GOVERNANCE AND THE EFFICIENCY OF ECONOMIC SYSTEMS
GESY

Discussion Paper No. 487

Pricing a Package of Services - When (not) to bundle

Felix Ketelaar *
Dezsö Szalay **

* University of Bonn
** University of Bonn and CEPR

December 2014

Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

Sonderforschungsbereich/Transregio 15 · www.sfbtr15.de
Universität Mannheim · Freie Universität Berlin · Humboldt-Universität zu Berlin · Ludwig-Maximilians-Universität München
Rheinische Friedrich-Wilhelms-Universität Bonn · Zentrum für Europäische Wirtschaftsforschung Mannheim

Speaker: Prof. Dr. Klaus M. Schmidt · Department of Economics · University of Munich · D-80539 Munich,
Phone: +49(89)2180 2250 · Fax: +49(89)2180 3510

PRICING A PACKAGE OF SERVICES - WHEN (NOT) TO BUNDLE

Felix Ketelaar and Dezső Szalay
University of Bonn — University of Bonn, CEPR

December 2014

Abstract. We study a tractable two-dimensional model of price discrimination. Consumers combine a rigid with a more flexible choice, such as choosing the location of a house and its quality or size. We show that the optimal pricing scheme involves no bundling if consumer types are affiliated. Conversely, if consumer types are negatively affiliated over some portion of types then some bundling occurs.

JEL classification: D42, D82, D86

Keywords: Price discrimination, Bundling, Monopoly, Multidimensional screening

1. Introduction

Many important decisions in our lives involve choices among bundles and trade-offs between several taste dimensions. Constructing a house is presumably one of the most important such instances. First and foremost, the location of the house has to be chosen. Several choices ranging from the number of floors, the number of rooms on each floor, construction materials, and on to the very last details of the interior decor follow the first decision. Some of these choices are extremely flexible and so involve marginal adjustments. Other choices are arguably more rigid; typically, only a very limited number of alternative locations are available at any given time. We are interested in the design of pricing schemes in such situations featuring a combination of flexible and rigid choices. Moreover, as in our leading example, the rigid choice has important consequences both in terms of utility and in terms of costs.

We thank the seminar audience at the SFB-TR15 conference 2014 (Mannheim) for comments, in particular David Martimort. This paper is a spin-off of an earlier working paper by Szalay (Szalay (2013)) that has benefitted from the comments of Johannes Hörner, Marco Ottaviani, Yossi Spiegel, and seminar audiences at HEC Lausanne, LBS, LSE, University of Bonn, University of Frankfurt, Northwestern University, PSE, ESEM Gothenburg, and ESSET Gerzensee. The authors acknowledge financial support through DFG grant SFB-TR15. Send communications to felixketelaar@gmail.com or to szalay@uni-bonn.de or by ordinary mail to Dezső Szalay, Institute for Microeconomics, University of Bonn, Adenauerallee 24-42, 53113 Bonn, Germany.

A natural concern one may have when making such important choices is "not to give away too much". Will a real estate developer adjust the price for constructing our house condition on whether the house is located in a posh or a middle class area? Intuition - at least ours - suggests that this would make a lot of sense.

We study this question in a stylized model involving choices along two margins only: the location and the quality (or equivalently, the size) of a house. The answer contradicts our naïve intuition. In the very case where consumers with a taste for living in the posh area are likely to be ones who appreciate higher quality houses more in the sense of affiliated taste parameters, the real estate developer does not "exploit" consumers in the posh area at the optimum. The optimal marginal price for increases in the quality of the house is exactly the same, whether the house is located in the posh or the middle class area. In the terminology of the literature, there is no bundling.

The intuition is as follows. The seller wishes to extract informational rents from the consumer. The consumer has two pieces of information, his marginal valuation for moving to a posh area and his marginal valuation for a slightly nicer house. Unfortunately, from the seller's perspective, extracting rents from poshness tastes comes at the cost of extracting rents from quality tastes. Clearly, if the consumer could be forced to choose the location of his house based on his tastes for poshness only, then the seller would definitely adjust the marginal prices for constructing the house upwards in the posh area. Given consumers in the posh area are more likely to have higher valuations for nicer houses, the usual trade-off between extracting rents from quality tastes versus efficiency of the quality allocation is resolved more in favour of rent extraction and hence marginal prices for quality are higher. However, when the tastes for the area are unobservable, consumers with a taste for poshness facing the above marginal prices for quality would have a strict incentive to live in the middle class area. To make moving into the posh area more attractive, the seller has to lower the marginal prices for increases in the quality of houses. The surprising element is that the optimal selling procedure involves *no flexibility at all* to condition marginal prices on location choice.

To understand the complete absence of flexilibity, consider any type who is just indifferent between living in the posh and the middle class area. Affiliation implies that in an optimal pricing scheme, all consumers with higher tastes for quality must also be indifferent between living in either area. This indifference condition directly forces the quality choices of these consumers to be equal and hence requires that they face the same marginal prices. Thus, facing affiliated types, the seller's flexibility to condition marginal prices for quality on the tastes for poshness is confined to consumers with low tastes for quality if at all. To gain any flexibility even with that portion of consumers, the seller must leave rents to consumers buying the lowest quality house in the posh area, which implies an increase in rents to *all* consumers locating in the posh area. The gain from the adjustment of marginal prices for qualities simply does not outweigh this cost. In short, our result can be understood as a characterization of the set of consumers who are indifferent between entering the market and staying out: all consumers with the lowest taste for quality, both among those with and without a taste for poshness, are indifferent with respect to participating. Given this indifference at the optimum, any attempt to bundle that is desirable from the seller's point of view would violate incentive compatibility with respect to the location choice, i.e. give consumers with a taste for poshness a strict incentive to live in the middle class area.

We also establish a converse to our no-bundling result. If consumers' tastes are locally negatively affiliated - that is, tastes for living in the middle class area and for nicer houses are affiliated - then the optimum will necessarily display some bundling, at least for a strictly positive mass of consumers. Indeed, the seller now would ideally want to (locally) decrease marginal prices for higher qualities for consumers in the posh area relative to the middle class area, a change that consumers with a taste for poshness would welcome. On the other hand, consumers who don't value living in a posh area have a strict incentive to buy a house in the middle-class area when marginal prices for nicer houses are the same in both areas. Hence, there is now some flexibility to adjust marginal prices suitably.

This is a paper on a model of multi-dimensional screening. The seminal references are Armstrong (1996) and Rochet and Choné (1998). Armstrong (1996) solves the multiproduct monopoly problem and shows that at the optimum typically some types are excluded if the type space is convex. Our problem involves convex types for the size of houses but large taste differences with respect to poshness of the area. As a result, we do not get exclusion. Rochet and Choné (1998) solve a very general problem and establish robust features of solutions; in particular, they confirm that optimal allocations generally feature exclusion of some types and show on top that optimal allocations also involve bunching over portions of types. Our problem is much simpler than these problems. The reason is that the second best optimal allocation of consumers to areas is immediate in our problem and so the name of the game is simply to choose optimal marginal prices for quality conditional on location choices. The optimum involves both bunching and separation: the allocation of consumers to areas separates consumers based on their tastes for poshness but is independent of their tastes for quality; the allocation of quality separates consumers with respect to their tastes for quality but is independent of their tastes for poshness.

Bundling was first analyzed by Adams and Yellen (1976), who showed by example that bundling can be profitable if tastes are negatively correlated. McAfee et al. (1989) establish sufficient conditions for bundling to increase profits in the Adams and Yellen (1976) model. Their conditions are consistent with weakly negative correlation, but they emphasize that the correlation of types is not the appropriate measure. In these approaches, the set of availabe mechanisms is restricted to prices of bundles, an approach that is extended by Manelli and Vincent (2007). Manelli and Vincent (2006) study more generally revenue maximizing mechanisms for a firm selling N objects. As in Thanassoulis (2004), the optimal mechanism may involve randomization. More recently, Hart and Reny (2014) show that revenue optimal pricing schemes may have surprising features; in particular, the seller's revenue can decrease if the buyer's multidimensional valuation increases - something that cannot happen in dimension one; moreover, they provide new examples where stochastic mechanisms are optimal. Armstrong and Rochet (1999) characterize the optimal mechanism in a model with two goods and two taste parameters on binary supports. They show that no bundling occurs for the case of strong positive correlation. We obtain no bundling even for slightly positive correlations - when types are affiliated. The reason lies precisely in the structure of our optimal location allocation, which is different from the one in Armstrong and Rochet (1999).

Our problem differs from the approaches taken in Manelli and Vincent (2006, 2007), Thanassoulis (2004), and Hart and Reny (2014) in two ways. First, these papers study revenue maximization whereas in our model there are substantial costs of production. In the presence of such costs - which seems reasonable in our introductory example - profit maximization and revenue maximization are different objectives. Though seemingly innocuous, this property allows us to reduce the dimensionality of our two-dimensional problem right away to one dimension. As a result, our optimum does not involve any randomization, which is not trivial to rule out in higher dimensional problems. The second difference is that we combine one inflexible choice (an either-or-choice) with a more flexible one. The inflexible choice is the same as the choices studied in the work by Manelli and Vincent (2006,2007) and, more recently, by Armstrong (2013).[1] We combine this choice of the seller with a more flexible one - any non-negative size of a house - as in the approaches of Armstrong (1996) and Rochet and Choné (1998).

Our model is so stylized that our problem becomes amenable to essentially uni-dimensional methods. In particular, our design problem boils down to choosing a pair of uni-dimensional schedules, where the consumer's choice of where to locate can be treated as a type dependent outside option. Type dependent outside options are studied in Jullien (2000). However, the difference to Jullien (2000) is that the outside option is endogenous, resulting from the optimal design of the scheme for consumers locating in the middle-class area. Similar techniques are also used in the countervailing incentives context by Maggi and Rodriguez-Clare (1995). Our approach is related to Kleven et al. (2009), who study the design of tax schemes for couples and singles and give conditions for bundling in the tax context: different marginal taxes at the same income level for couples and for singles. The institutional details as well as our approach and assumptions are quite different. [2] However, the common element is a combination of a discrete with a flexible choice. This gives rise to a design problem that remains nicely tractable despite its multidimensional nature.

2. The Model

A risk-neutral seller (she) wants to sell a package of two goods. The first good is divisible and its quantity (or quality) is labeled $x \in \mathbb{R}_{\geq 0}$. The second good is indivisible and we write $q \in [0,1]$ for the probability of selling the second good. The seller faces a risk-neutral buyer (he) whose valuation for the bundle of goods is given as

$$V(x, q, \theta, \eta) = \theta x + \eta q$$

where $\theta \in [\underline{\theta}, \overline{\theta}]$, $0 < \underline{\theta} < \overline{\theta} < \infty$ and $\eta \in \{\underline{\eta}, \overline{\eta}\}$, $0 < \underline{\eta} < \overline{\eta}$ are preference parameters that are private knowledge of the buyer. The seller only knows the distribution $F(\theta, \eta)$ of the buyer's preference parameters. We assume that F has a continuously differentiable probability density function $f(\theta, \eta)$ which is strictly positive everywhere on $[\underline{\theta}, \overline{\theta}] \times \{\underline{\eta}, \overline{\eta}\}$. To shorten notation we write $\beta = Pr(\eta = \underline{\eta})$

[1] Armstrong (2013) studies bundling when there are demand complementarities or substitutabilities and restricts his analysis to deterministic mechanisms. We allow for more flexible allocations and mechanisms but stick to additively separable valuations.

[2] The objective in the tax context (redistribution) differs from ours (profit maximization); while tax payers can be forced to participate, consumers cannot. Finally, we allow for statistical dependence among types and characterize direct mechanisms (also allowing for stochastic mechanisms).

and $1-\beta = Pr(\eta = \overline{\eta})$. Moreover, the buyer has an outside option of buying nothing which earns him a utility of zero.

For given values $x \geq 0$ and $q \in [0,1]$ the seller faces (expected) production costs

$$C(x,q) = C(x) + c \cdot q$$

where $C(x)$ denotes the costs for producing quantity x of good one and $c > 0$ denotes the constant marginal cost of producing the second good. We assume that $C(x)$ is increasing, twice continuously differentiable and strictly convex in x with $C(0) = 0$, $\lim_{x \searrow 0} \frac{\partial C}{\partial x}(x) = 0$ and $\lim_{x \nearrow \infty} \frac{\partial C}{\partial x}(x) = +\infty$. Production costs are known to the seller.

The seller aims to maximize his expected surplus from selling a bundle (or, more generally, a lottery over bundles) to the buyer given as

$$\Pi = \mathbb{E}\left[p(x,q) - C(x,q)\right]$$

where $p = p(x,q)$ denotes the (lottery over) prices for the bundle or the lottery (x,q).[3] Invoking the revelation principle (see e.g. Myerson (1982)) we can think of the seller's pricing problem as a direct mechanism where the buyer communicates his type (θ,η) and in return is offered a lottery over allocations $(x(\theta,\eta), q(\theta,\eta))$ together with a lottery over prices $p(\theta,\eta)$, subject to incentive compatibility and participation constraints.

A buyer (θ,η) who reports type $\left(\hat{\theta},\hat{\eta}\right)$ receives an expected utility of

$$U\left(\hat{\theta},\theta,\hat{\eta},\eta\right) = \theta x\left(\hat{\theta},\hat{\eta}\right) + \eta q\left(\hat{\theta},\hat{\eta}\right) - p\left(\hat{\theta},\hat{\eta}\right).$$

If reports coincide with true types we write $u(\theta,\eta) = U(\theta,\theta,\eta,\eta)$. Moreover, we write $\underline{u}(\theta)$, $\overline{u}(\theta)$, $\underline{x}(\theta)$, $\overline{x}(\theta)$ for $u\left(\theta,\underline{\eta}\right)$, $u(\theta,\overline{\eta})$, $x\left(\theta,\underline{\eta}\right)$, $x(\theta,\overline{\eta})$ respectively and $\underline{u} = \underline{u}(\underline{\theta})$, $\overline{u} = \overline{u}(\underline{\theta})$.

Writing

$$\Pi(\theta,\eta) = p(\theta,\eta) - C(x(\theta,\eta), q(\theta,\eta))$$

for her expected profit from type (θ,η), the seller seeks to maximize

$$\Pi = \mathbb{E}_{\theta,\eta}\Pi(\theta,\eta)$$

subject to incentive compatibility and participation[4]

$$u(\theta,\eta) \geq U\left(\hat{\theta},\theta,\hat{\eta},\eta\right) \quad \forall \theta,\hat{\theta} \in \left[\underline{\theta},\overline{\theta}\right], \eta,\hat{\eta} \in \left\{\underline{\eta},\overline{\eta}\right\}, \tag{1}$$

$$u(\theta,\eta) \geq 0 \quad \forall \theta \in \left[\underline{\theta},\overline{\theta}\right], \eta \in \left\{\underline{\eta},\overline{\eta}\right\}. \tag{2}$$

3. Analysis

To solve our problem, we proceed as follows. We begin with a discussion of the first-best and establish a connection between the first-best and the second-best that simplifies our problem dramatically. We then characterize implementable allocations and reformulate our problem in a more tractable way.

---

[3] Slightly abusing notation, we do not distinguish here notationswise between deterministic variables $x \geq 0$ and $p \geq 0$ and lotteries over these variables. As we shall prove in Lemma 1, only deterministic allocations will be relevant for our analysis.

[4] The seller may want to exclude certain types from participation. However, exclusion in the sense of buyers who prefer the outside option is outcome equivalent to offering the zero trade ($x = 0$, $q = 0$, $p = 0$) to these buyers and hence included in the optimization problem.

3.1. **First-best Efficiency and the Optimal q-Allocation.** When the buyer's preference parameters (θ, η) are common knowledge, the seller can perfectly discriminate between customers and would extract the full surplus by setting prices equal to

$$p(\theta, \eta) = \theta x(\theta, \eta) + \eta q(\theta, \eta)$$

where $x(\theta, \eta)$ and $q(\theta, \eta)$ are chosen efficiently, i.e.

$$C_x(x(\theta, \eta)) - \theta = 0$$

and

$$q(\theta, \eta) = \begin{cases} 0 & \eta < c \\ 1 & \eta \geq c. \end{cases}$$

In particular, $x(\theta, \eta) = x(\theta)$ is independent of η and $q(\theta, \eta) = q(\eta)$ is independent of θ.

In this paper we are interested in the case where production costs play a substantial role. We therefore assume that the efficient allocation involves allocating product q to the high preference type only and impose for the remainder of the paper

Assumption 1.

$$\underline{\eta} < c < \overline{\eta}.$$

The following lemma shows that with relevant production costs in the above sense the optimal mechanism is deterministic. Moreover, the efficient q-allocation determined by Assumption 3 will also be implemented in the optimal mechanism for the constrained problem with asymmetric information.

Lemma 1. *Under Assumption 1, the optimal mechanism is deterministic in x and q and separates η-types in q efficiently, that is $q(\theta, \overline{\eta}) = 1$, $q\left(\theta, \underline{\eta}\right) = 0$ for all $\theta \in \left[\underline{\theta}, \overline{\theta}\right]$.*

First best efficiency of the allocation in q translates into second best optimality for the following two reasons. Clearly, efficiency is desirable for the seller as it maximizes her profits for any given surplus of a buyer. At the same time the buyer's incentive constraints under the efficient allocation are as relaxed as they can possibly be. Indeed, the excess surplus of a buyer of type $\overline{\eta}$ mimicking type $\underline{\eta}$ compared to a buyer who truly is of type $\underline{\eta}$ is minimized, being equal to 0, while the excess loss of a buyer of type $\underline{\eta}$ mimicking type $\overline{\eta}$ compared to a buyer who truly is of type $\overline{\eta}$ is maximized, being equal to $\overline{\eta} - \underline{\eta}$. Clearly, the resulting allocation is deterministic in q. Since consumer valuations are linear in x and the seller has convex costs, the mechanism is deterministic in x by a standard result (cf. e.g. Fudenberg and Tirole (1991)).

Lemma 1 shows that with relevant production costs all potential distortions will occur in the x-dimension only. This finding has two important consequences. First, it makes the problem sufficiently tractable to derive explicit solutions, contrasting many other multidimensional setups. Secondly, it allows us to directly relate and compare all effects that arise from the presence of a second dimension to the well-known solution of the standard one-dimensional problem.

3.2. **Implementable Allocations.** In this section we bring the incentive constraints (1) into a more tractable form to solve the seller's problem. The key tool is the following lemma which allows us to split the two-dimensional incentive compatibility constraints into two one-dimensional constraints.

Lemma 2. *For any mechanism featuring* $q(\theta,\overline{\eta})=1$, $q\left(\theta,\underline{\eta}\right)=0$ *for all* $\theta\in\left[\underline{\theta},\overline{\theta}\right]$, *thus in particular for the optimal mechanism in our maximization problem, the incentive constraint*

$$u(\theta,\eta) \quad\geq\quad U\left(\hat{\theta},\theta,\hat{\eta},\eta\right) \qquad \forall\theta,\hat{\theta}\in\left[\underline{\theta},\overline{\theta}\right],\ \eta,\hat{\eta}\in\left\{\underline{\eta},\overline{\eta}\right\}$$

is equivalent to the pair of one-dimensional incentive constraints

$$u(\theta,\eta)\geq U\left(\hat{\theta},\theta,\eta,\eta\right) \quad \forall\theta,\hat{\theta}\in\left[\underline{\theta},\overline{\theta}\right],\ \eta\in\left\{\underline{\eta},\overline{\eta}\right\}, \tag{3}$$

$$u(\theta,\eta)\geq U(\theta,\theta,\hat{\eta},\eta) \quad \forall\theta\in\left[\underline{\theta},\overline{\theta}\right],\ \eta,\hat{\eta}\in\left\{\underline{\eta},\overline{\eta}\right\}. \tag{4}$$

The main insight behind Lemma 2 is that the buyer's incentive how to report his preference parameter θ does not depend on his preference parameter η. To see this, compare a buyer of type (θ,η) who considers reporting $\left(\hat{\theta},\hat{\eta}\right)$ with a buyer of type $(\theta,\hat{\eta})$ who considers the same report. The difference in utilities between type (θ,η) and type $(\theta,\hat{\eta})$ when reporting $\left(\hat{\theta},\hat{\eta}\right)$ is given as $q\left(\hat{\theta},\hat{\eta}\right)(\eta-\hat{\eta})$. But this term does not depend on $\hat{\theta}$ as $q\left(\hat{\theta},\hat{\eta}\right)=q(\hat{\eta})$. In particular, it takes the same value if $\hat{\theta}=\theta$, so it is non-positive by constraint (4). As a consequence, since type $(\theta,\hat{\eta})$ does not have an incentive to misreport as $\left(\hat{\theta},\hat{\eta}\right)$ by (3), neither does type (θ,η).

The optimal quality allocation characterized in Lemma 1 has the properties named in Lemma 2. We can therefore replace the general incentive compatibility constraint (1) in our maximization problem by the two one-dimensional constraints (3), (4). The following Lemma characterizes the implications of these two constraints for our maximization problem.

Lemma 3. *The incentive constraint* (3) *is satisfied if and only if*

$$u(\theta,\eta)=u\left(\underline{\theta},\eta\right)+\int_{\underline{\theta}}^{\theta}x(y,\eta)\,dy \tag{5}$$

and $x(\theta,\eta)$ *is non-increasing in* θ *for all* $\eta\in\left\{\underline{\eta},\overline{\eta}\right\}$. *The incentive constraint* (4) *is satisfied if and only if*

$$q\left(\theta,\underline{\eta}\right)\left(\overline{\eta}-\underline{\eta}\right)\leq\overline{u}(\theta)-\underline{u}(\theta)\leq q(\theta,\overline{\eta})\left(\overline{\eta}-\underline{\eta}\right)$$

for any $\theta\in\left[\underline{\theta},\overline{\theta}\right]$.

Lemma 3 is a standard result. For the reader's convenience, we give a full proof in the appendix. Note that Lemma 3 allows us to write prices as

$$\begin{aligned} p(\theta,\eta) &= \theta x(\theta,\eta)+\eta q(\theta,\eta)-u(\theta,\eta) \\ &= \theta x(\theta,\eta)+\eta q(\theta,\eta)-u\left(\underline{\theta},\eta\right)-\int_{\underline{\theta}}^{\theta}x(y,\eta)\,dy \end{aligned}$$

and thus to eliminate them from the optimization problem. Moreover, as an immediate consequence of Lemma 3 and non-negativity of $x\in\mathbb{R}_{\geq0}$, all participation constraints $u(\theta,\eta)\geq 0$, $(\theta,\eta)\in\left[\underline{\theta},\overline{\theta}\right]\times\left\{\underline{\eta},\overline{\eta}\right\}$, are implied by $\underline{u}=u\left(\underline{\theta},\underline{\eta}\right)\geq 0$ and incentive compatibility.

3.3. **The Pricing Problem.** The results derived in the previous section enable us to state the seller's problem in a tractable way. We write

$$B(x,q,\theta,\eta)=\theta x+\eta q-C(x)-cq-x\cdot\frac{1-F(\theta|\eta)}{f(\theta|\eta)}$$

for the seller's virtual surplus conditional on η and

$$\rho(\theta,\underline{u},\overline{u})=\overline{u}(\theta)-\underline{u}(\theta)=\overline{u}-\underline{u}+\int_{\underline{\theta}}^{\theta}\left[\overline{x}(y)-\underline{x}(y)\right]dy$$

for the excess rent of a type $(\theta,\overline{\eta})$ over a type $\left(\theta,\underline{\eta}\right)$. Substituting transfers, applying integration by parts and invoking Lemmas 1-3, the seller's optimization problem reads

$$\begin{aligned}\max_{\overline{x}(\cdot),\underline{x}(\cdot),\underline{u},\overline{u}}\quad & \Pi\left(\underline{x},\overline{x},\underline{u},\overline{u}\right)\\ &=\beta\int_{\underline{\theta}}^{\overline{\theta}}B\left(\underline{x}(\theta),0,\theta,\underline{\eta}\right)f\left(\theta|\underline{\eta}\right)d\theta-\beta\underline{u}\\ &\quad+(1-\beta)\int_{\underline{\theta}}^{\overline{\theta}}B\left(\overline{x}(\theta),1,\theta,\overline{\eta}\right)f\left(\theta|\overline{\eta}\right)d\theta-(1-\beta)\overline{u}\end{aligned}\tag{6}$$

subject to

$$\underline{u}\geq 0\tag{7}$$

$$\rho(\theta,\underline{u},\overline{u})\geq 0\quad\forall\theta\in\left[\underline{\theta},\overline{\theta}\right]\tag{8}$$

$$\rho(\theta,\underline{u},\overline{u})\leq\overline{\eta}-\underline{\eta}\quad\forall\theta\in\left[\underline{\theta},\overline{\theta}\right]\tag{9}$$

$$\overline{x}(\theta),\underline{x}(\theta)\qquad\text{non-decreasing in }\theta\tag{10}$$

$$\overline{x}(\theta),\underline{x}(\theta)\geq 0\quad\forall\theta\in\left[\underline{\theta},\overline{\theta}\right].\tag{11}$$

We refer to this problem as Problem P. An immediate observation is that setting $\underline{u}=0$ is optimal, avoiding any lump-sum rents for the low valuation types in the η-dimension. Indeed, at $\underline{\theta}$ constraint (8) implies $\overline{u}\geq\underline{u}$, so a positive $\underline{u}$ implies a positive $\overline{u}$. Since both constraints (8) and (9) only depend on $\overline{u}-\underline{u}$ and both, $\overline{u}$ and $\underline{u}$, are costly for the seller, choosing $\underline{u}$ as small as possible is optimal. Thus, to simplify notation, we write $\rho(\theta,\underline{u},\overline{u})=\rho(\theta,\overline{u})$ henceforth.

In what follows we will approach the above optimization problem with control-theoretic methods. As we demonstrate in the next section, under the additional assumption of affiliated preference types this approach allows us to reduce the problem to a simple one-dimensional optimization task, maximizing the objective $\Pi(\overline{u})$ as a function of the lump-sum rents $\overline{u}$ for high η-types only. The reader who, when reading first, is mainly interested in the rents-vs-bundling trade-off involved in this maximization may skip through the following discussion and jump directly to Section 6.

4. Solution: a Control-Theoretic Characterization

To make Problem P more tractable, we first show that the optimal x-schedules are reasonably regular.

Lemma 4. *The schedules* $(\underline{x}^*(\theta), \overline{x}^*(\theta))$ *that solve Problem P are continuous.*

The main intuition for Lemma 4 is ubiquitous in economics: $B(x, q, \theta, \eta)$ is strictly concave in x and concavity favors smoothing. A formal proof which ensures that smoothing near a putative discontinuity is compatible with constraints (8) and (9) is provided in the appendix.

To solve optimization problems like Problem P, it is a standard approach to consider a reduced problem with less constraints first and then impose (distributional) assumptions that guarantee the remaining constraints to be satisfied. Problem P without the monotonicity and non-negativity constraints (10) and (11) can be regarded as a control problem with state variables $\underline{u}(\theta)$, $\overline{u}(\theta)$, control variables $\underline{x}(\theta) = \dot{\underline{u}}(\theta)$, $\overline{x}(\theta) = \dot{\overline{u}}(\theta)$ and two inequality constraints that involve the state variables $\underline{u}(\theta)$, $\overline{u}(\theta)$ only. Call this reduced problem P'.

Problem P' is still relatively complex, yet solution techniques are available, e.g. from Seierstad and Sydsaeter (1987). Fixing some value $\overline{u}$ for the moment, Problem P' gives rise to a Hamiltonian

$$\begin{aligned} H &= H(\underline{u}(\theta), \overline{u}(\theta), \underline{x}(\theta), \overline{x}(\theta), \underline{\kappa}(\theta), \overline{\kappa}(\theta), \theta) \\ &= \beta B(\underline{x}(\theta), 0, \theta, \underline{\eta}) f(\theta|\underline{\eta}) + (1-\beta) B(\overline{x}(\theta), 1, \theta, \overline{\eta}) f(\theta|\overline{\eta}) \\ &\quad + \underline{\kappa}(\theta) \cdot \underline{x}(\theta) + \overline{\kappa}(\theta) \cdot \overline{x}(\theta). \end{aligned}$$

with two costate variables $\underline{\kappa}(\theta)$, $\overline{\kappa}(\theta)$ as well as an associated Lagrangian

$$L = H + \mu_1(\theta) \cdot (\overline{u}(\theta) - \underline{u}(\theta)) + \mu_2(\theta) \cdot \left((\overline{\eta} - \underline{\eta}) - (\overline{u}(\theta) - \underline{u}(\theta)) \right)$$

where $\mu_1(\theta)$ and $\mu_2(\theta)$ denote the multipliers associated to constraints (8) and (9).

A frequent issue in control-theoretic problems with constraints on the state variables lies in the fact that often neither costate variables nor control variables need to satisfy standard regularity conditions. However, since we know that the control schedules must be continuous to solve Problem P, we may restrict attention to continuous control schedules (which immediately implies continuous costate schedules) for Problem P' as well.

To proceed towards a solution of the control problem, it is helpful to make ourselves aware of some special features of constraints (8) and (9) as well as the Hamiltonian H. Indeed, note that H does not depend on $\overline{u}(\theta)$ and $\underline{u}(\theta)$ at all while (8) and (9) only depend on $\overline{u}(\theta) - \underline{u}(\theta)$ rather than each state variable individually. As a consequence, the Lagrange equations

$$\begin{aligned} \frac{\partial \underline{\kappa}(\theta)}{\partial \theta} &= -\frac{\partial L}{\partial \underline{u}}, \\ \frac{\partial \overline{\kappa}(\theta)}{\partial \theta} &= -\frac{\partial L}{\partial \overline{u}} \end{aligned}$$

rewrite as

$$\dot{\underline{\kappa}}(\theta) = \mu_1(\theta) - \mu_2(\theta) = -\dot{\overline{\kappa}}(\theta).$$

In the appendix we show that $\mu_1(\theta)$ and $\mu_2(\theta)$ are sufficiently regular to apply the fundamental theorem of calculus, so together with transversality $\underline{\kappa}(\overline{\theta}) = 0 = \overline{\kappa}(\overline{\theta})$ the above equations imply $\underline{\kappa}(\theta) = -\overline{\kappa}(\theta)$.[5]

The reduction from two costate variables to one is at the core of the following proposition which is based on a result in Seierstad and Sydsaeter (1987).

[5]See the proof of Proposition 1 in the appendix for the technical details.

Proposition 1. *An optimal continuous allocation for the reduced problem P' is characterized by the following pair of equations*

$$\left(-C_x\left(\underline{x}^*(\theta)\right)+\theta-\frac{1-F\left(\theta|\underline{\eta}\right)}{f\left(\theta|\underline{\eta}\right)}\right)\cdot\beta f\left(\theta|\underline{\eta}\right)+\kappa^*(\theta)=0 \tag{12}$$

$$\left(-C_x\left(\overline{x}^*(\theta)\right)+\theta-\frac{1-F\left(\theta|\overline{\eta}\right)}{f\left(\theta|\overline{\eta}\right)}\right)\cdot(1-\beta)f\left(\theta|\overline{\eta}\right)-\kappa^*(\theta)=0 \tag{13}$$

where $\kappa^*(\theta)$ denotes the optimal costate variable. The optimal costate variable has the following properties:

a) $\kappa^*(\theta)$ is continuous.

b) $\kappa^*(\theta)$ is locally constant around all $\theta\in\left[\underline{\theta},\overline{\theta}\right]$ at which neither (8) *nor* (9) *binds.*

c) At any $\theta\in\left(\underline{\theta},\overline{\theta}\right)$ where (8) *or* (9) *binds,*

$$\kappa^*(\theta)=\kappa_b(\theta)\equiv(1-\beta)f\left(\theta|\overline{\eta}\right)\cdot\left(\frac{1-F(\theta)}{f(\theta)}-\frac{1-F\left(\theta|\overline{\eta}\right)}{f\left(\theta|\overline{\eta}\right)}\right) \tag{14}$$

so that $\underline{x}^*(\theta)=\overline{x}^*(\theta)$.

d) $\kappa^*(\theta)$ is weakly increasing whenever (8) *binds and weakly decreasing whenever* (9) *binds.*

e) $\kappa^*\left(\overline{\theta}\right)=0$.

Most of the properties of the optimal costate variable $\kappa(\theta)$ in Proposition 1 are either standard in control theory or easy to see.[6] Part a) is a direct consequence of Lemma 4. Part b) is the usual type of result in a control-theoretic problem: the costate variable is the dynamic "integrated" equivalent to a Lagrange multiplier in a static optimization problem. Part c) follows directly from Lemma 3: marginal utilities must be equal for both η-types whenever (8) or (9) bind at an interior point as otherwise the constraints would be violated "slightly to the left" or "slightly to the right". Part e) is the standard transversality condition for control problems with free endpoints, implying the classical "no-distortion-at-the-top".

Part d), finally, constitutes a powerful tool to determine the areas of binding constraints.[7] Indeed, whenever (8) or (9) bind at an interior value θ we must have $\underline{x}^*(\theta)=\overline{x}^*(\theta)$ according to Part c). The corresponding $\kappa(\theta)$-schedule easily computes from equations (12) and (13) as $\kappa_b(\theta)$ where the subscript "b" stands for "bunching" of x in η. For any interval where $\kappa_b(\theta)$ is monotonic, Part d) of Proposition 1 leaves only one of the two constraints (8) and (9) as potentially binding. Moreover, neither of the two constraints can bind around a local extremum of $\kappa_b(\theta)$. Note that $\kappa_b\left(\overline{\theta}\right)=0$, so $\kappa_b(\theta)$ is compatible with transversality.

We find the following heuristic argument useful to understand the result. From the seller's point of view, the costate variable $\kappa(\theta)$ (or, more precisely, its absolute value) measures the additional distortion of the optimal x-allocation compared to the one-dimensional case with known η-types. This distortion is caused by the buyer's option to misreport his η-type, captured by constraints (8) and (9). The

[6] We omit stars for κ- and x-schedules as well as values of $\overline{u}$ whenever we discuss potential candidates for the optimal mechanism. The relation between κ- and x-schedules is nonetheless assumed to be defined through equations (12) and (13).

[7] Its proof goes back to Neustadt (1976).

seller clearly prefers $\kappa(\theta)$ to be equal to zero (and $\overline{u}=0$). In this case, both x-schedules maximize the values of the integrals in (6) pointwise and the schedules coincide with the solution of the two one-dimensional allocation problems in x conditional on $\eta=\underline{\eta}$ and $\eta=\overline{\eta}$, respectively. We therefore refer to this scenario as the case of (quasi-)observable η. It constitutes an upper bound on what is potentially achievable for the seller: she fully exploits the information over the θ-type contained in the η-type at no costs. Yet, the solution for observable η may not be incentive compatible as the resulting x- and u-schedules may violate constraints (8) or (9).

At the other extreme, consider the x-schedule that constitutes the solution to the one-dimensional problem unconditional on η defined by

$$-C_x(x(\theta))+\theta-\frac{1-F(\theta)}{f(\theta)}=0.$$

Setting $\underline{x}(\theta)=\overline{x}(\theta)=x(\theta)$ for all $\theta\in\left[\underline{\theta},\overline{\theta}\right]$ corresponds to setting $\kappa(\theta)=\kappa_b(\theta)$ everywhere. In this scenario, (8) and (9) are automatically satisfied as the utility of high and low η-types coincides for all θ and we may certainly set $\overline{u}=0$. However, this candidate solution comes at a cost: the seller does not exploit the information about θ that is contained in η at all. It thereby constitutes a lower bound on what is achievable for the seller.

We illustrate the previous discussion in Figure 1 and Figure 2. Both figures show a pair of x-schedules that corresponds to the case of observable η. In Figure 1, the schedules are clearly infeasible for our problem if $\overline{u}=0$. Indeed, on $[\underline{\theta},\theta_1]$ the x-schedule of the low η-type lies above the x-schedule of the high η-type, thereby violating (8) as

$$\rho(\theta,\overline{u}=0)=\int_{\underline{\theta}}^{\theta}\left[\overline{x}(y)-\underline{x}(y)\right]dy<0,$$

for any $\theta\in(\underline{\theta},\theta_1)$. Hence some distortion through $\kappa(\theta)$ or some positive lump-sum rent $\overline{u}$ is necessary. On the other hand, if the roles of both schedules are reversed as depicted in Figure 2, constraint 8 is clearly satisfied everywhere as

$$A=\int_{\underline{\theta}}^{\theta_1}\left[\overline{x}(\theta)-\underline{x}(\theta)\right]d\theta>0$$

and

$$A=\int_{\underline{\theta}}^{\theta_1}\left[\overline{x}(\theta)-\underline{x}(\theta)\right]d\theta>\int_{\theta_1}^{\overline{\theta}}\left[\underline{x}(\theta)-\overline{x}(\theta)\right]d\theta=B.$$

The allocation corresponding to the observable η-case is hence feasible if and only if constraint (9) holds everywhere, i.e. if

$$A=\rho(\theta_1,\overline{u}=0)\leq\overline{\eta}-\underline{\eta}.$$

The level and shape of the optimal x-allocation depends crucially on the joint distribution of types. In the next section, we impose more structure in that respect.

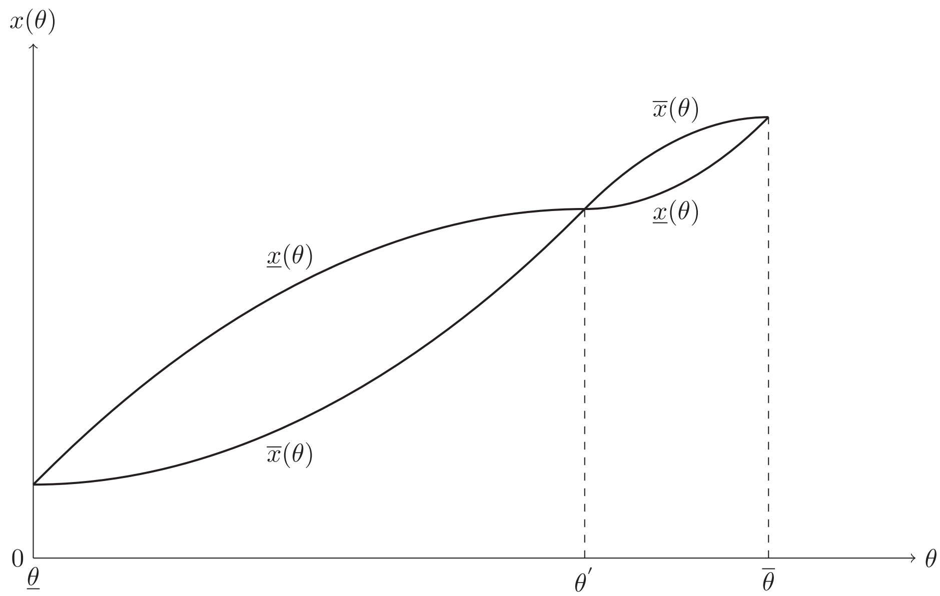

Figure 1: x-schedules for the observable-η-case are infeasible.

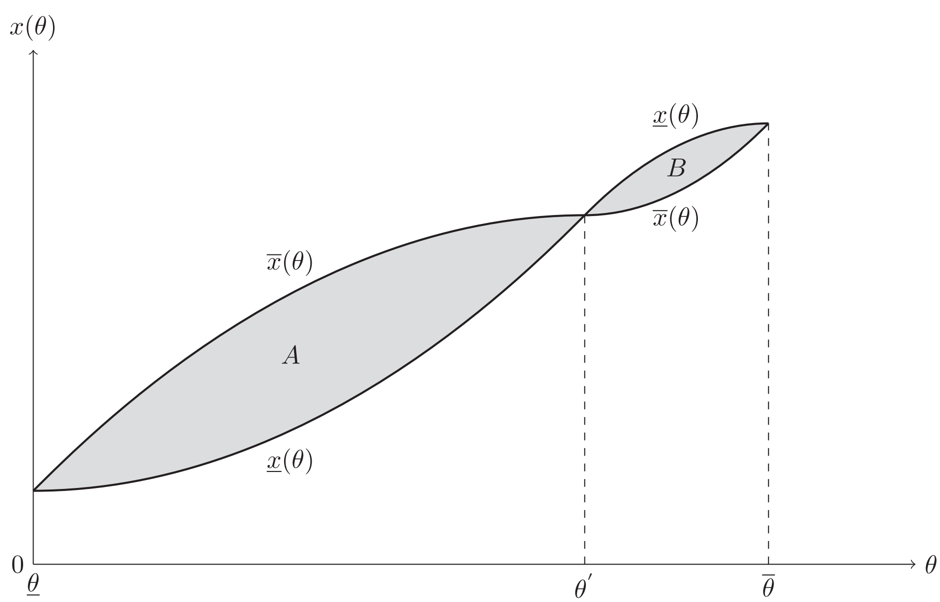

Figure 2: x-schedules for the observable-η-case are feasible iff $A \leq \overline{\eta} - \underline{\eta}$.

5. A Characterization for Affiliated Types

In this section we analyze the implications of Proposition 1 for the solution of Problem P' under the assumption that preference types are affiliated. We fully characterize the solution up to the choice of $\overline{u}$ which is studied in Section 6.

We consider continuously differentiable densities with full support. Hence affiliation is equivalent to

Assumption 2.

$$\frac{\partial}{\partial\theta}\left[\frac{f\left(\theta|\overline{\eta}\right)}{f\left(\theta|\underline{\eta}\right)}\right] > 0 \quad \forall\theta \in \left[\underline{\theta}, \overline{\theta}\right].$$

With affiliated types, it is easy to see that the seller cannot implement the schedules corresponding to observable η characterized by $\kappa\left(\theta\right) = 0$ and $\overline{u} = 0$. Affiliation implies the reversed hazard rate order

$$\frac{1 - F\left(\theta|\overline{\eta}\right)}{f\left(\theta|\overline{\eta}\right)} > \frac{1 - F\left(\theta|\underline{\eta}\right)}{f\left(\theta|\underline{\eta}\right)}$$

for any $\theta < \overline{\theta}$, cf. Shaked and Shanthikumar (2007). Setting $\kappa\left(\theta\right) = 0$ for all θ then implies that $\underline{x}\left(\theta\right) > \overline{x}\left(\theta\right)$ for any $\theta < \overline{\theta}$ and hence, together with $\overline{u} = 0$, a violation of the constraint $\rho\left(\theta, \overline{u}\right) \geq 0$ for any $\theta > \underline{\theta}$. The seller therefore faces a trade-off. To relax the constraint $\rho\left(\theta, \overline{u}\right) \geq 0$ she could either leave higher rents to the high η-types by increasing $\overline{u}$ or she could distort the optimal schedules away from the observable-η-case by choosing $\kappa\left(\theta\right)$ different from zero.

Affiliation allows us to pin down the optimal bunching region for x in η as a function of $\overline{u}$.

Proposition 2. *For given $\overline{u} \in \left[0, \overline{\eta} - \underline{\eta}\right]$, under affiliation there exists $\theta' \in \left[\underline{\theta}, \overline{\theta}\right]$ such that for all $\theta \geq \theta'$ the optimal schedules satisfy $\overline{x}^*\left(\theta\right) = \underline{x}^*\left(\theta\right) = x^*\left(\theta\right)$ where $x^*\left(\theta\right)$ is defined by*

$$-C_x^1\left(x^*\left(\theta\right)\right) + \theta - \frac{1 - F\left(\theta\right)}{f\left(\theta\right)} = 0, \tag{15}$$

implying

$$\kappa^*\left(\theta\right) = \kappa_b\left(\theta\right) \ \forall\theta \geq \theta'.$$

For all $\theta < \theta'$ the optimal schedules $\underline{x}^*\left(\theta\right)$ and $\overline{x}^*\left(\theta\right)$ are defined by

$$\left(-C_x^1\left(\underline{x}^*\left(\theta\right)\right) + \theta - \frac{1 - F\left(\theta|\underline{\eta}\right)}{f\left(\theta|\underline{\eta}\right)}\right) \cdot \beta f\left(\theta|\underline{\eta}\right) + \kappa^* = 0 \tag{16}$$

$$\left(-C_x^1\left(\overline{x}^*\left(\theta\right)\right) + \theta - \frac{1 - F\left(\theta|\overline{\eta}\right)}{f\left(\theta|\overline{\eta}\right)}\right) \cdot \left(1 - \beta\right) f\left(\theta|\overline{\eta}\right) - \kappa^* = 0 \tag{17}$$

for some constant $\kappa^* = \kappa^*\left(\overline{u}\right)$ such that

$$\overline{u} + \int_{\underline{\theta}}^{\theta'} \left[\overline{x}^*\left(\theta, \kappa^*\left(\overline{u}\right)\right) - \underline{x}^*\left(\theta, \kappa^*\left(\overline{u}\right)\right)\right] d\theta = 0, \tag{18}$$

and

$$\kappa^* = \kappa_b\left(\theta'\right) \tag{19}$$

by continuity of $\kappa^*\left(\theta\right)$.

Proposition 2 states that there is a single point θ' at which constraint (8) switches from slack to binding and the x-schedules switch from separation in η to bunching in η. Heuristically, the argument is very simple. Consider any point $\tilde{\theta}$ at which (8) is binding, implying that $\overline{x}^*\left(\tilde{\theta}\right) = \underline{x}^*\left(\tilde{\theta}\right)$ at that point. Such a point must exist, since otherwise the seller could increase profits by reducing $\overline{u}$. Then, (8) is satisfied slightly to the right of $\tilde{\theta}$, say for $\tilde{\theta} + \varepsilon$ for ε small but positive if and only if $\overline{x}^*(\theta) \geq \underline{x}^*(\theta)$ for $\theta = \tilde{\theta} + \varepsilon$. However, this requires that $\kappa^*(\theta)$ increases at that point. To see this, suppose κ were constant around $\tilde{\theta}$. Then, totally differentiating (16) and (17) with respect to θ reveals that $\underline{x}^*$ increases faster in θ than $\overline{x}^*$ at points $\tilde{\theta}$ where $\overline{x}^*\left(\tilde{\theta}\right) = \underline{x}^*\left(\tilde{\theta}\right)$ if and only if types are locally affiliated around $\tilde{\theta}$.

We now relate this heuristic argument more formally to our previous analysis. Taking derivatives of (14) yields

$$\text{sign}\frac{\partial \kappa_b(\theta)}{\partial \theta} = \text{sign}\frac{\partial}{\partial \theta}\left[\frac{f(\theta|\overline{\eta})}{f(\theta|\underline{\eta})}\right] \quad \forall \theta < \overline{\theta}, \tag{20}$$

so together with $\kappa_b\left(\overline{\theta}\right) = 0$ this shows that under affiliation $\kappa_b(\theta)$ is strictly increasing and non-positive everywhere as depicted in Figure 3.

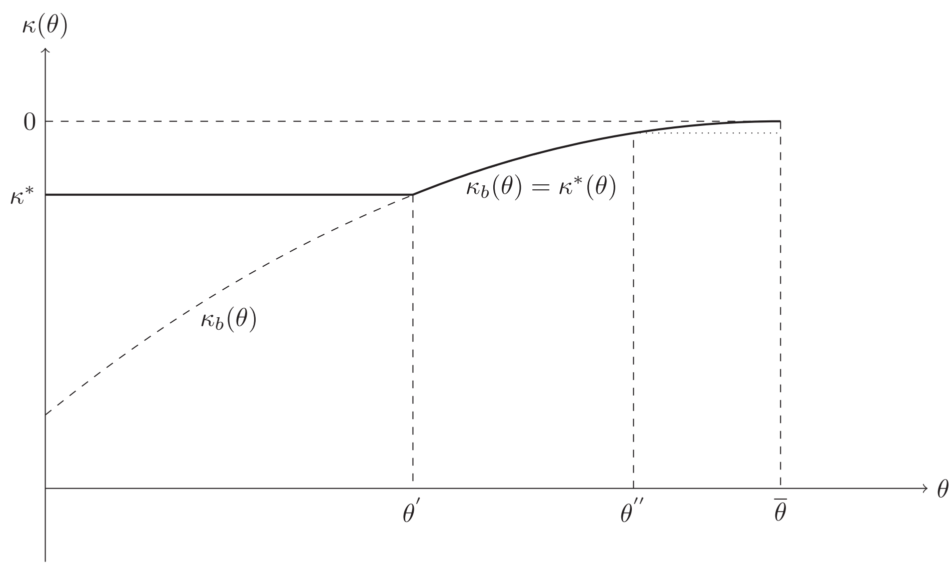

Figure 3: Following the schedule $\kappa_b(\theta)$ to the right of θ' minimizes distortions.

Leaving the schedule $\kappa_b(\theta)$ in favour of a constant κ-schedule at some $\theta'' > \theta'$ as indicated in Figure 3 through the dotted line is clearly suboptimal then. Informally speaking, it deliberately increases distortions through $\kappa(\theta)$ on the interval $\left[\theta'', \overline{\theta}\right]$ relative to following the schedule $\kappa_b(\theta)$. Formally, this is reflected in a violation of the transversality condition $\kappa^*(\theta) = 0$ as stated in Proposition 1, Part e). For all technical details, we refer to the appendix.

6. A No-Bundling Result

Proposition 2 boils the complex control problem P' down to a one-dimensional optimization problem in the choice parameter $\overline{u}$. Reformulating Proposition 2 in that spirit, we get

Proposition 2*. *Under affiliation, Problem P' is equivalent to the following Problem P". Maximize*

$$\begin{aligned}
\Pi\left(\overline{u}\right) = \beta & \int_{\underline{\theta}}^{\theta'\left(\overline{u}\right)} B\left(\underline{x}^*\left(\theta,\kappa^*\left(\overline{u}\right)\right),0,\theta,\underline{\eta}\right) f\left(\theta|\underline{\eta}\right) d\theta \\
& + \left(1-\beta\right) \int_{\underline{\theta}}^{\theta'\left(\overline{u}\right)} B\left(\overline{x}^*\left(\theta,\kappa^*\left(\overline{u}\right)\right),1,\theta,\overline{\eta}\right) f\left(\theta|\overline{\eta}\right) d\theta \\
& + \beta \int_{\theta'\left(\overline{u}\right)}^{\overline{\theta}} B\left(x^*\left(\theta\right),0,\theta,\underline{\eta}\right) f\left(\theta|\underline{\eta}\right) d\theta \qquad (21) \\
& + \left(1-\beta\right) \int_{\theta'\left(\overline{u}\right)}^{\overline{\theta}} B\left(x^*\left(\theta\right),1,\theta,\overline{\eta}\right) f\left(\theta|\overline{\eta}\right) d\theta \\
& - \left(1-\beta\right)\overline{u}.
\end{aligned}$$

subject to

$$\overline{u} \in \left[0, \overline{\eta} - \underline{\eta}\right]$$

where $\underline{x}^*\left(\theta\right), \overline{x}^*\left(\theta\right), x^*\left(\theta\right), \kappa^*\left(\overline{u}\right), \theta'$ *are defined by equations* (15)-(19).

The trade-off underlying the optimal choice of $\overline{u} \in \left[0, \overline{\eta} - \underline{\eta}\right]$ is depicted in Figure 4.

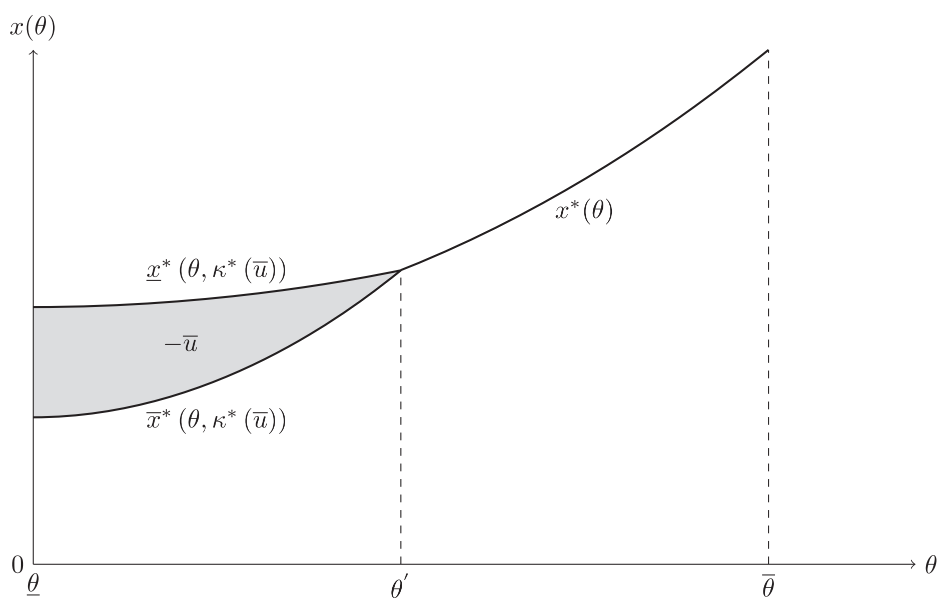

Figure 4: Raising $\overline{u}$ pushes θ' to the right and enables the seller to separate a larger portion of types.

Leaving higher rents $\overline{u}$ to the high η-types comes at a twofold gain. By moving κ^* upwards and hence closer towards zero, distortions of the x-schedules relative to the case of observable η in the separation region are reduced. By moving θ' to the right the separation region itself is enlarged. The first-order effect through marginally shifting θ', however, is zero by continuity of the optimal schedules at θ' together with an envelope argument and hence negligible.

By (18), the increase of κ^* in $\overline{u}$ is measured by

$$\frac{d\kappa^*}{d\overline{u}}\left(\overline{u}\right) = -\frac{1}{\int_{\underline{\theta}}^{\theta'}\left[\frac{\partial\overline{x}}{\partial\kappa^*}\left(\theta\right) - \frac{\partial\underline{x}}{\partial\kappa^*}\left(\theta\right)\right]d\theta} > 0. \tag{22}$$

Increasing κ^* shifts the x-schedules closer towards the case of observable η-types on $\left[\underline{\theta},\theta'\right]$ and reduces the excess rents of high η-types given by $\int_{\underline{\theta}}^{\theta'}\left[\overline{x}^*\left(\theta\right) - \underline{x}^*\left(\theta\right)\right]d\theta$. Using the above formula (22) as well as equations (12) and (13), we get

$$\frac{d\Pi}{d\overline{u}}\left(\overline{u}\right) \quad = \quad -\kappa^*\left(\overline{u}\right) - \left(1-\beta\right),$$

hence these conducive effects are measured precisely by $\left|\kappa^*\left(\overline{u}\right)\right| = -\kappa^*\left(\overline{u}\right)$. Moreover, we have just argued that

$$\frac{d^2\Pi}{d\overline{u}^2}\left(\overline{u}\right) = -\frac{d\kappa^*}{d\overline{u}}\left(\overline{u}\right) < 0,$$

so our problem is concave in $\overline{u}$. Therefore, setting $\overline{u}^* > 0$ is optimal if and only if increasing $\overline{u}$ away from zero is optimal. However, (14) implies that

$$-\kappa^*\left(\overline{u}=0\right) = -\kappa_b\left(\underline{\theta}\right) = \left(1-\beta\right)\cdot\left(1-\frac{f\left(\underline{\theta}|\overline{\eta}\right)}{f\left(\underline{\theta}\right)}\right) < 1-\beta.$$

Therefore the gain from moving the schedules closer to the observable-η-case can never compensate for the direct loss from leaving higher rents to high η-types through $\overline{u}$ measured by the share size $1-\beta$ of high η-types. Hence $\overline{u}^* = 0$ solves Problem P".

To ensure that the solution to Problem P' and P" is monotonic and non-negative, it suffices to impose standard assumptions on the inverse hazard rate:

Assumption 3. *The distribution $F = F\left(\theta\right)$ features strictly positive virtual valuations $\theta - \frac{1-F(\theta)}{f(\theta)}$ with strictly positive derivative $\frac{\partial}{\partial\theta}\left[\theta - \frac{1-F(\theta)}{f(\theta)}\right] > 0$ for all $\theta \in \left[\underline{\theta},\overline{\theta}\right]$.*[8]

Our main result is now a direct consequence of the previous analysis.

Theorem 1. *Under Assumptions 1-3, the optimal mechanism for the seller involves no lump-sum rents to high η-types, i.e. $\overline{u}^* = 0$. The optimal allocations are given as*

$$q^*\left(\theta,\eta\right) = q^*\left(\eta\right) = \begin{cases} 0 & \eta = \underline{\eta} \\ 1 & \eta = \overline{\eta} \end{cases}$$

[8]For Theorem 1, the slightly weaker condition of a non-negative and non-decreasing virtual valuation is sufficient. In the next section, however, it will be helpful to impose slightly stricter conditions as stated in Assumption 3.

and $x^*(\theta,\eta) = x^*(\theta)$, where $x^*(\theta)$ solves

$$-C_x^1(x^*(\theta)) + \theta - \frac{1-F(\theta)}{f(\theta)} = 0.$$

Prices are given as

$$p^*(\theta,\eta) = \theta x^*(\theta) + \eta q^*(\eta) - \int_{\underline{\theta}}^{\theta} x^*(y)\,dy.$$

Conditioning prices on allocations rather than buyers' types by setting $p^*(x,q) = p^*(x(\theta), q(\eta))$, optimal prices split into two additively separable price components

$$p_1^*(x) + p_2^*(q)$$

for the two goods. No bundling occurs.

The seller is not willing to leave rents to high η-types in order to buy the ability to condition x-allocations on η under affiliation. Rather, the optimal mechanism involves complete bunching of the x-schedules with respect to η. Just as the optimal q-allocation only depends on η by Lemma 1, the optimal x-allocation only depends on θ. The quantity schedule $x^*(\theta)$ has the familiar features. There is no distortion for buyers with valuation $\overline{\theta}$, there is a downward distortion for all types with valuation below $\overline{\theta}$, and there is no (lump-sum) rent at $\underline{\theta}$ for either η-type. The schedule coincides with the solution for the one-dimensional problem unconditional on η.

The reformulation at the end of Theorem 1 is a direct implication of the taxation principle (see e.g. Rochet (1985)). It allows us to rewrite prices p^* as conditional on allocations rather than preference types. Optimal prices are given as the buyer's valuation for quantity x plus the buyer's valuation for good q minus rents of the buyer. The rents of the buyer, however, do not depend on his η-type but only on his θ-type given that the excess rents $\rho^*(\theta, \overline{u} = 0)$ of high η-types over low η-types are equal to zero for all θ. Hence, rents of the buyer do not depend on his q-allocation but only on his x-allocation and

$$p^*(x, q = 1) - p^*(x, q = 0) = \overline{\eta}$$

for any quantity x. Optimal prices can therefore be split into two additively separable components

$$p^*(x,q) = p_1^*(x) + p_2^*(q)$$

where p_1^* denotes the price for good one and p_2^* denotes the price for good two, the latter being equal to zero for $q = 0$ by normalization and equal to $\overline{\eta}$ for $q = 1$.

7. Beyond Affiliation

The previous two sections have been devoted to the analysis of the optimal pricing scheme for affiliated preference types, showing that no bundling is optimal. In this final section, we show the converse result: whenever types are not (weakly) affiliated, there exists an interval of positive mass where in the optimum x-schedules are separated in η and hence bundling occurs.

Building on our previous analysis we directly state our result.

Theorem 2. *Under Assumptions 1 and 3, the solution of Problem P features no bundling if and only if types are weakly affiliated.*

The “if”-part of Theorem 2 has been covered in the two previous sections, noting that all proofs go through for weak affiliation as well. Showing the opposite direction consists of two steps.

First, we argue that the solution of the reduced problem P’ features bundling on an interval of positive mass whenever types are not weakly affiliated. To see this, note that no bundling implies $\kappa^*(\theta) = \kappa_b(\theta)$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$ as otherwise, by continuity of $\kappa^*(\theta)$ and Proposition 1c), there will be an interval of positive length where $\underline{x}^*(\theta) \neq \overline{x}^*(\theta)$. Clearly, in an optimal mechanism (8) must bind for at least one $\theta \in [\underline{\theta}, \overline{\theta}]$ as otherwise $\overline{u} > 0$ can be reduced without violating any constraints. However, together with $\kappa^*(\theta) = \kappa_b(\theta)$ and $\underline{x}^*(\theta) = \overline{x}^*(\theta)$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$ this implies that (8) must bind everywhere. But according to Proposition 1, Part d) this is only possible if $\kappa_b(\theta)$ is weakly increasing everywhere which cannot be the case unless types are weakly affiliated due to equation (20).

If the solution of Problem P’ is feasible for Problem P, we are done. However, Assumption 3 in general will not guarantee that the solution schedules of Problem P’ are non-negative and monotonic when $\kappa^*(\theta) \neq \kappa_b(\theta)$ on some interval. So suppose the solution to Problem P’ is not feasible for Problem P. Note that Assumption 3 guarantees the no bundling schedule $x^*(\theta)$ being strictly positive with strictly positive first derivative everywhere. In other words, the no-bundling schedule $\underline{x}(\theta) = \overline{x}(\theta) = x^*(\theta)$ is bounded away from the boundaries of the convex set of implementable allocations that are defined by monotonicity and feasibility constraints (11) and (10). This allows us to form a non-trivial convex combination of the no-bundling schedules $\underline{x}(\theta) = \overline{x}(\theta) = x^*(\theta)$ and the solution to the reduced Problem P’ that is feasible for Problem P and, due to concavity of the objective in x, strictly improves upon the no-bundling schedules.[9]

We qualitatively illustrate the reasoning of the previous paragraphs in Figure 5 and Figure 6 for the case of negatively affiliated preference types where

$$\frac{\partial}{\partial\theta}\left[\frac{f(\theta|\overline{\eta})}{f(\theta|\underline{\eta})}\right] < 0 \quad \forall\theta \in [\underline{\theta}, \overline{\theta}].$$

The solution for Problem P’ is easy to derive from Proposition 1. Negatively affiliated preference types imply

$$\frac{1 - F(\theta|\overline{\eta})}{f(\theta|\overline{\eta})} < \frac{1 - F(\theta|\underline{\eta})}{f(\theta|\underline{\eta})}$$

for all $\theta < \overline{\theta}$, so setting $\kappa^*(\theta) = 0$ would result in $\overline{x}(\theta) \geq \underline{x}(\theta)$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$ with equality only at $\overline{\theta}$. The relevant constraint hence is given by (9). As a consequence, the seller clearly has no incentive to leave lump-sum rents to high η-types; setting $\overline{u} = 0$ maximally relaxes (9) and simultaneously maximizes her profits. The seller hence separates x-schedules in η as long as this is possible without violating constraint (9) for $\overline{u} = 0$, that is, up to some point $\theta' > \underline{\theta}$. Correspondingly, the schedule $\kappa_b(\theta)$ is positive and decreasing everywhere and the optimal costate schedule $\kappa^*(\theta)$ for Problem P’ is shaped as in Figure 5, giving rise to schedules $\overline{x}^*(\theta)$ and $\underline{x}^*(\theta)$ such that

$$\int_{\underline{\theta}}^{\theta'} \left[\overline{x}^*(\theta) - \underline{x}^*(\theta)\right] d\theta = \overline{\eta} - \underline{\eta}.$$

---

[9] We refer to the appendix for all technical details.

This area corresponds to the grey-shaded area in Figure 6. However, the schedules $\overline{x}^*(\theta)$ and $\underline{x}^*(\theta)$ on $\left[\underline{\theta},\theta^{'}\right]$ do not necessarily constitute a feasible solution for Problem P, even though the bunching schedule $x^*(\theta)$ is positive and increasing everywhere. Indeed, for sufficiently small θ in Figure 6 the schedule $\overline{x}^*(\theta)$ is decreasing and the schedule $\underline{x}^*(\theta)$ becomes negative.[10] Yet, both these violations of constraints (10) or (11) can be resolved by forming a convex combination of $\overline{x}^*(\theta)$ or $\underline{x}^*(\theta)$, respectively, and the bunching schedule $x^*(\theta)$ as indicated by the dashed graphs in Figure 6 which still improves upon the full bunching schedule $x^*(\theta)$.

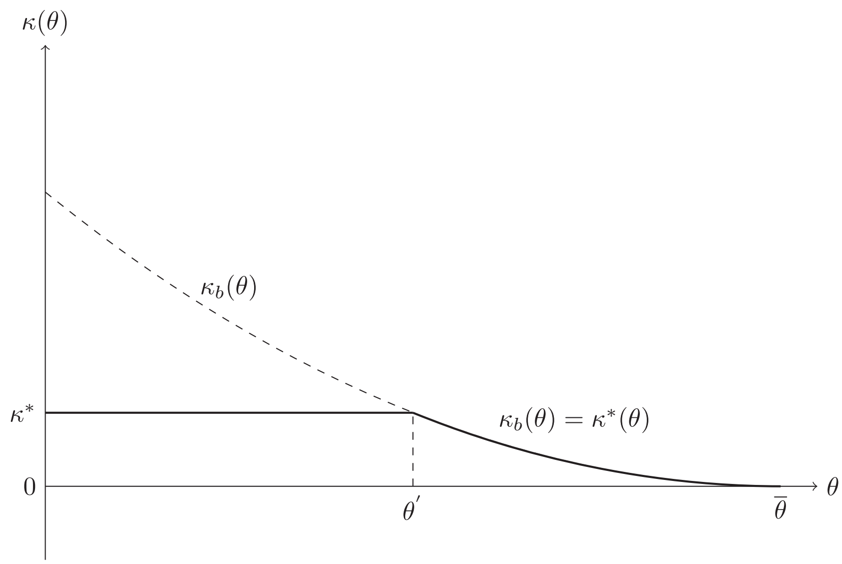

Figure 5: Optimal costate schedule for Problem P' if types are anti-affiliated.

[10]Schedules for $\overline{x}^*(\theta)$ and $\underline{x}^*(\theta)$ on $\left[\underline{\theta},\theta^{'}\right]$ are defined via (16) and (17). For positive κ^*, negative values of $\underline{x}^*(\theta)$ may occur if the optimal x-schedule for known $\eta = \underline{\eta}$ features exclusion of low θ-types while a (locally) decreasing schedule $\overline{x}^*(\theta)$ may occur if $f(\theta|\overline{\eta})$ is decreasing. Both phenomena are compatible with negative affiliation and Assumption 3.

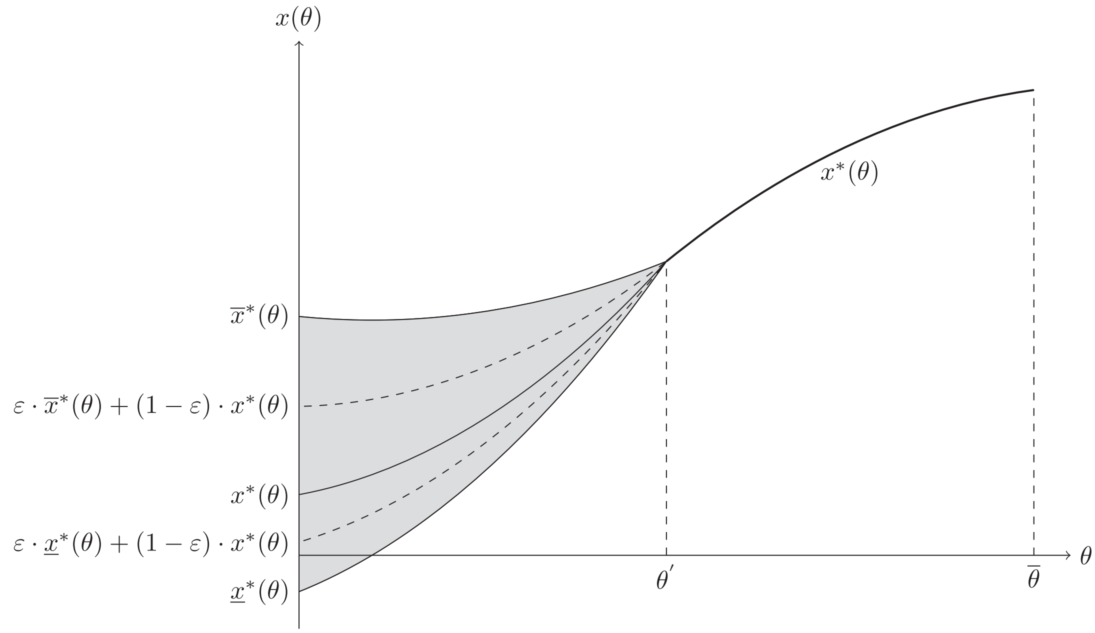

Figure 6: Convex combinations of schedules $\overline{x}^*(\theta)$ and $x^*(\theta)$ resp. $\underline{x}^*(\theta)$ and $x^*(\theta)$ render the solution feasible, that is non-negative and monotonic.

8. Conclusions

Adams and Yellen (1976) have shown that bundling increases profits if consumers' multidimensional tastes are negatively correlated. We study a related question in the context of a richer but still manageable allocation problem. We show that no bundling occurs if types are affiliated and conversely that some bundling does occur if they are not.

9. Appendix

Proof of Lemma 1: For any report (θ, η), a feasible stochastic mechanism is characterized by a distribution $H(\theta, \eta)(x, q, p)$ over allocations and transfers such that expectations $\mathbb{E}_{H(\theta,\eta)}[\cdot]$ over x, q, p exist. Clearly, due to quasilinear utilities, seller and buyer are indifferent between any lottery over prices and the expected value of this lottery, so we may assume that prices are deterministic. Start with an arbitrary Bayesian incentive compatible mechanism H with deterministic transfers and change the q-allocation to first best, i.e. $q(\theta, \overline{\eta}) = 1$, $q(\theta, \underline{\eta}) = 0$, while adjusting prices such that expected equilibrium profits of the buyer remain constant. For any θ at which the original mechanism featured $q(\theta, \overline{\eta}) < 1$ this increases revenues from type $(\theta, \overline{\eta})$ by $(1 - q(\theta, \overline{\eta})) \cdot [\overline{\eta} - c] > 0$ while for any θ at which the original mechanism featured $q(\theta, \underline{\eta}) > 0$ this increases revenues from type $(\theta, \underline{\eta})$ by

$q\left(\theta,\underline{\eta}\right)\cdot\left[c-\underline{\eta}\right]>0$. We need to show that the new mechanism is still incentive compatible. In the η-dimension, Bayesian incentive compatibility constraints of the original mechanism read

$$\begin{aligned} u^{old}\left(\theta,\overline{\eta}\right) &\geq U^{old}\left(\theta,\theta,\underline{\eta},\overline{\eta}\right)=u^{old}\left(\theta,\underline{\eta}\right)+q\left(\theta,\underline{\eta}\right)\cdot\left(\overline{\eta}-\underline{\eta}\right) \\ u^{old}\left(\theta,\underline{\eta}\right) &\geq U^{old}\left(\theta,\theta,\overline{\eta},\underline{\eta}\right)=u^{old}\left(\theta,\overline{\eta}\right)-q\left(\theta,\overline{\eta}\right)\cdot\left(\overline{\eta}-\underline{\eta}\right) \end{aligned}$$

at any $\theta\in\Theta$. They certainly imply the IC-constraints for the new mechanism given as

$$\begin{aligned} u^{new}\left(\theta,\overline{\eta}\right) &\geq U^{new}\left(\theta,\theta,\underline{\eta},\overline{\eta}\right)=u^{new}\left(\theta,\underline{\eta}\right), \\ u^{new}\left(\theta,\underline{\eta}\right) &\geq U^{new}\left(\theta,\theta,\overline{\eta},\underline{\eta}\right)=u^{new}\left(\theta,\overline{\eta}\right)-\left(\overline{\eta}-\underline{\eta}\right). \end{aligned}$$

since the RHS is smaller for the new mechanism while the LHS hasn't changed. Bayesian incentive compatibility then follows from

$$\begin{aligned} U^{new}\left(\hat{\theta},\theta,\hat{\eta},\eta\right) &= U^{new}\left(\hat{\theta},\hat{\theta},\hat{\eta},\eta\right)+\left(\theta-\hat{\theta}\right)\cdot\mathbb{E}^{new}_{H\left(\hat{\theta},\hat{\eta}\right)}[x] \\ &= U^{new}\left(\hat{\theta},\hat{\theta},\hat{\eta},\eta\right)+\left(\theta-\hat{\theta}\right)\cdot\mathbb{E}^{old}_{H\left(\hat{\theta},\hat{\eta}\right)}[x] \\ &\leq U^{old}\left(\hat{\theta},\hat{\theta},\hat{\eta},\eta\right)+\left(\theta-\hat{\theta}\right)\cdot\mathbb{E}^{old}_{H\left(\hat{\theta},\hat{\eta}\right)}[x] \\ &= U^{old}\left(\hat{\theta},\theta,\hat{\eta},\eta\right) \\ &\leq u^{old}\left(\theta,\eta\right) \\ &= u^{new}\left(\theta,\eta\right). \end{aligned}$$

Here we use that neither buyer's utilities nor x-allocations were changed (2^{nd} and 6^{th} line), we use incentive compatibility of the original mechanism (5^{th} line) and we use that $U^{new}\left(\hat{\theta},\hat{\theta},\hat{\eta},\eta\right)\leq U^{old}\left(\hat{\theta},\hat{\theta},\hat{\eta},\eta\right)$ as shown above (3^{rd} line).

As a consequence, the optimal mechanism is non-stochastic in q. To show that the optimal mechanism is non-stochastic in x, note that for any report profile $\left(\hat{\theta},\hat{\eta}\right)$ and type profile (θ,η) expected buyers' utilities

$$U\left(\hat{\theta},\theta,\hat{\eta},\eta\right) = \theta\cdot\mathbb{E}_{H\left(\hat{\theta},\hat{\eta}\right)}[x]+\eta\cdot\mathbb{E}_{H\left(\hat{\theta},\hat{\eta}\right)}[q]-p\left(\theta,\eta\right)$$

only depend on expected values of x (and q). Hence assigning $x\left(\hat{\theta},\hat{\eta}\right)=\mathbb{E}_{H\left(\hat{\theta},\hat{\eta}\right)}[x]$ with probability 1 does not alter the incentive problem of the buyer but increases expected equilibrium profits of the firm from type $\left(\hat{\theta},\hat{\eta}\right)$ due to Jensen's inequality by

$$\mathbb{E}_{H\left(\hat{\theta},\hat{\eta}\right)}[C(x)]-C\left(\mathbb{E}_{H\left(\hat{\theta},\hat{\eta}\right)}[x]\right)\geq 0.$$

Proof of Lemma 2: As $q\left(\theta,\eta\right)=q\left(\eta\right)$ is independent of θ for any $\eta\in\left\{\underline{\eta},\overline{\eta}\right\}$, the one-dimensional constraints imply

$$\begin{aligned} U\left(\hat{\theta},\theta,\hat{\eta},\eta\right) &= U\left(\hat{\theta},\theta,\hat{\eta},\hat{\eta}\right)+q\left(\hat{\theta},\hat{\eta}\right)\cdot\left(\eta-\hat{\eta}\right) \\ &\leq u\left(\theta,\hat{\eta}\right)+q\left(\theta,\hat{\eta}\right)\cdot\left(\eta-\hat{\eta}\right) \\ &= U\left(\theta,\theta,\hat{\eta},\eta\right) \\ &\leq u\left(\theta,\eta\right) \end{aligned}$$

for all $\theta, \hat{\theta} \in \left[\underline{\theta}, \overline{\theta}\right]$, $\eta, \hat{\eta} \in \left\{\underline{\eta}, \overline{\eta}\right\}$.

Proof of Lemma 3: Monotonicity of x in θ is necessary as

$$\begin{aligned} u(\theta,\eta) &\geq U\left(\hat{\theta},\theta,\eta,\eta\right) = u\left(\hat{\theta},\eta\right) + x\left(\hat{\theta},\eta\right)\left(\theta-\hat{\theta}\right), \\ u\left(\hat{\theta},\eta\right) &\geq U\left(\theta,\hat{\theta},\eta,\eta\right) = u(\theta,\eta) - x(\theta,\eta)\left(\theta-\hat{\theta}\right) \end{aligned}$$

imply

$$\left[x\left(\hat{\theta},\eta\right) - x(\theta,\eta)\right]\cdot\left(\theta-\hat{\theta}\right) \leq 0.$$

To show necessity for the first part, suppose without loss of generality that $\theta > \hat{\theta}$. Then

$$x\left(\hat{\theta},\eta\right) \leq \frac{u(\theta,\eta) - u\left(\hat{\theta},\eta\right)}{\left(\theta-\hat{\theta}\right)} \leq x(\theta,\eta).$$

But as $x(\theta,\eta)$ is non-decreasing in θ, it is continuous but for at most countably many points and at any point of continuity of $x(\theta,\eta)$ in θ taking limits $\hat{\theta}\to\theta$ yields $u_\theta(\theta,\eta) = x(\theta,\eta)$, so integrating over θ yields (5). For sufficiency, note that by the fact that the allocation is non-negative and monotonic we have

$$\begin{aligned} u(\theta,\eta) - U\left(\hat{\theta},\theta,\eta,\eta\right) &= u(\theta,\eta) - u\left(\hat{\theta},\eta\right) - x\left(\hat{\theta},\eta\right)\left(\theta-\hat{\theta}\right) \\ &= \int_{\hat{\theta}}^{\theta}\left[x(y,\eta) - x\left(\hat{\theta},\eta\right)\right]dy \\ &\geq 0 \end{aligned}$$

for all $\theta, \hat{\theta} \in \left[\underline{\theta}, \overline{\theta}\right]$. The second part of the Lemma follows from

$$\begin{aligned} U\left(\theta,\theta,\underline{\eta},\overline{\eta}\right) &= u\left(\theta,\underline{\eta}\right) + q\left(\theta,\underline{\eta}\right)\cdot\left(\overline{\eta}-\underline{\eta}\right), \\ U\left(\theta,\theta,\overline{\eta},\underline{\eta}\right) &= u\left(\theta,\overline{\eta}\right) - q\left(\theta,\overline{\eta}\right)\cdot\left(\overline{\eta}-\underline{\eta}\right). \end{aligned}$$

Proof of Lemma 4: Suppose one of the solution schedules is not continuous at some $\theta^{'} \in \left(\underline{\theta}, \overline{\theta}\right)$. By monotonicity of the optimal schedules, both schedules have left and right limits at $\theta^{'}$ so there must be a jump dicontinuity at $\theta^{'}$. Write $x_l^*(\theta)$ for the left limit points and. $x_r^*(\theta)$ for the right limit points. First, suppose that $\underline{x}^*$ is not continuous at $\theta^{'}$. For any $\delta > 0$ sufficiently small, define

$$\underline{x}_\delta(\theta) = \begin{cases} \underline{x}^*(\theta) & \theta \notin \left[\theta^{'}-\delta, \theta^{'}+\delta\right] \\ \underline{c}^*(\delta) = \frac{\int_{\theta^{'}-\delta}^{\theta^{'}+\delta}\underline{x}^*(\theta)d\theta}{2\delta} & \theta \in \left[\theta^{'}-\delta, \theta^{'}+\delta\right] \end{cases}.$$

Note that

$$\lim_{\delta\to 0}\underline{c}^*(\delta) = \frac{\left[\underline{x}_l^*\left(\theta^{'}\right) + \underline{x}_r^*\left(\theta^{'}\right)\right]}{2}$$

and that $\underline{x}_\delta(\theta)$ is non-decreasing and non-negative given that $\underline{x}^*$ is. By concavity of B in x, we have

$$\begin{aligned}
& \frac{\partial}{\partial\delta}|_{\delta=0}\left[\Pi\left(\underline{x}_\delta,\overline{x}^*,\overline{u}\right)-\Pi\left(\underline{x}^*,\overline{x}^*,\overline{u}\right)\right] \\
= & \frac{\partial}{\partial\delta}|_{\delta=0}\left[\beta\int_{\theta'-\delta}^{\theta'}\left[B\left(\underline{x}_\delta(\theta),0,\theta,\underline{\eta}\right)-B\left(\underline{x}^*(\theta),0,\theta,\underline{\eta}\right)\right]f\left(\theta|\underline{\eta}\right)d\theta\right] \\
& +\frac{\partial}{\partial\delta}|_{\delta=0}\left[\beta\int_{\theta'}^{\theta'+\delta}\left[B\left(\underline{x}_\delta(\theta),0,\theta,\underline{\eta}\right)-B\left(\underline{x}^*(\theta),0,\theta,\underline{\eta}\right)\right]f\left(\theta|\underline{\eta}\right)d\theta\right] \\
= & \beta\cdot f\left(\theta'|\underline{\eta}\right)\cdot\left[2B\left(\frac{\left[\underline{x}_l^*\left(\theta'\right)+\underline{x}_r^*\left(\theta'\right)\right]}{2},0,\theta',\underline{\eta}\right)-B\left(\underline{x}_l^*\left(\theta'\right),0,\theta',\underline{\eta}\right)-B\left(\underline{x}_r^*\left(\theta'\right),0,\theta',\underline{\eta}\right)\right] \\
> & 0.
\end{aligned}$$

Hence, for sufficiently small δ, the mechanism $(\underline{x}_\delta,\overline{x}^*)$ increases the value of the objective Π which would contradict optimality of $(\underline{x}^*,\overline{x}^*)$ if the mechanism $(\underline{x}_\delta,\overline{x}^*)$ were to satisfy all constraints. Note that $\rho(\theta,\overline{u})$ takes the same values for the original mechanism and for $(\underline{x}_\delta,\overline{x}^*)$ outside $\left[\theta'-\delta,\theta'+\delta\right]$. Hence it suffices to check contraints (8) and (9) on $\left[\theta'-\delta,\theta'+\delta\right]$ for the mechanism $(\underline{x}_\delta,\overline{x}^*)$. If neither (8) nor (9) binds at θ' for the original mechanism, then both constraints will also be satisfied for $(\underline{x}_\delta,\overline{x}^*)$ on $\left[\theta'-\delta,\theta'+\delta\right]$ if δ is chosen sufficiently small. Note that, at any θ,

$$\rho(\theta,\overline{u})=\overline{u}+\int_{\underline{\theta}}^{\theta}\left[\overline{x}(y)-\underline{x}(y)\right]dy$$

is (weakly) larger for the original mechanism than for $(\underline{x}_\delta,\overline{x}^*)$. Hence constraint (9) will never be violated by $(\underline{x}_\delta,\overline{x}^*)$ given that it was satisfied by $(\underline{x}^*,\overline{x}^*)$. So the only case in which $(\underline{x}_\delta,\overline{x}^*)$ is not feasible for any $\delta>0$ is when (8) binds at θ' for the original mechanism.

But then we must have $\overline{x}_l^*\left(\theta'\right)\leq\underline{x}_l^*\left(\theta'\right)$ as otherwise, if $\overline{x}_l^*\left(\theta'\right)>\underline{x}_l^*\left(\theta'\right)$, this inequality continues to hold on a small interval $\left[\theta'-\epsilon,\theta'\right]$ and hence $\rho\left(\theta'-\epsilon\right)<\rho\left(\theta'\right)=0$, contradicting incentive compatibility of the mechanism $(\underline{x}^*,\overline{x}^*)$. Similarly we must have $\underline{x}_r^*\left(\theta'\right)\leq\overline{x}_r^*\left(\theta'\right)$ as otherwise, if $\underline{x}_r^*\left(\theta'\right)>\overline{x}_r^*\left(\theta'\right)$, this inequality continues to hold on a small interval $\left[\theta',\theta'+\epsilon\right]$ and hence $\rho\left(\theta'+\epsilon\right)<\rho\left(\theta'\right)=0$, again contradicting incentive compatibility of the mechanism $(\underline{x}^*,\overline{x}^*)$. Thus we have $\overline{x}_l^*\left(\theta'\right)\leq\underline{x}_l^*\left(\theta'\right)<\underline{x}_r^*\left(\theta'\right)\leq\overline{x}_r^*\left(\theta'\right)$, so $\overline{x}^*$ also has a jump discontinuity at θ'.

Next, suppose that $\overline{x}^*$ is not continuous at θ'. Just as before, define the following schedule for $\delta > 0$ sufficiently small:

$$\overline{x}_\delta(\theta) = \begin{cases} \overline{x}^*(\theta) & \theta \notin \left[\theta' - \delta, \theta' + \delta\right] \\ \overline{c}^*(\delta) = \frac{\int_{\theta'-\delta}^{\theta'+\delta} \overline{x}^*(\theta) d\theta}{2\delta} & \theta \in \left[\theta' - \delta, \theta' + \delta\right] \end{cases}.$$

Note that

$$\lim_{\delta \to 0} \overline{c}^*(\delta) = \frac{\left[\overline{x}_l^*\left(\theta'\right) + \overline{x}_r^*\left(\theta'\right)\right]}{2}$$

and that $\overline{x}_\delta(\theta)$ is non-decreasing and non-negative given that $\overline{x}^*$ is. By concavity of B in x, we have

$$\begin{aligned} & \frac{\partial}{\partial \delta}|_{\delta=0} \left[\Pi\left(\underline{x}^*, \overline{x}_\delta, \overline{u}\right) - \Pi\left(\underline{x}^*, \overline{x}^*, \overline{u}\right)\right] \\ = & \frac{\partial}{\partial \delta}|_{\delta=0} \left[(1-\beta) \int_{\theta'-\delta}^{\theta'} \left[B\left(\overline{x}_\delta(\theta), 1, \theta, \overline{\eta}\right) - B\left(\overline{x}^*(\theta), 1, \theta, \overline{\eta}\right)\right] f(\theta|\overline{\eta}) d\theta\right] \\ & + \frac{\partial}{\partial \delta}|_{\delta=0} \left[(1-\beta) \int_{\theta'}^{\theta'+\delta} \left[B\left(\overline{x}_\delta(\theta), 1, \theta, \overline{\eta}\right) - B\left(\overline{x}^*(\theta), 1, \theta, \overline{\eta}\right)\right] f(\theta|\overline{\eta}) d\theta\right] \\ = & (1-\beta) \cdot f\left(\theta'|\overline{\eta}\right) \cdot \left[2B\left(\frac{\left[\overline{x}_l^*\left(\theta'\right) + \overline{x}_r^*\left(\theta'\right)\right]}{2}, 1, \theta', \overline{\eta}\right) - B\left(\overline{x}_l^*\left(\theta'\right), 1, \theta', \overline{\eta}\right) - B\left(\overline{x}_r^*\left(\theta'\right), 1, \theta', \overline{\eta}\right)\right] \\ > & 0. \end{aligned}$$

So again, for sufficiently small δ, the new mechanism $\left(\underline{x}^*, \overline{x}_\delta\right)$ increases the value of the objective Π compared to the original mechanism. As before, $\rho(\theta, \overline{u})$ takes the same values for the original mechanism and for $\left(\underline{x}^*, \overline{x}_\delta\right)$ outside $\left[\theta' - \delta, \theta' + \delta\right]$, so it suffices to check contraints (8 and (9) on $\left[\theta' - \delta, \theta' + \delta\right]$ for the mechanism $\left(\underline{x}^*, \overline{x}_\delta\right)$. If neither (8) nor (9) binds at θ' for the original mechanism, then both constraints will also be satisfied for $\left(\overline{x}_\delta, \overline{x}^*\right)$ on $\left[\theta' - \delta, \theta' + \delta\right]$ if δ is chosen sufficiently small. At any θ,

$$\rho(\theta, \overline{u}) = \overline{u} + \int_{\underline{\theta}}^{\theta} \left[\overline{x}(y) - \underline{x}(y)\right] dy$$

is (weakly) smaller for the original mechanism than for $\left(\underline{x}^*, \overline{x}_\delta\right)$. Hence constraint (8) will never be violated by $\left(\underline{x}^*, \overline{x}_\delta\right)$ given that it was not violated by $\left(\underline{x}^*, \overline{x}^*\right)$. So the only case in which $\left(\underline{x}^*, \overline{x}_\delta\right)$ is not feasible for any $\delta > 0$ is when (9) binds at θ' for the original mechanism. In that case, however, we must now have $\underline{x}_l^*\left(\theta'\right) \leq \overline{x}_l^*\left(\theta'\right) < \overline{x}_r^*\left(\theta'\right) \leq \underline{x}_r^*\left(\theta'\right)$, so $\overline{x}^*$ also has a jump discontinuity at θ'.

Together, both cases yield a contradiction and hence prove the result on $\left(\underline{\theta}, \overline{\theta}\right)$. Either schedule $\underline{x}^*$ is continuous or, at any point where $\underline{x}^*$ is not continuous, both schedules are discontinuous and constraint (9) binds. At the same time, either $\overline{x}^*$ is continuous or, at any point where $\overline{x}^*$is not continuous, both schedules are discontinuous and constraint (8) binds. As (8) and (9) cannot both bind at a given θ', the only possibility is that both schedules $\underline{x}^*$, $\overline{x}^*$ are continuous.

Finally, consider the the boundaries $\underline{\theta}$ and $\overline{\theta}$. By monotonicity, we have

$$\begin{aligned} \underline{x}^*(\underline{\theta}) &\leq \underline{x}_r^*(\underline{\theta}) \\ \underline{x}^*(\overline{\theta}) &\geq \underline{x}_l^*(\overline{\theta}) \\ \overline{x}^*(\underline{\theta}) &\leq \overline{x}_r^*(\underline{\theta}) \\ \overline{x}^*(\overline{\theta}) &\geq \overline{x}_l^*(\overline{\theta}) \end{aligned}$$

so $\underline{x}_r^*(\underline{\theta})$, $\underline{x}_l^*(\overline{\theta})$, $\overline{x}_r^*(\underline{\theta})$, $\overline{x}_l^*(\overline{\theta})$ exist. Setting

$$\begin{aligned} \underline{x}^*(\underline{\theta}) &= \underline{x}_r^*(\underline{\theta}) \\ \underline{x}^*(\overline{\theta}) &= \underline{x}_l^*(\overline{\theta}) \\ \overline{x}^*(\underline{\theta}) &= \overline{x}_r^*(\underline{\theta}) \\ \overline{x}^*(\overline{\theta}) &= \overline{x}_l^*(\overline{\theta}) \end{aligned}$$

neither changes Π or ρ nor violates monotonicity, and it guarantees continuity of $(\underline{x}^*, \overline{x}^*)$ at the boundaries.

Proof of Proposition 1: We invoke Seierstad and Sydsaeter (1987), Ch. 5, Thm. 2, p. 332 f.[11] For the optimal solution schedules $(\underline{x}^*(\theta), \overline{x}^*(\theta))$ and fixed values $\underline{u} = \underline{u}(\underline{\theta}) = 0$, $\overline{u} = \overline{u}(\underline{\theta})$ this theorem yields the existence of costate variables $(\underline{\kappa}(\theta), \overline{\kappa}(\theta))$ with $\underline{\kappa}(\overline{\theta}) = \overline{\kappa}(\overline{\theta}) = 0$ such that $(\underline{x}^*(\theta), \overline{x}^*(\theta))$ maximizes

$$\begin{aligned} H(\underline{u}^*(\theta), \overline{u}^*(\theta), \underline{x}(\theta), \overline{x}(\theta), \underline{\kappa}(\theta), \overline{\kappa}(\theta), \theta) = {} & \beta B(\underline{x}(\theta), 0, \theta, \underline{\eta}) f(\theta|\underline{\eta}) \\ & + (1-\beta) B(\overline{x}(\theta), 1, \theta, \overline{\eta}) f(\theta|\overline{\eta}) \qquad (23) \\ & + \underline{\kappa}(\theta) \cdot \underline{x}(\theta) + \overline{\kappa}(\theta) \cdot \overline{x}(\theta). \end{aligned}$$

In addition, there exist a componentwise non-decreasing function $\mu(\theta) = (\mu_1(\theta), \mu_2(\theta))$ with the following properties: μ_1 is constant on any interval where $\rho(\theta, \overline{u}) > 0$ and μ_2 is constant on any interval where $(\overline{\eta} - \underline{\eta}) - \rho(\theta, \overline{u}) > 0$. Moreover,

$$\varkappa(\theta) = (\underline{\varkappa}(\theta), \overline{\varkappa}(\theta)) \equiv (\underline{\kappa}(\theta), \overline{\kappa}(\theta)) + \mu^*(\theta) \cdot \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \qquad (24)$$

is continuous, and at any point where $(\underline{x}^*(\theta), \overline{x}^*(\theta))$ and $\mu(\theta)$ are continuous, $\varkappa(\theta)$ is differentiable with

$$\begin{aligned} \frac{\partial(\underline{\varkappa}(\theta), \overline{\varkappa}(\theta))}{\partial\theta} &= -\frac{\partial}{\partial(\overline{u}(\theta), \underline{u}(\theta))} \\ & \left[H(\underline{u}^*(\theta), \overline{u}^*(\theta), \underline{x}^*(\theta), \overline{x}^*(\theta), \underline{\kappa}(\theta), \overline{\kappa}(\theta), \theta) - \mu(\theta) \cdot \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} \underline{x}^*(\theta) \\ \overline{x}^*(\theta) \end{pmatrix} \right] \\ &= (0, 0). \end{aligned}$$

Since $\underline{x}^*(\theta)$ and $\overline{x}^*(\theta)$ are assumed to be continuous and $\mu(\theta)$ is non-decreasing and hence continuous everywhere except for at most countably many points, $\varkappa(\theta)$ is continuous and differentiable everywhere except for at most countably many points, with derivative being equal to 0. But then, by standard calculus (see e.g. Königsberger (2004)), $\varkappa(\theta)$ is Lipschitz continuous with Lipschitz constant 0 everywhere,

[11] Note that the roles of x and u are reversed in our paper compared to the notation in Seierstad and Sydsaeter (1987)

i.e. $\varkappa(\theta) = \varkappa$ is constant everywhere. Hence, evaluating (24) at $\theta = \overline{\theta}$ we get $\underline{\varkappa}(\overline{\theta}) = -\mu_1(\overline{\theta}) + \mu_2(\overline{\theta})$, $\overline{\varkappa} = +\mu_1(\overline{\theta}) - \mu_2(\overline{\theta})$ and hence

$$\begin{aligned}
\underline{\kappa}(\theta) &= \mu_1(\theta) - \mu_1(\overline{\theta}) - \mu_2(\theta) + \mu_2(\overline{\theta}), \\
\overline{\kappa}(\theta) &= -\mu_1(\theta) + \mu_1(\overline{\theta}) + \mu_2(\theta) - \mu_2(\overline{\theta}) = -\underline{\kappa}(\theta).
\end{aligned}$$

Define $\kappa^*(\theta) = \underline{\kappa}(\theta)$ and note the first-order conditions of (23) are precisely the equations charaterizing the optimal schedules. As H is concave in x the first-order conditions are sufficient. Moreover, as the constraints (8), (9) are linear (and hence quasi-concave) in $(\underline{u}(\theta), \overline{u}(\theta))$, the conditions in Seierstad and Sydsaeter (1987), Ch. 5, Thm. 2, p. 332 f. are sufficient, cf. Seierstad and Sydsaeter (1987), Ch. 5, Thm. 3, p. 337. The properties of $\kappa^*(\theta)$ are straightforward:

a) Continuity of $\kappa^*(\theta)$ follows directly from continuity of $(\underline{x}^*(\theta), \overline{x}^*(\theta))$.

b) This follows directly from $\kappa^*(\theta) = \underline{\kappa}(\theta) = \mu_1(\theta) - \mu_1(\overline{\theta}) - \mu_2(\theta) + \mu_2(\overline{\theta})$ and the respective properties of μ_1 and μ_2.

c) Continuity of $\kappa^*(\theta)$ and$(\underline{x}^*(\theta), \overline{x}^*(\theta))$ implies that for any $\theta \in (\underline{\theta}, \overline{\theta})$ where (8) or (9) binds it must hold that $\underline{x}^*(\theta) = \overline{x}^*(\theta)$ as otherwise the respective constraint would be violated either at $\theta + \delta$ or at $\theta - \delta$ for $\delta > 0$ sufficiently small.

d) This is implied by the monotonicity properties of μ_1 and μ_2 via $\kappa^*(\theta) = \underline{\kappa}(\theta) = \mu_1(\theta) - \mu_1(\overline{\theta}) - \mu_2(\theta) + \mu_2(\overline{\theta})$.

e) This follows from the transversality condition $\underline{\kappa}(\overline{\theta}) = 0$.

Proof of Proposition 2: Note first that from equations (12) and (13) we get

$$\begin{aligned}
\frac{\partial \underline{x}(\theta)}{\partial \kappa} &= \frac{1}{\beta f(\theta|\underline{\eta}) \cdot C^1_{xx}(\underline{x}(\theta))} > 0, \\
\frac{\partial \overline{x}(\theta)}{\partial \kappa} &= -\frac{1}{(1-\beta) f(\theta|\overline{\eta}) \cdot C^1_{xx}(\overline{x}(\theta))} < 0.
\end{aligned}$$

In an optimal mechanism, constraint (8) must bind for at least one $\theta \in [\underline{\theta}, \overline{\theta}]$. Otherwise the continuous function $\rho^*(\theta, \overline{u})$ attains its minimum $\epsilon > 0$ at some $\hat{\theta}$ on the compact interval $[\underline{\theta}, \overline{\theta}]$ and decreasing $\overline{u}$ by ϵ does not harm any constraints and simultaneously increases revenues. So assume $\rho^*(\hat{\theta}, \overline{u}) = 0$ for some $\hat{\theta} \in [\underline{\theta}, \overline{\theta}]$. We want to show that this implies $\rho^*(\hat{\theta}, \overline{u}^*) = 0$ for any $\theta \geq \hat{\theta}$. Suppose first that $\hat{\theta} = \underline{\theta}$. Since $\kappa^*(\theta)$ is continuous and locally either follows a constant schedule or $\kappa_b(\theta)$, and since $\kappa_b(\theta)$ is increasing, $\kappa^*(\underline{\theta}) < \kappa_b(\underline{\theta})$ would imply $\kappa^*(\theta) = \kappa^*(\underline{\theta}) < \kappa_b(\underline{\theta}) \leq \kappa_b(\overline{\theta}) = 0$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$, contradicting transversality $\kappa^*(\overline{\theta}) = 0$ as stated in Proposition 1, Part e). On the other hand, $\kappa^*(\underline{\theta}) > \kappa_b(\underline{\theta})$ by continuity of $\kappa^*(\theta)$ and $\kappa_b(\theta)$ would imply $\kappa^*(\theta) > \kappa_b(\theta)$ on some interval $[\underline{\theta}, \underline{\theta} + \epsilon]$ of positive length $\epsilon > 0$, violating (8) at any $\theta > \underline{\theta}$ within this interval. Hence for $\hat{\theta} = \underline{\theta}$ we must have $\kappa^*(\hat{\theta}) = \kappa_b(\hat{\theta})$ just as for any other $\hat{\theta} \in [\underline{\theta}, \overline{\theta}]$, following Proposition 1, Part c) and e).

As a consequence, note that $\kappa^*(\theta) \leq \kappa_b(\theta)$ for any $\theta > \hat{\theta}$ as $\kappa_b(\theta)$ is increasing and $\kappa^*(\theta)$ is continuous and follows either a constant schedule or $\kappa_b(\theta)$. Suppose $\kappa^*(\tilde{\theta}) < \kappa_b(\tilde{\theta})$ for some $\hat{\theta} < \tilde{\theta} \leq \overline{\theta}$. Then $\kappa^*(\theta) = \kappa^*(\tilde{\theta}) < \kappa_b(\tilde{\theta}) \leq \kappa_b(\overline{\theta}) = 0$ for any $\theta \geq \tilde{\theta}$ by continuity of $\kappa^*(\theta)$ and Proposition 1, Part b)+c), contradicting $\kappa^*\left(\overline{\theta}\right) = 0$ as required by Proposition 1, Part e). Defining $\theta' = \inf\left\{\theta \in \left[\underline{\theta}, \overline{\theta}\right] : \rho^*\left(\theta, \overline{u}^*\right) = 0\right\}$ completes the proof.

Proof of Theorem 1: By Proposition 2 we have

$$\overline{u} = \int_{\underline{\theta}}^{\theta'} \left[\overline{x}^*\left(y\right) - \underline{x}^*\left(y\right)\right] dy$$

which, as $\overline{x}^*\left(\theta'\right) = \underline{x}^*\left(\theta'\right) = x^*\left(\theta'\right)$, implies

$$\frac{d\kappa^*}{d\overline{u}}\left(\overline{u}\right) = -\frac{1}{\int_{\underline{\theta}}^{\theta'}\left[\frac{\partial \overline{x}^*}{\partial \kappa^*}\left(y\right) - \frac{\partial \underline{x}^*}{\partial \kappa^*}\left(y\right)\right] dy} > 0.$$

Using this as well as equations (12) and (13) we get

$$\begin{aligned}
&\frac{d\Pi}{d\overline{u}}\left(\overline{u}\right) \\
= &\int_{\underline{\theta}}^{\theta'}\left[\beta f\left(\theta|\underline{\eta}\right)\cdot\frac{\partial B}{\partial \underline{x}}\left(\underline{x}^*\left(\theta\right), 0, \theta, \underline{\eta}\right)\cdot\frac{\partial \underline{x}^*}{\partial \kappa^*}\left(\theta\right)\cdot\frac{d\kappa^*}{d\overline{u}}\left(\overline{u}\right)\right] d\theta \\
&+\int_{\underline{\theta}}^{\theta'}\left[\left(1-\beta\right) f\left(\theta|\overline{\eta}\right)\cdot\frac{\partial B}{\partial \overline{x}}\left(\overline{x}^*\left(\theta\right), 1, \theta, \overline{\eta}\right)\cdot\frac{\partial \overline{x}^*}{\partial \kappa^*}\left(\theta\right)\cdot\frac{d\kappa^*}{d\overline{u}}\left(\overline{u}\right)\right] d\theta - \left(1-\beta\right) \\
= &\frac{\kappa^*\left(\overline{u}\right)\cdot\int_{\underline{\theta}}^{\theta'}\frac{\partial \underline{x}^*}{\partial \kappa^*}\left(\theta\right) d\theta}{\int_{\underline{\theta}}^{\theta'}\left[\frac{\partial \overline{x}^*}{\partial \kappa^*}\left(y\right) - \frac{\partial \underline{x}^*}{\partial \kappa^*}\left(y\right)\right] dy} - \frac{\kappa^*\left(\overline{u}\right)\cdot\int_{\underline{\theta}}^{\theta'}\frac{\partial \overline{x}^*}{\partial \kappa^*}\left(\theta\right) d\theta}{\int_{\underline{\theta}}^{\theta'}\left[\frac{\partial \overline{x}^*}{\partial \kappa^*}\left(y\right) - \frac{\partial \underline{x}^*}{\partial \kappa^*}\left(y\right)\right] dy} - \left(1-\beta\right) \\
= &-\kappa^*\left(\overline{u}\right) - \left(1-\beta\right),
\end{aligned}$$

implying $\overline{u}^* = 0$ as demonstrated in the main text. The theorem then immediately follows from Proposition 2. The reformulation in terms of prices conditional on quantities rather than preference types is a direct implication of the taxation principle as explained in the main text.

Proof of Theorem 2: As demonstrated in the main text, the solution schedules $\left(\underline{x}^*\left(\theta\right), \overline{x}^*\left(\theta\right), \overline{u}^*\right)$ for Problem P' deviate from the full bunching schedule $\left(x^*\left(\theta\right), x^*\left(\theta\right), \overline{u} = 0\right)$ on an interval of positive mass. We are left to formally show that, under Assumption 3, there exists a convex combination

$$\left(\underline{x}_\lambda\left(\theta\right), \overline{x}_\lambda\left(\theta\right), \overline{u}_\lambda\right) = \lambda\cdot\left(\underline{x}^*\left(\theta\right), \overline{x}^*\left(\theta\right), \overline{u}^*\right) + \left(1-\lambda\right)\cdot\left(x^*\left(\theta\right), x^*\left(\theta\right), 0\right)$$

with $\lambda > 0$ such that the schedules $\left(\underline{x}_\lambda\left(\theta\right), \overline{x}_\lambda\left(\theta\right), \overline{u}_\lambda\right)$ satisfy constraints (8)-(11) and that

$$\Pi\left(\underline{x}_\lambda\left(\theta\right), \overline{x}_\lambda\left(\theta\right), \overline{\pi}_\lambda\right) > \Pi\left(x^*\left(\theta\right), x^*\left(\theta\right), 0\right).$$

As constraints (8) and (9) are linear in $\left(\overline{x}, \underline{x}, \overline{u}\right)$ and, by construction, satisfied by $\left(\underline{x}^*\left(\theta\right), \overline{x}^*\left(\theta\right), \overline{u}^*\right)$ and $\left(x^*\left(\theta\right), x^*\left(\theta\right), \overline{u} = 0\right)$, they are also satisfied by any convex combination of the two. Moreover, as $\left(\underline{x}^*\left(\theta\right), \overline{x}^*\left(\theta\right)\right)$ and $\left(x^*\left(\theta\right), x^*\left(\theta\right)\right)$ are continuous and hence bounded on $\left[\underline{\theta}, \overline{\theta}\right]$ and $x^*\left(\theta\right) \geq x^*\left(\underline{\theta}\right) > 0$ for all $\theta \in \left[\underline{\theta}, \overline{\theta}\right]$ by Assumption 3, $\left(\underline{x}_\lambda\left(\theta\right), \overline{x}_\lambda\left(\theta\right)\right)$ are positive for sufficiently small values $\lambda > 0$ as well and hence satisfy (11).

Concerning (10), note that the optimal costate variable $\kappa^*\left(\theta\right)$ that corresponds to $\left(\underline{x}^*\left(\theta\right), \overline{x}^*\left(\theta\right), \overline{u}^*\right)$ is bounded. Indeed, since $\kappa^*\left(\theta\right)$ is continuous on $\left[\underline{\theta}, \overline{\theta}\right]$ with $\kappa^*\left(\overline{\theta}\right) = \kappa_b\left(\overline{\theta}\right) = 0$ and is either locally constant or follows $\kappa_b\left(\theta\right)$, we have $\kappa^*\left(\theta\right) \in$

$\left[\min_{\theta\in\left[\underline{\theta},\overline{\theta}\right]}\kappa_b\left(\theta\right),\max_{\theta\in\left[\underline{\theta},\overline{\theta}\right]}\kappa_b\left(\theta\right)\right]$ which is bounded by compactness of $\left[\underline{\theta},\overline{\theta}\right]$ and continuity of $\kappa_b\left(\theta\right)$. Moreover, on any interval $\left[\theta^{'},\theta^{''}\right]\subset\left[\underline{\theta},\overline{\theta}\right]$, the costate variable $\kappa^*\left(\theta\right)$ is constant or lies within $\left[\min_{\theta\in\left[\theta^{'},\theta^{''}\right]}\kappa_b\left(\theta\right),\max_{\theta\in\left[\theta^{'},\theta^{''}\right]}\kappa_b\left(\theta\right)\right]$. Hence, for any $\theta\in\left[\underline{\theta},\overline{\theta}\right]$ the set of subdifferentials $\partial\kappa^*\left(\theta\right)$ is bounded by $\left[\min\left\{0,\frac{\partial\kappa_b(\theta)}{\partial\theta}\right\},\max\left\{0,\frac{\partial\kappa_b(\theta)}{\partial\theta}\right\}\right]$ and therefore the subdifferentials of $\left(\underline{x}^*\left(\theta\right),\overline{x}^*\left(\theta\right)\right)$ given as

$$\begin{aligned}\partial\underline{x}^*\left(\theta\right) &= \frac{1}{C_{xx}\left(\underline{x}^*\left(\theta\right)\right)}\cdot\left(\frac{\partial}{\partial\theta}\left[\theta-\frac{1-F\left(\theta|\underline{\eta}\right)}{f\left(\theta|\underline{\eta}\right)}\right]+\frac{\kappa^*\left(\theta\right)\cdot\frac{\partial f\left(\theta|\underline{\eta}\right)}{\partial\theta}-f\left(\theta|\underline{\eta}\right)\cdot\partial\kappa^*\left(\theta\right)}{\beta f\left(\theta|\underline{\eta}\right)^2}\right),\\ \partial\overline{x}^*\left(\theta\right) &= \frac{1}{C_{xx}\left(\overline{x}^*\left(\theta\right)\right)}\cdot\left(\frac{\partial}{\partial\theta}\left[\theta-\frac{1-F\left(\theta|\overline{\eta}\right)}{f\left(\theta|\overline{\eta}\right)}\right]-\frac{\kappa^*\left(\theta\right)\cdot\frac{\partial f\left(\theta|\overline{\eta}\right)}{\partial\theta}-f\left(\theta|\overline{\eta}\right)\cdot\partial\kappa^*\left(\theta\right)}{\left(1-\beta\right)f\left(\theta|\overline{\eta}\right)^2}\right)\end{aligned}$$

are bounded as well. Since

$$\frac{\partial x^*\left(\theta\right)}{\partial\theta}=\frac{1}{C_{xx}\left(\underline{x}^*\left(\theta\right)\right)}\cdot\left(\frac{\partial}{\partial\theta}\left[\theta-\frac{1-F\left(\theta|\underline{\eta}\right)}{f\left(\theta|\underline{\eta}\right)}\right]\right)$$

is continuous on $\left[\underline{\theta},\overline{\theta}\right]$ and strictly positive by Assumption 3, it is bounded away from zero by compactness of $\left[\underline{\theta},\overline{\theta}\right]$. Hence, again, for sufficiently small $\lambda>0$ we have

$$\begin{aligned}\partial\underline{x}_\lambda\left(\theta\right) &\subset \mathbb{R}_+,\\ \partial\overline{x}_\lambda\left(\theta\right) &\subset \mathbb{R}_+,\end{aligned}$$

showing (10).

Finally, from (6) and concavity of B in x together with

$$\Pi\left(\underline{x}^*\left(\theta\right),\overline{x}^*\left(\theta\right),\overline{u}^*\right)>\Pi\left(x^*\left(\theta\right),x^*\left(\theta\right),\overline{u}=0\right)$$

we get

$$\begin{aligned}\Pi\left(\underline{x}_\lambda\left(\theta\right),\overline{x}_\lambda\left(\theta\right),\overline{u}_\lambda\right) &= \Pi\left(\lambda\cdot\left(\underline{x}^*\left(\theta\right),\overline{x}^*\left(\theta\right),\overline{u}^*\right)+\left(1-\lambda\right)\cdot\left(x^*\left(\theta\right),x^*\left(\theta\right),0\right)\right)\\ &> \lambda\Pi\left(\underline{x}^*\left(\theta\right),\overline{x}^*\left(\theta\right),\overline{u}^*\right)+\left(1-\lambda\right)\Pi\left(x^*\left(\theta\right),x^*\left(\theta\right),0\right)\\ &> \lambda\Pi\left(x^*\left(\theta\right),x^*\left(\theta\right),0\right)+\left(1-\lambda\right)\Pi\left(x^*\left(\theta\right),x^*\left(\theta\right),0\right)\\ &= \Pi\left(x^*\left(\theta\right),x^*\left(\theta\right),0\right)\end{aligned}$$

which proves the theorem.

References

Adams, W. J. and Yellen, J. L. (1976). Commodity bundling and the burden of monopoly, *The Quarterly Journal of Economics* **90**(3): 475–498.

Armstrong, M. (1996). Multiproduct nonlinear pricing, *Econometrica* **64**: 51–76.

Armstrong, M. (2013). A more general theory of commodity bundling, *Journal of Economic Theory* **148**(2): 448–472.

Armstrong, M. and Rochet, J. C. (1999). Multi-dimensional screening: A user's guide, *European Economic Review* **43**(4): 959–979.

Fudenberg, D. and Tirole, J. (1991). *Game Theory*, The MIT Press.

Hart, S. and Reny, P. J. (2014). Maximal revenue with multiple goods: Nonmonotonicity and other observations. Theoretical Economics (forthcoming).

Jullien, B. (2000). Participation constraints in adverse selection models, *Journal of Economic Theory* **93**(1): 1–47.

Kleven, H. J., Kreiner, C. T. and Saez, E. (2009). The optimal income taxation of couples, *Econometrica* **77**(2): 537–560.

Königsberger, K. (2004). *Analysis 1*, Springer-Lehrbuch.

Maggi, G. and Rodriguez-Clare, A. (1995). On countervailing incentives, *Journal of Economic Theory* **66**(1): 238–263.

Manelli, A. M. and Vincent, D. R. (2006). Bundling as an optimal selling mechanism for a multiple-good monopolist, *Journal of Economic Theory* **127**(1): 1–35.

Manelli, A. M. and Vincent, D. R. (2007). Multidimensional mechanism design: Revenue maximization and the multiple-good monopoly, *Journal of Economic Theory* **137**(1): 153–185.

McAfee, R. P., McMillan, J. and Whinston, M. D. (1989). Multiproduct monopoly, commodity bundling, and correlation of values, *The Quarterly Journal of Economics* **104**(2): 371–383.

Myerson, R. B. (1982). Optimal coordination mechanisms in generalized principal-agent problems, *Journal of Mathematical Economics* **10**(1): 67–81.

Neustadt, L. W. (1976). *Optimization: a Theory of Necessary Conditions*, Princeton University Press.

Rochet, J. C. (1985). The taxation principle and multi-time hamiltonian-jacobi equations, *Journal of Mathematical Economics* **14**(2): 113–128.

Rochet, J.-C. and Choné, P. (1998). Ironing, sweeping, and multidimensional screening, *Econometrica* **66**(4): 783–826.

Seierstad, A. and Sydsaeter, K. (1987). *Optimal Control Theory with Economic Applications*, North-Holland.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*, Springer Series in Statistics.

Szalay, D. (2013). Regulating a multiproduct and multitype monopolist. Discussion Paper No. 397, GESY.

Thanassoulis, J. (2004). Haggling over substitutes, *Journal of Economic Theory* **117**(2): 217–245.