \in [0,t]$ is the uniform probability density function. The associated cumulative distribution function is $F(s) = (1-\pi(t))\,\frac{s}{t}$, which clearly approaches $1-\pi(t)$ as s approaches t.

Between time 0 and time t the consumer may learn her valuation is zero, but if she does not receive a signal, she will update her beliefs and expect to have a high valuation with greater and greater probability. At time s, using Bayes' rule, and conditional on not having already received a signal revealing that her valuation is 0 between time 0 and time s, the type t consumer believes she has a high valuation is $\pi(t)/(1-F(s))$. This is just the unconditional probability of a high valuation divided by the unconditional probability of receiving no bad news by time s, which approaches 1 as s approaches t. We can write this conditional probability as

$$\phi_1(s;t) = \frac{\pi(t)}{1-(1-\pi(t))\frac{s}{t}}$$

for all $s \in [0,t]$. Similarly, $\phi_0(s;t) = 0$ denotes the conditional probability of a high valuation conditional on learning that the valuation is low.

Note that this learning process satisfies the above definition of gradual learning. Clearly for all $\rho > \phi_1(s;t)$, $F(\rho|r,t) = F(\rho|s,t) = 1$, and for all $\rho < \phi_1(s;t)$, as long as r is sufficiently close to s then $\lim_{r\to s} F(\rho|r,t) = F(\rho|s,t)$ since the CDF is exactly equal to the mass at 0 which grows continuously in r. Note that $\lim_{r\to s} F(\rho|r,t) = F(\rho|s,t)$ does not hold at $\rho = \phi_1(s;t)$, but nevertheless holds almost everywhere so our definition is satisfied.

We now ask whether the prices $p(t) = V(t)$ are incentive compatible in this setting, and whether the firm still earns the complete information profits. As before, two conditions must be met. First, no consumer must prefer to wait and book her ticket at time t'' for any $t'' > t$, which is true as long as $V(t)$ is increasing. And second, no consumer must prefer to buy a ticket earlier at time $t' < t$.

In this setting, if a consumer deviates to purchase earlier (imitates a consumer of type $t' < t$), she can condition on the information she has about her valuation. A type t consumer's expected utility, or consumer surplus, function if she imitates type $t' \in [0,t]$, and purchases at time t', is $u_0(t';t) = 0$ if she already knows that her valuation is zero, and is

$$u_1(t';t) = \phi_1(t';t)V(t) - p(t'), \forall t' \in [0,t],$$

or

$$u_1(t';t) = \left(\frac{\pi(t)t}{t-(1-\pi(t))t'}\right)V(t) - p(t'), \tag{1}$$

if she has not yet received a bad signal. If instead she waits and purchases at time t, her expected utility is still 0 because $p(t) = V(t)$. So incentive compatibility requires that $u_0(t';t) \leq 0$, which clearly holds, and $u_1(t';t) \leq 0$ for all $t' \in [0,t]$. That is, the expected valuation for a type t consumer purchasing at time t', conditional on not yet learning her valuation is low, must be less than the price at time t'.

Notice that this condition reduces to $\pi(t)V(t) \leq V(t')$ when $t' = 0$, which holds, as before, when $\pi(t)$ is sufficiently small.

Equation (1) implies $u_1(t;t) = 0$, and $u_1(t';t)$ is continuously differentiable, so incentive compatibility clearly requires that

$$\frac{du_1(t';t)}{dt'} = \frac{\pi(t)(1-\pi(t))t}{(t-(1-\pi(t))t')^2}V(t) - V'(t') > 0 \tag{2}$$

in a neighborhood of t. Moreover, $du_1(t';t)/dt' > 0$ for all $t' \in [0,t]$ is clearly a sufficient condition for incentive compatibility. Differentiating $u_1(t';t)$ again yields

$$\frac{d^2u_1(t';t)}{(dt')^2} = \frac{2\pi(t)(1-\pi(t))^2 t}{(t-(1-\pi(t))t')^3}V(t) - V''(t').$$

which is clearly positive for all $t' \in [0,t]$ if $V''(t') \leq 0$. That is, if $V''(t') \leq 0$, then $u(t';t)$ is convex for $t' \in [0,t]$. Clearly $u_1(0;t) \leq 0$ (because $\pi(t)V(t) \leq V(0)$) and $u_1(t;t) = 0$ (because $p(t) = V(t)$), so $d^2u_1(t';t)/(dt')^2 > 0$ for all $t' \in [0,t]$ implies that $u_1(t';t) \leq 0$ for all $t' \in [0,t]$ and that the incentive compatibility constraint holds. So if $V''(t') \leq 0$ then Proposition 1 still holds. Formally,

Proposition 3 *If a type t consumer learns her valuation is 0 at a constant rate over time between time 0 and time t, and the valuation function, $V(t)$, is linear or concave then firm can implement the complete-information, or first-best, outcome with a sequence of increasing, non-refundable prices $p(t) = V(t)$ if and only if (i) $V(t)$ is nondecreasing, and (ii) $V(0) \geq \pi(t)V(t)$, for all t.*

Notice that time is merely an ordering and plays no role in profits or utility. So it is important to discuss the interpretation of $V''(t) \leq 0$ in Proposition 3. The assumption that $V''(t) \leq 0$ imposes a limit on the rate of change in $V'(t)$ relative to the rate of change in consumer information, which we assumed was constant over time. If the rate of change in $V'(t)$ relative to the rate of change in information is constant, then simply by reordering the units of time, we can satisfy the assumption that both rates of change are constant, and the proposition holds. If the rate of change in $V'(t)$ is growing relative to the rate of change in information than the proposition may fail to hold. If $V'(t)$ is growing to quickly, then price may be rising to quickly in a neighborhood of t and the incentive constraint may be locally violated.

Some additional intuition comes from analyzing the incentive constraint in the vicinity of t. The incentive compatibility condition holds near t if $du(t;t)/dt' \geq 0$, or

$$\frac{du(t;t)}{dt'} = \frac{1-\pi(t)}{\pi(t)}\frac{V(t)}{t} - V'(t) \geq 0,$$

which is easier to satisfy when $\pi(t)$ is sufficiently small. When $V'(t) = v$ is constant, this local incentive compatibility constraint holds if

$$(1-\pi(t))V(t) \geq \pi(t)tv = \pi(t)(V(t) - V(0)).$$

But $V(0) \geq \pi(t)V(t)$ and $V(t) \geq V(0)$ by assumption (conditions *(i)* and *(i)*), which implies

$$(1-\pi(t))V(t) = V(t) - \pi(t)V(t) \geq V(t) - V(0) \geq \pi(t)(V(t) - V(0)),$$

so the local incentive constraint holds under the condition that the consumer prefers not to purchase at time 0, or $\pi(t)V(t) \leq V(0)$ (condition *(ii)* of Proposition 1).

An important modeling assumption was that consumers learn whether or not their valuation is zero, not whether or not their valuation is high. If at any time t' before t a consumer learns her valuations is high, or $V(t)$, with probability one, she will immediately purchase at price $p(t') = V(t') < V(t)$. That is, incentive compatibility is violated. The modeling assumption placed a bound on the beliefs of the most optimistic consumer. That is, incentive compatibility requires that $\phi_1(t', t)V(t) \leq V(t')$, so incentive compatibility requires that $\phi_(t'; t)$ not approach one too quickly as t' goes to t. Moreover, if $V(t')$ is large, than $\partial\phi_1(t'; t)/\partial t'$ must also be large.

The model could easily be extended to more sophisticated learning dynamics. Incentive compatibility can easily hold in a richer model. But the crucial property is that consumers don't learn too quickly. Even with more general dynamics, it is sufficient to consider just the behavior of the type t consumer who at every stage receives the strongest possible signal that her valuation is high. As long as this consumer's conditional probability of having a high valuation approaches one sufficiently slowly at t' goes to t, then the incentive compatibility constraints will hold.

4 Conclusion

This paper considered a simple model of a monopolist selling to heterogeneous, privately informed consumers and found that if consumers learn their valuations at different times, and if when they learn is correlated with their expected valuation, then it may be possible to perfectly price discriminate with a sequence of increasing spot prices. The result relied strongly on the assumption that valuations have a binary distribution. With more general distributions the firm might still be able to implement the first-best, but only using advance sales with a return option (see Akan et al. 2012).

The paper also considered an extension in which consumers learn their valuations gradually as opposed to instantaneously. We showed then under mild additional assumptions, if the first best can be implemented with instantaneous learning, it can also be implemented with gradual learning. This implies that the first-best implementation result is not simply a consequence of the discontinuity in consumers' information.

References

Akan, M., Ata, B. & Dana, J. (2012), 'Revenue management by sequential screening'. Working Paper.

Baron, D. & Besanko, D. (1984), 'Regulation and information in a continuing relationship', *Information Economics and Policy* **1**, 267–302.

Bergemann, D. & Said, M. (2011), Dynamic auctions: A survey, *in* J. Cochrane, ed., 'Wiley Encyclopedia of Operations Research and Management Science', Vol. 2, John Wiley & Sons, pp. 1511–1522.

Board, S. (2011), 'Selling options', *Journal of Economic Theory* **136**, 324–340.

Board, S. & Skrzypacz, A. (2010), 'Revenue management with forward-looking buyers'. Working Paper.

Courty, P. (2003), 'Ticket pricing under demand uncertainty', *Journal of Law and Economic* **46**(2), 627–652.

Courty, P. & Li, H. (2000), 'Sequential screening', *Review of Economic Studies* **67**, 697–717.

Dana, J. (1998), 'Advance-purchase discounts and price discrimination in competitive markets', *Journal of Political Economy* **106**(2), 395–422.

Deb, R. & Said, M. (2013), 'Dynamic screening with limited commitment'. Working Paper.

Ely, J., Garrett, D. & Hinnosaar, T. (2013), 'Overbooking'. Working Paper.

Esö, P. & Szentes, B. (2007*a*), 'Optimal information disclosure in auctions and the handicap auction', *Review of Economic Studies* **74**, 705–731.

Esö, P. & Szentes, B. (2007*b*), 'The price of advice', *RAND Journal of Economics* **38**, 863–880.

Gale, I. & Holmes, T. (1992), 'The efiiciency of advance purchase discounts in the presence of aggregate demand uncertainty', *International Journal of Industrial Organization* **10**(3), 413–437.

Gale, I. & Holmes, T. (1993), 'Advance purchase discounts and monopoly allocation of capacity', *American Economic Review* **83**, 135–146.

Gallego, G. & Şahin, O. (2010), 'Revenue management with partially refundable fares', *Operations Research* **58**, 817–833.

Garrett, D. (2013), 'Incoming demand with private information'. Working Paper.

Gershkov, A. & Moldovanu, B. (2010), 'Efficient sequential assignment with incomplete information', *Games and Economic Behavior* **68**, 144–154.

Krähmer, D. & Strausz, R. (2011), 'Optimal procurement contracts with pre-project planning', *Review of Economic Studies* **78**, 1015–1041.

Nocke, V., Peitz, M. & Rosar, F. (2011), 'Advance-purchase discounts as a price discrimination device', *Journal of Economic Theory* **146**, 141–162.

Pavan, A., Segal, I. & Toikka, J. (forthcoming), 'Dynamic mechanism design: A Myersonian approach', *Econometrica* .

Riordan, M. & Sappington, D. (1987), 'Awarding monopoly franchises', *American Economic Review* **77**, 375–387.

Shen, Z. & Su, X. (2007), 'Customer behavior modeling in revenue management and auctions: a review and new research opportunities', *Production and Operations Management* **16**, 713–728.

Talluri, T. & van Ryzin, G. (2004), *The Theory and Practice of Revenue Management*, Kluwer Academic Publishers.

Vohra, R. (2012), 'Dynamic mechanism design', *Surveys in Operations Research and Management Science* **17**, 60–68.

Revenue Management by Sequential Screening[a]

Mustafa Akan[b] Barış Ata[c] James D. Dana Jr.[d]

Ocbober 24, 2014

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 achieves the complete information profits when consumers who learn later have the same expected valuation, but a more dispersed distribution of valuations (mean-preserving spread). We also show that the complete-information profits may be feasible when consumers who learn later have greater valuations (first-order stochastic dominance) and then derive the optimal pricing policy, and the associated distortions from efficiency, when perfect discrimination is not feasible. We find that when the correlation between when consumers learn and their valuation distribution is reversed, then time has no value as a screening device.

[a]We would like to thank Johannes Hörner, Hao Li, and Kathryn Spier, as well as the editor, Alessandro Pavan, an anonymous associate editor, and two referees for very help comments and suggestions on earlier versions of the paper. Thanks also 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, and Wharton, as well as participants at the Econometric Society Summer Meetings and the INFORMS Revenue Management and Pricing Conference. We also thank Booth School of Business, University of Chicago, Carnegie Mellon University, the Kellogg School of Management, Harvard Business School, and Northeastern University for financial support.

[b]Carnegie Mellon University, akan@andrew.cmu.edu

[c]University of Chicago, baris.ata@chicagobooth.edu

[d]Northeastern University, j.dana@neu.edu, correspondence author address: 301 LA, 360 Huntington Ave., Boston MA 02115, 617.373.3517

1 Introduction

This paper analyzes an optimal advance-purchase pricing, or revenue-management, problem when consumers learn their valuations for future consumption over time and are sophisticated, that is, when consumers are forward looking and choose when to buy. We assume that each consumer begins with private information about both the distribution of her valuation and 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 initial private information, or type, 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, different levels of refunds, and different times at which the refund 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. In simple models, screening on time refers to offering different prices at different times, but in our general dynamic mechanism design framework, screening on time refers to placing restrictions on when and for how much the consumer can return his or her purchases. For example, we find consumers who learn their valuations late may be deterred from imitating consumers who learn their valuations early if products sold to consumers who learn early cannot be returned late. The role of time as a screening device has received remarkably little attention in the dynamic mechanism design literature, in part because general analyses, like Pavan, Segal & Toikka (forthcoming), implicitly rule out our information assumptions by assuming the payoffs are continuous in the agents' ex ante private information. In our model, the agent's ex ante private information is when they will learn their valuations and clearly two consumers who learn their valuations at arbitrarily close times need not have arbitrarily close payoffs since the mechanism may nevertheless allow one consumer to act after observing her valuation while forcing the other to act before observing her valuation.

We first analyze the case in which consumers who learn later have more dispersed valuations in the sense that their distribution is a mean preserving spread of the distribution of valuations for all consumers who learn earlier. In this case, we show that the first best is always achievable. A menu of expiring refund contracts with the refunds equal to marginal cost is optimal, and consumers with less variable demand purchase tickets that become non-refundable sooner which deters imitation by consumers who learn later. We also show that when consumers who learn earlier have more dispersed valuations (the reverse correlation), time is not a useful screening device.

We also analyze the case in which consumers who learn their valuations later have higher valuations, in the sense of first-order stochastic dominance, and find that the first-best may or may not be feasible. When the first best is feasible, the firm offers a menu of expiring refund contracts with a refund equal to the marginal cost. When the first best is not feasible, the firm still offers a menu of expiring refund contracts, but lower-valuation consumers are offered contracts with a higher refund price (a downward distortion in allocation) while higher-valuation consumers capture positive rents and are offered contracts with a lower refund price (an upward distortion in allocation). The distortion for consumers who learn early deters imitation by the very highest type (a binding global incentive compatibility constraint) consumer because it increases what these consumers pay for the refund option which the imitator must pay for as well, but cannot use because they learn their valuations to late. The distortion for consumers who learn late deters imitation by consumers who learn earlier (an upward binding local incentive compatibility constraint) because lower types are less willing to pay up front for the increased allocation of the good. And finally, as in the analysis of mean-preserving spreads, we see that when consumer who learn their valuations earlier have higher valuations, i.e, the correlation is reversed, screening on the time that the option can be exercised does not increase profits.

Advance-purchase tickets with expiring refund options are commonly sold by airlines, hotels, theaters, and railroads. The US passenger service railroad, Amtrak, has a 90% of purchase price cancelation 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 Mariott, Hilton, Sheraton, and Westin, offer 21-day advance-purchase discounts that are non-refundable, but most other reservations at these hotels can be cancelled 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 cancelation fee of \$200 for cancelations 6 or more days before arrival and offers no refund for cancelations 5 or fewer days before arrival. Disney also offers bundled discount packages with less generous cancelation 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]

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 do not yet know their realized valuation. 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 \$200 with probability 1/2 and is \$0 with probability 1/2, and assume that a Type 2 consumer's valuation is \$400 with probability 1/2 and is \$0 with probability 1/2. Also, suppose that the firm's unit cost $c = \$100$ so that it is efficient for consumers to consume the good whenever their valuation is positive.

Suppose the firm sells in advance, before consumers learn. Suppose that consumers' types are common knowledge, and the firm can design a different contract for each type. The firm wants to create incentives for efficient ex post consumption, and one obvious way to extract all the surplus is to sell the good with an option to return later if the realized valuation is low (alternatively the firm could sell options that can be exercised after the consumers learn their valuations). To insure ex post efficient allocation, the firm wants to set the return price for both types equal to its unit cost, \$100, and in order to extract all the surplus the firm will charge Type 1 consumers an up-front payment of \$150 and Type 2 consumers an up-front payment of \$250.[2] Since each consumer receives a refund of \$100 with probability 1/2, this

[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.

[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

implies the consumer's expected total payment (after the refund and ignoring discounting) is equal to their expected valuations ($100 and $200 respectively), so the firm extracts all of the consumer surplus.

Now instead suppose that consumers' types are private information. Then the above contracts are not incentive compatible. Both consumer types clearly prefer the contract offered to the Type 1 consumer, so both consumer types would choose to pay $150 up-front. However using optimal contracts the firm can do better.

The firm has two instruments with which it can screen the Type 2 consumers. First, the firm can set an expiration time for the refund offers intended for Type 1 consumers. Second, the firm set a higher refund price for Type 1 consumers. The former discourages imitation by Type 2 consumers because they don't yet know their valuations, and the later discourages imitation by Type 2 consumers because higher refunds are associated with a higher upfront price and Type 2 consumers use the refund less frequently than Type 1 consumers.

If the firm uses only the first instrument, the firm can make it more costly for a Type 2 consumer to imitate a Type 1 consumer by requiring that consumers who choose the $150 contract decide whether or not they want the refund at time t_1, after the Type 1 consumers have learned their valuations, but before the Type 2 consumers have learned theirs. So it is incentive compatible for the firm to offer the Type 1 consumer a price of $150 with the option to return the good at time t_1 for a refund of $100 and offer the Type 2 consumer a price of $200 with the option to return the good at time t_2 for a refund of $200; Type 2 consumers weakly prefer to pay $200 when they consume the good and $100 when they do not consume to paying $150 whether or not they consume the good (they learn their valuations too late to be able to take advantage of the refund).

The firm can do even better using both instruments. In particular, it is incentive compatible for the firm to offer Type 1 consumers a price of $200 with the option to return for a full refund of $200 at time $t_!$, and to offer Type 2 consumers a price of $250 with the option to return for a refund of $100 at time t_2. Increasing the upfront price and simultaneously increasing the refund offered to Type 1 consumers has no effect on Type 1 consumers' surplus, but makes it more costly for Type 2 consumers to imitate Type 1 consumers. Type 2 consumers weakly prefer $250 with a return price of $100 to paying $200 whether or not they want the good (since they learn too late to ever exercise the refund option at time t_1).

These contracts are clearly optimal as they extract all of the consumer surplus without introducing any distortion in the allocation. In this binary distribution ex-

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.

ample, increasing the size of the refund offered to Type 1 consumers does not distort the allocation of the good, and it is profitable because it increases the effectiveness of the other instrument. Under more general distributions increasing the refund will be distortionary, but as long as Type 2's valuation distribution still first-order stochastically dominates Type 1's distribution, screening by increasing the size of the refund distorts downwards the allocation to Type 1 which reduces the incentive of Type 2 to imitate Type 1.

This numerical example illustrates that there are two elements to the firm's optimal screening strategy. First, the firm sets an expiration date for the refund option offered to Type 1 consumers that is before Type 2 consumers learn their valuations so the Type 2 consumers can't imitate the Type 1 consumers without paying for an option that they may never use. And second, 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. The firm makes imitation more costly by creating even more option value for Type 1 knowing that in order to imitate, Type 2 will have to purchase the more expensive option, but will never exercise it.

Of course, instead of offering a fully refundable, advance-purchase contract to Type 1 consumers, the firm could instead offer Type 1 consumers a non-refundable contract for \$200 that they purchase at time t_1, *after* they have learned their valuations, but before Type 2 consumers have learned theirs, and offer Type 2 consumers a non-refundable contract for \$400 that they purchase at time t_2, *after* they have learned their valuations. This equivalent 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 valuation, and then only if their valuation is \$200. Type 2 consumers represent business travelers who purchase at a higher price late, and only if their valuations is \$400. Extracting all of the surplus with spot prices is only feasible because of the binary distribution assumption (see Ata & Dana (2014)) for more analysis of pricing under this assumption).

Because consumers learn more over time (not less), time can be used to make it costly for consumers who learn late to imitate consumers who learn early, but not vice versa. So screening on time increases profits when consumers who learn late contribute more to firm profits than consumers who learn early.

While the numerical example illustrates the value of screening on when the refund can be exercised, it is restrictive in two 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 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 frequently the binding incentive constraints are not just the constraints preventing consumers from imitating other consumers who learn earlier. We show that 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 late than those that learn early.

In our general analysis consumer valuations are drawn from a continuous distribution. We focus mainly on two cases: when the distributions of consumers' valuations can be ordered with respect to a mean-preserving spread and when the distributions can be ordered with respect to first order stochastic dominance. We find that the firm can extract the entire surplus in the former case, and may be able to do so in the latter case. We analyze in detail the case in which the entire surplus cannot be extracted and show that this leads to rich set of distortions in the incomplete-information contracts (including both upward and downward distortions in the same mechanism) relative to the complete-information contracts.

The paper is structured as follows. Section 2 briefly reviews the related literature. Section 3 presents the general model, introduces the mechanism design problem, and derives the necessary and sufficient conditions for implementing the first-best solution using contingent contracts. Section 4 demonstrates that the first-best is achieved when consumers' valuations become more dispersed (in the sense of a mean-preserving spread) as a function of when they learn their valuations. In Section 5 we consider two cases in which screening on when refunds take place (that is, on when the refund option expires) is not useful: first, the case in which consumers' valuations are less dispersed when they learn later, and second, the case in which consumers' valuations decrease (in the sense of first order stochastic dominance) as a function of when they learn their valuations. In each of these cases the optimal mechanism is the equivalent to the optimal mechanism in a model in which consumers learn simultaneously (i.e., Courty & Li (2000)). Finally, in Section 6, we find the optimal mechanism in an interesting case in which screening on when refunds take place is a useful screening device, but the first best is not feasible. This is the case in which consumers' valuations increase as a function of when they learn their valuations. Section 7 discusses potential extensions of the model and offers some concluding remarks.

2 Related Literature

Our work is related to three closely related literatures: the applied 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 agents 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 & 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 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 & 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, Peitz & Rosar (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 the firm screens consumers that 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 finds firms are more likely to offer refunds when costs are high and when consumers are risk averse, though Che restricts attention to nonrefundable or

[3]See also Moller & Watanabe (2007) and Gale & 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, the optimal mechanism can always be implemented without selling refundable tickets.

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 & Li (2000) (see also Ringbom & Shy (2004)). In Courty & Li (2000), the firm screens consumers on the size of the refund, but not on when the consumer 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 & 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 & van Ryzin (2004) provides an extensive review of these papers and many other extensions of these capacity control models in the operations management literature.

Gallego & Şahin (2010) and other more recent papers in the operations management literature, model consumers as forward-looking and strategic (for a review see Shen & Su 2007). Gallego & Şahin (2010) (see also Su (2009)) consider a model with ex ante homogeneous consumers, but assume that consumers learn in a multiperiod,

[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 & 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.

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 most 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 & Besanko (1984) and Riordan & 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 the right to consume in the future after they have learned their private value of consumption. Esö & Szentes (2007*b*) consider the optimal dynamic mechanism in which an informed client can hire an expert to generated 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 & Szentes (2007*a*) 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. Krähmer & Strausz (2011) generalize both Courty & Li (2000) and Esö & Szentes (2007*b*) by adding endogenous information acquisition with moral hazard.

Board & 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 & 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.

Our work is related to other dynamic mechanism design models with stochastic arrivals of new consumers. Deb & Said (2013) consider a two-period model like Courty & Li (2000)'s, with additional consumers arriving in period two who cannot write contracts in period one, and with a firm that cannot commit in advance to the

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

[8]Gale & 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.

contracts it offers in period 2. They show that the seller can benefit from postponing ex ante contracting with buyers available in the first period, though this is never optimal with full commitment. Gershkov & Moldovanu (2010) consider the optimal mechanism for a patient firm selling a fixed number of heterogeneous objects to a stochastic sequence of impatient, privately informed buyers. Garrett (2013) considers a firm facing stochastic arrivals of privately informed buyers when the firm sells a durable good. Unlike our work, these papers assume consumers cannot write contracts before they arrive and don't learn additional information after they arrive. A paper that is more closely related to ours is Ely, Garrett & Hinnosaar (2013) which assumes that consumers observe only 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 learn their final demands at the same time, and, like all of these papers, consumers' valuations are not correlated with their arrival time, so firms do not screen on when consumers arrive or when they learn their valuations.

An important paper in this literature is Pavan et al. (forthcoming) 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, we carefully explain why their envelope theorem does not hold in our model in the next section of the paper.

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 their valuations are correlated with when they learn.

3 The Model

A single, risk neutral firm sells a single homogenous 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

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

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 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 $[\underline{v}, \overline{v}]$. It follows that

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

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 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 the 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. (forthcoming), Myerson (1986), and Green & 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

[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.

to the seller. Consider a consumer whose type is t and whose valuation is v. Her ex post utility (or surplus) is given by

$$u(v', t'; v, t) = vy(v', t') - x(v', t') \tag{1}$$

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,

$$u(v, t) = vy(v, t) - x(v, t). \tag{2}$$

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

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

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

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

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,

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

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

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

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

$$\max_{x(v,t), y(v,t)} \int_{v,t} f\left(v, t\right)\left[x\left(v, t\right) - cy\left(v, t\right)\right] dvdt \tag{P^0}$$

subject to

$$U(t) \geq 0, \forall t, \tag{IR}$$
$$u(v, t) \geq u(v', t; v, t), \forall v, v', t, \tag{IC_t}$$
$$U(t) \geq U(t'; t), \forall t, t', \tag{IC_0}$$
$$0 \leq y\left(v, t\right) \leq 1, \forall v, t, \tag{F}$$

and equations (1) through (4), the definitions of U and u, which are functions of x and y. For tractability, we restrict the firm to choose functions $y(\cdot, t)$ and $x(\cdot, t)$ that are piecewise continuous.

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 cancelled.

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. Thus, if a type t consumer pretends to be type $t' \geq t$, she will report the valuation that maximizes her ex post surplus when she is asked to report her valuation at time t'. In contrast, if a type t consumer reports her type as $t' < t$, when she is asked to report her valuation at time t', she will report the valuation that maximizes her expected ex post surplus given that she already reported her type as t'.

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

The following lemma characterizes how a consumer reports her valuation if she does not report her type truthfully at time zero and its proof is provided in Appendix A.

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 vacation at time t' (because $t' < t$), then it is optimal for her to report

her expected valuation, that is, $\mathbb{E}_t[v] \in \arg\max_{v'} \mathbb{E}_t[u(v', t'; v, t)]$.

Lemma 1(i) follows immediately from (IC_t) and holds because once the consumer learns v her payoff is independent of their 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 (IC_0) into two constraints, the seller's mechanism design problem becomes (P^1):

$$\max_{x(v,t),y(v,t)} \int_{v,t} f(v,t)\left[x(v,t) - cy(v,t)\right] dvdt \tag{P^1}$$

subject to

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

$$u(v,t) \geq u(v', t; v, t), \forall v, v', t, \tag{IC_t}$$

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

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

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

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

In this formulation, we separate the set of constraints (IC_0) 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 only, but nevertheless fundamental, difference between our the firm's problem when consumers learn sequentially, (P^1), and Courty & Li (2000)'s analysis of the firm's problem when consumers learn simultaneously. The ex ante incentive compatibility constraint in their paper is

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

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 other consumers who learn earlier than it is to imitate other 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 draws a greater valuation, 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 Screening on the Timing of Refunds Implements the First-Best

We now ask when can the firm implement the complete-information allocation and earn the complete-information profits even though consumers' types and valuations are privately observed. In the following proposition, we give necessary and sufficient conditions under the firm can implement the complete-information solution. Its proof is given in Appendix A.

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

$$\mathbb{E}_{t'}\left[\max(v-c,0)\right] \geq \mathbb{E}_{t}\left[\max(v-c,0)\right], \forall t' > t; \qquad \text{(Condition IEGT)}$$

and

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

Condition IEGT, short for *increasing expected gains from trade*, holds when the expected gains from trade are increasing in t.

The expected gains from trade can we written as

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

which is the difference between the expected valuation and the cost plus the value of the option to return the good for a refund equal to cost, c, when $v<c$. It follows that Condition SOV, short for *sufficient option value*, holds when

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

or equivalently, when the option value of being able to return the good when $v\leq c$ at any time t' is sufficiently large relative to the difference in the consumer's expected valuation between any type t and type t', for any $t>t'$.

Condition SOV is satisfied whenever the consumer's expected valuation is decreasing in t because the option value is always nonnegative. But Condition IEGT 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 IEGT is satisfied if and only if the expected valuation is increasing, so Conditions 1 and 2 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. 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 & 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., $\mathbb{E}_t\left[\max\left(v,c\right)\right]-c$ does not depend on t.[11]

An obvious implication of Proposition 1 is that when Condition IEGT and Condition SOV are both satisfied, then the complete-information solution can be implemented using a menu of expiring refund contracts: the initial upfront price is $p(t)=\mathbb{E}_t\left[\max\left(v,c\right)\right]$ 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 for the returned tickets, only consumers with valuations higher than the cost will consume, which implies that the allocation is ex

[11]This is because Courty & 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 does not distort ex post consumption (see Footnote 2).

post efficient. Moreover, the expected utility of all the consumers is equal to zero. Because Condition IEGT 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.

Before characterizing situations in which Condition IEGT and Condition SOV both hold, we want to emphasize that Proposition 1 is not consistent with the envelope theorem of Pavan et al. (forthcoming) which characterizes the information rents that privately informed agents earn in optimal dynamic mechanisms. Our assumptions differ in important ways from those of Pavan et al. (forthcoming). To see this, consider a slightly different description of our model. Suppose that the consumer learns her type t at time 0, and then at every period s observes v_s which is independently distributed with distribution $F(v|s)$. Of course only one of the variables in this sequence ($s = t$) corresponds to the consumer's valuation. Note that this formal structure fits within the general framework of Pavan et al. (forthcoming).[12]

In our model, and in this alternate characterization of our model, the buyer's payoff function is $v_t - p$ if she gets the good and $-p$ otherwise, where p is the net ex post payment to the seller. In other words, the valuation depends on only one of the signals in the sequence $\langle v_s \rangle_{s=0}^{T}$, and t identifies which of the sequence of signals determines the buyer's valuation. Note that the payoff function is well defined and is totally differentiable with respect to each v_s in the sequence, given the buyer's type t, but is not differentiable with respect to t. A small change in the buyer's type from t to t' implies that the buyer's valuation is v_t' not v_t, and these valuations are drawn from independent distributions: the difference between v_t' and v_t need not be small.[13] This means that our model violates both Condition U-D (differentiability of the utility function) and Condition U-ELC (equi-Lipschitz continuity of the utility function) of Pavan et al. (forthcoming).

The point is that Pavan et al. (forthcoming)'s framework encompasses a broad class of models in which agents have ex ante private information about the distribution of their ex post valuations, but does not encompass models in which agents have ex ante information about when they will acquire additional information, or more

[12]The sequence $(t, \langle v_s \rangle_{s=0}^{T})$ corresponds to the sequence $\langle \theta_t \rangle_{t=0}^{T}$ in Pavan et al. (forthcoming), as well as to the sequence $\langle \epsilon_t \rangle_{t=0}^{T}$, since we assume that t and each of the v_s's are independently distributed. And the restrictions on the impulse response functions in Pavan et al. (forthcoming) are trivially satisfied because of our independence assumption.

[13]Of course, Pavan et al. (forthcoming) analyze a discrete time model so continuity in time is not even well defined.

precisely, about which future signal is more informative, v_t or $v_{t'}$, because the later does not satisfy their continuity assumptions. Of course discontinuities of this sort do not necessarily represent a problem. In our model, the discontinuity in the payoff function matters because the buyer's private information about the distribution of their valuation is related to their private information about when they learn. That is, the discontinuity can potentially be exploited by the mechanism designer.

Intuitively, under Pavan et al. (forthcoming)'s assumptions, a consumer with slightly more good news at time 0 gets additional rents because the good news makes her better off holding all of her actions fixed. That is, because her payoff function is continuous, the mechanism designer cannot extract those rents because the buyer can always ask to be treated like the buyer without the better news. Because the payoff function is continuous in the private information, holding the mechanism fixed, the increase in the consumer's utility associated with better news must be equal to at least the increase in the consumer's expected utility holding the consumer's actions fixed. Pavan et al. (forthcoming) characterize a general environment in which this constraint holds with equality and the consumer is strictly better off when she receives better news.

However, our payoff function is discontinuous in the consumer's private information, even holding the mechanism fixed. Consumers with arbitrarily close ex ante types can have dramatically different expected utilities even if they are treated identically. Specifically, under the allocation mechanism that extracts all of the consumer surplus, a buyer of type t and a buyer of type $t' > t$ have remarkably different expected utilities when both are treated as a type t buyer. Regardless of how close t and t' are to each other, the type t' buyer, whose valuation first order stochastically dominates the type t buyer, is strictly worse off because the mechanism allocates the good based only on the signal v_t and not on the signal $v_{t'}$ which is equal to the type t' buyer's true valuation.

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. More precisely, suppose that the conditional valuation distributions, $F(v|t)$, can be ordered on $[0, T]$ with respect to a mean-preserving spread (MPS). That is, suppose that all consumers have the same expected valuation, but that consumers who learn their valuations later face greater uncertainty about their valuations. We call this forward MPS and in Section 5 contrast this with reverse MPS in which consumers who learn their valuations later face less uncertainty about their valuations.

We begin with a formal definition of a forward mean-preserving spread following the definition in Rothschild & Stiglitz (1970).

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

$$\int_{\underline{v}}^{v} [F(v'|t') - F(v'|t)]dv' \geq 0, \forall v, \tag{7}$$

with strict inequality for some v.

We now show that Condition IEGT and Condition SOV of Proposition 1 are satisfied under Assumption 1, and that 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), Assumption 1 clearly implies

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

for all $t' > t$, so Condition IEGT of Proposition 1 is satisfied. Assumption 1 also implies

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

for all t and t', so Condition SOV of Proposition 1 is satisfied. ■

The firm can implement the first-best solution with a menu of expiring refund contracts $\{(p(t), r(t)) : 0 \leq t \leq T\}$ where the refund size is equal to the marginal cost, i.e. $r(t) = c$ for $t \in [0, T]$ and the initial price is equal to $p(t) = \mathbb{E}_t[\max(v,c)]$, which implies:

Corollary 1 *Under Forward MPS, the optimal price* $p(t)$ *is increasing in type,* t*.*

By offering contracts with return options that expire when consumers learn their valuations, the firm makes it costly for consumers who learn their valuations later to purchase the contract intended for consumers who learn early. To these consumers purchasing early is equivalent to a nonrefundable contract since it is never worth it for them to exercise the refund option.

Note that unlike the standard sequential screening framework, 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 at which the option to claim the refund expires, and 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 the same as the consumer she imitates, but that consumer derives 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. In particular, when consumers who learn their valuations later have higher expected valuations (in the sense of first-order stochastic dominance) then Condition IEGT clearly holds, and Condition SOV might hold if c is sufficiently large so that the option value of being able to return the good is sufficiently large relative to the increase in expected valuation from time 0 to time T. However, in Section 6, we mainly focus on the case in which Condition SOV is not satisfied and show that screening on time increases the seller's profits even when the first best cannot be implemented.

5 Screening on the Timing of Refunds is not Profitable

In Section 4 we saw that the seller can sometimes achieve the first best by using the timing of refunds as a screening device. This section asks instead when screening on the timing of refunds is no longer effective, or equivalently, when the optimal mechanism with sequential learning is equivalent to the optimal mechanism with simultaneous learning.

The answer is quite simple. 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 so imitating types with higher t is simple, but it is costly for her to be forced to report her valuation before learning it, so imitating types with lower t is costly. So time is useful if the mechanism designer is trying to reduce imitation of lower types, but not when he is trying to reduce imitation of higher types. That is, when low types would like to imitate high types because the expected gains from trade are decreasing in t (and hence high types are offered a lower price), then time will not be valuable as a screening device.

Not surprisingly, we will show that under reverse MPS (when Assumption 1 is reversed), then screening on the timing of refunds is not useful, but in this section we analyze considerably more general environments in which the expected gains from trade are decreasing rather than increasing (as in forward MPS). Formally we consider distributions of v and t that satisfy the following assumption.

Assumption 2 *For all t,*

$$\int_{v}^{\overline{v}} \frac{\partial F(v'|t)}{\partial t} dv' \geq 0, \forall v, \qquad \text{(Condition SDEGT)}$$

with strict inequality for some v.

Assumption 2, which is different from second order stochastic dominance because of the range of integration, clearly implies that

$$\int_{v}^{\overline{v}} \left[F(v'|t) - F(v'|t')\right] dv' \geq 0, \forall v, t, t'; t > t'. \tag{8}$$

Condition SDEGT is short for *strong decreasing expected gains to trade*. This name is relevant because equation (8) is equivalent to

$$\mathbb{E}_t \left[\max\left(v, \alpha\right)\right] - \alpha \leq \mathbb{E}_{t'} \left[\max\left(v, \alpha\right)\right] - \alpha, \forall t < t', \forall \alpha. \tag{9}$$

with strict inequality for some α. So if we were to fix $\alpha = c$, then Condition SDEGT and Condition IEGT would be mutually exclusive.

Note that under Assumption 2 for $t > t'$,

$$U(t) - U(t') \leq U(t) - U(t;t'), \tag{10}$$

because of ($\overline{\mathrm{IC}}_0$), so Lemma 2 and integration by parts implies

$$\begin{aligned} U(t) - U(t;t') &= \int_{\underline{v}}^{\overline{v}} u(v,t) \left[f(v|t) - f(v|t')\right] dv \\ &= -\int_{\underline{v}}^{\overline{v}} y\left(v,t\right) \left[F(v|t) - F(v|t')\right] dv. \end{aligned} \tag{11}$$

Using integrations by parts again,

$$
\begin{aligned}
&-\int_{\underline{v}}^{\overline{v}} y(v,t)\left[F(v|t)-F(v|t')\right]dv \\
&= -y(\overline{v},t)\int_{\underline{v}}^{\overline{v}}\left[F(v'|t)-F(v'|t')\right]dv' + \int_{\underline{v}}^{\overline{v}} dy(v,t)\int_{\underline{v}}^{v}\left[F(v'|t)-F(v'|t')\right]dv' \\
&< -y(\overline{v},t)\int_{\underline{v}}^{\overline{v}}\left[F(v'|t)-F(v'|t')\right]dv' + \int_{\underline{v}}^{\overline{v}} dy(v,t)\int_{\underline{v}}^{\overline{v}}\left[F(v'|t)-F(v'|t')\right]dv' \\
&= 0, \qquad (12)
\end{aligned}
$$

where the inequality follows from Assumption 2, or more precisely (8), which implies

$$\int_{\underline{v}}^{v}\left[F(v'|t)-F(v'|t')\right]dv' \leq \int_{\underline{v}}^{\overline{v}}\left[F(v'|t)-F(v'|t')\right]dv', \qquad (13)$$

for all v, with strict inequality for some v.

Equations (10), (11), and (12) imply that $U(t)$ is monotone decreasing. Building on Lemma 2, the following lemma helps us characterizes 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, ($\overline{IC}_0$) implies that*

$$U'(t) \leq -\int_{\underline{v}}^{\overline{v}} y(v,t)\frac{\partial F(v|t)}{\partial t}dv \quad \text{for almost every } t. \qquad (14)$$

Using Lemmas 2 and 3, which are also used in the next section, we now consider the following relaxation of the firm's problem (P^1), which we call (P^2), in which $x(v,t)$ has been eliminated from the objective function (using the definitions of $U(t)$ and $u(v,t)$), in which (IC_t) is replaced (using Lemma 2), in which ($\underline{\mathrm{IC}}_0$) is dropped, and in which ($\overline{\mathrm{IC}}_0$) is replaced (using Lemma 3).

$$\max_{U(t),u(v,t),y(v,t)} \int_{v,t}(v-c)y(v,t)f(v,t)\,dvdt - \int_t U(t)h(t)dt \qquad (\mathrm{P}^2)$$

subject to

$$
\begin{aligned}
&U(t) \geq 0, \forall t, && (\mathrm{IR})\\
&y(v,t) = \frac{\partial u(v,t)}{\partial v} \text{ and } y(v,t) \text{ is non-decreasing in } v, && (\widetilde{\mathrm{IC}}_t)\\
&U'(t) \leq -\int_{\underline{v}}^{\overline{v}} y(v,t)\frac{\partial F(v|t)}{\partial t}dv, \forall t, && (\widetilde{\mathrm{IC}}_0)\\
&0 \leq y(v,t) \leq 1, \forall v,t. && (\mathrm{F})
\end{aligned}
$$

Note that $U(t)$ 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 relaxed problem (P^2) differs from the original mechanism design problem in several ways. First, the objective function is rewritten using the definitions of $u(v,t)$ and $U(t)$ to replace $x(v,t)$ so that the problem can be stated as one of choosing $U(t)$, $u(v,t)$ and $y(v,t)$. Next, the incentive compatibility constraints (IC_t) are replaced by their local counterparts ($\widetilde{\mathrm{IC}}_t$), which are equivalent by Lemma 2. Moreover, the incentive compatibility constraints regarding upward deviations ($\overline{\mathrm{IC}}_0$) are ignored, except for the local constraints ($\widetilde{\mathrm{IC}}_0$), which are derived from Lemma 3. And finally ($\underline{\mathrm{IC}}_0$) are dropped completely because we will see that they do not bind under Assumption 2.

The sequential learning problem can now be compared to the firm's problem when consumers learn their valuations simultaneously (as in Courty & Li (2000)). Clearly (P^2) also represents a relaxation of that problem. The only relevant difference between the simultaneous learning problem and the sequential learning problem is in the constraint ($\overline{\mathrm{IC}}_0$) which we will see doesn't bind under Assumption 2.

Using the lemmas we can now state and prove the main result of the section, which is a characterization of conditions under which the simultaneous learning model and the sequential learning model, i.e., (P^1), have the same solution.

The proof of the following proposition is in Appendix A.

Proposition 3 *Under Assumption 2, the optimal mechanism is equivalent to the optimal mechanism when consumers learn their valuations simultaneously.*

The proof verifies that under Assumption 2, the solution to (P^2) satisfies all of the omitted constraints in the sequential learning problem (P^1) and the simultaneous learning problem.

Assumption 2 generalizes versions reverse MPS and reverse FOSD. We say that F satisfies reverse MPS when $F(v|t)$ is a mean preserving spread of $F(v|t')$ for all $t' > t$. That is, if $\mathbb{E}_t[v] = \mu$ for all t and

$$\int_{\underline{v}}^{v} [F(v'|t) - F(v'|t')]dv' \geq 0, \forall v, t', t; t' > t, \tag{15}$$

with strict inequality for some v, or equivalently,

$$\int_{\underline{v}}^{v} \frac{\partial F(v'|t)}{\partial t} dv' \leq 0, \forall v, t, \tag{16}$$

with strict inequality for some v, and strict equality for $v = \overline{v}$. So reverse MPS clearly satisfies Condition SDEGT.

We say that F satisfies reverse FOSD when $F(v|t)$ first-order stochastic dominates $F(v|t')$ for all $t' > t$, and forward FOSD when the ordering is reversed. So reverse FOSD satisfies Condition SDEGT.

Corollary 2 *Under reverse FOSD and reverse MPS, the optimal mechanism is equivalent to the optimal mechanism when consumers learn their valuations simultaneously.*

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 contrast with forward first-order stochastic dominance in the next section. Under both reverse MPS and reverse FOSD we find that screening on when the refund option expires is not useful, and that the firm can only separate ex ante heterogenous consumers by screening on the size of the refund offered to consumers who subsequently return the good (a result first characterized by Courty & Li (2000)).

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. This is easy to see for reverse first-order stochastic dominance. The expected gains from trade are higher for lower types, all else equal, because valuations fall as t rises. And all else can be held equal, because each consumer can easily imitate a type T consumer and pay the same price conditional on their valuation.

Under reverse mean preserving spread, the expected gains from trade are higher for lower types because the expected valuation is unchanged, but conditional on $v < c$ valuations are lower, and conditional on $v > c$ valuations are higher. Again, the firm will extract all the surplus from type T and types $t < T$ can always earn strictly positive surplus by imitating type T.

6 Screening on the Timing of Refunds is Profitable, but cannot Implement the First-Best

So far we have considered cases with two extreme outcomes. First, cases in which screening on the expiration of refund option implements the first best, and second, cases in which screening on the expiration of the refund option is not useful and the firm treats all consumers as if they learned simultaneously. We now analyze cases in which screening on the expiration of the refund option 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 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. We call this case forward FOSD and contrast it with reverse FOSD considered in the previous section.

The optimal mechanism in this section has several interesting features. First, we show that global incentive compatibility constraints bind, or more precisely, exactly one global incentive constraint binds, not just the local incentive compatibility constraints. 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. This makes 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 upwards local incentive compatibility constraint binds for these consumers. For these consumers, the downward global incentive compatibility constraint does not bind because the cost of imitating type 0 is higher for them then the cost of imitating type T (and they also pay less than type T when they report truthfully), and they must give up even more option value by forgoing the option to return.

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

Third, we find that the optimal mechanism includes both upwards 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 upwards 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. (forthcoming)), 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 of 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 $(\tau, T]$.

And again, note that unlike the standard sequential screening framework, the firm is offering different initial type consumers contracts with the same strike price, or refund option, yet these consumers are not pooled. The firm can still separate initial types by offering the same refund, but varying the time at which the option to claim the refund expires.

Formally, our definition of forward FOSD is:

Assumption 3 (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 3 and equation (5) imply that $\mathbb{E}_t[\max(v,c)]$ is weakly increasing in t, so Condition IEGT of Proposition 1 is satisfied. This is because $\max(v,c)$ is an increasing function, and the expected value of any increasing function of v must increase when the random variable v increases in the sense of first order stochastic dominance.

When Condition SOV of Proposition 1 is also satisfied, then the firm allocates the good efficiently and extracts all the surplus. However under Assumption 3, Condition SOV need not be satisfied. For example, if $\underline{v} \geq c$, then Condition SOV cannot hold. Indeed, under Assumption 3, 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

$$\mathbb{E}_T[v] > \mathbb{E}_0[\max(v,c)], \tag{17}$$

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 3.

In this section we assume that equation (17) holds:

Assumption 4 *Condition SOV of Proposition 1 is violated.*

Condition SOV is satisfied if types have sufficiently close expected valuations. Under Assumption 3, if the firm tried to extract all of the surplus, then the price would be an increasing function of t, and, since Condition SOV is violated, type T would prefer type 0's contract. By deviating to a lower type, a higher type gives up the option to return, so the higher type will always imitate type 0, the type that pays the lowest ex ante price. And, under Assumption 3, if type t prefers to imitate type 0, then type T must also prefer to imitate type 0 since giving up the return option is always less costly for type T and the price savings is always greater. So when Condition SOV is violated, if the firm tries to extract all of the surplus, type T strictly prefers to report that her type is type 0. We will see shortly that this is in fact the only binding global (non-local) incentive constraint.

In what follows, we first solve a relaxed version of the seller's problem and then (in Proposition 6) 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 solutions of the form $y(v,t) \in \{0,1\}$ for all v, t.

Lemma 4 *The optimal expected utility schedule is non-decreasing, $U'(t) \geq 0$.*

Lemmas 3 and 4 imply

$$0 \leq U'(t) \leq -\int_{\underline{v}}^{\overline{v}} y(v,t) \frac{\partial F(v|t)}{\partial t} dv \quad \text{for almost every } t, \tag{18}$$

and together with Lemma 2, we use equation (18) to further relax the seller's problem, which is now P^3).

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

subject to

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

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

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

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

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

As in the previous section, 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 problem differs from (P^2) in that we do not entirely drop the ($\underline{IC}_0$) constraint and $\widetilde{\mathrm{IC}}_0$ places both an upper and lower bound on $U'(t)$.

Note that we are not restricting the firm to offering deterministic contracts (that is we are not assuming $y(v,t) \in \{0,1\}$), unlike some other recent papers in the dynamic mechanism design literature. Instead the optimality of deterministic contracts here is a result.

For the remainder of this section, we make the following assumptions:

Assumption 5

1. $\frac{\partial F(v|t)/\partial t}{f(v|t)}$ is non-increasing in v for all t; and
2. $\frac{\partial}{\partial v}\left[\frac{F(v|t)}{f(v|t)}\right] \leq \frac{1}{\sup_{v>c} \frac{(v-c)f(v|t)}{F(v|t)}}$ for all t.

The magnitude of the term $(\partial F(v|t)/\partial t)/f(v|t)$ can be interpreted (cf. Baron & Besanko (1984)) as a measure of how informative the consumer's private type knowledge is about his valuation. Since $\partial F(v|t)/\partial t < 0$ under first-order stochastic dominance, Assumption 5.1 corresponds to assuming that a consumer's type becomes more informative about her valuation as the valuation v increases. Finally, Assumption 5.2 is a technical condition used only in the proof of Proposition 4. As can be seen in the proof of Proposition 4, it is sufficient but not necessary.

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

Proposition 4 *Under Assumptions 3 and 5, there exists an optimal solution U, y to the relaxed problem P^3 such that $y(v,t) \in \{0,1\}$ for all v, t. 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*

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

Note that if Assumption 4 is violated, then the proposition still holds, but the first best is achieved, and $\tau = T$.

The following corollary of Proposition 4 will be instrumental in characterizing the optimal contract and is proved in Appendix B.

Corollary 3 *For $t \leq T$, we have that*

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

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

We can now further simplify the firm's relaxed problem, (P^3). First, using Proposition 4 and Corollary 3, 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 $(\widetilde{\mathrm{IC}}_0)$ 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 the consumer makes an ex ante payment for the good which is a function of her type, t, and then receives a refund of $r(t)$ if she returns the good at time t after learning their valuations (in which case she will return the good if and only if $v < r(t)$). These refund contracts clearly satisfy (F), so this constraint is dropped. Finally, we ignore $(\widetilde{\mathrm{IC}}_t)$ and then show in Proposition 6 that the solution to the relaxed problem satisfies the ignored constraints.

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

$$\max_{\{\tau, r(t)\}} \int_0^{\tau} \int_{r(t)}^{\overline{v}} f(v,t)(v-c)\,dvdt \tag{P^4}$$

$$+ \int_{\tau}^{T} \int_{r(t)}^{\overline{v}} f(v,t) \left(v - c + \frac{1 - H(t)}{h(t)} \frac{\partial F(v|t)/\partial t}{f(v|t)} \right) dvdt$$

subject to $U(T) \geq U(t';T)$, or equivalently

$$-\int_{\tau}^{T} \int_{r(t)}^{\overline{v}} \frac{\partial F(v|t)}{\partial t} dvdt \geq (\mathbb{E}_T[v] - r(t'))\,\mathbf{1}_{\{\mathbb{E}_T[v] - r(t') \geq 0\}} \tag{$\underline{\mathrm{IC}}_0^T$}$$

$$- \int_{\tau}^{t'} \int_{r(t)}^{\overline{v}} \frac{\partial F(v|t)}{\partial t} dvdt - \int_{r(t')}^{\overline{v}} (v - r(t')) f(v|t') dv, \forall t' \leq T.$$

The following assumption is used in the proof of Proposition 5. It is a sufficient condition for the optimal refund schedule to be everywhere decreasing in t.

Assumption 6 $\frac{\partial F(v|t)/\partial t}{f(v|t)}$ *is non-increasing in t for all $v \in [\underline{v}, c]$.*

Intuitively, Assumption 6 requires that the informativeness of a consumer's type about her valuation weakly increases with her type (for all valuations less than c). Note that Assumption 6 is sufficient and not necessary.

The next proposition characterizes the optimal solution to the problem (P^4).

Proposition 5 *Under Assumptions 3-6, the solution to* (P^4) *is unique and characterized by two thresholds σ and τ, such that $0 < \sigma < \tau < T$ and ($\underline{\mathrm{IC}}_0^T$) 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 ($\underline{\mathrm{IC}}_0^T$);*
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 relative to the complete information price) at and near $t = 0$ and upward distortions in allocation (a return price distorted lower relative to the complete information price) at and near $t = T$. Although Assumption 6 is sufficient to ensure that $r'(t) \leq 0$, it is not necessary. Indeed, the condition $r'(t) \leq 0$ is itself sufficient but not necessary for proving that ($\underline{\mathrm{IC}}_0^T$) is slack on $[\tau, T)$, cf. Lemma 6 in Appendix B. That is, the results of Proposition 5 can continue to hold even when Assumption 6 is not satisfied.

Given the cutoff points $\{r(t) : 0 \leq t \leq T\}$ characterized in Proposition 5, the payments can be written as follows:

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

Note that the expected surplus of type t consumer can be written as

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

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

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

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

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

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

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

We interpret $\{(p(t), r(t)) : 0 \leq t \leq T\}$ as a menu of expiring refund contracts where type t is charged the initial price $p(t)$ and is offered a refund of $r(t)$ if he chooses not to consume the good or use the service before time t. In other words, the refund $r(t)$ is only good before 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 Courty & Li (2000); it is included in Appendix B.

Proposition 6 *Under Assumptions 3-6, there exists a solution $x(v,t)$ and $y(v,t)$, 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 4 *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 $t \in [\sigma, \tau)$. Then, 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.*

When consumers who learn their valuations later have higher valuations, in the sense of first-order stochastic dominance, then the first best may or may not be feasible. When the firm can't implement the complete-information outcome, the firm can still increase its profits by offering a higher refund price to consumers who learn early (distorting their allocation downwards) This increases the price they pay conditional on getting the good, which is inefficient, but makes imitation less attractive (because the imitator always pays this price, even if they don't want the good) and allows the firm to charge more to consumers who learn late.

The type T has the strongest incentive to imitate a lower valuation consumer because she wants the good more often. As a consequence the global incentive compatibility constraint, ($\overline{\mathrm{IC}}_0$), binds for type T. However the firm cannot give the type T consumer rents without also giving rents to other consumers. In the optimal mechanism, consumers in a neighborhood of type T are offered rents to prevent the from imitating a higher type, so local incentive compatibility constraints bind in the opposite direction than expected. As a consequence, the refund price for type T consumers is set below marginal cost (distorting the allocation upwards) in order to reduce the information rents that the firm must give to other consumers.

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 consumers learn is correlated with the distribution of their valuations. The sequential learning by consumers gives the firm an additional instrument with which to screen consumers. In some cases this instrument enables the seller to implement the first-best allocation and extract all 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. Second, unlike most screening problems, imposing restrictions on when consumers purchase, or more precisely, on when they exercise their return option, need not be distortionary, so discriminating may be able to implement the first best.

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. That is, in more realistic settings, or empirical applications, consumers are likely to have multiple dimensions of heterogeneity (which could significantly increase the complexity of the optimal mechanism – see our discussion in Section 3). Third, we assumed that consumers learn instantaneously; it is more realistic to suppose that learning is gradual.

Allowing for gradual learning is difficult because the amount of information reporting increases significantly, but notice that intuitively 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 not binding, giving a type t consumer more information shortly before time t does violate the constraints.

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 transactions costs, competition with rivals, or because consumers arrive to late for initial contracting (though late arrival may itself require justification using transactions costs). Analysis of optimal pricing in more general environments is clearly needed.

References

Anderson, E. & Dana, J. (2009), ‘When is price discrimination profitable?’, *Management Science* **55**, 980–989.

Ata, B. & Dana, J. (2014), ‘Price discrimination on booking time’. Working Paper.

Baron, D. & Besanko, D. (1984), ‘Regulation and information in a continuing relationship’, *Information Economics and Policy* **1**, 267–302.

Bergemann, D. & Said, M. (2011), Dynamic auctions: A survey, *in* J. Cochrane, ed., ‘Wiley Encyclopedia of Operations Research and Management Science’, Vol. 2, John Wiley & Sons, pp. 1511–1522.

Board, S. (2011), ‘Selling options’, *Journal of Economic 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’, *Operations Research* **41**, 127–137.

Chen, Y.-J. (2011), ‘Optimal selling scheme for heterogeneous consumers with uncertain valuations’, *Mathematics of Operations Research* **36**, 695–720.

Courty, P. (2003), ‘Ticket pricing under demand uncertainty’, *Journal of Law and Economic* **46**(2), 627–652.

Courty, P. & Li, H. (2000), ‘Sequential screening’, *Review of Economic Studies* **67**, 697–717.

Curry, R. (1990), ‘Optimal airline seat allocation with fare classes nested by origin and destination’, *Transportation Science* **24**, 193–204.

Dana, J. (1998), ‘Advance-purchase discounts and price discrimination in competitive markets’, *Journal of Political Economy* **106**(2), 395–422.

Deb, R. & Said, M. (2013), ‘Dynamic screening with limited commitment’. Working Paper.

DeGraba, P. (1995), ‘Buying frenzies and seller induced excess demand’, *RAND Journal of Economics* **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. (2007*a*), ‘Optimal information disclosure in auctions and the handicap auction’, *Review of Economic Studies* **74**, 705–731.

Esö, P. & Szentes, B. (2007*b*), ‘The price of advice’, *RAND Journal of Economics* **38**, 863–880.

Gale, I. & Holmes, T. (1992), ‘The efiiciency of advance purchase discounts in the presence of aggregate demand uncertainty’, *International Journal of Industrial Organization* **10**(3), 413–437.

Gale, I. & Holmes, T. (1993), ‘Advance purchase discounts and monopoly allocation of capacity’, *American Economic Review* **83**, 135–146.

Gallego, G. & Şahin, O. (2010), ‘Revenue management with partially refundable fares’, *Operations Research* **58**, 817–833.

Garrett, D. (2013), ‘Incoming demand with private information’. Working Paper.

Gershkov, A. & Moldovanu, B. (2010), ‘Efficient sequential assignment with incomplete information’, *Games and Economic Behavior* **68**, 144–154.

Green, J. & Laffont, J.-J. (1986), ‘Partially verifiable information and mechanism design’, *Review of Economic Studies* **53**, 447–456.

Krähmer, D. & Strausz, R. (2011), ‘Optimal procurement contracts with pre-project planning’, *Review of Economic Studies* **78**, 1015–1041.

Littlewood, K. (1972), ‘Forecasting and control of passanger bookings’. In Proceedings of the Twelfth Annual AGIFORS Symposium, Nathanya, Isreal.

Moller, M. & Watanabe, M. (2007), ‘Should a monopolsit sell before or after buyers know their demands’. University Carlos III Madrid Working Paper.

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’, *Journal of Economic Theory* **146**, 141–162.

Pavan, A., Segal, I. & Toikka, J. (forthcoming), ‘Dynamic mechanism design: A Myersonian approach’, *Econometrica* .

Ringbom, S. & Shy, O. (2004), ‘Advance booking, cancellations, and partial refunds’, *Economics Bulletin* **13**, 1–7.

Riordan, M. & Sappington, D. (1987), ‘Awarding monopoly franchises’, *American Economic Review* **77**, 375–387.

Rothschild, M. & Stiglitz, J. (1970), ‘Increasing risk I: A definition’, *Journal of Economic Theory* **2**, 225–243.

Salant, S. (1989), ‘When is inducing self-selection suboptimal for a monopolist?’, *Quarterly Journal of Economics* **104**, 980–989.

Seierstad, A. & Sydsæter, K. (1987), *Optimal Control Theory with Economic Applications (Advanced Textbooks in Economics)*, 3rd edn, North Holland.

Shen, Z. & Su, X. (2007), ‘Customer behavior modeling in revenue management and auctions: a review and new research opportunities’, *Production and Operations Management* **16**, 713–728.

Stokey, N. (1979), ‘Intertemporal price discrimination’, *Quarterly Journal of Economics* **93**, 355–371.

Su, X. (2009), ‘Consumer returns policies and supply chain performance’, *Manufacturing & Service Operations Management* **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’, *Surveys in Operations Research and Management Science* **17**, 60–68.

A Proofs of the Results in Section 3 through 5

Proof of Lemma 1. Suppose that a type t consumer reports her type as t'. Then by (IC_t) for a consumer with type t' and valuation v, we write

$$vy(v,t') - x(v,t') \geq vy(v',t') - x(v',t') \text{ for all } v'.$$

Thus, $v \in \arg\max_{v'} u\left(v',t';v,t'\right)$. Since $u\left(v',t';v,t'\right) = u\left(v',t';v,t\right)$ for all v',v,t',t, we conclude that $v \in \arg\max_{v'} u\left(v',t';v,t\right)$.

To establish part (ii), first notice that $\mathbb{E}_t\left[u(v',t';v,t)\right] = u(v',t';\mathbb{E}_t\left[v\right],t)$. Then, by ($\mathrm{IC}_t$) for a consumer with type t' and valuation $\mathbb{E}_t\left[v\right]$, we write

$$\mathbb{E}_t\left[v\right] y\left(\mathbb{E}_t\left[v\right],t'\right) - x\left(\mathbb{E}_t\left[v\right],t'\right) \geq \mathbb{E}_t\left[v\right] y(v',t') - y(v',t') \text{ for all } v',$$

so that $\mathbb{E}_t\left[v\right] \in \arg\max_{v'} u\left(v',t';\mathbb{E}_t\left[v\right],t'\right)$. Since $u\left(v',t';v,t'\right) = u\left(v',t';v,t\right)$ for all v',v,t',t and $\mathbb{E}_t\left[u(v',t';v,t)\right] = u(v',t';\mathbb{E}_t\left[v\right],t)$, we conclude that

$$\mathbb{E}_t\left[v\right] \in \arg\max_{v'}\left\{\mathbb{E}_t\left[u(v',t';v,t)\right]\right\} = \arg\max_{v'}\left\{u(v',t';\mathbb{E}_t\left[v\right],t)\right\}. \quad \blacksquare$$

Proof of Proposition 1. Let the payments $\{x\left(v,t\right) : t \in [0,T],\ v \in [\underline{v},\overline{v}]\}$ and the allocation $\{y\left(v,t\right) : t \in [0,T],\ v \in [\underline{v},\overline{v}]\}$ implement the complete-information solution, or first best. Then, the allocation must be efficient, i.e., $y(v,t) = 1$ if $v \geq c$ and $y(v,t) = 0$ otherwise, and the payments, $x(v,t)$, must extract all of the ex-ante consumer surplus, so

$$\mathbb{E}_t\left[x(v,t)\right] = \mathbb{E}_t\left[vy(v,t)\right] = \int_c^{\overline{v}} vf\left(v|t\right)dv, \forall t. \tag{21}$$

Furthermore, given $y(v,t)$ and since by Lemma 2 $\partial u(v,t)/\partial v = y(v,t)$ almost everywhere, $x(v,t)$ satisfies (IC_t) only if for some $p(t)$ the payments $x(v,t)$ satisfy

$$x(v,t) = \begin{cases} p\left(t\right) & \text{if } v \geq c, \\ p\left(t\right) - c & \text{otherwise.} \end{cases} \tag{22}$$

for all v,t.

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 types $t'' > t'$ such that

$$\mathbb{E}_{t'}\left[\max\left(v-c,0\right)\right] > \mathbb{E}_{t''}\left[\max\left(v-c,0\right)\right] \tag{23}$$

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

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

where the strict inequality follows from (23) (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 (21). Thus, type t' has an incentive to misreport her type as t'' at time zero and ($\overline{\text{IC}}_0$) 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

$$\mathbb{E}_{t''}[v] > \mathbb{E}_{t'}[\max(v,c)]. \tag{24}$$

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

$$\mathbb{E}_{t''}[vy(\mathbb{E}_{t''}[v], t') - x(\mathbb{E}_{t''}[v], t')] = \mathbb{E}_{t''}[v] - x(\mathbb{E}_{t''}[v], t'), \tag{25}$$

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

$$\begin{aligned} U\left(t'; t''\right) &= \mathbb{E}_{t''}\left[v\right] - x\left(\mathbb{E}_{t''}\left[v\right], t'\right), \\ &= \mathbb{E}_{t''}\left[v\right] - p\left(t'\right), \\ &> \mathbb{E}_{t'}\left[\max\left(v, c\right)\right] - p\left(t'\right) \\ &= \int_{c}^{\overline{v}} v f\left(v|t'\right) dv - p\left(t'\right) + cF\left(c|t'\right), \\ &= \int_{c}^{\overline{v}} v f\left(v|t'\right) dv - \mathbb{E}_{t'}\left[x\left(v, t'\right)\right]. \\ &= 0, \end{aligned}$$

where the strict inequality follows from (24) and the last line follows from (21). Thus, Condition SOV is also necessary for implementing the first-best solution.

To prove the "if" part of Proposition 1, consider the following solution to the mechanism design problem: For all t, let $y\left(v, t\right) = 1$ if $v \geq c$ and $y\left(v, t\right) = 0$ if $v < c$. Define the payments $x\left(v, t\right)$ as follows: $x\left(v, t\right) = p\left(t\right)$ for $v \geq c$ and $x\left(v, t\right) = p\left(t\right) - c$ for $v < c$ where $p\left(t\right) = \mathbb{E}_{t}\left[v; v \geq c\right] + cF\left(c|t\right)$. Then, it is straightforward to check that this solution implements the first-best under Condition IEGT and Condition SOV of Proposition 1. ■

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

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

from which it follows that

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

where the last equality follows from Lemma 2 and integration by parts. ■

Proof of Proposition 3. The right-hand side of ($\widetilde{\text{IC}}_0$) in (P^2) is negative (see discussion of monotonicity in the text), and so using Lemma 3, (12) implies $U'(t) \leq 0, \forall t$. Therefore, under Assumption 2, it must be the case that $U(T) = 0$, since otherwise the firm could increase the transfer payment by a small amount from all consumers.

Furthermore, $U'(t) \leq 0, \forall t$ implies that the solution to (P^2) must satisfy constraint ($\widetilde{\text{IC}}_0$) with equality almost everywhere because otherwise

$$U'(t) < -\int_{\underline{v}}^{\overline{v}} y(v,t) \frac{\partial F(v|t)}{\partial t} dv, \forall t \in [t_1, t_2], \tag{26}$$

for some interval $[t_1, t_2]$ and the firm could earn a strictly higher payoff choosing $\tilde{U}, y$ where $\tilde{U}'(t) = U'(t), \forall t \in [0,T] \setminus [t_1, t_2]$ and $\tilde{U}'(t) = -\int_{\underline{v}}^{\overline{v}} y(v,t) \frac{\partial F(v|t)}{\partial t} dv$, $\forall t \in [t_1, t_2]$. Profit is clearly higher because the transfers are smaller for all $t > t_1$ and none of the constraints would be violated.

So the solution to (P^2) satisfies $U(T) = 0$ and

$$U'(t) = -\int_{\underline{v}}^{\overline{v}} y(v,t) \frac{\partial F(v|t)}{\partial t} dv \quad \text{for almost every } t. \tag{27}$$

It is easy to check that (27) implies that ($\overline{\text{IC}}_0$) and ($\underline{\text{IC}_0}$) are both satisfied, so under either assumption the solution to (P^1) is the same as the solution to (P^2). Moreover, it is easy to see that this is also the solution to Courty & Li's (2000) problem. So screening on time has no added value under Assumption 2, and the optimal contract is the same as the contract when consumers learn at the same time. ■

B Proofs of the Results in Section 6

Proof of Lemma 4. To prove that $U'(t) \geq 0$ for almost every t, we argue by contradiction. Suppose U is an expected utility schedule 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

$$\begin{aligned} U(t+\varepsilon) < U(t) = \\ \mathbb{E}_t [u(\mathbb{E}_t [v], t'; v, t)] \leq \mathbb{E}_{t+\varepsilon} [u(\mathbb{E}_{t+\varepsilon} [v], t'; v, t+\varepsilon)] = U(t'; t+\varepsilon), \end{aligned} \quad (28)$$

implying that ($\underline{\text{IC}}_0$) constraint is violated for type $t+\varepsilon$, which contradicts the supposition that U is an expected utility schedule resulting from an optimal mechanism. Note that the strict inequality in (28) is true since $U'(t) < 0$ for $t \in (\tau_1, \tau_2)$. The weak inequality follows from Assumption 3 and the fact that $u(\cdot, t'; v, t)$ is nondecreasing 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 3 is satisfied and $u(\cdot, t'; v, t)$ is nondecreasing, contradicting U being an expected utility schedule 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 ($\underline{\text{IC}}_0$) and ($\overline{\text{IC}}_0$) 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 schedule is not only feasible (i.e. satisfies all IC and IR constraints) but also strictly improves the objective. Contradiction to U being a utility schedule resulting from an optimal mechanism. Hence, $U'(t) \geq 0$ for almost every t. This combined with Lemma 3 gives (18). ■

Preparation for Proof of Proposition 4. The following auxiliary optimal control problem, (P_A), its necessary and sufficient optimality conditions, and the bounds on the adjoint variables will facilitate the proof of Proposition 4.

Choose a piecewise continuous control $\{y(v) : v \in [\underline{v}, \bar{v}]\}$ and the associated continuously differentiable state vector function $\mathbf{z}(v) = (z_1(v), z_2(v))$, defined on the fixed interval $[\underline{v}, \bar{v}]$ that will

$$\max_{y(v),\mathbf{z}(v)} \int_{\underline{v}}^{\bar{v}} (v - c) f(v|t) y(v) \, dv \tag{P_A}$$

subject to

$$\begin{aligned}
&0 \leq y(v) \leq 1, \\
&\dot{z}_1(v) = -F(v|t) y(v), \\
&\dot{z}_2(v) = -\frac{\partial F(v|t)}{\partial t} y(v), \\
&z_i(\bar{v}) \geq \bar{z}_i, i = 1, 2, \\
&z_i(\underline{v}) = 0, i = 1, 2,
\end{aligned}$$

where $\bar{z}_2 > 0$. To facilitate the analysis, also define the value function $V(\bar{\mathbf{z}})$ as follows:

$$V(\bar{\mathbf{z}}) = \sup \left\{ \int_{\underline{v}}^{\bar{v}} (v - c) f(v|t) y(v) \, dv \right\}$$

where the supremum is over the admissible controls in (P_A) and $\bar{\mathbf{z}} = (\bar{z}_1, \bar{z}_2)$.

Lemma 5 *A feasible control $\{y(v) : v \in [\underline{v}, \bar{v}]\}$ for P_A with the corresponding state trajectory $\{\mathbf{z}(v) : v \in [\underline{v}, \bar{v}]\}$ is optimal if and only if there exist multipliers $\lambda_1, \lambda_2 \geq 0$ such that for all continuity points of $y(v)$*

$$y(v) \in \arg\max_{0 \leq w \leq 1} \left\{ w \left(v - c - \frac{F(v|t)}{f(v|t)} \lambda_1 - \frac{\partial F(v|t) / \partial t}{f(v|t)} \lambda_2 \right) \right\}, \tag{29}$$

and

$$\lambda_i (z_i(\bar{v}) - \bar{z}_i) = 0, i = 1, 2. \tag{30}$$

Moreover,

$$\lambda_i = \frac{\partial V(\bar{\mathbf{z}})}{\partial \bar{z}_i}. \tag{31}$$

Proof of Lemma 5. First, the existence of a feasible solution is by construction because $y_*(v,t)$ is a feasible control. Then existence of an optima control $y(\cdot)$ and the associated state trajectory $\mathbf{z}(\cdot)$ follows from Filippov-Cesari Theorem; see Theorem 2.8 of Seierstad & Sydsæter (1987).

Next, following Theorem 2.2 of Seierstad & Sydsæter (1987), we look for adjoint variables λ_0 and $\lambda_1(v), \lambda_2(v)$ for $v \in [\underline{v}, \bar{v}]$ that satisfy the necessary conditions below. The Hamiltonian is defined by

$$H(\mathbf{z}, y, \lambda, v) = \lambda_0(v-c)f(v|t)y - \lambda_1(v)F(v|t)y - \lambda_2(v)\frac{\partial F(v|t)}{\partial t}y.$$

That is,

$$H(\mathbf{z}, y, \lambda, v) = \left[\lambda_0(v-c)f(v|t) - \lambda_1(v)F(v|t) - \lambda_2(v)\frac{\partial F(v|t)}{\partial t}\right] y$$

Then by Theorem 2.2 of Seierstad & Sydsæter (1987), any optimal control must satisfy the following (at any continuity point of y):

$$y(v) \in \arg\max_{0 \le w \le 1} \left\{ w \left(\lambda_0(v-c)f(v|t) - \lambda_1(v)F(v|t) - \lambda_2(v)\frac{\partial F(v|t)}{\partial t} \right) \right\}. \tag{32}$$

Also note that $\dot{\lambda}_1(v) = \dot{\lambda}_2(v) = 0$, because the Hamiltonian is independent of the state variable, z. Thus, we conclude that

$$\lambda_i(v) = \lambda_i, \text{ for } i = 1, 2,$$

and $\lambda_0 \in 0, 1$. We also argue that $\lambda_0 = 1$. Otherwise, the optimal allocation probabilities are independent of the objective function, and in particular, independent of the unit cost. Then the "necessary" part of Lemma 5 follows from Theorem 2.8 of Seierstad & Sydsæter (1987). Moreover, Arrow's sufficiency theorem (Theorem 2.5 of Seierstad & Sydsæter (1987)) ensures that any y satisfying the necessary conditions is optimal provided that

$$\hat{H}(\mathbf{z}, \lambda, v) = \left[\lambda_0(v-c)f(v|t) - \lambda_1(v)F(v|t) - \lambda_2(v)\frac{\partial F(v|t)}{\partial t}\right]^+$$

is concave in $\mathbf{z}$ for all v. This holds trivially and proves the "sufficiency" part of Lemma 5.

Lastly, applying Theorem 3.9 of Seierstad & Sydsæter (1987) gives

$$\lambda_i = \frac{\partial V(\bar{z}_1, \bar{z}_2)}{\partial \bar{z}_i}$$

which requires the uniqueness of λ_1, λ_2 and is verified below.

For all $P_1, P_2 \geq 0$, define $\Theta(P_1, P_2)$ as

$$\Theta(P_1, P_2) = \left\{ v \in [\underline{v}, \bar{v}] : (v - c)f(v|t) - P_1 F(v|t) - P_2 \frac{\partial F(v|t)}{\partial t} > 0 \right\}.$$

Define $A \subset \mathbf{R}_+^2$ as the set of (P_1, P_2) such that $\Theta(P_1, P_2)$ has strictly positive measure. Also define the functions

$$g_1(P_1, P_2) = \int_{\Theta(P_1,P_2)} -F(v|t)dv,$$

and

$$g_2(P_1, P_2) = \int_{\Theta(P_1,P_2)} -\frac{\partial F(v|t)}{\partial t} dv.$$

Then the definition of (P_A) and equation (32) imply that:

$$z_1(\bar{v}) = g_1(\lambda_1, \lambda_2) \geq \bar{z}_1, \text{ and} \tag{33}$$

$$z_2(\bar{v}) = g_2(\lambda_1, \lambda_2) \geq \bar{z}_2. \tag{34}$$

First note that for (33) and (34) to hold we must have $(\lambda_1, \lambda_2) \in A$. Otherwise $g_2(\lambda_1, \lambda_2) = 0$, yet $\bar{z}_2 > 0$, so equation (34) is violated. Therefore, without loss of generality, we can restrict g_1, g_2 to A and view A as its domain. Also note that on A, $g_1(P_1, P_2)$ is strictly decreasing in P_1 and strictly increasing in P_2. Similarly, on A, $g_2(P_1, P_2)$ is strictly increasing in P_1 and strictly decreasing in P_2.

To show the uniqueness of the adjoint variables, consider the following four cases:

Case 1. $z_i(\bar{v}) > \bar{z}_i$ for $i = 1, 2$, which implies that $\lambda_1 = \lambda_2 = 0$ by equation (31).

Case 2. $z_1(\bar{v}) = \bar{z}_1$ and $z_2(\bar{v}) > \bar{z}_2$, which implies that $\lambda_2 = 0$ by equation (31), and λ_1 is defined by $g_1(\lambda_1, 0) = \bar{z}_1$. Since $g_1(\cdot, 0)$ is strictly decreasing, the solution is unique.

Case 3. $z_1(\bar{v}) > \bar{z}_1$ and $z_2(\bar{v}) = \bar{z}_2$, which implies that $\lambda_1 = 0$ by equation (31), and λ_2 is defined by $g_2(0, \lambda_2) = \bar{z}_2$. Since $g_2(0, \cdot)$ is strictly decreasing, the solution is unique.

Case 4. $z_1(\bar{v}) = \bar{z}_1$ and $z_2(\bar{v}) = \bar{z}_2$. We argue this case by contradiction. Suppose there are two solutions, $(\lambda_1, \lambda_2) \neq (\hat{\lambda}_1, \hat{\lambda}_2)$, both in A. Then it must be true that

$$g_1(\lambda_1, \lambda_2) = g_1(\tilde{\lambda}_1, \tilde{\lambda}_2) \tag{35}$$

and

$$g_2(\lambda_1, \lambda_2) = g_2(\tilde{\lambda}_1, \tilde{\lambda}_2) \tag{36}$$

Consider the following three sub cases of *Case 4*:

Case 4(i). $\lambda_1 > \tilde{\lambda}_1$. Then by (35), it must be that $\tilde{\lambda}_2 > \lambda_2$. Then by the monotonicity of g_2 we have that

$$g_2(\lambda_1, \lambda_2) > g_2(\tilde{\lambda}_1, \lambda_2) > g_2(\tilde{\lambda}_1, \tilde{\lambda}_2),$$

which contradicts (36).

Case 4(ii). $\lambda_1 < \tilde{\lambda}_1$. Then by (35), it must be that $\tilde{\lambda}_2 < \lambda_2$. Then by the monotonicity of g_2 we have that

$$g_2(\lambda_1, \lambda_2) < g_2(\tilde{\lambda}_1, \lambda_2) < g_2(\tilde{\lambda}_1, \tilde{\lambda}_2),$$

which contradicts (36).

Case 4(iii). $\lambda_1 = \tilde{\lambda}_1$. Then by (35), it must be that $\tilde{\lambda}_2 = \lambda_2$, which is a contradiction.

Therefore the adjoint variables (λ_1, λ_2) are unique, and Theorem 3.9 of Seierstad & Sydsæter (1987) can be applied to yield (31).

Corollary 5 *Lemma 5 implies*

$$\lambda_1 \leq \sup_{v>c} \frac{(v-c)f(v|t)}{F(v|t)}.$$

Proof of Corollary 5. From (31), it follows that

$$V(\bar{z}_1 + \epsilon, \bar{z}_2) - V(\bar{z}) \leq \epsilon \sup_{v>c} \frac{(v-c)f(v|t)}{F(v)} + o(\epsilon).$$

Hence,

$$\frac{\partial V(\bar{z})}{\partial \bar{z}_1} = \lim_{\epsilon \downarrow 0} \frac{V(\bar{z}_1 + \epsilon, \bar{z}_2) - V(\bar{z})}{\epsilon} \leq \sup_{v>c} \frac{(v-c)f(v|t)}{F(v)}.$$

Proof of Proposition 4. Let U_*, y_* (and u_* be an optimal solution to (P^2) with the corresponding ex-post utility function denoted by u_*. We will proceed by modifying this solution appropriately to obtain another solution U_*, y, which is of the desired form. In particular, we will keep the expected utility function U_* unchanged, while modifying only the allocation probabilities. This will show 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, assume y_* is not identically zero (otherwise y_* is of the desired form). Then because we restrict attention to piecewise continuous allocation probabilities, this implies $y_*(v,t) > 0$ on a subset of $[\underline{v}, \bar{v}]$ that has positive measure.

Define the modified allocation $y(v,t)$ as the solution to the following problem, denoted by $(\widetilde{\mathrm{P}}_A)$ (for notational brevity, we suppress the dependence of y and u on t):

Choose the control $\{y(v) : v \in [\underline{v}, \bar{v}]\}$ and u_0, and hence the associated ex-post utility $\{u(v) : v \in [\underline{v}, \bar{v}]\}$, to solve

$$\max_{y(v),u_0} \int_{\underline{v}}^{\bar{v}} (v-c)\, f(v|t)\, y(v)\, dv \tag{$\widetilde{\mathrm{P}}_A$}$$

subject to

$$\begin{aligned}
&0 \le y(v) \le 1, \forall v,\\
&u(\underline{v}) = u_0,\\
&\dot{u}(v) = y(v), \forall v,\\
&\int_{\underline{v}}^{\bar{v}} f(v|t)\, u(v)\, dv \ge U_*(t),\\
&\int_{\underline{v}}^{\bar{v}} \left[\frac{-\partial F(v|t)}{\partial t}\right] y(v)\, dv \ge \int_{\underline{v}}^{\bar{v}} \left[\frac{-\partial F(v|t)}{\partial t}\right] y_*(v)\, dv,\\
&u(\bar{v}) \le u_*(\bar{v}, t),
\end{aligned}$$

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 fourth constraint can be rewritten as

$$-\int_{\underline{v}}^{\bar{v}} F(v|t)\, y(v)\, dv \ge U_*(t) - u(\bar{v}). \tag{37}$$

It is crucial to observe that because u_0 is a free variable, we can combine the third, fourth, and sixth constraints in $(\widetilde{\mathrm{P}}_A)$ and replace them with the following constraint

$$-\int_{\underline{v}}^{\bar{v}} F(v|t)\, y(v)\, dv \ge U_*(t) - u_*(\bar{v}, t). \tag{38}$$

which establishes that $(\widetilde{\mathrm{P}}_A)$ is equivalent to the auxiliary problem (P_A) with

$$z_1(v) = -\int_{\underline{v}}^{v} F\left(v|t\right) y\left(v\right) dv, \tag{39}$$

$$\bar{z}_1 = U_*\left(t\right) - u_*\left(\bar{v}, t\right), \tag{40}$$

$$z_2(v) = -\int_{\underline{v}}^{v} \frac{\partial F\left(v|t\right)}{\partial t} y\left(v\right) dv, \tag{41}$$

and

$$\bar{z}_2 = -\int_{\underline{v}}^{v} \frac{\partial F\left(v|t\right)}{\partial t} y_*\left(v\right) dv. \tag{42}$$

So it follows from Lemma 5 that there exist multipliers $\lambda_1, \lambda_2 \geq 0$ such that for all continuity points of y we have the following:

$$y\left(v\right) \in \arg\max_{0\leq w\leq 1} \left\{ w\left(v - c - \frac{F\left(v|t\right)}{f\left(v|t\right)}\lambda_1 - \frac{\partial F\left(v|t\right)/\partial t}{f\left(v|t\right)}\lambda_2 \right) \right\}.$$

If the expression in parentheses is increasing, then $y(v)$ is non-decreasing and $y(v,t) \in \{0,1\}$ almost everywhere. In particular, this implies that 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., $v < r(t)$).

This expression is non-decreasing if

$$-\frac{\partial F\left(v|t\right)/\partial t}{f\left(v|t\right)}\lambda_2$$

is non-decreasing, which follows from Assumption 5.1, and if

$$v - c - \frac{F\left(v|t\right)}{f\left(v|t\right)}\lambda_1$$

is increasing, or

$$\frac{\partial}{\partial v}\left[v - \frac{F\left(v|t\right)}{f\left(v|t\right)}\lambda_1 \right] > 0.$$

So it is sufficient to show that

$$1 > \lambda_1 \frac{\partial}{\partial v}\left[\frac{F\left(v|t\right)}{f\left(v|t\right)} \right]. \tag{43}$$

By Assumption 5.2

$$\frac{\partial}{\partial v}\left[\frac{F(v|t)}{f(v|t)}\right] < \frac{1}{\sup_{v>c}\frac{(v-c)f(v|t)}{F(v|t)}},$$

and by Corollary 4 we have that

$$\lambda_1 \leq \sup_{v>c} \frac{(v-c)f(v|t)}{F(v|t)},$$

so (43) holds, and the allocation probabilities $y(v,t)$ of the desired form.

As argued above, these allocation probabilities, $\underset{\sim}{y}$, also constitute an optimal solution to $(\widetilde{\mathrm{P}}_A)$. Also note that for each solution to $(\widetilde{\mathrm{P}}_A)$, since the initial value, u_0, is free, we can decrease u_0 and make the constraint (37) bind. Hence, without loss of generality we will consider only solutions in which (37) 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^2). To establish that it is indeed an optimal solution to (P^2), we only need to check the constraint $(\underline{\mathrm{IC}}_0^T)$. To check this, note that $u(\bar{v},t) \leq u_*(\bar{v},t)$ (for all t) by the last constraint of $(\widetilde{\mathrm{P}}_A)$ 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 $(\widetilde{\mathrm{P}}_A)$, 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 $(\underline{\mathrm{IC}}_0^T)$ can be rewritten in a more transparent format as follows:

$$U_*(T) \geq \max_t \left\{u\left(\mathbb{E}_T[v], t\right)\right\}, \tag{44}$$

where we obtained (44) by rewriting the constraint ($\underline{\text{IC}}_0^T$) 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 ($\underline{\text{IC}}_0^T$) 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 ($\underline{\text{IC}}_0^T$) still holds since

$$\begin{aligned}
U_*(T) &\geq \max_t \{u_*(\mathbb{E}_T[v], t)\}, && (45)\\
&\geq \max_{t \in \{\tau : u_*(\mathbb{E}_T[v], \tau) \geq 0\}} \{u_*(\mathbb{E}_T[v], t)\}, && (46)\\
&\geq \max_{t \in \{\tau : u_*(\mathbb{E}_T[v], \tau) \geq 0\}} \{u(\mathbb{E}_T[v], t)\}, && (47)\\
&\geq \max_{t \in \{\tau : u(\mathbb{E}_T[v], \tau) \geq 0\}} \{u(\mathbb{E}_T[v], t)\}, && (48)\\
&= \max_t \{u(\mathbb{E}_T[v], t)\}, && (49)
\end{aligned}$$

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 (47) and (48) follow from the fact that $u(v, t) \leq u_*(v, t)$ for all v such that $u(v, t) \geq 0$. Finally, (49) 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^2) 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

$$U'(t) \in \left\{0, -\int_{r(t)}^{\overline{v}} \frac{\partial F(v|t)}{\partial t} dv\right\} \text{ for almost every } t. \tag{50}$$

Lemma 3 and Lemma 4 and the first half of the proof imply that

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

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

$$\begin{aligned}
\int_{v,t} f(v,t)\left[x(v,t) - cy(v,t)\right] dvdt \quad = \\
\int_{v,t} f(v,t)(v - c)\, y(v,t)\, dvdt - \int_0^{\bar{t}} (1 - H(t))\, dU(t) \quad (51)
\end{aligned}$$

To prove (50), we argue by contradiction. Suppose that instead there exists some interval $[\tau_1, \tau_2]$ such that

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

Then the objective function can be improved by setting $U'(t) = 0$ for $t \in [\tau_1, \tau_1 + \varepsilon]$ for $\varepsilon > 0$ sufficiently small and appropriately increasing $U'(t)$ (which can be done since $U'(t) < -\int_{r(t)}^{\overline{v}} \frac{\partial F(v|t)}{\partial t} dv$ for $t \in [\tau_1, \tau_2]$) such that $U(\tau_2)$ remains unchanged. This modification improves the objective in (51) since $\int_{\tau_1}^{\tau_2} (1 - H(t))\, dU(t)$ decreases; and it can be achieved by changing $u(\underline{v}, t)$ on the interval $[\tau_1, \tau_2]$ appropriately so that $y(v, t)$ remains the same. Then, the constraints of (P^2) including ($\underline{\mathrm{IC}}_0^T$) 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, (50) 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

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

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

Proof of Corollary 3. It follows from Lemma 2 and Proposition 4 that

$$y(v, t) = \begin{cases} 1 & \text{if} \quad v \geq r(t), \\ 0 & \text{if} \quad v < r(t), \end{cases} \quad \text{and} \quad \frac{\partial u(v, t)}{\partial t} = \begin{cases} 1 & \text{if} \quad v \geq r(t), \\ 0 & \text{if} \quad v < r(t), \end{cases} \tag{52}$$

from which it is straightforward to conclude that

$$x(v, t) = \begin{cases} x(\bar{v}, t) & \text{if} \quad v \geq r(t), \\ x(\underline{v}, t) & \text{if} \quad v < r(t), \end{cases} \tag{53}$$

where $x(\underline{v},t) = -u(\underline{v},t)$ and $x(\bar{v},t) = \bar{v} - u(\bar{v},t)$. Moreover, it follows from (52) 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

$$x(v,t) = \begin{cases} p(t) & \text{if} \quad v \geq r(t), \\ p(t) - r(t) & \text{if} \quad v < r(t). \end{cases} \tag{54}$$

Then, for all t,

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

Solving for $p(t)$, we obtain

$$p(t) = \mathbb{E}_t[\max\{v, r(t)\}] - U(t). \tag{55}$$

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 (54) and (55) 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 (52) implies $y(\mathbb{E}_T[v], t) = \mathbf{1}_{\{\mathbb{E}_T[v] - r(t) \geq 0\}}$, so it follows that

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

Preparation for Proof of Proposition 5. The following lemma facilitates the proof of Proposition 5.

Lemma 6 *If $r(t)$ is decreasing over $[\tau, T]$, then*

$$\begin{aligned} -\int_{\tau}^{T}\int_{r(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t}\,dvdt \quad &> \quad \mathbb{E}_T[v] - \int_{\tau}^{t'}\int_{r(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t}\,dvdt \\ &\qquad - \int_{r(t')}^{\bar{v}} (v - r(t'))f(v|t')dv - r(t') \quad \text{for } t' \in [\tau, T). \end{aligned}$$

In particular, ($\underline{\mathrm{IC}}_0^T$) is slack for all $t' \in [\tau, T)$.

Proof of Lemma 6. To see this, let $\hat{t} \in [\tau, T)$ and assume without loss of generality that $\mathbb{E}_T[v] \geq r(\hat{t})$. Note that

$$\begin{aligned} U(T) - U(\hat{t}; T) &\geq -\int_{\tau}^{T} \int_{r(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv dt + \int_{\tau}^{\hat{t}} \int_{r(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv dt \\ &\quad + \int_{r(\hat{t})}^{\bar{v}} (v - r(\hat{t})) f(v|\hat{t}) dv - \left(\mathbb{E}_T[v] + r(\hat{t})\right) \mathbf{1}_{\{\mathbb{E}_T[v] - r(t) \geq 0\}} \\ &= -\int_{\hat{t}}^{T} \int_{r(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv dt + \int_{r(\hat{t})}^{\bar{v}} (v - r(\hat{t})) f(v|\hat{t}) dv \\ &\quad - \left(\mathbb{E}_T[v] + r(\hat{t})\right) \mathbf{1}_{\{\mathbb{E}_T[v] - r(t) \geq 0\}}. \end{aligned}$$

If $\mathbb{E}_T[v] - r(\hat{t}) < 0$, then the first and second terms are positive and the third term is zero, so the constraint is slack for all $\hat{t} \in [\tau, T)$. If $\mathbb{E}_T[v] \geq r(\hat{t})$ then we can rewrite the above equation as

$$\begin{aligned} U(T) - U(\hat{t}; T) &\geq -\int_{\hat{t}}^{T} \int_{r(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv dt - \int_{\underline{v}}^{\bar{v}} (v - r(\hat{t})) f(v|T) dv \\ &\quad + \int_{r(\hat{t})}^{\bar{v}} (v - r(\hat{t})) f(v|\hat{t}) dv \\ &= -\int_{\hat{t}}^{T} \int_{r(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv dt - \int_{r(\hat{t})}^{\bar{v}} (v - r(\hat{t}))(f(v|T) - f(v|\hat{t})) dv \\ &\quad - \int_{\underline{v}}^{r(\hat{t})} (v - r(\hat{t})) f(v|T) dv. \end{aligned}$$

Integrating the second term on the right-hand side by parts, we have

$$\begin{aligned} U(T) - U(\hat{t}; T) &\geq -\int_{\hat{t}}^{T} \int_{r(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv dt + \int_{r(\hat{t})}^{\bar{v}} [F(v|T) - F(v|\hat{t})] dv \\ &\quad - \int_{\underline{v}}^{r(\hat{t})} (v - r(\hat{t})) f(v|T) dv \\ &= -\int_{\hat{t}}^{T} \int_{r(t)}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv dt + \int_{\hat{t}}^{T} \int_{r(\hat{t})}^{\bar{v}} \frac{\partial F(v|t)}{\partial t} dv dt \\ &\quad - \int_{\underline{v}}^{r(\hat{t})} (v - r(\hat{t})) f(v|T) dv > 0. \end{aligned}$$

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, so the constraint is slack for all $\hat{t} \in [\tau, T)$. ■

Proof of Proposition 5. We further relax the firm's problem, P^4, by assuming that the constraint does not bind for all $t \in [\tau, T]$. We show that this implies 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 6 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 on each. Denoting $r(t)$ by $r_1(t)$ on $[0, \tau]$ and by $r_2(t)$ on $[\tau, T]$, we relax the constraint by assuming that ($\underline{\mathrm{IC}}_0^T$) does not bind on $[\tau, T]$. Moreover, from Assumption 4, $\mathbb{E}_T[v] > c$ and hence $\mathbb{E}_T[v] \geq r(t)$ for $t \in [0, \tau)$ since otherwise decreasing $r(t)$ improves the objective and does not violate ($\underline{\mathrm{IC}}_0^T$).

So the firm's problem can be written as

$$\max_{\{\tau \in [0,T], r_1(t), r_2(t)\}} \int_0^\tau \int_{r_1(t)}^{\bar{v}} f(v,t)(v-c)dvdt + \int_\tau^T \int_{r_2(t)}^{\bar{v}} f(v,t)\phi(v,t)dvdt \tag{56}$$

subject to

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

where

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

Note that relaxing the constraint is without loss of generality by Lemma 6 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

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

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

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

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

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

Clearly $\sigma^*(\tau, r_2)$ is unique, given τ and r_2, since Assumption 3 (FOSD) implies that $\int_c^{\bar{v}}(v-c)f\left(v|t\right)dv = \bar{v} - c - \int_c^{\bar{v}} F\left(v|t\right)dv$ is strictly increasing in t. It follows that for all $t \in (\sigma^*(\tau, r_2), \tau)$, ($\underline{\text{IC}}_0^T$) does not bind and that $r_1(t) = r_1^*(t; \sigma^*(\tau, r_2)) = c$. Moreover, for all $t \le \sigma^*(\tau, r_2)$, ($\underline{\text{IC}}_0^T$) binds and $r_1(t) = r_1^*(t; \sigma^*(\tau, r_2))$ is uniquely defined by

$$\begin{aligned}\mathbb{E}_T[v] - \int_{r_1^*(t;\sigma^*(\tau,r_2))}^{\bar{v}} (v - r_1^*(t;\sigma^*(\tau,r_2)))f\left(v|t\right)dv - r_1^*(t;\sigma^*(\tau,r_2)) \\ = \mathbb{E}_T[v] - \int_c^{\bar{v}}(v-c)f(v|\sigma^*(\tau,r_2))dv - c, \quad (57)\end{aligned}$$

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

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

$$\left.\frac{\partial \sigma^*(\tau, r_2+\epsilon z)}{\partial \epsilon}\right|_{\epsilon=0} = -\left[\int_c^{\bar{v}}(v-c)\frac{\partial f(v|\sigma^*(\tau,r_2))}{\partial t}dv\right]^{-1}\int_{\tau}^{T}\frac{\partial F(r_2(t)|t)}{\partial t}z(t)dt, \tag{58}$$

and

$$\frac{\partial \sigma^*(\tau, r_2)}{\partial \tau} = -\left[\int_c^{\bar{v}}(v-c)\frac{\partial f(v|\sigma^*(\tau,r_2))}{\partial t}dv\right]^{-1}\int_{r_2(\tau)}^{T}\frac{\partial F(v|\tau)}{\partial t}dv > 0, \tag{59}$$

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 (57) we have

$$\frac{\partial r_1^*(t;\sigma)}{\partial \sigma} = \frac{\int_c^{\bar{v}}(v-c)\frac{\partial f(v|\sigma)}{\partial t}dv}{F(r_1^*(t;\tau,r_2)|t)}$$

which is clearly positive.

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:

$$\max_{\{\tau, r_2(t)\}} \int_0^{\sigma^*(\tau,r_2)} \int_{r_1^*(t;\sigma^*(\tau,r_2))}^{\bar{v}} f(v,t)(v-c)dvdt \\ + \int_{\sigma^*(\tau,r_2)}^{\tau} \int_c^{\bar{v}} f(v,t)(v-c)dvdt + \int_\tau^T \int_{r_2(t)}^{\bar{v}} f(v,t)\phi(v,t)dvdt.$$

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

$$- \int_0^{\sigma^*(\tau,r_2)} f(r_1^*(t';\sigma^*(\tau,r_2)),t')(r_1^*(t';\sigma^*(\tau,r_2)) - c) \\ \frac{\partial r_1^*(t';\sigma^*(\tau,r_2))}{\partial \sigma} \left. \frac{\partial \sigma^*(\tau, r_2 + \epsilon z)}{\partial \epsilon} \right|_{\epsilon=0} dt' \\ - \int_\tau^T f(r_2(t),t)\phi(r_2(t),t)z(t)dt = 0, \forall z(t),$$

and

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

The first-order conditions imply that $0 < \tau^* < T$. To see this, first notice that $\tau^* = 0$ implies that $\sigma^*(\tau,r_2) = 0$, so the first term in the second condition is zero and hence the second condition fails to hold with equality, so clearly $\tau^* > 0$. 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 condition fails to hold with equality, so clearly $\tau^* < T$.

Also, note that if $\sigma^*(\tau,r_2) = \tau$, then the first condition implies $\phi(r_2^*(t),t) = 0$. From the implicit function theorem, $dr_2^*(t)/dt = -\phi_t(r_2^*(t),t)/\phi_v(r_2^*(t),t)$. From Assumption 6, it follows that $r_2^*(t)$ is decreasing, and from the Lemma 6 above, this implies the constraint is slack at τ^* which implies $\sigma^* < \tau^*$, which is a contradiction. So the solution to the first order conditions must satisfy $\sigma^*(\tau,r_2) \in (0,\tau)$.

So using (58) and (59) the first-order conditions can be written as:

$$\int_0^{\sigma^*(\tau,r_2)} \frac{f(r_1^*(t';\sigma^*(\tau,r_2)),t')(r_1^*(t';\sigma^*(\tau,r_2))-c)}{F(r_1^*(t';\sigma^*(\tau,r_2))|t')} \int_\tau^T \frac{\partial F(r_2(t)|t)}{\partial t} z(t)dtdt' - \int_\tau^T f(r_2(t),t)\phi(r_2(t),t)z(t)dt = 0, \forall z(t),$$

or

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

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

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

or

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

or simply $\Phi(r_2(t),t;\tau,r_2)=0$ for all $t\in[\tau,T]$ where

$$\Phi(v,t;\tau,r_2) = v-c+(1-H(t)-\Psi(\tau,r_2))\frac{\partial F(v|t)/\partial t}{f(v,t)}. \tag{60}$$

Substituting for $\phi(v,\tau)$, 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 $\int_{r_2(\tau)}^{c}(v-c)f(v,\tau)dv \leq 0$, 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 (60), clearly, $1 - H(\tau) - \Psi(\tau, r_2) \leq 0$ combined with $\Phi(r_2(\tau), \tau; \tau, r_2) = 0$ implies that $r_2^*(\tau) \leq c$.

Also, note that $1 - H(\tau) - \Psi(\tau, r_2) \leq 0$ implies $1 - H(t) - \Psi(\tau, r_2) < 0$ for all $t > \tau$, so $\Phi(r_2(t), t; \tau, r_2) = 0$ implies that $r_2^*(t) < c$ for all $t > \tau$.

Finally, we prove that $dr_2^*(t)/dt < 0$. Using the implicit function theorem, $\Phi(r_2(t), t; \tau, r_2) = 0$ implies that for any solution $r_2^*(t)$,

$$\frac{dr_2(t)}{dt} = -\frac{\Phi_t(r_2(t), t; \tau, r_2)}{\Phi_v(r_2(t), t; \tau, r_2)} \tag{61}$$

where

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

and

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

Under Assumption 5.1, the denominator in (61) is positive. Under Assumption 6, the first term in the numerator of (61) is positive and under Assumption 3 (FOSD) the second term is positive, so the numerator of (61) is positive and it follows that $dr_2^*(t)/dt < 0$. So from Lemma 6, 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 $\hat{U}(\tau + \epsilon) = U(\tau + \epsilon)$ is unchanged, but so that $\hat{U}$ increases more slowly and over a longer interval than U. Specifically, for some $\nu < \tau$, $U(\nu) = 0$ and

$$\hat{U}'(t) = -\int_{c}^{\overline{v}} \frac{\partial F(v|t)}{\partial t} dv \tag{62}$$

while $U(\tau) = 0$ and

$$U'(t) = -\int_{r_1(t)}^{\overline{v}} \frac{\partial F(v|t)}{\partial t} dv, \tag{63}$$

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

$$\int_{\nu}^{\tau+\epsilon} \left(\hat{U}(t) - U(t)\right) h(t) dt, \tag{64}$$

but total surplus increases by

$$\int_{\tau}^{\tau+\epsilon} \int_{r_2(t)}^{c} (c - v) f(v, t) dv dt, \tag{65}$$

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

It also must be true that $r_1(\sigma) = c$. Suppose not. Since $U(t) = 0, \forall t \in [0, \tau]$, and $U(T) = U(t; T), \forall t \in [0, \sigma]$, that is, the global incentive compatibility constraint binds, and $U(T) > U(t; T)$, $\forall t \in (\sigma, \tau)$, that is, the global incentive compatibility constraint does not bind. So $U(T) = U(\sigma; T)$, and using $U(t) = 0$ and the definition of $U(t; T)$ in Corollary (3), this implies

$$U(T) = (\mathbb{E}_T[v] - r(\sigma)) \mathbf{1}_{\{\mathbb{E}_T[v] - r(\sigma) \geq 0\}} - \int_{\sigma}^{\overline{v}} (v - r(\sigma)) f(v|\sigma) dv, \tag{66}$$

But then clearly

$$\lim_{\epsilon \downarrow 0} U(\sigma + \epsilon; T) = (\mathbb{E}_T[v] - c) \mathbf{1}_{\{\mathbb{E}_T[v] - c \geq 0\}} - \int_{c}^{\overline{v}} (v - c) f(v|\sigma) dv, \tag{67}$$

is greater than $U(T)$, since $r(\sigma) > c$, which implies that the right hand side of (67) is greater than the that of (66). But this is a contradiction (for t slightly larger than σ the incentive compatibility constraint is violated). ■

Proof of Proposition 6. 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, cf. Proposition 5, the individual rationality constraints (IR) are satisfied. The feasibility constraints (F) are readily satisfied. Moreover, from Lemma 2, the incentive compatibility constraint ($\widetilde{\mathrm{IC}}_t$) 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

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

The incentive compatibility constraints ($\underline{\mathrm{IC}}_0$) 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 ($\underline{\mathrm{IC}}_0$) 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

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

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

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

as well and $U(t';t) \leq U(t)$ for all $t > t'$. Hence ($\underline{\mathrm{IC}}_0$) is satisfied.

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

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

We can rewrite this as

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

So

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

since $r'(t') \leq 0$ and $U'(t') \leq -\int_{r(t')}^{\overline{v}} \frac{\partial F(v|t')}{\partial t'} dv$, and since Assumption 3 holds. Integrating, we get $U(t';t) \leq U(t)$ for all $t' > t$ since $U(t;t) = U(t)$, which shows that $(\overline{\mathrm{IC}}_0)$ is satisfied.

Hence, the menu of expiring refund contracts $\{(p(t), r(t)) : 0 \leq t \leq T\}$ is optimal for (P^1). ■

Proof of Corollary 4. For all t,

$$
U(t) = \int_{r(t)}^{\bar{v}} v f(v|t)\, dv - p(t)(1 - F(r(t)|t)) + (p(t) - r(t)) F(r(t)|t). \tag{68}
$$

Solving for $p(t)$ in (68) we get for all t,

$$
p(t) = \mathbb{E}_t [v; v \geq r(t)] + r(t) F(r(t)|t) - U(t) = \mathbb{E}_t [\max\{v, r(t)\}] - U(t). \tag{69}
$$

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 Corollary 3, $\mathbb{E}_t[\max\{v, r(t)\}]$ 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 3 (FOSD) we obtain

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

For $t \geq \tau$, taking the derivative of both sides of (69) and using Proposition 4 we get

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

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

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

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 (69), $U(T) = \mathbb{E}_T[\max\{v, r(T)\}] - p(T)$ 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. ■