Regulating a multiproduct and multitype monopolist*

Dezsö Szalay

University of Bonn

preliminary and incomplete

October 8, 2010

Abstract

I study the optimal regulation of a firm producing two goods. The firm has private information about its cost of producing either of the goods. I explore the ways in which the optimal allocation differs from its one dimensional counterpart. With binding constraints in both dimensions, the allocation involves distortions for the most efficient producers and features overproduction for some less efficient types.

JEL: D82, L21, Asymmetric Information, Multi-dimensional Screening, Regulation.

1 Introduction

When duplication of fixed costs is wasteful, a service is efficiently provided by a natural monopoly. To keep the service provider from abusing its monopoly power, the pricing of the firm is regulated. If the regulator had access to the firm's information, regulation would be a trivial matter. As is well known, the firm should follow a marginal cost pricing rule and should be subsidized for the losses it makes on a lump-sum basis. However, the problem is precisely that the firm has better information than the regulator has about payoff relevant circumstances. Baron and Myerson [1982]

*This paper is one spin-off of an earlier version, which was joint with Charles Blarckorby and entitled "Regulating a Monopolist with unknown costs and unknown quality capacity"; the other spin-off is now entitled "Quality requirements and optimal regulation". Many thanks for comments and conversations to Chuck Blackorby, Yeon-Koo Che, Erik Eyster, Leonardo Felli, Claudio Mezzetti, Benny Moldovanu, Rudolf Muller, and seminar participants at HEC Lausanne, LBS, LSE, the University of Bonn, University of Frankfurt, ESSET Gerzensee, and the conference on "Multidimensional Mechanism Design" at HCM Bonn. Correspondence can be sent to Dezso Szalay, Chair of Economic Theory I, University of Bonn, Adenauerallee 24-42, 53113 Bonn, Germany, or to szalay@uni-bonn.de

first analyzed this problem when the firm's costs are unknown to the regulator but known to the firm. Marginal cost pricing is no longer optimal as this rule gives firms with relatively low marginal costs incentives to exaggerate their costs in order to get larger subsidies. To make such exaggeration unattractive, prices are distorted upwards for all but the most efficient firms.

In many industries regulation is more complex than deciding how much to produce at what price. In my extension of the Baron and Myerson [1982] model, a monopolist produces two goods instead of one and has private information about his costs of producing each of the goods. E.g. a railway company typically offers both cargo and passenger services; telephone networks can be used for communication and data transfer. How does this simple extension of the standard model affect its predictions? I develop a model that is both stylized and rich to answer this question. It is stylized in that it allows me to answer the question in the first place. It is rich because it allows both for the case where the optimal allocation coincides with the one which is well known from the analysis of onedimensional models and for cases that are markedly different from the standard case.

The technology of the regulated firm is described by the cost of producing two goods. Good one is divisible and can be produced in any amount; the firm has private information about its marginal cost of producing this good, a realization of a random variable supported on an interval. In addition the firm has also private information about its cost of producing good two, which can be either high or low; good two can be produced in two different versions (or quantities). The interesting case to analyze is when it is optimal to separate firms according to their costs of producing good two, so in equilibrium firms with high costs of producing good two produce the less expensive variant, and firms with low costs of producing good two produce the lean version of the good. In that case the production schedules for good one production can be conditioned on the amount of good two delivered, so there is a potential for separating firms further in the good one production dimension. These schedules differ from their counterparts from the case where private information is one dimensional exactly when the incentive constraints are binding in both dimensions. Moreover, the way these schedules differ from the known counterparts depends crucially on why incentive constraints to mimick another firm with a different cost of producing the discrete good are binding. First, they can be binding because increasing the amount of good two production increases the marginal cost of producing good one. Second, they can bind because observing high production of good two allows the regulator to update his beliefs about good one costs of production.

Consider first the case where incentive constraints are binding because there are real differences in marginal costs. In order to keep a producer with low costs of producing good two from mimicking a producer with high costs of producing good two, depending on properties of the type distribution, it becomes optimal to distort the production schedule for good one upwards. Instead of no distortion for the most efficient producer and too little production for less efficient producers, the good-one-production schedule for a low-cost-good-two producer features no distortion for the most inefficient producer and excessive production for more efficient producers (among those with low costs of producing good two). In particular, there is a distortion for the most efficient producer within that class. The good-one-production schedule of a high-cost-good-two producer has more standard, but not quite completely standard features; there is too little production for all such firms, including the most efficient one among them.

If incentive constraints are binding due to inference about costs of producing good one, then it is never optimal to actually make use of this information; the optimal allocation for good one production is the same, irrespective of the firm's cost of producing good two. Although separation would be feasible, the rents required to achieve this separation would be too high relative to the benefit of separation. The common quantity schedule is optimal against the marginal distribution of marginal costs of producing good one and has the standard features of no distortion at the top, too little production below the top, and no rent at the bottom.

Finally, if both reasons for binding constraints interact, then the optimal allocation is a hybrid of the two cases. If the schedules involve bunching, then it is for producers with low costs of producing good one. If that is the case, then there is no distortion for the most efficient producers too little production slightly below that. However, for producers with higher costs, good-one-production of the low-cost-good-two producer is again distorted upwards so as to inhibit this firm from mimicking its high-cost-good-two producer counterpart.

These results confirm findings by Lewis and Sappington [1988] and Armstrong [1999], although I obtain mine using very different verifiability assumptions and techniques. Lewis and Sappington [1988] and Armstrong [1999] assume that the firm knows the intercept of a linear demand function and the value of its marginal cost parameter, while the regulator does not have any of this information. Thus, there are two parameters of private information, but the regulator has only one instrument, the marginal price, to screen firms. This problem is amenable to techniques developed in Laffont, Maskin and Rochet [1987] and McAfee and MacMillan [1988]. Armstrong [1999] proves the optimality of exclusion in the Lewis and Sappington model[1]. Compared to the approach of Lewis and Sappington and Armstrong, my model is more tractable and allows for closed form solutions in a wide range of cases; in particular, I prove the optimality of prices below marginal costs, while the earlier work only suggests that this may happen. Moreover, due to the added tractability my model allows me to pin down precisely how the pattern of distortions depends on the features of the model as described above.

Rochet and Choné [1998], Armstrong [1996], and Wilson [1993] offer the most general treatment of and results on multidimensional incentive model. Rochet and Stole [2003] explain the obstacles researchers face when solving multidimensional models. In part this is due to the techniques required to tackle the problems. To an another part this is due to the ignorance of what one is actually looking for: it usually becomes much easier to prove the optimality of an allocation once it is known how it looks like. Armstrong and Rochet [1999] study a two-by-two model in detail and show under what conditions constraints are binding in which direction. This paper continues along these lines and studies the pattern of binding constraints and the resulting allocations in detail. Relatively simple solution techniques are shown to apply, which hopefully prove useful in general for different problems.

Some of the techniques I am using here were introduced in Beaudry, Blackorby, and Szalay [2009] in the context of a taxation model where workers possibilities to mimick others were limited in one dimension. In contrast, this paper makes no restriction on the feasible deviations. However, I obtain tractability by assuming a particular cost function. So, in some sense this model is in between the Rochet and Choné [1998] model and models with restrictions on feasible deviations.

The paper is organized as follows. In Section two I lay out the model and explain the regulator's allocation choice and its solution in the first-best. In Section three, I describe the set of implementable allocations and derive the regulator's control problem. In Section four, I lay out a benchmark case where constraints are binding in only one dimension. In Section five, I treat the multidimensional problem and discuss how binding constraints relate to bunching. Section six contains the first set of closed form solutions for the case of a fully separating solution (with binding constraints in both dimensions). Section seven contains closed form solutions for the case where bunching occurs everywhere. Section eight discusses hybrid cases. Section nine concludes. Long proofs have been relegated to the appendix.

[1] Technically, exclusion is optimal because the density of the sum of two random variables goes to zero at the bounds of its support. Armstrong [1996] shows that exclusion is robust in these kind of settings under more general assumptions.

2 The model and the main assumptions

There are two goods. Consumers' valuations for these two goods are given by the function $V(x,q)$, where x is the quantity of the first good and q the quantity of the second good. Consumer's valuation for good one is independent of the valuation for good two, so $V(x,q) = V^1(x) + V^2(q)$. Letting $P^1(x)$ denote the inverse demand function for good one and $P^2(q)$ denote the inverse demand function for good two, I have

$$V^1(x) \equiv \int_0^x P^1(z)\,dz$$

and

$$V^2(x) \equiv \int_0^q P^1(z)\,dz.$$

I assume that the inverse demand functions are differentiable and decreasing in x, so the valuations are twice differentiable and concave; obviously, the valuations are increasing in their argument. Good one is perfectly divisible, so x can be any non-negative amount. Good two can be produced in two different quantites q_1, and q_2, - or more generally, in two different variants - where $V^2(q_2) > V^2(q_1) > V^2(0) = 0$.

The goods are produced by a monopoly firm subject to price regulation. The firm's cost of producing the goods in quantities x and q is

$$C(x,q,\theta,\eta) = K + x\theta + q\underline{\eta} + \gamma\left[\eta - \underline{\eta}\right][q - q_1] + \delta xq,$$

where $\gamma > 0$, $\delta \geq 0$, and $K > 0$ are constants known to both regulator and firm, and x and q are verifiable so that contracts can be written on these variables; θ and η are parameters that are known to the firm but not to the regulator. The regulator knows only the joint distribution of these variables. θ and η are distributed on a product set $\mathbf{\Theta} \times \mathbf{H}$ with probability density function $f(\theta,\eta) > 0$ for all θ, η. The set $\mathbf{\Theta}$ is taken as the interval $\left[\underline{\theta}, \overline{\theta}\right]$, where $\underline{\theta} > 0$. The set $\mathbf{H}$ is taken as $\left\{\underline{\eta}, \overline{\eta}\right\}$ where $\overline{\eta} > \underline{\eta} > 0$. The marginal probability that $\eta = \underline{\eta}$ is equal to β. Let $G(\eta)$ denote the cdf of η. Given the full support assumption, for each η that has $dG(\eta) > 0$, the conditional distribution of θ given η has full support. The density and cdf of this distribution are denoted $f(\theta\,|\eta)$ and $F(\theta\,|\eta)$, respectively. Let $\mathcal{E}$ denote the expectation operator and let $f(\theta) \equiv \mathcal{E}_H\left[f(\theta\,|\eta)\right]$ and $F(\theta)$ denote the density and the cdf of the marginal distribution, respectively.

The cost and valuation functions are simple and rich enough at the same time. They are simple, because the firms knowledge about its cost of producing q matters only for equilibrium costs of production when a firm with parameter $\overline{\eta}$ is asked to produce $q = q_2$ but not at all if that firm is asked to produce $q = q_1$. Let

$$\Delta \equiv \gamma \left[q_2 - q_1\right] \left[\overline{\eta} - \underline{\eta}\right]. \tag{1}$$

Δ is the extra cost of production if a firm with parameter $\overline{\eta}$ is asked to produce q_2 instead of q_1. I assume throughout that paper that Δ is sufficiently large, so that a firm with parameter $\overline{\eta}$ produces q_1 at the optimum. On the other hand, the problem is rich enough in the sense that the optimal allocation in the multi-product model is substantially different from the optimal allocation in the one-product case. Notice that the cost function satisfies the standard Spence-Mirrlees conditions in x,θ and q, η, respectively.

The firm is subject to price regulation. However, it is equivalent and notationally much more convenient to analyze the model directly in terms of quantity regulation. If the firm produces quantities x and q then it receives a subsidy t and its profit is

$$t + P^1 (x)\, x + P^2 (q)\, q - C (x, q, \theta, \eta).$$

Define the sum of consumer and producer surplus as

$$S (x, q, \theta, \eta) \equiv V_1 (x) + V_2 (q) - C (x, q, \theta, \eta).$$

Notice that the surplus function is concave in x and q jointly; moreover, the function satisfies $S_{xq} (x, q, \theta, \eta) < 0$, so x and q are net substitutes in the surplus function[2].

2.1 The regulator's problem

I think of the regulator's problem in term's of a direct revelation mechanism, which is a triple of functions $\{q (\theta, \eta), x (\theta, \eta), t (\theta, \eta)\}$ for all $(\theta, \eta) \in \mathbf{\Theta} \times \mathbf{H}$ that satisfy incentive compatibility constraints. The regulator maximizes a weighted sum of net consumer surplus and producer surplus. If a firm announces parameters $\hat{\theta}$ and $\hat{\eta}$, then its profits are given by

$$\Pi \left(\hat{\theta}, \theta, \hat{\eta}, \eta\right) \equiv t \left(\hat{\theta}, \hat{\eta}\right) + P_1 \left(x \left(\hat{\theta}, \hat{\eta}\right)\right) x \left(\hat{\theta}, \hat{\eta}\right) + P_2 \left(q \left(\hat{\theta}, \hat{\eta}\right)\right) q \left(\hat{\theta}, \hat{\eta}\right) - C \left(x \left(\hat{\theta}, \hat{\eta}\right), q \left(\hat{\theta}, \hat{\eta}\right), \theta, \eta\right).$$

Under a truthful mechanism, the weighted joint surplus for a given pair (θ, η) is equal to

$$W (\theta, \eta) \equiv V_1 (x (\theta, \eta)) + V_2 (q (\theta, \eta)) - P_1 (x (\theta, \eta))\, x (\theta, \eta) + P_2 (q (\theta, \eta))\, q (\theta, \eta) - t (\theta, \eta) + \alpha \Pi (\theta, \theta, \eta, \eta)$$

where $\alpha \in (0, 1)$. Since α is kept constant throughout the paper, I suppress the dependence of the welfare function on α in what follows. I let Θ and H denote the random variables with typical

[2] See, e.g., Laffont and Tirole [1993].

realizations θ and η, respectively, and let $\mathcal{E}_{\Theta H}$ denote the expectation operator taken over the random variables Θ and H. The regulator solves problem following problem, which I denote as problem P:

$$\max_{x(\cdot,\cdot),q(\cdot,\cdot),t(\cdot,\cdot)} \mathcal{E}_{\Theta H} W(\theta,\eta) \tag{2}$$

s.t. for all θ, η and all $\hat{\theta}, \hat{\eta}$

$$\Pi(\theta,\theta,\eta,\eta) \geq \Pi\left(\hat{\theta},\theta,\hat{\eta},\eta\right), \tag{3}$$

and for all θ, η

$$\Pi(\theta,\theta,\eta,\eta) \geq 0, \tag{4}$$

(3) is the incentive compatibility condition, requiring that a firm of type θ, η must have no incentive to mimic any other type of firm;(4) requires that each firm in equilibrium obtain a non-negative profit.

2.2 The first-best

Since $\alpha < 1$, the regulator allocates all surplus to the consumer in the first-best allocation; the participation constraint is binding for each type, $\Pi(\theta,\theta,\eta,\eta) = 0$, so

$$t(\theta,\eta) = C(x(\theta,\eta), q(\theta,\eta), \theta,\eta) - P_1(x(\theta,\eta))\, x(\theta,\eta) - P_2(q(\theta,\eta))\, q(\theta,\eta).$$

Substituting for $t(\theta,\eta)$ into the regulator's objective function, I obtain

$$\max_{x(\cdot,\cdot),q(\cdot,\cdot)} \mathcal{E}_{\Theta H}\left(V_1(x(\theta,\eta)) + V_2(q(\theta,\eta)) - C(x(\theta,\eta), q(\theta,\eta), \theta,\eta)\right)$$

I assume throughout the paper that producing good two is valuable; moreover, if $V_x^1(0)$ is sufficiently large, the solution for x is interior as well. Given these assumptions, the first-best optimal policy for good on, $x^{fb}(\theta,\eta)$, satisfies equality of marginal benefits and costs, so

$$V_x^1\left(x^{fb}(\theta,\eta)\right) - C_x\left(x^{fb}(\theta,\eta), q^{fb}(\theta,\eta), \theta,\eta\right) = 0.$$

Consider now the first-best optimal policy for q. Assuming $V^2(q_2)$ is sufficiently large and δ and $\underline{\eta}$ sufficiently small, it is optimal to have firms with $\eta = \underline{\eta}$ good two in a large quantity, so $q^{fb}\left(\theta,\underline{\eta}\right) = q_2$. Assuming that $V^2(q_1)$, γ and $\overline{\eta}$ are sufficiently large, it is optimal to have firms with $\eta = \overline{\eta}$ produce quantity $q = q_1$, so $q^{fb}(\theta,\overline{\eta}) = q_1$. It is straightforward to characterize the optimal policy for q also in different cases; however, the present one is the most interesting one for the analysis that follows.

3 Statement of the problem

3.1 Implementable allocations

To solve the regulator's problem I begin by bringing the incentive and participation constraints, (3) and (4) into a more tractable form. Obviously, the set of implementable allocations for good one production depend on the implemented allocation for good two. However, by design the first-best allocation rule for good two continues to be optimal even when the firm has private information. The reason is the special way in which the parameter η enters the problem - of course, this is precisely why I assumed this particular form of cost function in the first place. To demonstrate this fact, I characterize optimal policies for good one assuming q is set according to the first-best rule; later on I verify that the conjectured form of policy for good q is indeed optimal.

Lemma 1 *If the regulator implements the first-best allocation rule for good two, then the incentive constraint* (3) *is equivalent to the pair of one-dimensional constraints*

$$\Pi(\theta,\theta,\eta,\eta) \geq \Pi\left(\hat{\theta},\theta,\eta,\eta\right) \tag{5}$$

and

$$\Pi(\theta,\theta,\eta,\eta) \geq \Pi(\theta,\theta,\hat{\eta},\eta). \tag{6}$$

Proof of Lemma 1. Clearly, (5) and (6) are necessary for (3). So, I need to show that they are sufficient as well.

Suppose that the regulator follows a good two allocation rule $q(\theta,\overline{\eta}) = q_1$ and $q\left(\theta,\underline{\eta}\right) = q_2$. Then, the profit of a type $(\theta,\overline{\eta})$ firm mimicking a type $\left(\hat{\theta},\underline{\eta}\right)$ firm is equal to

$$\begin{aligned}\Pi\left(\hat{\theta},\theta,\underline{\eta},\overline{\eta}\right) &= t\left(\hat{\theta},\underline{\eta}\right) + P_1\left(x\left(\hat{\theta},\underline{\eta}\right)\right)x\left(\hat{\theta},\underline{\eta}\right) + P_2(q_2)\,q_2 - C\left(x\left(\hat{\theta},\underline{\eta}\right),q_2,\theta,\overline{\eta}\right)\\ &= t\left(\hat{\theta},\underline{\eta}\right) + P_1\left(x\left(\hat{\theta},\underline{\eta}\right)\right)x\left(\hat{\theta},\underline{\eta}\right) + P_2(q_2)\,q_2 - C\left(x\left(\hat{\theta},\underline{\eta}\right),q_2,\theta,\underline{\eta}\right) - \Delta,\end{aligned}$$

so

$$\Pi\left(\hat{\theta},\theta,\underline{\eta},\overline{\eta}\right) = \Pi\left(\hat{\theta},\theta,\underline{\eta},\underline{\eta}\right) - \Delta \tag{7}$$

Likewise, the profit of a type $\left(\theta,\underline{\eta}\right)$ firm mimicking a type $\left(\hat{\theta},\overline{\eta}\right)$ firm is equal to

$$\Pi\left(\hat{\theta},\theta,\overline{\eta},\underline{\eta}\right) = t\left(\hat{\theta},\overline{\eta}\right) + P_1\left(x\left(\hat{\theta},\overline{\eta}\right)\right)x\left(\hat{\theta},\overline{\eta}\right) + P_2(q_1)\,q_1 - C\left(x\left(\hat{\theta},\overline{\eta}\right),q_1,\theta,\underline{\eta}\right),$$

so

$$\Pi\left(\hat{\theta},\theta,\overline{\eta},\underline{\eta}\right) = \Pi\left(\hat{\theta},\theta,\overline{\eta},\overline{\eta}\right) \tag{8}$$

By (5) , it follows that

$$\max_{\hat{\theta}} \Pi\left(\hat{\theta}, \theta, \underline{\eta}, \overline{\eta}\right) = \Pi\left(\theta, \theta, \underline{\eta}, \underline{\eta}\right) - \Delta,$$

that is, the best feasible report in the θ dimension is to report the truth, even conditional on having misrepresented the η dimension. On the other hand,

$$\Pi\left(\theta, \theta, \underline{\eta}, \underline{\eta}\right) - \Delta = \Pi\left(\theta, \theta, \underline{\eta}, \overline{\eta}\right)$$

So, by (6) for type $(\theta, \overline{\eta})$,

$$\Pi\left(\theta, \theta, \overline{\eta}, \overline{\eta}\right) \geq \Pi\left(\theta, \theta, \underline{\eta}, \overline{\eta}\right)$$

the best feasible deviation is suboptimal for type $(\theta, \overline{\eta})$.

By the (5) for type $\left(\theta, \underline{\eta}\right)$ it follows that

$$\max_{\hat{\theta}} \Pi\left(\hat{\theta}, \theta, \overline{\eta}, \underline{\eta}\right) = \Pi\left(\theta, \theta, \overline{\eta}, \overline{\eta}\right)$$

so (6) for type $\left(\theta, \underline{\eta}\right)$, that is

$$\Pi\left(\theta, \theta, \underline{\eta}, \underline{\eta}\right) \geq \Pi\left(\theta, \theta, \overline{\eta}, \underline{\eta}\right),$$

even the most profitable deviation available to type $\left(\theta, \underline{\eta}\right)$ is suboptimal. ■

The intuition for this result is very simple. A firm's incentive to report its cost parameter θ do not depend on what the firm reported about its cost parameter η, and vice versa. To see this, suppose a firm with cost parameters (θ, η) announces $\hat{\eta} \neq \eta$. Its profit differs from the profit of a firm with cost parameters $(\theta, \hat{\eta})$ by the amount $\left[q\left(\theta, \overline{\eta}\right) - q\left(\theta, \underline{\eta}\right)\right] \underline{\eta}$. However, as long as the functions $q\left(\theta, \overline{\eta}\right)$ and $q\left(\theta, \underline{\eta}\right)$ are independent of θ, the difference in profits is an additive constant. Hence, the firm's optimal report in the θ dimension is not affected by its report in the η dimension. Hence, the two-dimensional constraint breaks down into a pair of one dimensional constraints. This is a crucial difference to the multi-dimensional problem of Rochet and Choné (2003), where the reduction of incentive compatibility conditions is not possible.

It is the specific form of the cost function that implies that the constraints fall apart; it is essential that the optimal allocation rule for good two production is independent of θ. A second implication of this cost function is that knowledge of η does not give rise to informational rents. The reason is that, at the optimal allocation, the costs of a firm with parameter $\underline{\eta}$ and a firm with parameter $\overline{\eta}$ (both having the same cost parameter θ) differ only by factors that are observable (that is, x and q). However, observable cost differences are just reimbursed by the regulator, given that $\alpha < 1$. This insight allows me to prove the following result. Let $\pi\left(\theta, \eta\right) \equiv \max_{\hat{\theta}, \hat{\eta}} \Pi\left(\hat{\theta}, \theta, \hat{\eta}, \eta\right)$.

Lemma 2 *i) The incentive constraint* (5) *is satisfied if and only if*

$$t(\theta,\eta) = C(x(\theta,\eta), q(\theta,\eta), \theta, \eta) + \pi(\overline{\theta},\eta) + \int_{\theta}^{\overline{\theta}} x(y,\eta)\,dy - x(\theta,\eta)\,P_1(x(\theta,\eta), q(\theta,\eta)) - P_2(q(\theta,\eta))\,q(\theta,\eta) \tag{9}$$

and $x(\theta,\eta)$ is non-increasing in θ for all η;

ii) The incentive constraint (6) *is satisfied if and only if*

$$\Delta \geq \pi(\theta,\underline{\eta}) - \pi(\theta,\overline{\eta}) \geq 0 \tag{10}$$

where Δ is defined in (1)*;*

iii) The participation constraints (4) *are met if $\pi(\overline{\theta},\eta) \geq 0$ for all η.*

The proof of part i) in the Lemma is standard and therefore omitted. I compute the rent of a firm of type (θ,η) by summing the rent of the most inefficient cost type for a given quality capacity, $\pi(\overline{\theta},\eta)$, and the marginal changes of the firm's rent with respect to changes in its cost parameter θ. Notice that (9) allows for the case where $\pi(\overline{\theta},\eta) > 0$, so some high cost types may receive rents. Apart from allowing for rents for the most inefficient types, I can essentially use the standard procedure as in the one-dimensional case. Part ii) follows directly from Lemma 1. In the proof of Lemma 1, I have shown that

$$\Pi(\theta,\theta,\underline{\eta},\overline{\eta}) = \Pi(\theta,\theta,\underline{\eta},\underline{\eta}) - \Delta$$

and

$$\Pi(\theta,\theta,\overline{\eta},\underline{\eta}) = \Pi(\theta,\theta,\overline{\eta},\overline{\eta}).$$

In words, differences in profits when mimicking a firm with a different cost of producing good two are captured entirely by differences in "fixed costs". Condition (10) merely restates this finding. Finally, part iii) is obvious by the usual argument in one-dimensional models implying that the single-crossing condition (in x and θ) implies that the participation constraint can only bind at one end.

4 The control problem

I can ease notation letting $\overline{x}(\theta) \equiv x(\theta,\overline{\eta})$ and $\underline{x}(\theta) \equiv x(\theta,\underline{\eta})$, and likewise for the quantity schedules of good two, $\overline{q}(\theta)$ and $\underline{q}(\theta)$, for the transfer schedules $\overline{t}(\theta)$ and $\underline{t}(\theta)$, and the rent schedules $\overline{\pi}(\theta)$ and $\underline{\pi}(\theta)$. I let $\overline{\pi} \equiv \overline{\pi}(\overline{\theta})$ and $\underline{\pi} \equiv \underline{\pi}(\overline{\theta})$. Define the virtual surplus

$$B(x,q,\theta,\eta) \equiv V_1(x) + V_2(q) - C(x,q,\theta,\eta) - (1-\alpha)\,x\frac{F(\theta\,|\eta)}{f(\theta\,|\eta)}. \tag{11}$$

For future reference, I also define the excess rent of a type $\left(\theta, \underline{\eta}\right)$ over a type $(\theta, \overline{\eta})$ as

$$\rho\left(\theta, \overline{\pi}, \underline{\pi}\right) \equiv \underline{\pi} + \int_{\theta}^{\overline{\theta}} \underline{x}\left(y\right) dy - \overline{\pi} - \int_{\theta}^{\overline{\theta}} \overline{x}\left(y\right) dy.$$

Using (9) to substitute out transfers from the regulator's problem, and integrating by parts, I obtain obtain the following representation of the regulator's problem, which, for future reference, I denote as problem P'

$$\max_{\overline{x}(\cdot), \underline{x}(\cdot), \overline{\pi}, \underline{\pi}} \left\{ \begin{array}{c} \beta \int_{\underline{\theta}}^{\overline{\theta}} B\left(\underline{x}\left(\theta\right), q_2, \theta, \underline{\eta}\right) f\left(\theta \left| \underline{\eta}\right.\right) d\theta - \beta\left(1-\alpha\right) \underline{\pi} \\ + \left(1-\beta\right) \int_{\underline{\theta}}^{\overline{\theta}} B\left(\overline{x}\left(\theta\right), q_1, \theta, \overline{\eta}\right) f\left(\theta \left| \overline{\eta}\right.\right) d\theta - \left(1-\beta\right)\left(1-\alpha\right) \overline{\pi} \end{array} \right\} \tag{12}$$

$$s.t.$$

$$\underline{\pi} + \int_{\theta}^{\overline{\theta}} \underline{x}\left(y\right) dy - \overline{\pi} - \int_{\theta}^{\overline{\theta}} \overline{x}\left(y\right) dy \geq 0 \tag{13}$$

$$\underline{\pi} + \int_{\theta}^{\overline{\theta}} \underline{x}\left(y\right) dy - \overline{\pi} - \int_{\theta}^{\overline{\theta}} \overline{x}\left(y\right) dy \leq \Delta, \text{ and} \tag{14}$$

$$\overline{x}\left(\theta\right), \underline{x}\left(\theta\right) \text{ non-increasing in } \theta. \tag{15}$$

Problem P' has the following structure. If the monotonicity constraints on $\overline{x}\left(\theta\right)$ and $\underline{x}\left(\theta\right)$ are nonbinding, the problem can be viewed as a control problem with two control variables, $\overline{x}\left(\theta\right)$ and $\underline{x}\left(\theta\right)$, and two state variables, $-\int_{\theta}^{\overline{\theta}} \overline{x}\left(y\right) dy$ and $-\int_{\theta}^{\overline{\theta}} \underline{x}\left(y\right) dy$. Moreover, the state variables enter the problem through inequality constraints. This is a relatively complex problem, but solution techniques are available in the literature (see, e.g., Kamien and Schwartz (1981) or Seyerstad and Sydsaeter (1999)). If the monotonicity constraints are binding for some θ, the problem involves second derivatives. This case becomes extremely difficult to analyze. Therefore my approach is to impose assumptions that guarantee that the monotonicity constraints are slack at the solution to problem P'.

5 A benchmark

Before I dive into the main analysis of this problem, it is useful to look into a benchmark case where all the constraints are automatically satisfied.

Suppose Δ is sufficiently large, so at the optimum a firm with parameter $\overline{\eta}$ produces q_1. Suppose further that $\delta = 0$, so that goods one and two are neither net substitutes nor net complements.

Moreover, suppose that θ conditional on $\eta = \overline{\eta}$ is smaller than θ conditional on $\eta = \underline{\eta}$ in the reversed hazard rate order:

Assumption Bi: $\frac{f(\theta|\overline{\eta})}{F(\theta|\overline{\eta})} \leq \frac{f(\theta|\underline{\eta})}{F(\theta|\underline{\eta})}$ for all θ.

This implies that θ conditional on $\eta = \overline{\eta}$ is smaller in the usual stochastic order $\eta = \underline{\eta}$ (that is, First Order Stochastic Dominance) than θ conditional on $\eta = \overline{\eta}$ (see Shaked and Shantikumar (2007)). Finally, suppose that the conditional reversed hazard rates are monotonic in θ :

Assumption Bii: For all θ $\frac{\partial}{\partial\theta}\frac{F(\theta|\eta)}{f(\theta|\eta)} \geq 0$ for $\eta \in \{\underline{\eta}, \overline{\eta}\}$.

In this case the optimal allocation is very easy to characterize:

Proposition 1 *Suppose $\delta = 0$, Δ is sufficiently large, and Assumptions **Bi** and **Bii** hold. Then, the optimal allocation for good two is $\overline{q}^*(\theta) = q_1$ and $\underline{q}^*(\theta) = q_2$ for all θ. The optimal allocation for good one satisfies*

$$V_x^1(\underline{x}^*(\theta)) = C_x\left(\underline{x}^*(\theta), q_2, \theta, \underline{\eta}\right) + (1-\alpha)\frac{F(\theta|\underline{\eta})}{f(\theta|\underline{\eta})} \tag{16}$$

and

$$V_x^1(\overline{x}^*(\theta)) = C_x(\overline{x}^*(\theta), q_1, \theta, \overline{\eta}) + (1-\alpha)\frac{F(\theta|\overline{\eta})}{f(\theta|\overline{\eta})} \tag{17}$$

and thus $\underline{x}^*(\theta) \geq \overline{x}^*(\theta)$. The optimal transfer schedules satisfy (9). There is no rent at the bottom, that is, $\overline{\pi}^* = \underline{\pi}^* = 0$.

The proof is very simple and uses only well known arguments, so I give only a heuristic sketch. If Δ is sufficiently large, it is obvious that $\overline{q}^*(\theta) = q_1$. As long as good two is sufficiently valuable, it is optimal to set $\underline{q}^*(\theta) = q_2$. Suppose now that problem P' is solved neglecting all its constraints. The resulting quantity schedules are precisely the ones given in the proposition. The point is that these schedules satisfy all the constraints under the conditions given in the proposition. In particular, the conditions imply that $\underline{x}^*(\theta) \geq \overline{x}^*(\theta)$ for all θ, so (13) is automatically satisfied for $\overline{\pi} = \underline{\pi} = 0$. Moreover, if feasible, then it is of course always optimal to extract all the rents from high θ types. If Δ is sufficiently large, then (14) is satisfied as well. Finally, under the maintained assumption of monotonic reverse hazard rates, the schedules $\underline{x}^*(\theta)$ and $\overline{x}^*(\theta)$ are monotonic.

The intuition is simple too. The schedules $\underline{x}^*(\theta)$ and $\overline{x}^*(\theta)$ are optimal resolutions of rents versus efficiency trade-offs. The higher is $x(\theta)$ the higher are the rents the regulator has to leave to types with costs smaller than θ. On the other hand, the regulator wishes to set an efficient quantity so as to raise surplus. The reverse hazard rate determines the weight given to each of the two motives. If the reversed hazard rate is smaller for $\eta = \overline{\eta}$, then raising η increases the regulator's concern for efficiency relative to rent extraction. Hence, truthfully revealing a low type $\underline{\eta}$ is in the firm's interest. A high $\overline{\eta}$ firm has no incentive to mimick a low $\underline{\eta}$ firm because Δ, the cost of this deviation is too high. Hence, only the incentive constraints in the θ dimension are binding and the problem can be solved as a family of one dimensional problems.

6 The multi-dimensional case

I now solve the more interesting case where incentive constraints in both dimensions are binding. To make progress beyond an "anything goes" characterization, I maintain the assumption that the reversed hazard rate can be ordered. In particular, I impose for the remainder of the article

Assumption M: $\frac{f(\theta|\underline{\eta})}{F(\theta|\underline{\eta})} \leq \frac{f(\theta|\overline{\eta})}{F(\theta|\overline{\eta})}$ for all θ.

As explained above, assumption I implies that a firm has stochastically lower costs of producing good one conditional on having low costs of producing good two. It is easy to see that the allocation given in the last section can no longer be optimal as it now would violate condition (13).

I continue to assume that Δ is sufficiently large, which implies that (14) is never binding at the optimum. In that case it is also immediate that $\overline{\pi}^* = 0$; that is, the firm with the highest cost in both dimensions must have no rent at the optimum. The reason is that reducing $\overline{\pi}$ relaxes constraint (13) and raises the regulator's objective. Hence, the excess rent of a low η type over his high η counterpart can be written as

$$\rho(\theta, \underline{\pi}) \equiv \underline{\pi} + \int_{\theta}^{\overline{\theta}} \underline{x}(y)\, dy - \int_{\theta}^{\overline{\theta}} \overline{x}(y)\, dy. \tag{18}$$

Consider a "reduced" version of the regulator's problem, which for future reference is denoted problem P":

$$\max_{\overline{x}(\cdot), \underline{x}(\cdot), \underline{\pi}} \left\{ \begin{array}{c} \beta \int_{\underline{\theta}}^{\overline{\theta}} B\left(\underline{x}(\theta), \underline{\eta}, \theta, \underline{\eta}\right) f\left(\theta \left| \underline{\eta}\right.\right) d\theta - \beta(1-\alpha)\underline{\pi} \\ + (1-\beta) \int_{\underline{\theta}}^{\overline{\theta}} B\left(\overline{x}(\theta), \overline{\eta}, \theta, \overline{\eta}\right) f\left(\theta \left| \overline{\eta}\right.\right) d\theta \end{array} \right\} \tag{19}$$

s.t.

$$\rho(\theta, \underline{\pi}) \geq 0. \tag{20}$$

The problem is reduced in the sense that (15) is omitted from the problem; as I have argued above, omitting (14) is without loss of generality as this constraint is never binding for Δ large.

In now solve problem P". If the solution to this reduced problem satisfies all the constraints of the full problem, then the solution to the full problem has been found. This is the case under restrictions on the distribution of types.

6.1 The pattern of binding constraints

I begin with a few general observations about the problem and its solution. First, I observe that constraint (13) is binding for some θ at the optimum. To see this, suppose (13) were non-binding for all θ. Then the solution to the problem would involve the quantity schedules $\underline{x}^*(\theta)$ and $\overline{x}^*(\theta)$ defined by (16) and (17) in Proposition 1. But this requires that the regulator leaves at least a rent $\underline{\tilde{\pi}}$ to all firms with low costs of producing good two, where

$$\underline{\tilde{\pi}} \equiv \int_{\theta}^{\overline{\theta}} \overline{x}^*(y)\, dy - \int_{\theta}^{\overline{\theta}} \underline{x}^*(y)\, dy > 0.$$

$\underline{\tilde{\pi}}$ is strictly positive because, due to Assumption **M**, the schedules defined by the conditions (16) and (17) in Proposition 1 satisfy $\underline{x}^*(\theta) \leq \overline{x}^*(\theta)$. However, setting $\underline{\pi} \geq \underline{\tilde{\pi}}$ cannot be optimal. Around $\underline{\pi} = \underline{\tilde{\pi}}$, the marginal cost of increasing $\underline{\pi}$ is equal to $-\beta(1-\alpha)$; a fraction of firms β has low costs of producing good two and rents left to firms enter the regulators payoff function with a weight of $-(1-\alpha)$. On the other hand, the benefit of increasing $\underline{\pi}$ around $\underline{\pi} = \underline{\tilde{\pi}}$ is zero, as the regulator is already unconstrained by condition (13) for $\underline{\pi} = \underline{\tilde{\pi}}$. Hence, at the optimum I must have $0 \leq \underline{\pi}^* < \underline{\tilde{\pi}}$.[3]

It is possible to locate where (20) is necessarily binding.

Lemma 3 *Consider the "reduced"* (19) *subject to* (20). *At the solution to this problem, constraint* (20) *is binding at* $\theta = \underline{\theta}$.

The proof is a simple argument by contradiction. Let $\overline{x}^*(\theta)$ and $\underline{x}^*(\theta)$ denote the optimal quantity schedules solving the reduced problem (19) subject to (20). If constraint (20) were slack at $\theta = \underline{\theta}$, I could use the transversality conditions of the problem (and do so in the appendix) to conclude that $\overline{x}^*(\theta)$ and $\underline{x}^*(\theta)$ satisfy conditions (16) and (17) defined in Proposition 1 for all $\theta \leq \theta'$, where θ' is the smallest θ where (20) is binding. By definition of the point θ', the excess rent of type $(\theta', \underline{\eta})$ is zero, that is $\rho(\theta', \underline{\pi}) = 0$. However, from the discussion above we know that $\underline{x}^*(\theta) \leq \overline{x}^*(\theta)$ given assumption I, where the inequality is strict for $\delta > 0$.Hence, I have

$$\int_{\underline{\theta}}^{\theta'} \underline{x}^*(y)\, dy - \int_{\underline{\theta}}^{\theta'} \overline{x}^*(y)\, dy < 0.$$

But this shows that

$$\rho(\underline{\theta}, \underline{\pi}) = \rho(\theta', \underline{\pi}) + \int_{\underline{\theta}}^{\theta'} \underline{x}^*(y)\, dy - \int_{\underline{\theta}}^{\theta'} \overline{x}^*(y)\, dy < 0,$$

[3] This heuristic argument is made more formally in the proof of Proposition 2 below.

for any $\theta' > \underline{\theta}$. Hence, I must have $\theta' = \underline{\theta}$, that is, constraint (20) is binding at $\underline{\theta}$. The complete statement of the problem and the details of the argument can be found in the formal proof in the appendix.

6.2 Binding Constraints and Bunching

Suppose there is an interval $[\theta', \theta'']$ such that constraint (20) binds over that interval, that is for all $\theta \in [\theta', \theta'']$

$$\underline{\pi} + \int_{\theta}^{\overline{\theta}} \underline{x}(y)\, dy - \int_{\theta}^{\overline{\theta}} \overline{x}(y)\, dy = 0.$$

Differentiating this condition with respect to θ, I have

$$\overline{x}(\theta) = \underline{x}(\theta) \text{ for all } \theta \in [\theta', \theta''].$$

Thus, when the incentive constraint in the $\eta-$dimension is binding for θ in some nonempty interval, then the solution schedules $\overline{x}(\theta)$ and $\underline{x}(\theta)$ involve bunching.

Thus, unless constraint (20) is binding on a set of isolated points, the solution schedules $\overline{x}(\theta)$ and $\underline{x}(\theta)$ must be identical (i.e., independent of η) for some realizations of costs θ. Of course there could be a second sort of bunching in the θ dimension, that is at least one of the schedules $\overline{x}(\theta)$ or $\underline{x}(\theta)$ could have flat parts. I rule this case out by imposing assumptions on the joint distribution of θ and η. There are two reasons to do this. First, bunching in one dimension is well understood by now, so I concentrate on the novelties here. Second, the problem becomes quickly untractable when the two sorts of bunching are present together and interact with each other.

It proves useful to organize the analysis along the predictions arising from it and the assumptions needed to get them. I start with the simplest and perhaps most illuminating case where the solution is fully separating between all types. I obtain this when the θ and η are statistically independent of each other so that the only reason why the problem is truly multidimensional is that the two goods are strict net substitutes (that is $\delta > 0$). In that case, condition (20) binds at the optimal allocation only for $\theta = \underline{\theta}$ if the distribution of θ has a nonincreasing density. For densities that are increasing over some range, even if independent, the analysis becomes more involved, but remains tractable, as condition (20) binds over a single interval at the optimum. To study that case it is instructive to analyze first a case that leads to complete bunching. That is the case when the two goods are neutral (that is $\delta = 0$), so the only reason that the problem is multidimensional is that there is some correlation of types. If θ and η are affiliated[4], then the problem involves bunching

[4]This is consistent with assumption I as affiliation implies reverse hazard rate dominance (see Shaked and

over the whole support. Finally, I characterize the solution if types are affiliated and the goods strict net substitutes. In that case, bunching occurs again for types at the lower end of costs.

7 Strict net substitutes, independence, and full separation

In this section, I focus on the case where there is a real, that is a non-information based reason for bunching. If $\delta > 0$, then the two goods are strict net substitutes in the regulator's payoff function. Raising production from q_1 to q_2 raises the marginal cost of producing good one. To isolate the role of this substitutability, I assume in this section that knowing η does not provide any additional information about θ :

Assumption Ii: $f(\theta\,|\overline{\eta}) = f\left(\theta\,|\underline{\eta}\right) = f(\theta)$.

To eliminate bunching from the problem, I impose:

Assumption Iii: $f_\theta(\theta) \leq 0$ and $\frac{\partial}{\partial\theta}\frac{1-F(\theta)}{f(\theta)} \leq 0.$

This assumption serves two purposes. First, guarantees that the monotonicity constraints (15) are automatically satisfied. Second, it guarantees that at the solution to problem P", (20) binds only at $\theta = \underline{\theta}$.

Consider thus a "doubly reduced" problem where the regulator maximizes (19) subject to

$$\rho\left(\underline{\theta}, \underline{\pi}\right) = 0. \tag{21}$$

For future reference, I denote this problem as P"'. Under assumptions **Ii** and **Iii**, solving the doubly reduced problem picks up the full optimum.

Let k denote the Lagrange multiplier attached to the constraint (21). I can now characterize the solution to my problem.

Proposition 2 *Suppose that Δ is sufficiently large and δ is strictly positive. Then, under Assumptions **Ii** and **Iii**, the solution to the regulation problem is given by the quality schedules $\overline{q}^*(\theta) = q_1$ and $\underline{q}^*(\theta) = q_2$, the quantity schedules $\overline{x}^*(\theta)$ and $\underline{x}^*(\theta)$ defined by the conditions*

$$\left(V_x^1\left(\underline{x}^*(\theta)\right) - C_x\left(\underline{x}^*(\theta), q_2, \theta, \underline{\eta}\right) - (1-\alpha)\frac{F(\theta)}{f(\theta)}\right)\beta f(\theta) + k^* = 0 \tag{22}$$

and

$$\left(V_x^1\left(\overline{x}^*(\theta)\right) - C_x\left(\overline{x}^*(\theta), q_1, \theta, \overline{\eta}\right) - (1-\alpha)\frac{F(\theta)}{f(\theta)}\right)(1-\beta) f(\theta) - k^* = 0, \tag{23}$$

Shantikumar (2007)).

and the associated transfer schedules $\overline{t}^*(\theta)$ *and* $\underline{t}^*(\theta)$ *defined by* (9). *Moreover,* $\underline{\pi}^* < \tilde{\underline{\pi}}$ *and either*

$$k^* \leq \beta(1-\alpha) \text{ and } \pi^* = 0, \text{ or} \tag{24}$$

$$k^* = \beta(1-\alpha) \text{ and } \pi^* > 0. \tag{25}$$

The proof of the Proposition simply consists in verifying that all the constraints are met; the details are in the appendix. It is easy to verify that the solution schedules (22) and (23) satisfy

$$\underline{x}(\theta) = \overline{x}(\theta) \Longrightarrow \frac{d\underline{x}(\theta)}{d\theta} \geq \frac{d\overline{x}(\theta)}{d\theta}, \tag{26}$$

a single crossing condition. Since these schedules are continuous, they do indeed cross at most once. Furthermore, it cannot be the case that $\underline{x}(\theta) > \overline{x}(\theta)$ for all θ, since that would imply that condition (20) is be slack, contradicting the lemma above. Thus, two cases can arise. First, it is possible that $\underline{x}(\theta) < \overline{x}(\theta)$ for all θ. Since

$$\rho_\theta(\theta, \underline{\pi}) = \overline{x}(\theta) - \underline{x}(\theta),$$

in this case $\rho(\theta, \underline{\pi})$ is minimized at $\theta = \underline{\theta}$, which is consistent with (20) binding only at $\underline{\theta}$. Second, it is possible that $\underline{x}(\theta) < \overline{x}(\theta)$ for small θ and $\underline{x}(\theta) > \overline{x}(\theta)$ for large θ. In that case $\rho(\theta, \underline{\pi})$ is increasing for small θ and decreasing for large θ, which is again consistent with (20) binding only at $\underline{\theta}$. It is also straightforward to verify monotonicity of the schedules defined by (22) and (23), but I refer the reader to the appendix for this.

The solution is a remarkably simple pair of first-order conditions. Up to the optimal choice of π^* Proposition 2 provides a complete characterization of the optimum. The solution schedules (22) and (23) differ markedly from their counterparts for the case where η is known. In particular, $\underline{x}^*(\theta)$ is distorted upwards relative to the case where the cost of producing good two is known, and $\overline{x}^*(\theta)$ is distorted downwards relative to that case. The reason is of course that the schedules one obtains for known costs of producing good two violate constraint (20). A particular implication of constraint (20) being binding is that the schedules (22) and (23) violate the "no distortion at the top" condition both within firms with the same cost of producing good two and overall. A firm with cost parameters $(\underline{\theta}, \underline{\eta})$ produces more than the efficient quantity of good one, a firm with cost parameters $(\underline{\theta}, \overline{\eta})$ produces less than the efficient quantity of good one. These distortions are largest when k^* the equilibrium value of the multiplier attached to constraint (20) is as large as possible. I now show that the upper bound on k^* can be attained.

7.1 The complete optimum for the uniform case

As shown in the proof of Proposition 2, the problem is concave in the choice variables. Hence, $\pi^* > 0$ if and only if $k(0) \equiv k(\pi)|_{\pi=0} > \beta(1-\alpha)$. It is straightforward to compute $k(0)$ from conditions (22), (23), and (21). To do so, one needs to assume a particular density. I now compute the full optimum for the case of a uniform type distribution. I obtain

$$\underline{x}^*(\theta) = \left(V_x^1\right)^{-1}\left(\theta + (1-\alpha)\frac{(\theta-\underline{\theta})}{\overline{\theta}-\underline{\theta}} + \delta q_2 - \frac{k}{\beta}\left(\overline{\theta}-\underline{\theta}\right)\right) \tag{27}$$

and

$$\overline{x}^*(\theta) = \left(V_x^1\right)^{-1}\left(\theta + (1-\alpha)\frac{(\theta-\underline{\theta})}{\overline{\theta}-\underline{\theta}} + \delta q_1 + \frac{k}{1-\beta}\left(\overline{\theta}-\underline{\theta}\right)\right) \tag{28}$$

From (21) with $\underline{\pi} = 0$, I conclude that $k(0)$ satisfies the condition $\delta q_2 - \frac{k(0)}{\beta}\left(\overline{\theta}-\underline{\theta}\right) = \delta q_1 + \frac{k(0)}{1-\beta}\left(\overline{\theta}-\underline{\theta}\right)$, for otherwise I would have either $\overline{x}^*(\theta) > \underline{x}^*(\theta)$ for all θ or $\overline{x}^*(\theta) < \underline{x}^*(\theta)$ for all θ, and both possibilities are inconsistent with condition (21). Hence, I have $k(0) = \frac{(1-\beta)\beta}{\overline{\theta}-\underline{\theta}}\delta(q_2-q_1)$. The following result is now immediate:

Proposition 3 *For a uniform distribution of* θ, *the optimal quantity schedules are given by* (27) *and* (28), *where* $\underline{\pi}^* > 0$ *and* $k^* = \beta(1-\alpha)$ *for* $\frac{1-\beta}{\overline{\theta}-\underline{\theta}}\delta(q_2-q_1) > (1-\alpha)$.

The conditions in the Proposition are easy to meet - and consistent with the requirement that Δ be sufficiently large. Thus, leaving a rent to type $\left(\overline{\theta},\underline{\eta}\right)$ is natural, rather than a pathological outcome. The intuition for the result is that increasing $\underline{\pi}$ allows the regulator to tailor the quantity schedules $\overline{x}^*(\theta)$ and $\underline{x}^*(\theta)$ better to the marginal costs - which are higher for types with low cost of producing good two, because they are asked to produce the larger amount of good two. It becomes optimal to do so if, all else equal, the weight attached to firm profits becoms larger, if the fraction of firms with a low cost of producing good two becomes smaller, and if the difference in marginal costs of producing good one become larger.

7.2 Optimal pricing with rents for inefficient producers

For the case where the regulator leaves a positive rent to all high quality producers, the solution has quite unconventional features. Substituting $k^* = \beta(1-\alpha)$ (from (25)) into (22) and (23), I obtain the optimal quantity schedules for the case where $\pi^* > 0$:

$$V_x^1\left(\underline{x}^*(\theta)\right) = C_x\left(\underline{x}^*(\theta), q_2, \theta, \underline{\eta}\right) - (1-\alpha)\frac{1-F(\theta)}{f(\theta)} \tag{29}$$

and

$$V_x^1\left(\overline{x}^*\left(\theta\right)\right) = C_x\left(\overline{x}^*\left(\theta\right), q_1, \theta, \overline{\eta}\right) + \left(1-\alpha\right)\frac{\frac{\beta}{1-\beta} + F\left(\theta\right)}{f\left(\theta\right)}. \tag{30}$$

Since $V_x^1(x) = P^1(x)$, these conditions relate marginal prices to marginal costs and statistics of the type distribution. The optimal marginal prices of the producer with low costs of producing good two are below marginal costs for all but the the least efficient type $\left(\overline{\theta}, \underline{\eta}\right)$. This is to implement a quantity allocation that is distorted upwards relative to the first-best quantity for all but the least efficient type $\left(\overline{\theta}, \underline{\eta}\right)$. The least efficient type produces the efficient quantity. For producers with high costs of producing good two, the pricing scheme and the implemented allocation for good one has almost but not quite the traditional features: these firms all price above marginal costs, even the most efficient of these producers. Obviously, these distortions arise due to the constraint that prevents the firms with the low cost of producing good two from producing a low amount of good two, that is from mimicking the high-good-two-cost-producer with the same cost of producing good one.

It is possible to characterize the solution also for the case of an increasing density. However, it is most convenient to present this case after the analysis of correlated types, because the result is basically a corollary of that analysis.

8 Neutral goods, affiliated types, and complete bunching

In this section I focus on information based reasons for binding constraints in both dimensions. In particular, I assume that $\delta = 0$, that is, the goods are neutral. Moreover, I impose:

Assumption Ai: θ and η are affiliated, i.e. $\frac{f(\theta|\overline{\eta})}{f(\theta|\underline{\eta})}$ is increasing in θ.

The reason I assume affiliation[5] is that it allows me to pin down the bunching region.

Lemma 4 *i) If the schedules $\underline{x}(\theta)$ and $\overline{x}(\theta)$ defined by*

$$\left(V_x^1\left(\underline{x}\left(\theta\right)\right) - C_x\left(\underline{x}\left(\theta\right), q_2, \theta, \underline{\eta}\right) - \left(1-\alpha\right)\frac{F\left(\theta|\,\underline{\eta}\right)}{f\left(\theta|\,\underline{\eta}\right)}\right)\beta f\left(\theta|\,\underline{\eta}\right) + k = 0 \tag{31}$$

and

$$\left(V_x^1\left(\overline{x}\left(\theta\right)\right) - C_x\left(\overline{x}\left(\theta\right), q_1, \theta, \overline{\eta}\right) - \left(1-\alpha\right)\frac{F\left(\theta|\,\overline{\eta}\right)}{f\left(\theta|\,\overline{\eta}\right)}\right)\left(1-\beta\right) f\left(\theta|\,\overline{\eta}\right) - k = 0, \tag{32}$$

satisfy for any $k \geq 0$

$$\overline{x}\left(\theta\right) = \underline{x}\left(\theta\right) \Longrightarrow \frac{d\overline{x}\left(\theta\right)}{d\theta} > \frac{d\underline{x}\left(\theta\right)}{d\theta}, \tag{33}$$

[5] Affiliation is consistent with the reverse hazard rate order; more precisely, affiliation implies the reverse hazard rate order but is not implied by it. (See Shaked and Shantikumar (2007)).

then at the solution to problem P", constraint (20) *is binding on a set* $\left[\underline{\theta}, \theta'\right]$ *for some* $\theta' \geq \underline{\theta}$.

ii) For $\delta = 0$ *and under assumption A,* $\underline{x}(\theta)$ *and* $\overline{x}(\theta)$ *defined by* (31) *and* (32) *satisfy* (33).

The proof of the Lemma is straightforward. On any subinterval of the type space between any two points where constraint (20) binds, the incremental changes of the extra rent type a firm with low cost of producing good two receives relative to a firm with high costs of producing good two and the same cost of producing good one, must sum to zero. But then, I face on this subinterval a problem that is essentially identical to problem P"'. Hence, I get solution schedules that are analogous to (22) and (23) on that subinterval. If these schedules satisfy a single crossing condition, then on any subinterval where constraint (20) is slack, the schedules $\overline{x}(\theta)$ and $\underline{x}(\theta)$ cross at most once. They do have to cross at least once so as to guarantee the incremental rents of high versus low η types add up to zero. Suppose now there were two points θ_1 and θ_2 such that (20) binds for all $\theta \leq \theta_1$ and for all $\theta \geq \theta_2$, but not for any θ in between θ_1 and θ_2. As I have just argued, this would imply that $\underline{x}(\theta_1)$ is above $\overline{x}(\theta_1)$ so that $\rho_\theta(\theta_1, \pi) = \overline{x}(\theta_1) - \underline{x}(\theta_1) < 0$. However, since (13) binds at θ_1, I have $\rho(\theta_1, \pi) = 0$, so $\rho(\theta, \pi) < 0$ for θ close to but larger than θ_1.

To rule out problems of bunching in the θ dimension, I assume that

Assumption Aii: $\frac{F(\theta)}{f(\theta)}$ is nondecreasing in θ.

Although the solution procedure is quite lengthy, the solution to the regulator´s problem takes a very simple form:

Proposition 4 *For* $\delta = 0$, Δ *sufficiently large, and under Assumptions Ai and Aii, the solution to problem P" is given by the quality schedules* $\overline{q}^*(\theta) = q_1$ *and* $\underline{q}^*(\theta) = q_2$ *in conjunction with* $\underline{\pi}^* = 0$, *and quantity schedules* $\underline{x}^*(\theta) = \overline{x}^*(\theta) \equiv x^*(\theta)$ *where* $x^*(\theta)$ *satisfies*

$$V_x^1(x^*(\theta)) = \theta + (1-\alpha)\frac{F(\theta)}{f(\theta)}, \qquad (34)$$

and the associated transfer schedules $\overline{t}^*(\theta)$ *and* $\underline{t}^*(\theta)$ *defined by* (9)

There is complete bunching of types, that is the solution schedules become independent of η altogether - except for the allocation of good two. The quantity schedule has the familiar features: there is no distortion at the top, there is a downward distortion for all types with cost larger than the minimum, and there is no rent at the bottom.

9 Hybrid cases: bunching for low θ, separation for high θ

The previous analysis suggests that strict substitutability (that is $\delta > 0$) is necessary for rents at the bottom and separating firms, at least if one is willing to accept affiliation as a good assumption. In this section, I analyze the case where there is strict subsitutability between goods and at the same time the schedules defined by (31) and (32) satisfy (33). To add new insights relative to the two cases already studied, I impose a restriction on the paramters:

Assumption Hi: $\delta\left(q_2 - q_1\right) > (1-\alpha)\frac{1}{f\left(\overline{\theta}|\underline{\eta}\right)}$

If the parameters fail to satisfy the condition in Assumption Hi, then the solution would coincide with one of the cases I have already discussed.

Second, to rule out problems of bunching, I assume

Assumption Hii: $\frac{\partial}{\partial\theta}\frac{F(\theta)}{f(\theta)} \geq 0$.

Moreover, I require a set of quite restrictive assumptions on the distribution of types to guarantee that the monotonicity requirement on the solution schedules is always met and to organize the bunching regions in a tractable way. The conditions needed for this are either

Assumption HAi: $f_\theta\left(\theta|\,\underline{\eta}\right) \leq 0$ and $\frac{\partial}{\partial\theta}\frac{1-F\left(\theta|\underline{\eta}\right)}{f\left(\theta|\underline{\eta}\right)} \leq 0$, and

Assumption HAii: $f_\theta\left(\theta|\,\overline{\eta}\right) \geq 0$ and $\frac{\partial}{\partial\theta}\frac{\frac{\beta}{1-\beta}+F\left(\theta|\overline{\eta}\right)}{f\left(\theta|\overline{\eta}\right)} \geq 0$.[6]

Or, alternatively, **Assumption Ii** (independence) and in addition

Assumption Iiii: $f_\theta\left(\theta\right) \geq 0$.

The solution is a mixture between the case of complete separation and of complete bunching:

Proposition 5 *Suppose δ is positive but close to zero, α is smaller, but sufficiently close to one, Δ is sufficiently large, β is sufficiently small, and Assumptions Hi and Hii hold. Then, if the the type distribution satisfies in addition either assumption HAi and HAii or alternatively assumption Ii and Iiii the solution to problem P" is given by the quality schedules $\overline{q}^*\left(\theta\right) = q_1$ and $\underline{q}^*\left(\theta\right) = q_2$ in conjunction with $\underline{\pi}^* > 0$ and quantity schedules that satisfy for $\theta \geq \theta'$*

$$V_x^1\left(\underline{x}\left(\theta\right)\right) = C_x\left(\underline{x}\left(\theta\right), q_2, \theta, \underline{\eta}\right) - (1-\alpha)\frac{1-F\left(\theta|\,\underline{\eta}\right)}{f\left(\theta|\,\underline{\eta}\right)} \tag{35}$$

and

$$V_x^1\left(\overline{x}\left(\theta\right)\right) = C_x\left(\overline{x}\left(\theta\right), q_1, \theta, \overline{\eta}\right) + (1-\alpha)\frac{\frac{\beta}{1-\beta}+F\left(\theta|\,\overline{\eta}\right)}{f\left(\theta|\,\overline{\eta}\right)}. \tag{36}$$

θ' is the unique point of intersection of these schedules if such an intersection exists. For $\theta < \theta'$,

[6] Notice that $f_\theta\left(\theta|\,\overline{\eta}\right) \geq 0 \geq f_\theta\left(\theta|\,\underline{\eta}\right)$ is a special case of affiliated random variables.

the quantity schedules for good one satisfy $\underline{x}^*(\theta) = \overline{x}^*(\theta) \equiv x^*(\theta)$ *where* $x^*(\theta)$ *satisfies*

$$V_x^1(x^*(\theta)) = \theta + \delta\left[g\left(\underline{\eta}\middle|\theta\right)q_2 + g\left(\overline{\eta}\middle|\theta\right)q_1\right] + (1-\alpha)\frac{F(\theta)}{f(\theta)}. \tag{37}$$

The associated transfer schedules $\overline{t}^*(\theta)$ *and* $\underline{t}^*(\theta)$ *are defined by* (9)*.*

There is bunching of different η types who have the same marginal cost parameter θ at the low end of the θ support; at the high end there is separation of such types. Moroever, at the point where the regime changes from bunching to separation of different η types, the solution schedules switch continuously from one regime to the other. The idea to show this is the following. The value of the regulator's payoff function at an optimum should be invariant to small changes in the switch-point θ'. This requires that, conditional on $\theta = \theta'$, the expected value of the objective at θ' - where there is bunching - should be the same as the expected value of the objective just after the switch point, that is at $\theta = \theta' + \varepsilon$ for ε positive but arbitrarily small. This value matching condition essentially boils down to requiring continuity of the solution schedules.

Consider now the regulator's interest, to leave a positive rent to type $\left(\overline{\theta}, \underline{\eta}\right)$. While this would never happen in the case of known $\eta-$types, the advantage in the current context is that a higher $\underline{\pi}$ shifts the switch point θ' to the left. In other words, there is a new trade-off between rent extraction and efficiency. The solution schedules are constrained efficient given the informational asymmetry about θ. Raising $\underline{\pi}$ allows the regulator to get closer to the optimal solution schedules that are not constrained by the informational asymmetry about η. Let $\widehat{\overline{x}}^*\left(\overline{\theta}\right)$ denote the optimal quantity that the regulator would require a firm with type $\left(\overline{\theta}, \overline{\eta}\right)$ to produce if that firms cost of producing good two were known; $\widehat{\overline{x}}^*\left(\overline{\theta}\right)$ is defined by condition (16) in Proposition 1. Moreover, let $\underline{x}^{fb}\left(\overline{\theta}\right)$ denote the first-best efficient quantity allocation for type $\left(\overline{\theta}, \underline{\eta}\right)$. $\pi^* > 0$ for β small enough if and only if $\widehat{\overline{x}}^*\left(\overline{\theta}\right) > \underline{x}^{fb}\left(\overline{\theta}\right)$, which is equivalent to Assumption Hi being met.

Using standard envelope theorems I observe that the effect of a small increase in π on the regulator's payoff is equal to $-(1-\alpha)\beta + k(\pi)$. The marginal cost of setting $\pi > 0$ is that an additional rent of π has to be left to all firms with a high quality capacity. There is a measure β of such firms and the cost enters the regulator's objective with a weight of $(1-\alpha)$. On the other hand, there is a benefit to raising π which is measured by $k(\pi)$, the value of the multiplier attached to constraint (13) over the separation region $\left[\theta', \overline{\theta}\right]$. Clearly, $k(\pi)$ is the higher the more the presence of firms with quality capacity $\overline{\eta}$ impinges the regulator from pursuing an optimal regulation policy for firms with a lower quality capacity. In particular, if the regulator would, when he knew η but not θ, have the firm of type $\left(\overline{\theta}, \overline{\eta}\right)$ produce more than the first-best efficient amount for a firm of type $\left(\overline{\theta}, \underline{\eta}\right)$, then the value of the multiplier is larger than the shadow cost of raising π. Since the regulator's objective is concave in π, this argument shows that the unique optimal π is positive under the conditions given in the proposition.

10 Concluding remarks

I have solved a regulation problem featuring two dimensional asymmetric information about the costs of production of two goods in some detail. The optimal allocation differs markedly from its one-dimensional counterpart, except in special cases. Most interestingly it can be optimal to distort production upwards instead of downwards. The rationale for this result is a trade-off between efficiency and rent extraction that involves the second dimension of asymmetric information, and this trade-off feeds back into the efficiency-rent extraction trade-off in the first dimension. Moreover, it can be optimal to leave rents to the most inefficient producer among those with a low cost of producing the second good. The rationale is again that increasing this rent allows the regulator to better resolve the standard trade-off between efficiency and rent-extraction within groups of producers with the same cost of producing good two (but different and privately known costs of producing good one).

11 Appendix

Proof of Lemma 3. Let $\beta = \Pr\left[\eta = \underline{\eta}\right]$. The "reduced" problem where I neglect the monotonicity constraints on $\overline{x}(\theta)$ and $\underline{x}(\theta)$ can be written as follows:

$$\Gamma(\pi) = \max_{\overline{u}(\theta), \underline{u}(\theta)} \left[\begin{array}{c} \beta \int_{\underline{\theta}}^{\overline{\theta}} B\left(\underline{x}(\theta), \underline{\eta}, \theta, \underline{\eta}\right) f\left(\theta \left| \underline{\eta}\right.\right) d\theta - \beta(1-\alpha)\underline{\pi} \\ +(1-\beta) \int_{\underline{\theta}}^{\overline{\theta}} B\left(\overline{x}(\theta), \overline{\eta}, \theta, \overline{\eta}\right) f\left(\theta \left| \overline{\eta}\right.\right) d\theta \end{array} \right]$$

$$s.t.\ \underline{\pi} + \int_{\theta}^{\overline{\theta}} \underline{x}(y)\, dy - \int_{\theta}^{\overline{\theta}} \overline{x}(y)\, dy \leq 0$$

Letting $\underline{z} \equiv -\int_{\theta}^{\overline{\theta}} \underline{x}(y)\, dy$ and $\overline{z} \equiv -\int_{\theta}^{\overline{\theta}} \overline{x}(y)\, dy$ I can note further that $\underline{x} = \underline{z}_\theta$ and $\overline{x} = \overline{z}_\theta$.

I can view this as a control problem with Hamiltonian of the following form:

$$\begin{aligned} H &= B\left(\underline{x}(\theta), \underline{\eta}, \theta, \underline{\eta}\right) \beta f\left(\theta \left| \underline{\eta}\right.\right) + B\left(\overline{x}(\theta), \overline{\eta}, \theta, \overline{\eta}\right)(1-\beta) f\left(\theta \left| \underline{\eta}\right.\right) \\ &\quad + \overline{\kappa x} + \underline{\kappa x} + \mu\left(\underline{\pi} - \left(\underline{z} - \overline{z}\right)\right) \end{aligned}$$

Differentiating with respect to state variables, I get the conditions of optimality

$$\begin{aligned}\frac{\partial H}{\partial \overline{z}} &= \mu = -\overline{\kappa}_\theta \\ \frac{\partial H}{\partial \underline{z}} &= -\mu = -\underline{\kappa}_\theta;\end{aligned}$$

differentiating with respect to the controls I get

$$\begin{aligned}\frac{\partial H}{\partial \overline{x}} &= \left(V^1_{\overline{x}}\left(\overline{x}\left(\theta\right)\right) - C_{\overline{x}}\left(\overline{x}\left(\theta\right), q_1, \overline{\eta}, \theta\right) - \left(1-\alpha\right)\frac{F\left(\theta\left|\overline{\eta}\right.\right)}{f\left(\theta\left|\overline{\eta}\right.\right)}\right)\left(1-\beta\right)f\left(\theta\left|\overline{\eta}\right.\right) + \overline{\kappa} = 0 \quad (38) \\ \frac{\partial H}{\partial \underline{x}} &= \left(V^1_{\underline{x}}\left(\underline{x}\left(\theta\right)\right) - C_{\underline{x}}\left(\underline{x}\left(\theta\right), q_2, \underline{\eta}, \theta\right) - \left(1-\alpha\right)\frac{F\left(\theta\left|\underline{\eta}\right.\right)}{f\left(\theta\left|\underline{\eta}\right.\right)}\right)\beta f\left(\theta\left|\underline{\eta}\right.\right) + \underline{\kappa} = 0\end{aligned}$$

The Kuhn-Tucker conditions are

$$\underline{\pi} - \left(\underline{z} - \overline{z}\right) \leq 0, \ \ \mu \geq 0, \ \text{ and } \mu\left(\underline{\pi} - \left(\underline{z} - \overline{z}\right)\right) = 0.$$

For the transversality conditions, I have to distinguish two cases. If $\mu\left(\underline{\theta}\right) = 0$, then $\overline{z}\left(\underline{\theta}\right)$ and $\underline{z}\left(\underline{\theta}\right)$ are both free and the transversality conditions are

$$\overline{\kappa}\left(\underline{\theta}\right) = \underline{\kappa}\left(\underline{\theta}\right) = 0.$$

If $\mu\left(\underline{\theta}\right) > 0$, then $\overline{z}\left(\underline{\theta}\right)$ is fully determined once $\underline{z}\left(\underline{\theta}\right)$ is given and vice versa. Hence, I do not impose any transversality condition in this case.

Suppose that $\mu\left(\underline{\theta}\right) = 0$ and that $\mu\left(\theta\right) = 0$ on a set of positive measure $\left[\underline{\theta}, \theta'\right]$. From conditions (38) it is clear that $\overline{\kappa}$ and $\underline{\kappa}$ are continuously differentiable in θ whenever $\overline{u}$ and $\underline{u}$ are continuously differentiable in θ. Using the conditions of optimality for the state variables, $\overline{\kappa}_\theta = -\mu$ and $\underline{\kappa}_\theta = \mu$, and the transversality conditions - which must hold if $\mu\left(\underline{\theta}\right) = 0$ - I have for $\theta \leq \theta'$

$$\overline{\kappa}\left(\theta\right) = \overline{\kappa}\left(\underline{\theta}\right) + \int_{\underline{\theta}}^{\theta} \overline{\kappa}_\tau d\tau = -\int_{\underline{\theta}}^{\theta} \mu\left(\tau\right) d\tau = 0$$

and

$$\underline{\kappa}\left(\theta\right) = \underline{\kappa}\left(\underline{\theta}\right) + \int_{\underline{\theta}}^{\theta} \mu\left(\tau\right) d\tau = 0.$$

Hence, for $\theta \in \left[\underline{\theta}, \theta'\right]$, I have

$$V^1_{\overline{x}}\left(\overline{x}\left(\theta\right)\right) - C_{\overline{x}}\left(\overline{x}\left(\theta\right), q_1, \overline{\eta}, \theta\right) - \left(1-\alpha\right)\frac{F\left(\theta\left|\overline{\eta}\right.\right)}{f\left(\theta\left|\overline{\eta}\right.\right)} = 0$$

and

$$V^1_{\underline{x}}\left(\underline{x}\left(\theta\right)\right) - C_{\underline{x}}\left(\underline{x}\left(\theta\right), q_2, \underline{\eta}, \theta\right) - \left(1-\alpha\right)\frac{F\left(\theta\left|\underline{\eta}\right.\right)}{f\left(\theta\left|\underline{\eta}\right.\right)} = 0.$$

Of course these conditions are equivalent to those obtained in Proposition 1. However, under assumption M, it is easy to see that for $\theta \leq \theta'$

$$\overline{x}(\theta) > \underline{x}(\theta).$$

The remainder of the argument, showing that $\mu(\theta) = 0$ over a nonempty interval $\theta \in [\underline{\theta}, \theta']$ leads into a contradiction, is in the text. ■

Proof of Proposition 2. The proof is split into two parts. In the first part, (after spelling out the Lagrangian of the problem), I prove that the first-order conditions are sufficient for an optimum. In the second part, I show that the solution is not only a solution to the reduced problem but satisfies also the neglected constraints.

The Lagrangian of the problem is

$$L = \max_{\overline{x}(\theta), \underline{x}(\theta), \underline{\pi}} \left[\begin{array}{c} \beta \int_{\underline{\theta}}^{\overline{\theta}} B\left(\underline{x}(\theta), q_2, \theta, \underline{\eta}\right) f\left(\theta \left| \underline{\eta}\right.\right) d\theta - \beta(1-\alpha)\underline{\pi} \\ + (1-\beta) \int_{\underline{\theta}}^{\overline{\theta}} B\left(\overline{x}(\theta), q_1, \theta, \overline{\eta}\right) f\left(\theta \left| \overline{\eta}\right.\right) d\theta \end{array} \right] \tag{39}$$

$$+k\left(\int_{\underline{\theta}}^{\overline{\theta}} \left(\underline{x}(\theta) - \overline{x}(\theta)\right) d\theta + \underline{\pi}\right)$$

i) The objective is strictly concave in $\overline{x}$ and $\underline{x}$ and the constraint is linear in these variables. By a standard theorem, the first-order conditions in $\overline{x}$ and $\underline{x}$, respectively, are also sufficient for an optimum. Consider now the derivatives with respect to $\underline{\pi}$. I have

$$\frac{\partial L}{\partial \underline{\pi}} = -\beta(1-\alpha) + k$$

and

$$\frac{\partial^2 L}{\partial \underline{\pi}^2} = \frac{dk}{d\underline{\pi}}.$$

I can compute $\frac{dk}{d\underline{\pi}}$ from a total differentiation of the constraint. I obtain

$$\frac{dk}{d\underline{\pi}} = \frac{1}{\int_{\underline{\theta}}^{\overline{\theta}} \left(\overline{x}_k(\theta; k) - \underline{x}_k(\theta; k)\right) d\theta}.$$

From a total differentiation of (22) and (23), I get

$$\frac{d\underline{x}^*}{dk} = -\frac{1}{B_{xx}\left(\underline{x}^*(\theta), q_2, \theta, \underline{\eta}\right) \beta f\left(\theta \left| \underline{\eta}\right.\right)} > 0 \tag{40}$$

and

$$\frac{d\overline{x}^*}{dk} = \frac{1}{B_{xx}\left(\overline{x}^*(\theta), q_1, \theta, \overline{\eta}\right)(1-\beta) f\left(\theta \left| \overline{\eta}\right.\right)} < 0. \tag{41}$$

This shows that $\frac{dk}{d\underline{\pi}} < 0$,and hence the problem is concave in the choice variable $\underline{\pi}$. It follows that the condition of optimality is that either $\underline{\pi}^* = 0$ and

$$k\left(\underline{\pi}^*\right) \leq \beta\left(1-\alpha\right)$$

or $\pi^* > 0$ and

$$k\left(\overline{\pi}^*\right) = \beta\left(1-\alpha\right).$$

Moreover, as $k\left(\tilde{\underline{\pi}}\right) = 0$, I note that $\underline{\pi}^* < \tilde{\underline{\pi}}$.

ii) Consider first incentive compatibility in the θ dimension. The schedules (22) and (23) are continuous; hence they are differentiable everywhere and if they satisfy $\frac{d\overline{x}(\theta)}{d\theta} \leq 0$ and $\frac{d\underline{x}(\theta)}{d\theta} \leq 0$, they are monotonic. Consider first the schedule $\overline{x}\left(\theta\right)$. From a total differentiation of (22), I obtain

$$\left(V_x^1\left(\underline{x}^*\left(\theta\right)\right) - C_x\left(\underline{x}^*\left(\theta\right), q_2, \theta, \underline{\eta}\right) - \left(1-\alpha\right)\frac{F\left(\theta\right)}{f\left(\theta\right)}\right)\beta f\left(\theta\right) + k^* = 0 \tag{42}$$

and

$$\left(V_x^1\left(\overline{x}^*\left(\theta\right)\right) - C_x\left(\overline{x}^*\left(\theta\right), q_1, \theta, \overline{\eta}\right) - \left(1-\alpha\right)\frac{F\left(\theta\right)}{f\left(\theta\right)}\right)\left(1-\beta\right) f\left(\theta\right) - k^* = 0, \tag{43}$$

$$\frac{d\underline{x}^*}{d\theta} = \frac{\left(1+\left(1-\alpha\right)\frac{\partial}{\partial\theta}\frac{F(\theta)}{f(\theta)}\right) + \frac{k}{\beta}\frac{f_\theta(\theta)}{f(\theta)^2}}{V_{xx}^1\left(\underline{x}^*\left(\theta\right)\right)}. \tag{44}$$

From Assumption **Iii,** $f_\theta\left(\theta\right) \leq 0$, so the numerator is minimized for k as large as possible, that is for $k = \beta\left(1-\alpha\right)$. Substituting back into (44), the numerator is bounded below by

$$1+\left(1-\alpha\right)\frac{\partial}{\partial\theta}\frac{F\left(\theta\right)}{f\left(\theta\right)} + \left(1-\alpha\right)\frac{f_\theta\left(\theta\right)}{f\left(\theta\right)^2} = 1-\left(1-\alpha\right)\frac{\partial}{\partial\theta}\frac{1-F\left(\theta\right)}{f\left(\theta\right)} \geq 0,$$

where the final inequality uses again Assumption **Iii.**

A total differentiation of (23) gives

$$\frac{d\overline{x}^*}{d\theta} = \frac{\left(1+\left(1-\alpha\right)\frac{\partial}{\partial\theta}\frac{F(\theta)}{f(\theta)}\right) - \frac{k^*}{1-\beta}\frac{f_\theta(\theta)}{f(\theta)^2}}{V_{xx}^1\left(\overline{x}^*\left(\theta\right)\right)}, \tag{45}$$

which clearly satisfies $\frac{d\overline{x}^*(\theta)}{d\theta} \leq 0$ for all θ since $f_\theta\left(\theta\right) \leq 0$.

To demonstrate incentive compatibility in the η dimension, I show that the assumptions imply that the schedules $\overline{x}\left(\theta\right)$ and $\underline{x}\left(\theta\right)$ satisfy (26). Indeed it is easy to verify that $f_\theta\left(\theta\right) \leq 0$ implieas that $\underline{x}\left(\theta\right) = \overline{x}\left(\theta\right) \Longrightarrow \frac{d\underline{x}(\theta)}{d\theta} \geq \frac{d\overline{x}(\theta)}{d\theta}$. ■

Proof of Lemma 4. To prove the lemma, I show the bunching region is convex. Together with Lemma 3, this demonstrates that bunching occurs only at the low end of the support, and does so over an interval.

The bunching region is convex if and only if there cannot exist two points θ' and θ'' with $\theta'' > \theta'$ such that (20) binds at θ' and θ'' but not in between. Suppose the contrapositive were true, and two such points did exist. Then, I know that

$$\int_{\theta'}^{\theta''} \overline{x}(y)\, dy - \int_{\theta'}^{\theta''} \underline{x}(y)\, dy = 0, \tag{46}$$

because (20) is binding at θ' and θ''. But then, I can split the problem into three subproblems, where each subproblem is to choose optimal schedules $\overline{x}(\theta)$ and $\underline{x}(\theta)$ on the (possibly empty) subintervals, $[\underline{\theta}, \theta')$, $[\theta', \theta'']$, and $(\theta'', \overline{\theta}]$. This is possible, because I solve a "reduced" absent monotonicity constraints on $\overline{x}(\theta)$ and $\underline{x}(\theta)$. On the subinterval $[\theta', \theta'']$, the problem is identical to the isoperimetric problem solved in the proof of lemma 3 with θ' replacing $\underline{\theta}$ and θ'' replacing $\overline{\theta}$, and where $\underline{\pi} = 0$. Hence, the optimal schedules satisfy conditions (22) and (23).

Suppose (33) holds, so $\underline{x}(\theta) = \overline{x}(\theta) \Longrightarrow \frac{d\overline{x}(\theta)}{d\theta} \geq \frac{d\underline{x}(\theta)}{d\theta}$. Then, the schedules $\overline{x}(\theta)$ and $\underline{x}(\theta)$ can only satisfy condition (46) if they cross at least once. Moreover, schedule $\underline{x}(\theta)$ must cross schedule $\overline{x}(\theta)$ from above. Hence, I must have $\underline{x}(\theta') > \overline{x}(\theta')$. But then

$$\rho_\theta(\theta', \pi) = \overline{x}(\theta') - \underline{x}(\theta') < 0$$

contradicting the supposition that (20) is non-binding for $\theta \in (\theta', \theta'')$. ■

Proof of Proposition 4. The proof is organized in two parts. In the first part, I show that the solution schedules are continuous at the switching point between the bunching and the separation region. This argument holds for arbitrary value of $\underline{\pi}$. In the second part I choose $\underline{\pi}$ optimally and demonstrate that $\underline{\pi}^* = 0$ under the assumptions of the proposition. This implies that the bunching region extends over the whole domain of θ.

Part i): continuity of the solution schedules

Consider again the control problem spelled out in the proof of Lemma 3. I first show that for $\theta \leq \theta'$, the optimal schedule satisfies $\overline{x}(\theta) = \underline{x}(\theta) = x^*(\theta)$ and $x^*(\theta)$ solves

$$V_x^1(x^*(\theta)) - \theta - \delta\left[g(\underline{\eta}|\theta)\, q_2 + g(\overline{\eta}|\theta)\, q_1\right] - (1-\alpha)\frac{F(\theta)}{f(\theta)} = 0. \tag{47}$$

To see this, I can use the conditions for $\overline{\kappa}(\theta)$ and $\underline{\kappa}(\theta)$ and the equations of motion for these costate variables to get

$$\overline{\kappa}(\theta) = \overline{\kappa}(\underline{\theta}) - \int_{\underline{\theta}}^{\theta} \mu(\tau)\, d\tau$$

and

$$\underline{\kappa}(\theta) = \underline{\kappa}(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \mu(\tau)\, d\tau.$$

Substituting back into (38)

$$\begin{aligned}\left(V_{\overline{x}}^{1}\left(\overline{x}\left(\theta\right)\right)-\theta-\delta q_{1}-\left(1-\alpha\right)\frac{F\left(\theta\left|\overline{\eta}\right.\right)}{f\left(\theta\left|\overline{\eta}\right.\right)}\right)\left(1-\beta\right)f\left(\theta\left|\overline{\eta}\right.\right) &= -\overline{\kappa}\left(\underline{\theta}\right)+\int_{\underline{\theta}}^{\theta}\mu\left(\tau\right)d\tau \\ \left(V_{\underline{x}}^{1}\left(\underline{x}\left(\theta\right)\right)-\theta-\delta q_{2}-\left(1-\alpha\right)\frac{F\left(\theta\left|\underline{\eta}\right.\right)}{f\left(\theta\left|\underline{\eta}\right.\right)}\right)\beta f\left(\theta\left|\underline{\eta}\right.\right) &= -\underline{\kappa}\left(\underline{\theta}\right)-\int_{\underline{\theta}}^{\theta}\mu\left(\tau\right)d\tau\end{aligned}$$

Recall that $\beta f\left(\theta\left|\underline{\eta}\right.\right)=f\left(\theta,\underline{\eta}\right)$ and $\left(1-\beta\right)f\left(\theta\left|\overline{\eta}\right.\right)=f\left(\theta,\overline{\eta}\right)$. Letting $g\left(\eta\left|\theta\right.\right)$ denote the probability density function for η conditional on θ, I can substitute for $f\left(\theta,\eta\right)=g\left(\eta\left|\theta\right.\right)f\left(\theta\right)$. Moreover, note that $\overline{x}=\underline{x}$ as $\mu>0$ for $\theta\leq\theta'$. Adding the two conditions of optimality for the control variables, and dividing by $f\left(\theta\right)$ I get

$$V_{x}^{1}\left(x^{*}\left(\theta\right)\right)-\theta-\delta\left[g\left(\overline{\eta}\left|\theta\right.\right)q_{1}+g\left(\underline{\eta}\left|\theta\right.\right)q_{2}\right]-\left(1-\alpha\right)\frac{F\left(\theta\right)}{f\left(\theta\right)}=\frac{-\underline{\kappa}\left(\underline{\theta}\right)-\overline{\kappa}\left(\underline{\theta}\right)}{f\left(\theta\right)} \tag{48}$$

where I have used the fact that $\beta F\left(\theta\left|\underline{\eta}\right.\right)+\left(1-\beta\right)F\left(\theta\left|\overline{\eta}\right.\right)=F\left(\theta\right).$

To complete the argument I now argue that $-\underline{\kappa}\left(\underline{\theta}\right)-\overline{\kappa}\left(\underline{\theta}\right)=0$. Since $\overline{x}\left(\theta\right)=\underline{x}\left(\theta\right)=x^{*}\left(\theta\right)$ for $\theta\leq\theta'$, any solution of (48) for given $-\underline{\kappa}\left(\underline{\theta}\right)-\overline{\kappa}\left(\underline{\theta}\right)$ satisfies constraint $\left(20\right)$. Moreover, $\underline{\kappa}\left(\underline{\theta}\right)$ and $\overline{\kappa}\left(\underline{\theta}\right)$ have no influence on the value of the objective for $\theta>\theta'$, because the costate variables are allowed to jump at points where the state variable constraint switches from binding to non-binding. Moreover, $\underline{\kappa}\left(\underline{\theta}\right)$ and $\overline{\kappa}\left(\underline{\theta}\right)$ have no influence on the location of the switching point θ' either. Hence, at the optimum $\underline{\kappa}\left(\underline{\theta}\right)$ and $\overline{\kappa}\left(\underline{\theta}\right)$ must be such that, conditional on θ, the expected value of the objective is maximized. Hence, $\overline{\kappa}\left(\underline{\theta}\right)=-\underline{\kappa}\left(\underline{\theta}\right)$, and I obtain the expression in the Proposition.

For $\theta>\theta'$, $\mu\left(\theta\right)=0$, so that $\overline{\kappa}\left(\theta\right)=\overline{k}$ and $\underline{\kappa}\left(\theta\right)=\underline{k}$ for $\theta>\theta'$. A priori it is neither clear how $\overline{k}$ relates to $\underline{k}$, nor is it clear how the values of the costate variables relate to $\overline{\kappa}\left(\theta'\right)$ and $\underline{\kappa}\left(\theta'\right)$. That is, there may be jumps in the costate variables at θ'.

I first show that $\overline{k}+\underline{k}=0$. To see this, consider a candidate pair of schedules that give rise to a switch point θ'. Clearly, for the subinterval $\left[\theta',\overline{\theta}\right]$, constraint (20) is binding only at θ'. But then, choosing the optimal schedules $\overline{x}\left(\theta\right)$ and $\underline{x}\left(\theta\right)$ on the subinterval $\left[\theta',\overline{\theta}\right]$ is equivalent to the isoperimetric problem (39) with θ' replacing $\underline{\theta}$. Hence, $k\equiv\overline{k}=-\underline{k}$.

Finally, I show that the solution schedules are continuous at the switch point θ'.

I can write the value of the objective as

$$\Gamma\left(\theta'\right)=W^{1}\left(\theta'\right)+W^{2}\left(\theta'\right)-\beta\left(1-\alpha\right)\underline{\pi}, \tag{49}$$

where

$$W^{1}\left(\theta'\right)\equiv\beta\int_{\underline{\theta}}^{\theta'}B\left(x^{*}\left(\theta\right),q_{2},\theta,\underline{\eta}\right)f\left(\theta\left|\underline{\eta}\right.\right)d\theta+\left(1-\beta\right)\int_{\underline{\theta}}^{\theta'}B\left(x^{*}\left(\theta\right),q_{1},\theta,\overline{\eta}\right)f\left(\theta\left|\overline{\eta}\right.\right)d\theta$$

and

$$W^2\left(\theta'\right) \equiv \beta \int_{\theta'}^{\overline{\theta}} B\left(\underline{x}\left(\theta\right), q_2, \theta, \underline{\eta}\right) f\left(\theta \left| \underline{\eta}\right.\right) d\theta + \left(1-\beta\right) \int_{\theta'}^{\overline{\theta}} B\left(\overline{x}\left(\theta\right), q_1, \theta, \overline{\eta}\right) f\left(\theta \left| \overline{\eta}\right.\right) d\theta$$
$$+k\left(\underline{\pi} + \int_{\theta'}^{\overline{\theta}} \underline{x}\left(y\right) dy - \int_{\theta'}^{\overline{\theta}} \overline{x}\left(y\right) dy\right).$$

Clearly θ' must pass the following test: the value of the objective, $\Gamma\left(\theta'\right)$, should not increase through a small change in θ'. Invoking the envelope theorem, the effect of a marginal change in θ' is

$$\Gamma_{\theta'}\left(\theta'\right) = W^1_{\theta'}\left(\theta'\right) + W^2_{\theta'}\left(\theta'\right)$$

where

$$W^1_{\theta'}\left(\theta'\right) \equiv \beta B\left(x^*\left(\theta'\right), q_2, \theta', \underline{\eta}\right) f\left(\theta' \left| \underline{\eta}\right.\right) d\theta + \left(1-\beta\right) B\left(x^*\left(\theta'\right), q_1, \theta', \overline{\eta}\right) f\left(\theta' \left| \overline{\eta}\right.\right)$$

and

$$W^2_{\theta'}\left(\theta'\right) = -\beta B\left(\underline{x}\left(\theta'\right), q_2, \theta', \underline{\eta}\right) f\left(\theta' \left| \underline{\eta}\right.\right) - \left(1-\beta\right) B\left(\overline{x}\left(\theta'\right), q_1, \theta', \overline{\eta}\right) f\left(\theta' \left| \overline{\eta}\right.\right) \qquad (50)$$
$$-k\left(\underline{x}\left(\theta'\right) - \overline{x}\left(\theta'\right)\right)$$

Clearly, at the optimum I must have $W^1_{\theta'}\left(\theta'\right) + W^2_{\theta'}\left(\theta'\right) = 0$, so the values of the objectives evaluated at the bound θ' must match. One solution is clearly reached when $\overline{x}\left(\theta'\right) = \underline{x}\left(\theta'\right) = x^*\left(\theta'\right)$. I now show this solution is unique. To make the dependence of $\overline{x}\left(\theta'\right)$ and $\underline{x}\left(\theta'\right)$ on k explicit, I write these schedules as $\overline{x}\left(\theta'; k\right) = \underline{x}\left(\theta'; k\right)$, respectively. A total differentiation of the integral constraint

$$\underline{\pi} + \int_{\theta'}^{\overline{\theta}} \underline{x}\left(y; k\right) dy - \int_{\theta'}^{\overline{\theta}} \overline{x}\left(y; k\right) dy = 0$$

delivers

$$\frac{dk}{d\theta'} = \frac{\underline{x}\left(\theta'; k\right) - \overline{x}\left(\theta'; k\right)}{\int_{\theta'}^{\overline{\theta}} \left(\underline{x}_k\left(y; k\right) - \overline{x}_k\left(y; k\right)\right) dy}$$

Recall from (40) and (41) that in the proof of Proposition 2 that $\frac{d\underline{x}}{dk} > 0$ and $\frac{d\overline{x}}{dk} < 0$. Hence, the denominator of the expression for $\frac{dk}{d\theta'}$ is positive. Hence, I have $\frac{dk}{d\theta'} < 0$ for $\underline{x}\left(\theta'; k\right) < \overline{x}\left(\theta'; k\right)$ and $\frac{dk}{d\theta'} > 0$ for $\underline{x}\left(\theta'; k\right) > \overline{x}\left(\theta'; k\right)$. Thus k is minimized when θ' is such that $\overline{x}\left(\theta'; k\right) = \underline{x}\left(\theta'; k\right)$. Again using (40) and (41), for any other value of θ', I will have $\underline{x}\left(\theta'; k\right) > \overline{x}\left(\theta'; k\right)$. However, it is easy to see that the values

$$\overline{\hat{x}} \equiv \arg\max_x \left\{\left(V^1\left(x\right) - C\left(x, q_1, \theta', \overline{\eta}\right)\right) f\left(\theta' \left| \overline{\eta}\right.\right) - \left(1-\alpha\right) F\left(\theta' \left| \overline{\eta}\right.\right)\right\}$$

and

$$\underline{\hat{x}} \equiv \arg\max_{x} \left\{ \left(V^1(x) - C\left(x, q_2, \theta', \underline{\eta}\right)\right) f\left(\theta' \left| \underline{\eta}\right.\right) - (1-\alpha) F\left(\theta' \left| \underline{\eta}\right.\right) \right\}$$

satisfy $\overline{\hat{x}} \geq \underline{\hat{x}}$ due to Assumption M. Hence, the sum of the terms in the first line in (50) decreases by an increase in k. Moreover, $-k\left(\underline{x}\left(\theta'\right) - \overline{x}\left(\theta'\right)\right)$ becomes negative. Hence, there can be no other solution.

Part ii): the optimal value of $\underline{\pi}$

It is easy to see that the derivative of (49) with respect to π is still given by

$$\Gamma_{\underline{\pi}}\left(\underline{\pi}\right) = -\beta(1-\alpha) + k; \tag{51}$$

and the second derivative is still given by $\Gamma_{\underline{\pi}\underline{\pi}} = \frac{dk}{d\underline{\pi}}$ whenever this is well defined. Letting $\overline{x}(\theta; k)$ and $\underline{x}(\theta; k)$ denote the functions defined by (22) and (23), and using the fact that (20) is binding at $\theta'(k)$, I have for any $\underline{\pi} > 0$

$$\frac{dk}{d\underline{\pi}} = \frac{1}{\int_{\theta'(k)}^{\overline{\theta}} \left(\overline{x}_k(y; k) - \underline{x}_k(y; k)\, dy\right) dy} < 0,$$

where I have used the fact that effects of k on $\theta'(k)$ exactly cancel out because $\underline{x}\left(\theta'(k); k\right) = \overline{x}\left(\theta'(k); k\right)$. Hence, the optimum features $\underline{\pi}^* = 0$ if $k\left(\underline{\pi}\right)\big|_{\underline{\pi}=0} \leq \beta(1-\alpha)$ and $\underline{\pi}^* > 0$ if $k\left(\underline{\pi}\right)\big|_{\underline{\pi}=0} > \beta(1-\alpha)$.

Due to the single-crossing condition I must have $\theta' = \overline{\theta}$ for $\underline{\pi}= 0$. This follows directly from substituting $\theta'' = \overline{\theta}$ into condition (46) in the proof of Lemma 4. For $\underline{\pi} = 0$, constraint (20) is binding at $\overline{\theta}$; by convexity of the bunching region, the constraint is binding for all θ.

Hence, when evaluating the derivative $\Gamma_{\underline{\pi}}\left(\underline{\pi}\right)$ at $\underline{\pi} = 0$, I can use the fact that $\overline{x}\left(\overline{\theta}\right) = \underline{x}\left(\overline{\theta}\right) = x^*(\theta)$ for $\underline{\pi} = 0$. Hence, for $\theta = \overline{\theta}$ and $\delta = 0$, I can write (47) in explicit form as

$$\left(V_x^1\left(x^*\left(\overline{\theta}\right)\right) - \overline{\theta}\right) f\left(\overline{\theta}\right) = (1-\alpha) \tag{52}$$

From (23), $k(0)$ satisfies

$$\left(V_x^1\left(\overline{x}^*\left(\overline{\theta}\right)\right) - \overline{\theta}\right)(1-\beta) f\left(\overline{\theta}\left|\overline{\eta}\right.\right) - (1-\alpha) = -\beta(1-\alpha) + k(0),$$

Substituting from (52), we have

$$\left(V_x^1\left(\overline{x}^*\left(\overline{\theta}\right)\right) - \overline{\theta}\right)(1-\beta) f\left(\overline{\theta}\left|\overline{\eta}\right.\right) - (1-\alpha) = -\left(V_x^1\left(\overline{x}^*\left(\overline{\theta}\right)\right) - \overline{\theta}\right)\beta f\left(\overline{\theta}\left|\underline{\eta}\right.\right)$$

and thus

$$\Gamma_{\underline{\pi}}(0) = -\left(V_x^1\left(\overline{x}^*\left(\overline{\theta}\right)\right) - \overline{\theta}\right)\beta f\left(\overline{\theta}\left|\underline{\eta}\right.\right) < 0.$$

$\blacksquare$

Proof of Proposition 5. The proof is organized in two parts. In the first part, I give conditions for $\underline{\pi}^* > 0$. In the second part I show that the assumptions listed in the proposition imply the desired monotonicity conditions, which in turn imply incentive compatibility.

Part i: the optimal level of $\underline{\pi}$

It is easy to see that the derivative of (49) with respect to π is still given by

$$\Gamma_{\underline{\pi}}(\underline{\pi}) = -\beta(1-\alpha) + k; \tag{53}$$

and the second derivative is still given by $\Gamma_{\underline{\pi}\underline{\pi}} = \frac{dk}{d\underline{\pi}}$ whenever this is well defined. Letting $\overline{x}(\theta; k)$ and $\underline{x}(\theta; k)$ denote the functions defined by (22) and (23), and using the fact that (20) is binding at $\theta'(k)$, I have for any $\underline{\pi} > 0$

$$\frac{dk}{d\underline{\pi}} = \frac{1}{\int_{\theta'(k)}^{\overline{\theta}} \left(\overline{x}_k(y;k) - \underline{x}_k(y;k)\, dy\right) dy} < 0,$$

where I have used the fact that effects of k on $\theta'(k)$ exactly cancel out because $\underline{x}\left(\theta'(k); k\right) = \overline{x}\left(\theta'(k); k\right)$. Hence, the optimum features $\underline{\pi}^* = 0$ if $k(\underline{\pi})|_{\underline{\pi}=0} \leq \beta(1-\alpha)$ and $\underline{\pi}^* > 0$ if $k(\underline{\pi})|_{\underline{\pi}=0} > \beta(1-\alpha)$.

Due to the single-crossing condition I must have $\theta' = \overline{\theta}$ for $\underline{\pi}= 0$. This follows directly from substituting $\theta'' = \overline{\theta}$ into condition (46) in the proof of Lemma 4. For $\underline{\pi} = 0$, constraint (20) is binding at $\overline{\theta}$; by convexity of the bunching region, the constraint is binding for all θ.

Hence, when evaluating the derivative $\Gamma_{\underline{\pi}}(\underline{\pi})$ at $\underline{\pi} = 0$, I can use the fact that $\overline{x}\left(\overline{\theta}\right) = \underline{x}\left(\overline{\theta}\right) = x^*(\theta)$ for $\underline{\pi} = 0$. Hence, for $\theta = \overline{\theta}$, I can write (47) in explicit form as

$$\left(V_x^1(x^*(\theta)) - \theta\right) f(\theta) - \delta\left[\beta f\left(\overline{\theta}\,|\underline{\eta}\right) q_2 + (1-\beta) f\left(\overline{\theta}\,|\overline{\eta}\right) q_1\right] = (1-\alpha) \tag{54}$$

From (23) (allowing for dependence), $k(0)$ satisfies

$$\left(V_x^1\left(\overline{x}^*\left(\overline{\theta}\right)\right) - \theta\right)(1-\beta) f\left(\overline{\theta}\,|\overline{\eta}\right) - \delta q_1 (1-\beta) f\left(\overline{\theta}\,|\overline{\eta}\right) - (1-\alpha) = -\beta(1-\alpha) + k(0),$$

Substituting from (54) for $(1-\alpha)$, I can write the left-hand side of this equation as

$$\begin{aligned}&\left(V_x^1\left(\overline{x}^*\left(\overline{\theta}\right)\right) - \theta\right)(1-\beta) f\left(\overline{\theta}\,|\overline{\eta}\right) - \delta q_1(1-\beta) f\left(\overline{\theta}\,|\overline{\eta}\right)\\ &-\left(V_x^1(x^*(\theta)) - \theta\right) f(\theta) + \delta\left[\beta f\left(\overline{\theta}\,|\underline{\eta}\right) q_2 + (1-\beta) f\left(\overline{\theta}\,|\overline{\eta}\right) q_1\right]\end{aligned}$$

Simplifying, and using $x^*\left(\overline{\theta}\right) = \overline{x}^*\left(\overline{\theta}\right)$, I obtain

$$\Gamma_{\pi}(0) = -\left(V_x^1\left(x^*\left(\overline{\theta}\right)\right) - \overline{\theta} - \delta q_2\right)\beta f\left(\overline{\theta}\,|\underline{\eta}\right)$$

Again using (47) and subsituting for $\frac{\beta f(\overline{\theta}|\underline{\eta})}{f(\overline{\theta})} = g\left(\underline{\eta}\right|\theta\right)$ and $\frac{(1-\beta)f(\overline{\theta}|\overline{\eta})}{f(\overline{\theta})} = g\left(\overline{\eta}\right|\theta)$ I obtain

$$V_x^1\left(x^*\left(\overline{\theta}\right)\right) - \overline{\theta} = \delta\left[\frac{\beta f\left(\overline{\theta}\left|\underline{\eta}\right.\right)}{f\left(\overline{\theta}\right)}q_2 + \frac{(1-\beta) f\left(\overline{\theta}\left|\overline{\eta}\right.\right)}{f\left(\overline{\theta}\right)}q_1\right] + (1-\alpha)\frac{1}{f\left(\overline{\theta}\right)}$$

and so

$$V_x^1\left(x^*\left(\overline{\theta}\right)\right) - \overline{\theta} - \delta q_2 = \delta\left[-\frac{(1-\beta) f\left(\overline{\theta}\left|\overline{\eta}\right.\right)}{f\left(\overline{\theta}\right)}\left(q_2 - q_1\right)\right] + (1-\alpha)\frac{1}{f\left(\overline{\theta}\right)}.$$

Thus,

$$\Gamma_\pi\left(0\right) = -\left(\delta\left[-\frac{(1-\beta) f\left(\overline{\theta}\left|\overline{\eta}\right.\right)}{f\left(\overline{\theta}\right)}\left(q_2 - q_1\right)\right] + (1-\alpha)\frac{1}{f\left(\overline{\theta}\right)}\right)\beta f\left(\overline{\theta}\left|\underline{\eta}\right.\right)$$

Hence, $\Gamma_\pi\left(0\right) > 0$ iff

$$\delta\left(1-\beta\right) f\left(\overline{\theta}\left|\overline{\eta}\right.\right)\left(q_2 - q_1\right) - (1-\alpha) > 0.$$

Part ii: incentive compatibility of the schedules

The monotonicity follows mostly directly from the assumptions. The schedule $x^*\left(\theta\right)$ is monotonic in θ if

$$\frac{\partial}{\partial\theta}\left[\delta\left[g\left(\underline{\eta}\right|\theta\right) q_2 + g\left(\overline{\eta}\right|\theta) q_1\right] + (1-\alpha)\frac{F\left(\theta\right)}{f\left(\theta\right)}\right] \geq 0.$$

Differentiating out, using $g\left(\underline{\eta}\right|\theta\right) = \frac{\beta f\left(\theta|\underline{\eta}\right)}{f(\theta)} = \frac{\beta f\left(\theta|\underline{\eta}\right)}{\beta f\left(\theta|\underline{\eta}\right)+(1-\beta)\beta f\left(\theta|\overline{\eta}\right)}$ and so on, we have

$$\frac{\partial}{\partial\theta}\delta\left[g\left(\underline{\eta}\right|\theta\right) q_2 + g\left(\overline{\eta}\right|\theta) q_1\right] = \frac{\partial}{\partial\theta}\delta\left[\frac{1}{1+\frac{(1-\beta)f\left(\theta|\overline{\eta}\right)}{\beta f\left(\theta|\underline{\eta}\right)}}q_2 + \frac{1}{\frac{\beta f\left(\theta|\underline{\eta}\right)}{(1-\beta)f\left(\theta|\overline{\eta}\right)}+1}q_1\right] \leq 0$$

where the inequality follows from HAi:$\left(f_\theta\left(\theta\right|\underline{\eta}\right) \leq 0\right)$ and HAii: $\left(f_\theta\left(\theta\right|\overline{\eta}\right) \geq 0\right)$. For δ sufficiently small and α sufficiently close to unity, this effect is dominated by $\frac{\partial}{\partial\theta}\frac{F(\theta)}{f(\theta)} \geq 0$. ■

REFERENCES

Armstrong, M., "Multiproduct Nonlinear Pricing", *Econometrica* **64**, 1996, 51-76.

Armstrong, M. ,"Optimal Regulation with Unknown Demand and Cost Functions", *Journal of Economic Theory* **99**, 1999, 196-215.

Armstrong, M. and J. Rochet ,"Multi-dimensional Screening: A User's Guide", *European Economic Review* **43**, 1999, 959-979.

Baron, D. and R. Myerson, "Regulating a Monopolist with Unknown Costs", *Econometrica* **50**, 1982, 911-930.

Beaudry, P., C. Blackorby and D. Szalay, "Taxes and Employment Subsidies in Optimal Redistribution Programs ", *American Economic Review* **99**(1), 2009, 216-242.

Blackorby, C. and D. Szalay, “Regulating a Monopolist with unknown costs and unknown quality capacity”, *The Warwick Economic Research Papers Series wp 858* , 2008

Che, Y. and I. Gale, “Standard Auctions with Financially Constrained Bidders”, *Review of Economic Studies* **65**, 1998, 1-21.

Che, Y. and I. Gale, “The Optimal Mechanism for Selling to a Budget-Constrained Buyer”, *Journal of Economic Theory* **92**, 2000, 198-233.

Fudenberg, D. and J. Tirole, *Game Theory*, MIT Press, Cambridge, MA., 1991.

Iyengar, G. and Kumar, A. “Optimal Procurement Mechanisms for Divisible Goods with Capacitated Suppliers” working paper Columbia University, 2006.

Jehiel, P., Moldovanu, B, and E. Stacchetti, “Multidimensional Mechanism Design for Auctions with Externalities”, *Journal of Economic Theory* **85**, 1999, 258-293.

Kamien, M., and N. Schwartz, “Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management”, *Elsevier Science Publishers B.V. North Holland*, 1981

Laffont, J., E. Maskin, and J. Rochet, “Optimal Nonlinear Pricing with Two-Dimensional Characteristics”, in *Information, Incentives and Economic Mechanisms* edited by T. Groves, R. Radner, and S. Reiter, Minneapolis University of Minnesota Press, 1987, 256-266.

Lewis, T. and D. Sappington, “Regulating a Monopolist with Unknown Demand and Cost Function”, *Rand Journal of Economics* **19**, 1988, 438-457.

McAfee, R.P. and J. MacMillan, “Multidimensional Incentive Compatibility and Mechanism Design”, *Journal of Economic Theory* **46**, 1988, 335-354.

Malakow, A. and R. Vohra, “An Optimal Auction for Capacitated Bidders - a Network Perspective”, mimeo University of North Carolina and Northwestern University, 2005a.

Malakow, A. and R. Vohra, “Optimal Auctions for Asymmetrically Budget Constrained Bidders”, mimeo University of North Carolina and Northwestern University, 2005b.

Matthews, S. and Moore, J. H., Optimal Provision of Quality and Warranties: An Exploration in the Theory of Multidimensional Screening, *Econometrica*, **55**, 1997, 441-467.

Milgrom, P., and R. Weber, “A Theory of Auctions and Competitive Bidding”, *Econometrica*, **50**, 1982, 1089-1122.

Mirrlees, J.A., An exploration of the Theory of Optimal Taxation, *The Review of Economic Studies*, **38**, 1971, 175-208.

Mirrlees, J.A., Information and Incentives: The Economics of Carrots and Sticks, *Economic Journal*, **107**, 1997, 1311-1329.

Rochet, J.-C. and Choné, P., Ironing, Sweeping, and Multidimensional Screening, *Econometrica*, **66**, 1998, 783-826.

Rochet, J.-C. and L. A. Stole, "The Economics of Multidimensional Screening" Chapter 5 in *Advances in Economics and Econometrics, Theory and Applications, Eighth World Congress, Vol. I* edited by M. Dewatripont, L. P. Hansen, and S. Turnovsky, Cambridge University Press, Cambridge, 2003, 150-197.

Seyerstad, K. and A. Sydsaeter "Optimal Control Theory with Economic Applications; Second reprint", *North Holland,* 1999

Shaked, M. and Shantikumar, J.G., "Stochastic Orders", *Springer, New York,* 2007

Topkis, D. M., "Minimizing a Submodular function on a Lattice, *Operations Research* **26**, 1978, 305-321

Topkis, D. M., Supermodularity and Complementarity, *Frontiers of Economic Research; Princeton University Press* 1998.

Vives, X., *Oligopoly Pricing: Old Ideas and New Tools* The MIT Press, Cambridge, Massachusetts, 1999.

Wilson, R., *Nonlinear Pricing* Oxford University Press, Oxford, 1993.