



GOVERNANCE AND THE EFFICIENCY  
OF ECONOMIC SYSTEMS  
**GESY**

Discussion Paper No. 397

## Regulating a multiproduct and multitype monopolist

Dezső Szalay\*

\* University of Bonn

March 2013

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

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

---

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

# Regulating a multiproduct and multitype monopolist\*

Dezső Szalay

University of Bonn

March 15, 2013

## 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 knew what the firm knows, then optimal regulation would simply entail pricing at marginal costs and reimbursing the firm for the losses it makes on a lump-sum basis. However, typically firms are better informed about cost and/or demand conditions than the regulator is. This problem was first addressed by Loeb and Magat [1979] and studied rigorously using the tools of mechanism design by Baron and Myerson [1982] and Sappington [1983]. If cost conditions

---

\*This paper is a substantially generalized version of an earlier paper that was joint with Charles Blackorby (see Blackorby and Szalay (2008)). I am indebted to Chuck for many insightful discussions. Many thanks also in particular to Felix Ketelaar, to Yeon-Koo Che, Erik Eyster, Leonardo Felli, Claudio Mezzetti, Benny Moldovanu, Rudolf Muller, Marco Ottaviani, Tracy Lewis, and seminar participants at HEC Lausanne, LBS, LSE, the University of Bonn, University of Frankfurt, Northwestern University, ESSET Gerzensee, and the conference on "Multidimensional Mechanism Design" at HCM Bonn. Dirk Belger provided excellent research assistance. Correspondence can be sent to Dezső Szalay, Institute for Microeconomics, University of Bonn, Adenauerallee 24-42, 53113 Bonn, Germany, or to szalay@uni-bonn.de

are known to the firm but not to the regulator, then marginal cost pricing is no longer optimal. Marginal cost pricing 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.

Much of our theoretical understanding of the regulation problem is built around one-dimensional models where the firm's informational advantage is captured by the realization of either a cost or a demand shifter. This paper wishes to shed light on the the optimal pricing of a firm that knows several parameters that escape the regulator. This is a natural and important problem. Firms typically produce multiple goods; e.g. a railway company can transport cargo and passengers, a telephone company can use its wires to transmit voice and data, water utilities deliver fresh water and provide sewerage services. Moreover, serving different customer groups that are differentiated either by usertype (household versus non-household) or geographical location may be viewed as supplying different goods. It is analytically convenient but otherwise rather special to assume that the firm's cost conditions along the various dimensions of production are perfectly correlated - which is essentially what we do if we assume that the cost shifter is a one-dimensional parameter.

The minimal complexity needed to address this question in some generality is a model featuring two dimensions of production and two dimensions of asymmetric information. Moreover, the natural extension of one-dimensional models with a continuum of types as, e.g., in Baron and Myerson [1982] is to assume that types are drawn from a rectangle. Rochet and Choné [1998] analyze such a model and show that two properties are robust in the multidimensional problem. Firstly, confirming Armstrong [1996], a fraction of types is excluded at the optimum. Secondly, the optimal contractual arrangement displays bunching at the low end of the type support. Unfortunately, these two properties make the problem so hard to analyze that it basically becomes impossible to gain any further insights. The known alternative models that are more accessible, surveyed below, are based on reducing the dimension of the design problem back to one dimension. This paper proposes an alternative to this approach where the design problem remains two-dimensional and yet the problem remains tractable. Essentially, the idea is to identify assumptions that make the second best amount of production in one dimension easy to characterize. Once this part of the allocation is known, the solution to the remainder of the allocation problem can be characterized in closed form as well.

Armstrong and Rochet [1999] solve the regulation problem in a two-by-two model, that is, in a model of a firm producing two goods and knowing the realization of two cost parameters on

binary supports. In the present paper, I extend their analysis to the case where one parameter is drawn from a continuum while the other parameter still has a binary support. The regulated firm produces two goods - say, fresh water and sewerage services - and privately observes two parameters shifting its cost of production. The first parameter, drawn from a continuum, affects the marginal cost of producing good one - fresh water - only. The second parameter, drawn from a binary distribution, affects the marginal cost of producing good two - the sewerage service - only. In addition, the marginal costs of producing each of the goods are allowed to depend on the amount of the other good. The crucial assumption that makes the model tractable is that the difference between the realizations of the binary parameter are relatively large. Under this assumption, the second best allocation features a good two allocation that depends only on the realization of the binary parameter. As a result, the population of firms is essentially split into two groups according to the amount of good two they produce. Within these groups, firms face the same incentives as firms in the Baron Myerson [1982] model do. However, in addition, firms can also mimic firms in the other group by producing a different amount of good two. Consequently, the pricing of good one needs to be adjusted so as to keep firms from engaging in such behaviour.

The model is designed to understand one of the various effects in the multidimensional problem in detail. In particular, the model sheds light on how the additional information through the observed amount of good two feeds back into the pricing of good one. The model is both stylized and rich. It is stylized in that I obtain a complete solution as a function of the model's primitives. It is rich because it allows for a large but manageable variety of optimal allocations and offers clearcut predictions as to what primitives give rise to which allocation.

The optimal allocation depends crucially on two factors. Firstly, the nature of interaction between the two goods in the social surplus function and secondly, the statistical dependence between the cost parameters. Each of these factors can make it unattractive for a firm to signal its binary type through the amount of good two it produces. For concreteness, suppose that the cost parameters are statistically independent but assume that the goods are net substitutes in the sense of Laffont and Tirole [1993]. Then, consuming a larger amount of good two makes it desirable to consume a smaller amount of good one. Since the firm's rents arising from good one production are related to the quantity of good one the firm can sell, all else equal, firms find it particularly attractive to produce the smaller amount of good two. Statistical dependence can have a similar effect: in equilibrium, the amount of good two indicates the firm's cost of producing good two which in turn provides information about its cost of producing good one. If costs are positively

correlated, then the regulator becomes relatively more concerned with extracting rents from the firm that produces the larger amount of good two and hence sets prices such that this firm produces a smaller amount of good one. In both cases, the firms that are efficient at producing good two need to be given an explicit incentive to produce the large amount of good two and the optimal allocation reflects this constraint being binding.

The optimal allocation is strikingly different in the case of surplus interactions and the case of purely statistical interaction. In the case of pure surplus interactions it is optimal to set marginal prices for good one below marginal costs for all but the most inefficient firm supplying a large amount of good two. In contrast, good one prices for firms producing the smaller amount of good two are set above marginal cost, over and above the level that would be optimal if only one quantity of good two could be produced to begin with. Moreover, I provide conditions making it optimal to leave unusually large rents to all firms that produce the larger amount of good two, including the most inefficient producer within that group. Finally, this allocation features complete separation between all types.

In the case of purely statistical interaction, it is optimal to set the marginal price of good one independently of what amount of good two is produced; in other words, in this case the good one allocation features bunching in the parameter relating to production costs of good two. Moreover, it is never optimal to leave any rents to the inefficient firms in both groups.

Among the qualitative features of the solution, presumably the most interesting finding is the pricing below marginal costs. As documented by Sawkins and Reid [2007], there is evidence of some prices below marginal cost in the Scottish Water industry, so this does not seem to be a mere theoretical curiosity but rather something worth understanding. There are various explanations for such observations. It is well known that a multiproduct monopolist may find it optimal to set some prices below marginal costs (see Tirole [1988]) in order to stimulate the demand for some of its other products. This explanation crucially relies on demand complementarities. The present approach offers an explanation that relies exclusively on incentive concerns, so the model can rationalize prices below marginal costs even in the absence of demand complementarities.

Results closest in the literature are Armstrong and Rochet [1999], Lewis and Sappington [1988] and Armstrong [1999]. All three of these papers discuss prices below marginal costs. In particular, Armstrong and Rochet [1999] show that prices below marginal costs arise at the optimum under particular conditions. As explained above, the present model enriches the one by Armstrong and Rochet [1999] towards a more general yet still tractable version of the two-dimensional problem.

Hence, by intention, the main contribution of this paper is to identify the effects that survive in the richer context. Since the solution techniques remain manageable, there are reasons to hope that the present results can be extended to even richer models, bridging the gap between the two-by-two and the continuum-by-continuum model even further.<sup>1</sup>

Lewis and Sappington [1988] and Armstrong [1999] also note the possibility of prices below marginal costs, even though in quite different models. In their approaches, there are two parameters of private information - marginal costs and the level of demand - but the regulator has only one instrument, the marginal price, to screen firms. As a result, the dimensionality of the design problem is reduced to one, making the problem amenable to techniques developed by Laffont, Maskin and Rochet [1987] and McAfee and McMillan [1988]. While there is bunching by design in these models, there are as many parameters of private information as screening instruments in the present paper and so the design problem remains two-dimensional<sup>2</sup>. While both Lewis and Sappington [1988] and Armstrong [1999] note the possibility that prices can be set below marginal costs, Armstrong [1999] proves the optimality of exclusion in the Lewis and Sappington model, due to, essentially the same reasons as in Armstrong [1996]. One reason why the present approach remains tractable is that exclusion, as in Armstrong [1996, 1999], does not occur here. As a result, I am able to turn the possibility result into a definite taxonomy of model primitives<sup>3</sup>, and delineate the precise circumstances which feature marginal prices below marginal costs. It is left for future work - some of which is described in the final section - to see whether the qualitative features of the optimal allocation survive in even richer contexts.<sup>4</sup>

The paper is organized as follows. In Section two I lay out the model and explain the regu-

---

<sup>1</sup>Formally, the reason the present model remains manageable is that double-deviations - simultaneous lies in both dimensions - can be handled. While such double-deviations can be analyzed explicitly in the two-by-two model, they are the reason why the multidimensional problem with a richer type space is hardly tractable. Another approach that overcomes the double-deviation issue differently is Kleven et al. (2009) in the context of the taxation of couples. The optimal mechanism in the present context is very different from theirs. See also Beaudry et al. [2009] for yet a different approach where double-deviations are not optimal. In contrast to Beaudry et al. [2009] the present approach makes no restrictions on the available deviation strategies.

<sup>2</sup>See also Rochet and Stole [2003] for an overview of different approaches to multidimensional screening problems in the literature.

<sup>3</sup>Formally, exclusion does not occur since the type space in the current model is binary $\times$ continuum, while Armstrong [1996, 1999] assumes a continuum $\times$ continuum type space.

<sup>4</sup>A main difficulty with richer models is the techniques required to solve the models; see Rochet and Stole [2003] for an overview. However, a further difficulty arises from the ignorance of what one is actually looking for. Since it usually becomes much easier to prove the optimality of an allocation once it is known how it looks like, I hope the current results prove to be a useful guide in the search of solutions to richer models.

lator'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 closed form solutions for the case of a fully separating solution (with binding constraints in both dimensions) and the case of full separation in one dimension and complete bunching along the other dimension. Section seven discusses extensions and offers some conclusions. Long proofs have been relegated to the appendix.

## 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. Consumers' valuation for good one is independent<sup>5</sup> of the valuation for good two, so  $V(x, q) = V^1(x) + V^2(q)$ . Good one is perfectly divisible and consumers decide how much to consume. Letting  $P^1(x)$  denote the inverse demand function for good one, I have

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

I assume that the inverse demand function is differentiable and decreasing in  $x$ , so the valuation is twice differentiable and concave. Obviously, the valuation is increasing in  $x$ .  $V_x^1(0)$  is sufficiently large to make all solutions interior. Good two can be produced in discrete quantities, or more generally variants,  $q \in \{q_0 \equiv 0, q_1, q_2, \dots, q_n\}$ , where  $q_i > q_{i-1}$  for  $i = 1, \dots, n$ .  $V^2(q)$  is increasing with  $q$ ,  $V^2(0) = 0$ , and  $\frac{V^2(q_i) - V^2(q_{i-1})}{q_i - q_{i-1}}$  is decreasing in  $i$ . Given the discreteness of the good two allocation problem, there is a range of prices that induces consumers to consume  $q_i$  units of good two, if that is the desired amount. Assuming  $q_0 = 0$  is a normalization that has two interpretations. Variant  $q_0$  can be understood as shutting down the second dimension, that is, consuming zero units of that good; or it can be understood as a baseline version whose costs are known to the regulator. The normalization is not used until section 4. It is not essential that the variants are discrete. What matters is that there is a maximum quantity  $q_n$ .

The goods are produced by a monopoly firm subject to price regulation. The firm's cost of

---

<sup>5</sup>Assuming independent demands shifts all the interactions into the firm's cost function, which allows me to clearly trace back reasons for binding constraints. The assumption can be relaxed.

producing the goods in quantities  $x$  and  $q$  is

$$C(x, q, \theta, \eta) = K + x\theta + q\eta + \delta xq,$$

where  $K > 0$  and  $\delta$  are constants known to both regulator and firm, and  $x$  and  $q$  are verifiable so that contracts can be written on these variables;  $\theta$  and  $\eta$  are parameters that are known to the firm but not to the regulator.  $\delta$  is a parameter that captures the sign and strength of interactions between good one and good two in the firm's cost function.

The regulator knows only the joint distribution of the variables  $\theta$  and  $\eta$ ; these parameters are distributed on a product set  $\Theta \times \mathbf{H}$  with probability density function  $f(\theta, \eta) > 0$  for all  $\theta, \eta$ . The set  $\Theta$  is taken as the interval  $[\underline{\theta}, \bar{\theta}]$ , where  $\bar{\theta} > \underline{\theta} > 0$ . The set  $\mathbf{H}$  is taken as  $\{\underline{\eta}, \bar{\eta}\}$  where  $\bar{\eta} > \underline{\eta} > 0$ . The marginal probability that  $\eta = \underline{\eta}$  is equal to  $\beta$ . Given the full support assumption, the conditional distribution of  $\theta$  given  $\eta$  has full support. The density and cdf of this distribution are denoted  $f(\theta|\eta)$  and  $F(\theta|\eta)$ , respectively. Let  $\mathbb{E}$  denote the expectation operator and let  $f(\theta) \equiv \mathbb{E}_{\mathbf{H}}[f(\theta|\eta)]$  and  $F(\theta)$  denote the density and the cdf of the marginal distribution, respectively.

The cost function satisfies the standard Spence-Mirrlees conditions in  $x, \theta$  and  $q, \eta$ , respectively. The allocation problem is rich and simple at the same time. It is simple in the sense that the good two allocation is discrete; this makes the problem solvable. It is rich in the sense that the solution to the problem differs substantially from the one where good two is absent.

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 (that is, as a procurement problem). If the firm produces quantities  $x$  and  $q$  then it receives a payment  $t$  and its profit is

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

The properties of the optimal incentive scheme can be traced back into the context of price regulation using the well known fact that

$$V_x^1(x) = P^1(x).$$

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  $x$ ; moreover,  $S_x(x, q, \theta, \eta) = -(\theta + \delta q)$  is non-increasing in  $q$  for  $\delta \geq 0$  and increasing in  $q$  for  $\delta < 0$ . In the former case  $x$  and  $q$  are net



substitutes in the surplus function while they are net complements in the latter case<sup>6</sup>.

## 2.1 The regulator's problem

By the revelation principle, I can think of the regulator's problem in terms of a direct mechanism, which is a triple of functions  $\{q(\theta, \eta), x(\theta, \eta), t(\theta, \eta)\}$  for all  $(\theta, \eta) \in \Theta \times \mathbf{H}$ , 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

$$\Pi(\hat{\theta}, \theta, \hat{\eta}, \eta) \equiv t(\hat{\theta}, \hat{\eta}) - C(x(\hat{\theta}, \hat{\eta}), q(\hat{\theta}, \hat{\eta}), \theta, \eta).$$

Under a truthful mechanism, the weighted joint surplus for a given pair  $(\theta, \eta)$  is equal to

$$W(\theta, \eta) \equiv V^1(x(\theta, \eta)) + V^2(q(\theta, \eta)) - t(\theta, \eta) + \alpha \Pi(\theta, \theta, \eta, \eta)$$

where  $\alpha \in (0, 1)$ . Since  $\alpha$  is kept constant throughout the paper, I suppress the dependence of the welfare function on  $\alpha$  in what follows. I let  $\Theta$  and  $H$  denote the random variables with typical realizations  $\theta$  and  $\eta$ , respectively, and let  $\mathbb{E}_{\Theta H}$  denote the expectation operator taken over the random variables  $\Theta$  and  $H$ . The regulator solves the following problem, which I denote as problem P:

$$\max_{x(\cdot, \cdot), q(\cdot, \cdot), t(\cdot, \cdot)} \mathbb{E}_{\Theta H} W(\theta, \eta) \tag{1}$$

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

$$\Pi(\theta, \theta, \eta, \eta) \geq \Pi(\hat{\theta}, \theta, \hat{\eta}, \eta), \tag{2}$$

and for all  $\theta, \eta$

$$\Pi(\theta, \theta, \eta, \eta) \geq 0, \tag{3}$$

(2) is the incentive compatibility condition, requiring that a firm of type  $\theta, \eta$  must have no incentive to mimic any other type of firm; (3) 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 consumers in the first-best allocation; the participation constraint is binding for each type,  $\Pi(\theta, \theta, \eta, \eta) = 0$ , so

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

---

<sup>6</sup>See, e.g., Laffont and Tirole [1993].

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

$$\max_{x(\cdot, \cdot), q(\cdot, \cdot)} \mathbb{E}_{\Theta H} (V^1(x(\theta, \eta)) + V^2(q(\theta, \eta)) - C(x(\theta, \eta), q(\theta, \eta), \theta, \eta)).$$

The first-best optimal policy for good one,  $x^{fb}(\theta, \eta)$ , satisfies equality of marginal benefits and costs, so

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

I assume throughout the paper that  $\bar{\eta} - \underline{\eta}$  is sufficiently large and  $\delta$  sufficiently small in the sense of the following assumption:

**Assumption 1:** For all  $i$  and all  $x$  in the relevant range  $\underline{\eta} + \delta x \leq \frac{V^2(q_i) - V^2(q_{i-1})}{q_i - q_{i-1}} \leq \bar{\eta} + \delta x$ .

Assumption 1 implies that, regardless of the quantity  $x$ , for a firm with a high cost parameter  $\bar{\eta}$ , the increase in costs due to an increase in  $q_i$  outweighs the increase in surplus, while for a firm with cost parameter  $\underline{\eta}$ , the reverse is true. Hence, the first-best policy entails  $q^{fb}(\theta, \bar{\eta}) = q_0$  for all  $\theta$  and  $q^{fb}(\theta, \underline{\eta}) = q_n$ .

### 3 Statement of the problem

#### 3.1 Implementable allocations

To solve the regulator's problem I begin by bringing the incentive and participation constraints, (2) and (3), into a more tractable form. Obviously, the set of implementable allocations for good one production depends on the implemented allocation for good two. However, Assumption 1 pins down the optimal allocation for good two also in the second best, which is of course precisely the reason to impose it in the first place.

**Lemma 1** *Under Assumption 1, a second-best allocation entails  $q(\theta, \bar{\eta}) = q_0$  and  $q(\theta, \underline{\eta}) = q_n$  for all  $\theta$ .*

If the cost differences in the  $\eta$  dimension are large, the first-best allocation rule for good two production continues to be optimal also with asymmetric information. The reason is that surplus is maximal for this allocation and at the same time incentive constraints are as relaxed as they can be. Focussing on this case allows me to concentrate on the distortions relative to the case where there is only one good that is to be produced in quantity  $x$ . By design, all the distortions occur in the good one dimension.

**Lemma 2** *Under Assumption 1, for the second-best optimal allocation rule for good two,  $q(\theta, \bar{\eta}) = q_0$  and  $q(\theta, \underline{\eta}) = q_n$  for all  $\theta$ , the incentive constraint (2) is equivalent to the pair of one-dimensional constraints*

$$\Pi(\theta, \theta, \eta, \eta) \geq \Pi(\hat{\theta}, \theta, \eta, \eta) \quad (4)$$

and

$$\Pi(\theta, \theta, \eta, \eta) \geq \Pi(\theta, \theta, \hat{\eta}, \eta). \quad (5)$$

The intuition for this result is very simple. A firm's incentive to report its cost parameter  $\theta$  do not depend on what the firm reported about its cost parameter  $\eta$ , and vice versa. To see this, suppose a firm with cost parameters  $(\theta, \eta)$  announces  $\hat{\eta} \neq \eta$ . Its profit differs from the profit of a firm with cost parameters  $(\theta, \hat{\eta})$  by the amount  $[q(\theta, \hat{\eta}) - q(\theta, \eta)]\eta$ . However, as long as the functions  $q(\theta, \hat{\eta})$  and  $q(\theta, \eta)$  are independent of  $\theta$ , the difference in profits is an additive constant. Hence, the firm's optimal report in the  $\theta$  dimension is not affected by its report in the  $\eta$  dimension. Hence, the two-dimensional constraint breaks down into a pair of one-dimensional constraints.<sup>7</sup>

This insight allows me to state the incentive and participation constraints in a tractable manner. Let  $\pi(\theta, \eta) \equiv \max_{\hat{\theta}} \Pi(\hat{\theta}, \theta, \eta, \eta)$ .

**Lemma 3** *i) The incentive constraint (4) is satisfied if and only if*

$$\pi(\theta, \eta) = \pi(\bar{\theta}, \eta) + \int_{\theta}^{\bar{\theta}} x(y, \eta) dy \quad (6)$$

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

ii) The incentive constraint (5) is satisfied if and only if

$$q_n(\bar{\eta} - \underline{\eta}) \geq \pi(\theta, \underline{\eta}) - \pi(\theta, \bar{\eta}) \geq q_0(\bar{\eta} - \underline{\eta}). \quad (7)$$

iii) The participation constraints (3) are met if  $\pi(\theta, \bar{\eta}) \geq 0$ .

The proof of the Lemma is standard and therefore omitted<sup>8</sup>. To prove part i) one applies the well known envelope arguments to compute changes in the firm's rents when  $\theta$  changes but  $\eta$  is held constant. Therefore, the rent of a firm of type  $(\theta, \eta)$  is equal to the sum of the rent of the most inefficient firm within firms with cost parameter  $\eta$ ,  $\pi(\bar{\theta}, \eta)$ , and the marginal changes of the firm's rent with respect to changes in its cost parameter  $\theta$ . Notice that (6) allows for the

<sup>7</sup>This is the crucial difference to the multi-dimensional problems of Armstrong and Rochet (1999) and Rochet and Choné (2003), where the reduction of incentive compatibility conditions is not possible.

<sup>8</sup>See, e.g., Laffont and Tirole (1993).

case where  $\pi(\bar{\theta}, \eta) > 0$ , so some high cost types may receive rents. Part ii) follows directly from Lemma 2. In the proof of Lemma 2, I have shown that differences in profits when mimicking a firm with a different cost of producing good two are captured entirely by differences in “fixed costs”. Condition (7) 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  $\theta$ ) implies that the participation constraint can only bind at one end. Condition (7) implies, that type  $(\bar{\theta}, \underline{\eta})$  will automatically participate if type  $(\bar{\theta}, \bar{\eta})$  does and hence the result.

### 3.2 The control problem

I ease notation henceforth letting  $\bar{x}(\theta) \equiv x(\theta, \bar{\eta})$  and  $\underline{x}(\theta) \equiv x(\theta, \underline{\eta})$ , and likewise for the rent schedules  $\bar{\pi}(\theta)$  and  $\underline{\pi}(\theta)$ . Moreover, I let  $\bar{\pi} \equiv \bar{\pi}(\bar{\theta})$  and  $\underline{\pi} \equiv \underline{\pi}(\bar{\theta})$ . Define the virtual surplus

$$B(x, q, \theta, \eta) \equiv V^1(x) + V^2(q) - C(x, q, \theta, \eta) - (1 - \alpha)x \frac{F(\theta|\eta)}{f(\theta|\eta)}.$$

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

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

Using (6) to substitute out transfers from the regulator’s problem, and integrating by parts, I obtain the following representation of the regulator’s problem, which, for future reference, I denote as problem P’

$$\max_{\bar{x}(\cdot), \underline{x}(\cdot), \bar{\pi}, \underline{\pi}} \left\{ \begin{array}{l} \beta \int_{\underline{\theta}}^{\bar{\theta}} B(\underline{x}(\theta), q_n, \theta, \underline{\eta}) f(\theta|\underline{\eta}) d\theta - \beta(1 - \alpha)\underline{\pi} \\ + (1 - \beta) \int_{\underline{\theta}}^{\bar{\theta}} B(\bar{x}(\theta), q_0, \theta, \bar{\eta}) f(\theta|\bar{\eta}) d\theta - (1 - \beta)(1 - \alpha)\bar{\pi} \end{array} \right\} \quad (8)$$

*s.t.*

$$\rho(\theta, \bar{\pi}, \underline{\pi}) \geq q_0(\bar{\eta} - \underline{\eta}) \quad (9)$$

$$\rho(\theta, \bar{\pi}, \underline{\pi}) \leq q_n(\bar{\eta} - \underline{\eta}), \text{ and} \quad (10)$$

$$\bar{x}(\theta), \underline{x}(\theta) \text{ non-increasing in } \theta. \quad (11)$$

$$\bar{\pi} \geq 0 \quad (12)$$

Problem P’ has the following structure. If the monotonicity constraints on  $\bar{x}(\theta)$  and  $\underline{x}(\theta)$  are nonbinding, the problem can be viewed as a control problem with two control variables,  $\bar{x}(\theta)$  and

$\underline{x}(\theta)$ , and two state variables,  $-\int_{\theta}^{\bar{\theta}} \bar{x}(y) dy$  and  $-\int_{\theta}^{\bar{\theta}} \underline{x}(y) dy$ . 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  $\theta$ , 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'.

The presence of the constraints (9) and (10) alters the problem substantially. However, the conceptual difference to the standard problem requires only one of these constraints being nontrivially present. Therefore, to streamline the exposition, I assume that  $q_n(\bar{\eta} - \underline{\eta})$  is sufficiently large. This is consistent with Assumption 1 and guarantees that (10) is satisfied automatically. So, the relevant constraint is (9). Normalizing  $q_0 = 0$ , this constraint simplifies to  $\rho(\theta, \bar{\pi}, \underline{\pi}) \geq 0$ .<sup>9</sup>

## 4 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 I neglected all the constraints in problem P'. Clearly, it would then be optimal to extract all the rents from the least efficient producers within firms with the same parameter  $\eta$ , so  $\bar{\pi} = \underline{\pi} = 0$ . Moreover, the quantity schedules for good one would be chosen to maximize virtual surplus, so these schedules would satisfy

$$V_x^1(\underline{x}^\dagger(\theta)) = \theta + \delta q_n + (1 - \alpha) \frac{F(\theta | \underline{\eta})}{f(\theta | \underline{\eta})} \quad (13)$$

and

$$V_x^1(\bar{x}^\dagger(\theta)) = \theta + (1 - \alpha) \frac{F(\theta | \bar{\eta})}{f(\theta | \bar{\eta})}. \quad (14)$$

Direct inspection of these schedules reveals the following results:

---

<sup>9</sup>The formulation above reveals an interesting connection between the multidimensional model of screening and models with type dependent participation constraints, as in the literature on countervailing incentives (Lewis and Sappington [1989] and Maggi and Rodriguez-Clare [1995]) and type dependent participation constraints in general (Jullien [2000]). Formally, one solves a family of problems that are constrained by the between groups incentive constraints, where firms are grouped or stratified along the observed choice of production of good two they make. When concentrating on the menu offered to firms making the same choice, the designer solves a mechanism design problem with a type dependent constraint given by the menu offered to the group of firms making the other choice. In contrast to this literature, the type dependent constraint is endogenous in the present context, while the outside option is exogenous in the literature on participation constraints.

**Proposition 1** Suppose that  $\bar{\eta} - \underline{\eta}$  is sufficiently large and that  $\frac{\partial}{\partial \theta} \frac{F(\theta|\underline{\eta})}{f(\theta|\underline{\eta})}, \frac{\partial}{\partial \theta} \frac{F(\theta|\bar{\eta})}{f(\theta|\bar{\eta})} \geq 0$ . If  $\delta \leq 0$  and  $\frac{F(\theta|\bar{\eta})}{f(\theta|\bar{\eta})} \geq \frac{F(\theta|\underline{\eta})}{f(\theta|\underline{\eta})}$ , then the optimal quantity schedules for good one are given by (13) and (14). Within the class of firms with the same parameter  $\eta$ , all but the lowest cost firm produce less than the first-best amount; there is no rent for the highest cost producer, that is  $\bar{\pi}^* = \underline{\pi}^* = 0$ .

The point is that the “unconstrained” solution that would obtain if I neglected the multi-dimensional nature of the problem happens to satisfy all the neglected constraints. Clearly, if such an “unconstrained” solution is feasible, then it is also optimal.

Constraint (10) is satisfied whenever  $q_n(\bar{\eta} - \underline{\eta})$  is sufficiently large, which I assume, consistently with Assumption 1. To develop an intuition why the other two constraints are satisfied as well, it is useful to begin with the case where  $\theta$  and  $\eta$  are statistically independent and the distribution of  $\theta$  has a monotonic inverse reversed hazard rate. Firms with a low  $\eta$  parameter produce higher amounts of good two. For  $\delta < 0$  this reduces their marginal cost of producing good one, so firms with low  $\eta$  parameter produce higher amounts of good one than their counterparts with high  $\eta$  parameters do. In turn, higher amounts of production of good one create higher rents for firms. Hence, a firm that has a cost advantage in the production of good two has every incentive to announce this truthfully. The intuition is depicted graphically in figure 1:

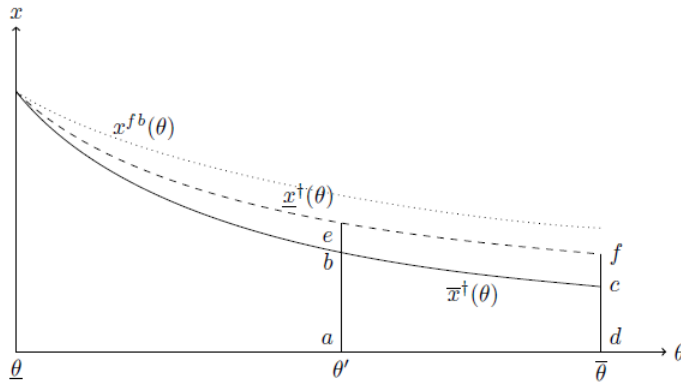


Figure 1: The rent of a type  $(\theta', \underline{\eta})$  is given by the area  $aefd$ , which is - by monotonicity of the allocation in  $\eta$  - larger than the area  $abcd$ , the rent of a type  $(\theta', \bar{\eta})$ .

A similar intuition underlies the case of statistical dependence. If  $\frac{F(\theta|\bar{\eta})}{f(\theta|\bar{\eta})} \geq \frac{F(\theta|\underline{\eta})}{f(\theta|\underline{\eta})}$ , then  $\theta$  conditional on  $\eta = \bar{\eta}$  is said to be smaller than  $\theta$  conditional on  $\eta = \underline{\eta}$  in the reversed hazard

rate order.<sup>10</sup> The reversed hazard rate measures for given  $\theta$  the relative importance the regulator attaches to obtaining an efficient quantity and to extracting rents from firms with costs lower than  $\theta$ , respectively. In turn, reversed hazard rate dominance in the sense of the proposition implies that rent extraction is relatively more important for firms with high  $\eta$  parameters, so the optimum features  $\underline{x}^\dagger(\theta) \geq \bar{x}^\dagger(\theta)$  for any  $\delta \leq 0$ , again generating higher rents for firms with low parameter  $\eta$ .

In sum, truthfully revealing a low type  $\underline{\eta}$  is in the firm's interest because that increases the firm's rents. A high  $\bar{\eta}$  firm has no incentive to mimic a low  $\underline{\eta}$  firm because  $q_n(\bar{\eta} - \underline{\eta})$ , the cost of this deviation, is too high. Hence, only the incentive constraints in the  $\theta$  dimension are binding and the problem can be solved as a pair of one-dimensional problems. Replacing marginal utilities by marginal prices, (13) and (14) pin down the marginal prices of good one.

## 4.1 The Agenda

Proposition 1 lists conditions under which the solution of a multidimensional problem can be obtained using entirely one-dimensional methods; hence everything is exactly as in the one-dimensional world. The agenda for the remainder of the paper is to drop these assumptions in a controlled fashion. If  $\delta$  is strictly positive, then the marginal cost of producing good one is increased for firms with parameter  $\underline{\eta}$ . Increases in marginal costs reduce the schedule  $\underline{x}^\dagger(\theta)$  point-wise and thus reduce rents of types with parameter  $\underline{\eta}$ . Hence, for  $\delta$  strictly positive, constraint (9) must be binding for  $\underline{\pi} = 0$ , unless the statistical effect through the reversed inverse hazard rate counteracts this effect and is sufficiently large. Similarly, even if there are no direct interactions in the cost function,  $\delta = 0$ , constraint (9) becomes binding for  $\underline{\pi} = 0$  due to statistical inference when the distributions satisfy  $\frac{F(\theta|\bar{\eta})}{f(\theta|\bar{\eta})} < \frac{F(\theta|\underline{\eta})}{f(\theta|\underline{\eta})}$  for  $\theta > \underline{\theta}$ .

To isolate these effects, I proceed as follows. I first analyze the effect of direct cost interactions ( $\delta > 0$ ), assuming away any interactions due to statistical inference, i.e. assuming that the distribution of  $\theta$  is independent of  $\eta$ . Secondly, I analyze the case where the only interaction is statistical inference, thus assuming that  $\delta = 0$  and that  $\theta$  conditional on  $\eta = \bar{\eta}$  is strictly higher than  $\theta$  conditional on  $\eta = \underline{\eta}$  in the reversed hazard rate order. Finally, I discuss what can be said about cases where there is both real interaction in the sense that  $\delta \neq 0$  and interaction due to statistical inference.

---

<sup>10</sup>This implies that  $\theta$  conditional on  $\eta = \bar{\eta}$  is smaller in the usual stochastic order  $\eta = \underline{\eta}$  (that is, First Order Stochastic Dominance) than  $\theta$  conditional on  $\eta = \bar{\eta}$  (see Shaked and Shantikumar (2007)).

Throughout this paper, I will maintain two important assumptions. Firstly, I maintain the assumption that cost differences in the production of good two are sufficiently large. If this assumption is dropped, then the allocation for good two identified in Lemma 1 may no longer be optimal. I study this case in companion work. Putting everything together would overload the present paper. Secondly, I impose assumptions on the distribution of types that allow me to neglect constraint (11) without loss of generality. Without such assumptions I would face problems of bunching in the  $\theta$ -dimension of my problem, and would face these problems already in the benchmark case, where the problem reduces to one-dimensional subproblems. By assuming such issues away, I can focus on bunching along the  $\eta$ -dimension, which seems more novel.

## 5 Interactions between the dimensions

It is immediate that the firm with the highest cost in both dimensions must have no rent at the optimum, that is  $\bar{\pi}^* = 0$ . The reason is that reducing  $\bar{\pi}$  relaxes constraint (9) and raises the regulator's objective. Hence, with a slight abuse of notation, I write the excess rent of a low  $\eta$  type over his high  $\eta$  counterpart as

$$\rho(\theta, \underline{\pi}) \equiv \underline{\pi} + \int_{\underline{\theta}}^{\bar{\theta}} \underline{x}(y) dy - \int_{\underline{\theta}}^{\bar{\theta}} \bar{x}(y) dy. \quad (15)$$

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

$$\max_{\bar{x}(\cdot), \underline{x}(\cdot), \underline{\pi}} \left\{ \begin{array}{l} \beta \int_{\underline{\theta}}^{\bar{\theta}} B(\underline{x}(\theta), q_n, \theta, \underline{\eta}) f(\theta | \underline{\eta}) d\theta - \beta(1 - \alpha) \underline{\pi} \\ + (1 - \beta) \int_{\underline{\theta}}^{\bar{\theta}} B(\bar{x}(\theta), q_0, \theta, \bar{\eta}) f(\theta | \bar{\eta}) d\theta \end{array} \right\} \quad (16)$$

s.t. for all  $\theta$

$$\rho(\theta, \underline{\pi}) \geq 0. \quad (17)$$

The problem is reduced in the sense that (11) and (10) are omitted from the problem. However, as I have argued above, given that  $\bar{\eta} - \underline{\eta}$  is sufficiently large, omitting (10) is without further loss of generality. Dropping (11) from the problem is justified under assumptions on the conditional distributions of  $\theta$  given  $\eta$ , which I will make explicit as I go along.

Compared to the original problem, maximizing (16) under constraint (17) is relatively simple. However, the problem remains quite nasty. The reason is that the costate variables of control



problems with inequality constraints on state variables can display jumps at points where constraint (17) switches from being binding to being slack. Therefore, one needs an educated guess as to where precisely the constraint is binding.

## 5.1 The nature of incentive spillover-effects

The way the regulator resolves the traditional efficiency versus rent extraction trade-off within groups of firms with the same cost parameter  $\eta$  impacts on these firms' incentive to mimic firms with a different  $\eta$  parameter. This can be seen from a differentiation of (17) with respect to  $\theta$ . We have  $\rho_\theta(\theta, \underline{\pi}) = \bar{x}(\theta) - \underline{x}(\theta)$ . So, if  $\bar{x}(\theta) > \underline{x}(\theta)$ , then increasing  $\theta$  marginally eases the firm's incentive to overstate  $\eta$ ; if  $\bar{x}(\theta) < \underline{x}(\theta)$ , then the reverse happens. Vice-versa, constraint (17) may force the regulator to bunch firms with different  $\eta$  parameters but the same  $\theta$  parameter together. In particular, if (17) is binding over an interval  $[\theta', \theta'']$ , then  $\rho_\theta(\theta, \underline{\pi}) = 0$  over that interval and hence  $\bar{x}(\theta) = \underline{x}(\theta)$  for all  $\theta \in [\theta', \theta'']$ .

Suppose that the schedules  $\underline{x}^\dagger(\theta)$  and  $\bar{x}^\dagger(\theta)$  defined by (13) and (14) violate constraint (17) by an amount  $\underline{\pi}^\dagger \equiv \int_{\underline{\theta}}^{\bar{\theta}} \bar{x}^\dagger(y) dy - \int_{\underline{\theta}}^{\bar{\theta}} \underline{x}^\dagger(y) dy > 0$ . Then at the solution of problem P'', constraint (17) must necessarily be binding for some  $\theta$ . If (17) were non-binding for all  $\theta$ , then  $\underline{x}^\dagger(\theta)$  and  $\bar{x}^\dagger(\theta)$  would be optimal. However, this requires that the regulator leaves at least a rent  $\underline{\pi}^\dagger$  to all firms with low costs of producing good two. However, setting  $\underline{\pi} \geq \underline{\pi}^\dagger$  cannot be optimal. Around  $\underline{\pi} = \underline{\pi}^\dagger$ , the marginal cost of increasing  $\underline{\pi}$  is equal to  $-\beta(1 - \alpha)$ ; a fraction of firms  $\beta$  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  $\underline{\pi} = \underline{\pi}^\dagger$  is zero, as the regulator is already unconstrained by condition (17) for  $\underline{\pi} = \underline{\pi}^\dagger$ . Hence, at the optimum I must have  $0 \leq \underline{\pi}^* < \underline{\pi}^\dagger$ .<sup>11</sup>

In the Appendix, I formulate the optimal control version of problem P''. The following Lemma follows immediately from the optimality conditions at the low end of the support:

**Lemma 4** *Suppose that the schedules  $\underline{x}^\dagger(\theta)$  and  $\bar{x}^\dagger(\theta)$  defined by (13) and (14) satisfy*

$$i) \int_{\underline{\theta}}^{\bar{\theta}} \bar{x}^\dagger(y) dy - \int_{\underline{\theta}}^{\bar{\theta}} \underline{x}^\dagger(y) dy > 0.$$

*If in addition either*

$$ii) \underline{x}^\dagger(\theta) < \bar{x}^\dagger(\theta) \text{ for all } \theta, \text{ or}$$

$$iii) \underline{x}^\dagger(\underline{\theta}) < \bar{x}^\dagger(\underline{\theta}) \text{ and } \underline{x}^\dagger(\theta) \text{ crosses } \bar{x}^\dagger(\theta) \text{ exactly once, then}$$

<sup>11</sup>This heuristic argument is made more formally in the proof of Proposition 2 below.

at the solution to problem P", constraint (17) is binding at  $\theta = \underline{\theta}$ .

In particular, conditions i) and ii) are met if  $\delta \geq 0$  and  $\frac{F(\theta|\bar{\eta})}{f(\theta|\bar{\eta})} \leq \frac{F(\theta|\underline{\eta})}{f(\theta|\underline{\eta})}$  for all  $\theta$ , and either of these inequalities is strict; condition iii) is met for  $\delta > 0$  but sufficiently close to zero and  $\frac{\partial}{\partial \theta} \frac{F(\theta|\bar{\eta})}{f(\theta|\bar{\eta})} > \frac{\partial}{\partial \theta} \frac{F(\theta|\underline{\eta})}{f(\theta|\underline{\eta})}$  for all  $\theta$ .

The proof is a simple argument by contradiction. Let  $\bar{x}^*(\theta)$  and  $\underline{x}^*(\theta)$  denote the optimal quantity schedules solving problem P". If constraint (17) were slack at  $\theta = \underline{\theta}$ , then I could use the transversality conditions of the problem to conclude that  $\bar{x}^*(\theta) = \bar{x}^\dagger(\theta)$  and  $\underline{x}^*(\theta) = \underline{x}^\dagger(\theta)$  for all  $\theta \leq \theta'$ , where  $\theta'$  is the smallest  $\theta$  where (17) is binding. If  $\underline{x}^\dagger(\theta)$ ,  $\bar{x}^\dagger(\theta)$  satisfy condition i), then there is indeed  $\theta'$  such that (17) is binding at  $\theta'$ . Using conditions ii) and iii), it is easy to see that we must have  $\underline{x}^\dagger(\theta') < \bar{x}^\dagger(\theta')$ .<sup>12</sup> However, this implies that  $\rho_\theta(\theta', \underline{x}) = \bar{x}^\dagger(\theta') - \underline{x}^\dagger(\theta') > 0$  and hence that  $\rho(\theta, \underline{x}) < 0$  for  $\theta$  close to but smaller than  $\theta'$ . In fact, under conditions ii) or iii), we have  $\underline{x}^\dagger(\theta) < \bar{x}^\dagger(\theta)$  for all  $\theta < \theta'$  hence (17) would be violated for all  $\theta < \theta'$ .

Even though one needs to invoke optimal control theory to solve the problem, the intuition for the structure of the solution can be grasped using simpler methods. Essentially, this is because on intervals where constraint (17) is slack, the costate variables of my control problem are constants. So, consider a candidate optimal pair of schedules such that (17) is binding at  $\theta_1$  and slack on a set  $(\theta_1, \theta_2]$ . The quantity schedules for good one then take the form

$$V_x^1(\underline{x}^*(\theta)) = \theta + \delta q_n + (1 - \alpha) \frac{F(\theta)}{f(\theta)} - \frac{\kappa}{\beta f(\theta)} \quad (18)$$

and

$$V_x^1(\bar{x}^*(\theta)) = \theta + (1 - \alpha) \frac{F(\theta)}{f(\theta)} + \frac{\kappa}{(1 - \beta) f(\theta)} \quad (19)$$

on the interval  $(\theta_1, \theta_2]$  for some value  $\kappa \geq 0$ . Obviously, the difficulty is that the value of  $\kappa$  is not known; it is only determined as part of the solution. Moreover, it is not clear a priori, whether at the optimum there exists such an interval at all. Define  $\underline{x}^*(\theta; \kappa)$  and  $\bar{x}^*(\theta; \kappa)$  by conditions (18) and (19) for arbitrary, not just the optimal, value of  $\kappa \geq 0$ .

**Lemma 5** Suppose that  $\underline{x}^*(\theta; \kappa)$  and  $\bar{x}^*(\theta; \kappa)$  defined by (18) and (19) satisfy

$$\underline{x}^*(\theta; \kappa) = \bar{x}^*(\theta; \kappa) \implies \frac{d\underline{x}^*(\theta; \kappa)}{d\theta} \geq \frac{d\bar{x}^*(\theta; \kappa)}{d\theta}, \quad (20)$$

---

<sup>12</sup>Under condition ii) this is immediate. Under condition iii), note that  $\int_{\theta'}^{\bar{\theta}} \bar{x}^\dagger(y) dy - \int_{\theta'}^{\bar{\theta}} \underline{x}^\dagger(y) dy = 0$ . Since  $\bar{x}^\dagger(\theta)$  crosses  $\underline{x}^\dagger(\theta)$  exactly once, and  $\bar{x}^\dagger(\bar{\theta}) < \underline{x}^\dagger(\bar{\theta})$ , this implies the conclusion.



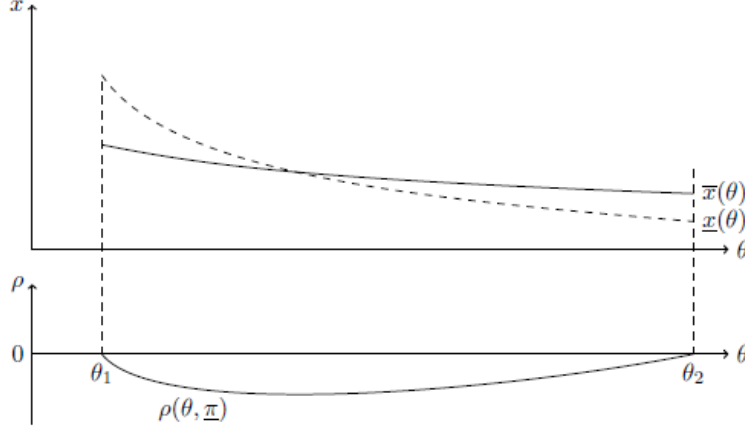


Figure 3: Under condition (21), it cannot be the case that (17) is binding over an interval, then nonbinding over some interval and then binding again.

When condition (21) holds, then there cannot be two points  $\theta_1, \theta_2$  such that constraint (17) is binding at these points and slack in between. Letting  $\kappa$  adjust to a value such that  $\int_{\theta_1}^{\theta_2} \underline{x}^*(y; \kappa^*) dy - \int_{\theta_1}^{\theta_2} \bar{x}^*(y; \kappa^*) dy = 0$ , it would have to be the case that the schedule  $\underline{x}^*(\theta; \kappa)$  crosses the schedule  $\bar{x}^*(\theta; \kappa)$  exactly once from above. However, that would imply that  $\rho_\theta(\theta_1, \underline{\pi}) = \bar{x}^*(\theta_1; \kappa^*) - \underline{x}^*(\theta_1; \kappa^*) < 0$ , so (17) would be violated for  $\theta$  larger than but close to  $\theta_1$ . The intuition is depicted graphically in figure 3.

Note once again that the schedules  $\underline{x}^*(\theta)$  and  $\bar{x}^*(\theta)$  are endogenous, so I cannot make direct assumptions on these schedules. However, Lemma 5 is nevertheless extremely useful, because it makes my problem accessible to an “educated guessing and verifying” solution procedure, where I search for conditions on the conditional distributions of  $\theta$  such that the solution schedules  $\underline{x}^*(\theta)$  and  $\bar{x}^*(\theta)$  endogenously inherit single-crossing conditions. Then, Lemma 5 allows me to pin down the regions where constraint (17) is binding.

## 6 The Optimal Allocation

I can now present the solution of my problem. The following section is organized along conditions on the primitives that give rise to the cases stated in Lemma 5.

### 6.1 Strict net substitutes and independent types

In this section, I focus on the case where  $\delta > 0$ , so raising the amount of production of good two raises the marginal cost of producing good one. To isolate the role of this net-substitutability, I assume in this section that knowing  $\eta$  does not provide any additional information about  $\theta$  :

**Assumption 2a:**  $f(\theta|\bar{\eta}) = f(\theta|\underline{\eta}) = f(\theta)$ .

In addition to that, I impose:

**Assumption 2b:**  $f_\theta(\theta) \leq 0$  and  $\frac{\partial}{\partial \theta} \frac{1-F(\theta)}{f(\theta)} \leq 0$ .<sup>14</sup>

The purpose of Assumption 2 is twofold<sup>15</sup>. First, it guarantees that the monotonicity constraints (11) are automatically satisfied. Second, it guarantees that at the solution to problem P", (17) binds only at  $\theta = \underline{\theta}$ .

With constraint (17) binding only at  $\underline{\theta}$ , the solution is very easy to obtain. Formally, if (17) is replaced by

$$\rho(\underline{\theta}, \underline{\pi}) = 0, \tag{22}$$

then the regulator's problem no longer involves constraints on the state variables but turns into an isoperimetric problem. Hence, the optimum can be found by simple pointwise maximization. It is useful to split the maximization into two steps. In the first step,  $\underline{\pi}$  is taken as given and the quantity schedules for good one production are chosen optimally against the given level of  $\underline{\pi}$ . In the second step, I optimize over the choice of  $\underline{\pi}$ .

Let  $k(\underline{\pi})$  denote the multiplier attached to constraint (22). For the isoperimetric problem,  $k(\underline{\pi})$  is independent of  $\theta$ . The shadow cost attached to the constraint is the smaller the higher is  $\underline{\pi}$ . Moreover, let  $\underline{x}^*(\theta; k)$  and  $\bar{x}^*(\theta; k)$  denote the optimal schedules for given  $\underline{\pi}$ . The first-order

<sup>14</sup>As is well known, distributions with logconcave densities have non-decreasing hazard rates. The restriction to decreasing densities amounts obviously to a subset of these distributions. Examples include the uniform but many more. As logconcave densities are unimodal, we can create a distribution that satisfies the restriction from any distribution with a mode in the interior of the support by truncating the distribution to the part to the right of the mode.

<sup>15</sup>Throughout the paper, distributions that satisfy both Assumptions 2a and 2b will be referred to as distributions satisfying Assumption 2.

conditions for the optimal schedules are

$$V_x^1(\underline{x}^*(\theta; k)) = \theta + \delta q_n + (1 - \alpha) \frac{F(\theta)}{f(\theta)} - \frac{k(\underline{\pi})}{\beta f(\theta)} \quad (23)$$

and

$$V_x^1(\bar{x}^*(\theta; k)) = \theta + (1 - \alpha) \frac{F(\theta)}{f(\theta)} + \frac{k(\underline{\pi})}{(1 - \beta) f(\theta)}. \quad (24)$$

The intuition for these expressions is straightforward. If I were to solve my problem simply as a pair of independent regulation problems, then I would obtain conditions (23) and (24) with  $k = 0$ . However, by the conditions in Lemma 4, these schedules violate constraint (22). So, to satisfy the constraint,  $\underline{x}^*(\theta; k)$  is adjusted downwards and  $\bar{x}^*(\theta; k)$  is adjusted upwards so as to increase the rents of firms with low costs of producing good two relative to their counterparts with high costs of producing good two.

Consider now the optimal choice of  $\underline{\pi}$  and let  $\Gamma(\underline{\pi})$  denote the value of the regulator's objective as a function of  $\underline{\pi}$ . Invoking the envelope theorem, I have

$$\Gamma_{\underline{\pi}}(\underline{\pi}) = k(\underline{\pi}) - \beta(1 - \alpha).$$

The marginal benefit of increasing  $\underline{\pi}$  is that the regulator becomes less constrained when choosing the production schedules  $\underline{x}^*(\theta; k)$  and  $\bar{x}^*(\theta; k)$ ; this is measured by the shadow cost  $k(\underline{\pi})$ . The marginal cost of increasing  $\underline{\pi}$  is the additional rents that are left to firms with cost parameter  $\underline{\eta}$ . As  $k(\underline{\pi})$  is decreasing in  $\underline{\pi}$ , the regulator's problem is concave in  $\underline{\pi}$ , so it is optimal to leave a strictly positive rent to firms with a cost parameter  $\underline{\eta}$  if and only if  $k(0) > \beta(1 - \alpha)$ . Moreover, if  $\underline{\pi}^* > 0$ , then we know that marginal benefits and costs of increasing  $\underline{\pi}$  must be equal at the optimum, so the value of the multiplier in this case is  $k(\underline{\pi}^*) = \beta(1 - \alpha)$ . Supposing for the sake of the argument that  $\underline{\pi}^* > 0$ , I can substitute this value into (23) and (24); I obtain the following schedules  $\underline{x}^{\dagger\dagger}(\theta)$  and  $\bar{x}^{\dagger\dagger}(\theta)$ :

$$V_x^1(\underline{x}^{\dagger\dagger}(\theta)) = \theta + \delta q_n - (1 - \alpha) \frac{1 - F(\theta)}{f(\theta)} \quad (25)$$

and

$$V_x^1(\bar{x}^{\dagger\dagger}(\theta)) = \theta + (1 - \alpha) \frac{\frac{\beta}{1 - \beta} + F(\theta)}{f(\theta)}. \quad (26)$$

I can now characterize the solution to my problem.

**Proposition 2** *Suppose that  $\bar{\eta} - \underline{\eta}$  is sufficiently large to satisfy Assumption 1. Moreover, suppose that  $\delta$  is strictly positive and the distribution of  $\theta$  satisfies Assumption 2. Then, the optimal schedules of good one production are given by  $\underline{x}^{\dagger\dagger}(\theta)$  and  $\bar{x}^{\dagger\dagger}(\theta)$  (defined in (25) and (26), respectively)*

together with  $\underline{\pi}^* > 0$  if and only if  $\int_{\underline{\theta}}^{\bar{\theta}} \underline{x}^{\dagger\dagger}(\theta) d\theta < \int_{\underline{\theta}}^{\bar{\theta}} \bar{x}^{\dagger\dagger}(\theta) d\theta$ . Otherwise, the optimal schedules are

given by (23) and (24) together with  $\underline{\pi}^* = 0$ , where  $k^*$  solves  $\int_{\underline{\theta}}^{\bar{\theta}} \underline{x}^*(\theta; k^*) d\theta = \int_{\underline{\theta}}^{\bar{\theta}} \bar{x}^*(\theta; k^*) d\theta$ .

In particular, for the uniform distribution,  $\underline{\pi}^* > 0$  if and only if  $(1 - \beta) \delta q_n (\bar{\theta} - \underline{\theta}) > (1 - \alpha)$ .

The proof of the Proposition simply consists in verifying that all the constraints are met and that  $\underline{\pi}^* > 0$  under the said conditions.

It is easy to see that the constraints are all satisfied. Assumption **2** implies that the solution schedules (23) and (24) satisfy the single crossing condition (20) for any  $k \leq \beta(1 - \alpha)$ , so imposing (22) rather than (17) is justified. Moreover, the schedules are monotonic, so they satisfy constraint (11). To understand the conditions for  $\underline{\pi}^* > 0$ , consider the function

$$\Phi(k) \equiv \int_{\underline{\theta}}^{\bar{\theta}} (\underline{x}^*(\theta; k) - \bar{x}^*(\theta; k)) d\theta.$$

I show in the appendix that  $\Phi(k)$  is an increasing function of the multiplier,  $k$ . We know that  $\Phi(0) < 0$ , since otherwise the constraint would not be binding at all.  $k(0)$  is the value of  $k$  solving  $\Phi(k)|_{k=k(0)} = 0$ . Since  $\Phi(k)$  is increasing in  $k$ , we have  $k(0) \leq \beta(1 - \alpha)$  if and only if

$$0 = \Phi(k)|_{k=k(0)} \leq \Phi(k)|_{k=\beta(1-\alpha)}$$

Substituting  $k = \beta(1 - \alpha)$  into the first-order conditions we get the schedules (25) and (26); in turn, substituting these schedules into  $\Phi(k)$ , we get the condition in the Proposition.

For the uniform distribution, the condition can be expressed completely in terms of parameters. (23) and (24) simplify to

$$\underline{x}^*(\theta) = (V_x^1)^{-1} \left( \theta + (1 - \alpha) \frac{(\theta - \underline{\theta})}{\bar{\theta} - \underline{\theta}} + \delta q_n - \frac{k}{\beta} \frac{1}{\bar{\theta} - \underline{\theta}} \right)$$

and

$$\bar{x}^*(\theta) = (V_x^1)^{-1} \left( \theta + (1 - \alpha) \frac{(\theta - \underline{\theta})}{\bar{\theta} - \underline{\theta}} + \frac{k}{1 - \beta} \frac{1}{\bar{\theta} - \underline{\theta}} \right).$$

$k(0)$  satisfies the condition  $\delta q_n - \frac{k(0)}{\beta} \frac{1}{\bar{\theta} - \underline{\theta}} = \frac{k(0)}{1 - \beta} \frac{1}{\bar{\theta} - \underline{\theta}}$ .<sup>16</sup> Hence, I have  $k(0) = (1 - \beta) \beta \delta q_n (\bar{\theta} - \underline{\theta})$  and so  $k(0) \geq \beta(1 - \alpha)$  if and only if  $(1 - \beta) \delta q_n (\bar{\theta} - \underline{\theta}) > (1 - \alpha)$ .

Leaving a rent to type  $(\bar{\theta}, \eta)$  seems to be a natural, rather than a pathological outcome. These conditions are easy to meet and consistent with the requirement that  $\bar{\eta} - \underline{\eta}$  be sufficiently large.

<sup>16</sup>To see this, observe that for any value of  $k(0)$  different from this one, I would have either  $\bar{x}^*(\theta) > \underline{x}^*(\theta)$  for all  $\theta$  or  $\bar{x}^*(\theta) < \underline{x}^*(\theta)$  for all  $\theta$ , and both possibilities are inconsistent with condition (22) for  $\underline{\pi} = 0$ .

The intuition for the result is that increasing  $\underline{\pi}$  allows the regulator to tailor the quantity schedules  $\bar{x}^*(\theta)$  and  $\underline{x}^*(\theta)$  for good one better to the marginal costs of production for good one; these costs are higher for types with low cost of producing good two, because they are asked to produce the larger amount of good two. It is optimal to leave a rent to type  $(\bar{\theta}, \underline{\eta})$  if, all else equal, the weight attached to firm profits becomes 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.

The solution has some remarkable properties. Firstly, the production schedules are distorted away from first-best for the most efficient producer. The least cost producer of good one is the firm with cost parameters  $(\underline{\theta}, \bar{\eta})$ . Even though the firm with parameters  $(\underline{\theta}, \underline{\eta})$  has the lowest cost parameters, this firm is asked to produce a higher amount of good two, and this raises its marginal cost of producing good one. However, regardless of how the most efficient firm is defined, we have  $\underline{x}^*(\underline{\theta}; k) > \underline{x}^{fb}(\underline{\theta})$  and  $\bar{x}^*(\underline{\theta}; k) < \bar{x}^{fb}(\underline{\theta})$ .

Secondly, the direction of distortions away from the first-best allocation differs from the usual features obtained in one-dimensional models. Whereas my model features downward distortions if production schedules are designed when the regulator and the firm both know  $\eta$ , asymmetric information about  $\eta$  causes upwards distortions in the production schedule for firms with cost parameter  $\underline{\eta}$  and downward distortions for firms with cost parameter  $\bar{\eta}$ . The downward distortions for the latter group of firms is more pronounced than in the one-dimensional model. The upward distortion in the production schedule for firms with cost parameter  $\underline{\eta}$  is most pronounced if  $k$  is as large as possible, that is if  $\underline{\pi}^* > 0$ . In this case,  $\underline{x}^{\dagger\dagger}(\theta)$  reveals that, all but the highest cost producer in this group produce more than the first-best amount; the highest cost producer in this group produces an efficient amount. Note that, in order to induce consumers to buy an amount  $\underline{x}^{\dagger\dagger}(\theta)$ , the firm has to set prices below marginal cost. Formally, this can be seen noting that  $V_x^1(x) = P^1(x)$ . So, this model can explain below-marginal-cost-pricing in the absence of competition or demand complementarities.

## 6.2 Neutral goods and affiliated types: the case of 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 3:**  $\theta$  and  $\eta$  are affiliated, i.e.  $\frac{\partial}{\partial \theta} \frac{f(\theta|\bar{\eta})}{f(\theta|\underline{\eta})} > 0$ .

The reason I assume affiliation<sup>17</sup> is that it allows me to pin down the bunching region.

<sup>17</sup>Affiliation is consistent with the reverse hazard rate order; more precisely, affiliation implies the reverse hazard



**Lemma 6** *If  $\theta$  and  $\eta$  are affiliated, then the schedules  $\underline{x}(\theta)$  and  $\bar{x}(\theta)$  defined by*

$$\left( V_x^1(\underline{x}(\theta)) - \theta - (1 - \alpha) \frac{F(\theta|\underline{\eta})}{f(\theta|\underline{\eta})} \right) \beta f(\theta|\underline{\eta}) + \kappa = 0 \quad (27)$$

and

$$\left( V_x^1(\bar{x}(\theta)) - \theta - (1 - \alpha) \frac{F(\theta|\bar{\eta})}{f(\theta|\bar{\eta})} \right) (1 - \beta) f(\theta|\bar{\eta}) - \kappa = 0, \quad (28)$$

satisfy (21) for any  $\kappa \geq 0$ .

Given Lemmas 5 and 6, the optimum can be found among solution schedules that involve a regime of bunching in the  $\eta$ -dimension for values of  $\theta$  below some value  $\theta'$  and a regime of separation in the  $\eta$ -dimension for values of  $\theta$  higher than  $\theta'$ . The point  $\theta'$  is endogenously determined by the optimal choice of the quantity schedules for good one and the level of  $\underline{\pi}$ .

To understand the trade-offs involved it is again useful to approach the problem sequentially. I first take as given any value of  $\underline{\pi} \in [0, \underline{\pi}^\dagger)$ , and solve for the optimal production schedules for good one. In the second step I endogenize the choice of  $\underline{\pi}$ .

**Lemma 7** *For a given  $\underline{\pi} \in [0, \underline{\pi}^\dagger)$ , the optimal production schedules  $\underline{x}^*(\theta, k(\underline{\pi}))$  and  $\bar{x}^*(\theta, k(\underline{\pi}))$  are determined as follows:*

for  $\theta \leq \theta'$ , the optimal schedules satisfy  $\underline{x}^*(\theta, \underline{\pi}) = \bar{x}^*(\theta, \underline{\pi}) = x^*(\theta)$  where  $x^*(\theta)$  satisfies

$$V_x^1(x^*(\theta)) = \theta + (1 - \alpha) \frac{F(\theta)}{f(\theta)}. \quad (29)$$

For  $\theta > \theta'$ ,  $\underline{x}^*(\theta, k(\underline{\pi}))$  satisfies (27) and  $\bar{x}^*(\theta, k(\underline{\pi}))$  satisfies (28).

$\theta'(\underline{\pi})$  and  $k(\underline{\pi})$  are jointly determined by the conditions

$$\underline{\pi} + \int_{\theta'}^{\bar{\theta}} \underline{x}(\theta, k) d\theta - \int_{\theta'}^{\bar{\theta}} \bar{x}(\theta, k) d\theta = 0$$

and

$$\underline{x}(\theta', k) = \bar{x}(\theta', k) = x^*(\theta').$$

The solution is depicted in figure 4.

There is bunching of different  $\eta$  types who have the same marginal cost parameter  $\theta$  at the low end of the  $\theta$  support; at the high end there is separation of such types. Moreover, at the point where the regime changes from bunching to separation of different  $\eta$  types, the solution schedules switch continuously from one regime to the other. The idea to show this is the following. The effect rate order but is not implied by it. (See Shaked and Shantikumar (2007)).

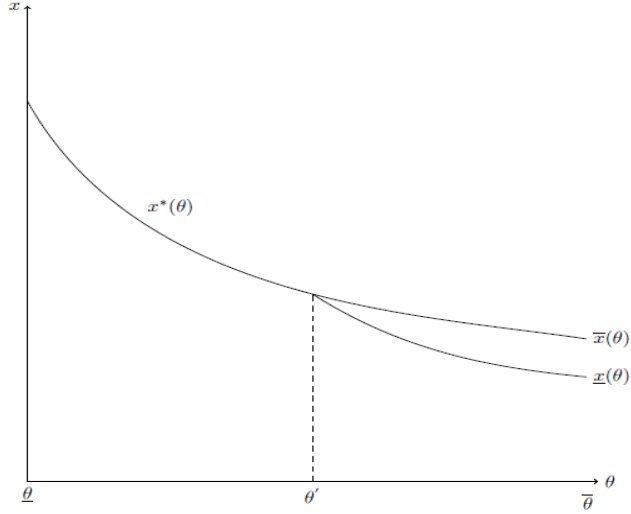


Figure 4: Raising  $\underline{\pi}$  “pushes the point  $\theta'$  to the left”; a higher  $\underline{\pi}$  enables the regulator to separate a larger portion of types.

of a marginal change of  $\theta'$  on the value of the regulator’s payoff function should be zero around the optimal value of the switch-point  $\theta'$ . This requires that, conditional on  $\theta = \theta'$ , the expected value of the objective at  $\theta'$  - 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  $\varepsilon$  positive but arbitrarily small. This value matching condition essentially boils down to requiring continuity of the solution schedules. Thus,  $\theta'$  is the unique intersection of the schedules defined by (27), (28), and (29) and the value of the multiplier,  $k(\underline{\pi})$ , adjusts so that the rents of types  $(\theta', \underline{\eta})$  and  $(\theta', \bar{\eta})$  are exactly equal.

Consider now the optimal choice of  $\underline{\pi}$ . Figure 4 illustrates the trade-off the regulator faces when choosing  $\underline{\pi}$ . The higher is  $\underline{\pi}$ , the larger is the separating region at the high end of the support. So, at the cost of giving up rents to all types with a low parameter  $\eta$ , the regulator can solve the efficiency versus rent extraction trade-off within groups of agents with the same parameter  $\eta$  better.

The following Proposition completely describes the optimum. To rule out problems of bunching in the  $\theta$  dimension, I assume that

**Assumption 4:**  $\frac{F(\theta)}{f(\theta)}$  is nondecreasing in  $\theta$ .

**Proposition 3** For  $\delta = 0$ ,  $\bar{\eta} - \underline{\eta}$  sufficiently large, and under Assumptions 3 and 4, the optimum

involves  $\underline{\pi}^* = 0$  and quantity schedules  $\underline{x}^*(\theta) = \bar{x}^*(\theta) \equiv x^*(\theta)$  where  $x^*(\theta)$  satisfies

$$V_x^1(x^*(\theta)) = \theta + (1 - \alpha) \frac{F(\theta)}{f(\theta)}. \quad (30)$$

There is complete bunching of types, that is the solution schedules become independent of  $\eta$  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.

It is never optimal to leave rents to inefficient producers among firms with low cost parameter  $\eta$ . Since  $\delta = 0$ , the only motive to offer two different schedules  $\underline{x}^*(\theta)$  and  $\bar{x}^*(\theta)$  to producers with different cost of producing good two is to extract more rents from producers within each group. If  $\underline{\pi}$  is increased marginally, then the regulator can extract some more rents over a small additional interval of types in the  $\theta$  dimension. However, the cost of this change outweighs the benefits by far, as the regulator leaves an additional rent  $d\underline{\pi}$  to *all* producers with cost parameter  $\underline{\eta}$ .

## 7 Extensions and Conclusions

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).

The presentation in this paper is streamlined around cases that display interesting economic findings. However, the solution procedure readily extends to other cases that satisfy the conditions in Lemma 5. In particular, one can rationalize allocations that feature separation at the high end and bunching at the low end of the  $\theta$ -range by allowing for the right kind of statistical dependence. Likewise, the paper is focussed around cases where the most tempting deviations in the  $\eta$ -dimension relate to the incentive constraints of types with low costs of producing good two.

It is straightforward to analyze cases where the firms with high  $\eta$ -parameter must be kept from mimicking firms with a low  $\eta$ -parameter. This is analytically straightforward and not performed here for reasons of space.

It has been observed that bunching is a robust feature in the multidimensional problem. This paper clearly agrees with that view; with affiliated types the solution displays complete bunching in the added dimension. However, the appeal of the present approach is that the solution techniques still remain manageable even though there is bunching, as long as there is no bunching in the  $\theta$ -dimension. This feature encourages further extensions, such as introducing richer trade-offs in the good two allocation problem and enriching the type space further towards the double continuum case. These extensions are pursued in ongoing work.

## 8 Appendix

### 8.1 Implementable Allocations

**Proof of Lemma 1.** Consider a candidate optimal allocation. It is useful to organize the incentive constraints any such candidate needs to fulfill into categories. First, one-dimensional deviations must be suboptimal; for all  $\theta, \eta$  it must be true that

$$\Pi(\theta, \theta, \eta, \eta) \geq \Pi(\hat{\theta}, \theta, \eta, \eta) \text{ for all } \hat{\theta} \quad (31)$$

and

$$\Pi(\theta, \theta, \eta, \eta) \geq \Pi(\theta, \theta, \hat{\eta}, \eta) \text{ for all } \hat{\eta}. \quad (32)$$

Moreover, two-dimensional deviations must be suboptimal too. Define the sets  $\underline{\Theta}_i \equiv \{\theta : q(\theta, \underline{\eta}) = q_i\}$  and  $\bar{\Theta}_i \equiv \{\theta : q(\theta, \bar{\eta}) = q_i\}$  for  $i = 0, 1, \dots, n$ . two-dimensional deviations are suboptimal for any type with cost parameters  $(\theta, \underline{\eta})$  if

$$\Pi(\theta, \theta, \underline{\eta}, \underline{\eta}) \geq \Pi(\hat{\theta}, \theta, \bar{\eta}, \underline{\eta}) \text{ for all } \hat{\theta}.$$

The right-hand side of this condition can be rewritten as

$$\Pi(\hat{\theta}, \theta, \bar{\eta}, \underline{\eta}) = \Pi(\hat{\theta}, \theta, \bar{\eta}, \bar{\eta}) + q_i(\bar{\eta} - \underline{\eta}) \text{ for } \hat{\theta} \in \bar{\Theta}_i.$$

Likewise, two-dimensional deviations are suboptimal for any type with cost parameters  $(\theta, \bar{\eta})$  if

$$\Pi(\theta, \theta, \bar{\eta}, \bar{\eta}) \geq \Pi(\hat{\theta}, \theta, \underline{\eta}, \bar{\eta}) \text{ for all } \hat{\theta}.$$

Again, the right-hand side of this condition can be rewritten as

$$\Pi(\hat{\theta}, \theta, \underline{\eta}, \bar{\eta}) = \Pi(\hat{\theta}, \theta, \underline{\eta}, \underline{\eta}) - q_i(\bar{\eta} - \underline{\eta}) \text{ for } \hat{\theta} \in \underline{\Theta}_i.$$

So, any candidate optimal allocation fulfills in addition the following sets of dimensional constraints.

For all  $(\theta, \underline{\eta})$  and all  $i$

$$\Pi(\theta, \theta, \underline{\eta}, \underline{\eta}) \geq \Pi(\hat{\theta}, \theta, \bar{\eta}, \bar{\eta}) + q_i(\bar{\eta} - \underline{\eta}) \text{ for } \hat{\theta} \in \bar{\Theta}_i. \quad (33)$$

and

$$\Pi(\theta, \theta, \bar{\eta}, \bar{\eta}) \geq \Pi(\hat{\theta}, \theta, \underline{\eta}, \underline{\eta}) - q_i(\bar{\eta} - \underline{\eta}) \text{ for } \hat{\theta} \in \underline{\Theta}_i \quad (34)$$

The proof consists in showing that, starting from an incentive compatible allocation for which any of the sets  $\bar{\Theta}_j$  for  $j \neq 0$  and  $\underline{\Theta}_k$  for  $k \neq n$  are non-empty, one can create a new, incentive compatible allocation and increase surplus. Hence, the initial candidate allocation cannot be optimal.

Suppose we adjust the allocation to  $q(\theta, \bar{\eta}) = q_0$  and  $q(\theta, \underline{\eta}) = q_n$  for all  $\theta$ . While doing this, we adjust payments to keep the equilibrium profits  $\Pi(\theta, \theta, \underline{\eta}, \underline{\eta})$  and  $\Pi(\theta, \theta, \bar{\eta}, \bar{\eta})$  unchanged for all  $\theta$ . Keeping the production of good one constant, the transfer to the firm with type  $(\theta, \underline{\eta})$ , where  $\theta \in \underline{\Theta}_k$ , needs to increase by  $(\underline{\eta} + \delta x(\theta, \underline{\eta}))(q_n - q_k)$ . Likewise, the transfer to a firm with type  $(\theta, \bar{\eta})$  can be decreased by the amount  $(\bar{\eta} + \delta x(\theta, \bar{\eta}))(q_j - q_0)$  for  $\theta \in \bar{\Theta}_j$ . By Assumption 1, both changes result in an increase in surplus.

It remains to be shown that the new allocation satisfies all the incentive constraints. Notice that

$$\Pi(\hat{\theta}, \theta, \eta, \eta) = \Pi(\hat{\theta}, \hat{\theta}, \eta, \eta) + (\hat{\theta} - \theta)x(\hat{\theta}, \eta).$$

Hence, the change of allocation does not affect  $\Pi(\hat{\theta}, \theta, \eta, \eta)$  in any sense. Therefore, (31) continues to hold after the change of allocation. Consider now (34) and (33). After the change of allocation, these constraints take the form

$$\Pi(\theta, \theta, \underline{\eta}, \underline{\eta}) \geq \Pi(\hat{\theta}, \theta, \bar{\eta}, \bar{\eta}) + q_0(\bar{\eta} - \underline{\eta}) \text{ for all } \hat{\theta}$$

and

$$\Pi(\theta, \theta, \bar{\eta}, \bar{\eta}) \geq \Pi(\hat{\theta}, \theta, \underline{\eta}, \underline{\eta}) - q_n(\bar{\eta} - \underline{\eta}) \text{ for all } \hat{\theta}.$$

Compared to (34) and (33), the right-hand side of these inequalities are reduced, so the old incentive constraints imply the new ones.

Finally, note that the incentive constraints in the  $\eta$  dimension alone are just a special case of these two-dimensional ones. Since both  $\Pi(\hat{\theta}, \theta, \bar{\eta}, \underline{\eta})$  and  $\Pi(\hat{\theta}, \theta, \underline{\eta}, \bar{\eta})$  are (weakly) reduced for any  $\hat{\theta}$ , this holds in particular true for  $\hat{\theta} = \theta$ .

Hence, if the initial allocation is incentive compatible, we can create a new allocation, by adjusting good two production as claimed, and increase surplus. ■

**Proof of Lemma 2.** Clearly, (4) and (5) are necessary for (2). So, I need to show that they are sufficient as well.

Given that the regulator follows a good two allocation rule  $q(\theta, \bar{\eta}) = q_0$  and  $q(\theta, \underline{\eta}) = q_n$  for all  $\theta$ , the profit of a type  $(\theta, \bar{\eta})$  firm mimicking a type  $(\hat{\theta}, \underline{\eta})$  firm is equal to

$$\Pi(\hat{\theta}, \theta, \underline{\eta}, \bar{\eta}) = \Pi(\hat{\theta}, \theta, \underline{\eta}, \underline{\eta}) - q_n(\bar{\eta} - \underline{\eta}). \quad (35)$$

Since  $\Pi(\hat{\theta}, \theta, \underline{\eta}, \underline{\eta})$  and  $\Pi(\hat{\theta}, \theta, \underline{\eta}, \bar{\eta})$  differ only by an additive constant (which is independent of  $\hat{\theta}$ ), we have

$$\arg \max_{\hat{\theta}} \Pi(\hat{\theta}, \theta, \underline{\eta}, \bar{\eta}) = \arg \max_{\hat{\theta}} \Pi(\hat{\theta}, \theta, \underline{\eta}, \underline{\eta}).$$

By (4),  $\theta = \arg \max_{\hat{\theta}} \Pi(\hat{\theta}, \theta, \underline{\eta}, \underline{\eta})$ , so

$$\Pi(\theta, \theta, \bar{\eta}, \bar{\eta}) \geq \Pi(\theta, \theta, \underline{\eta}, \bar{\eta}) \text{ for all } \theta$$

implies that for all  $\theta$

$$\Pi(\theta, \theta, \bar{\eta}, \bar{\eta}) \geq \Pi(\hat{\theta}, \theta, \underline{\eta}, \bar{\eta}) \text{ for all } \hat{\theta}.$$

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

$$\Pi(\hat{\theta}, \theta, \bar{\eta}, \underline{\eta}) = \Pi(\hat{\theta}, \theta, \bar{\eta}, \bar{\eta}) + q_0(\bar{\eta} - \underline{\eta}) \quad (36)$$

so clearly again

$$\arg \max_{\hat{\theta}} \Pi(\hat{\theta}, \theta, \bar{\eta}, \underline{\eta}) = \arg \max_{\hat{\theta}} \Pi(\hat{\theta}, \theta, \bar{\eta}, \bar{\eta}).$$

Again by (4),  $\theta = \arg \max_{\hat{\theta}} \Pi(\hat{\theta}, \theta, \bar{\eta}, \bar{\eta})$ , so

$$\Pi(\theta, \theta, \underline{\eta}, \underline{\eta}) \geq \Pi(\theta, \theta, \bar{\eta}, \underline{\eta}) \text{ for all } \theta$$

implies that for all  $\theta$

$$\Pi(\theta, \theta, \underline{\eta}, \underline{\eta}) \geq \Pi(\hat{\theta}, \theta, \bar{\eta}, \underline{\eta}) \text{ for all } \hat{\theta}.$$

■

**Proof of Lemma 4.** Let  $\beta = \Pr[\eta = \underline{\eta}]$ . The “reduced” problem where I neglect the monotonicity constraints on  $\bar{x}(\theta)$  and  $\underline{x}(\theta)$  can be written as follows:

$$\Gamma(\underline{\pi}) = \max_{\bar{x}(\theta), \underline{x}(\theta)} \left[ \begin{array}{l} \beta \int_{\underline{\theta}}^{\bar{\theta}} B(\underline{x}(\theta), q_n, \theta, \underline{\eta}) f(\theta | \underline{\eta}) d\theta - \beta(1 - \alpha) \underline{\pi} \\ + (1 - \beta) \int_{\underline{\theta}}^{\bar{\theta}} B(\bar{x}(\theta), q_0, \theta, \bar{\eta}) f(\theta | \bar{\eta}) d\theta \end{array} \right]$$

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

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

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

$$H = B(\underline{x}(\theta), q_n, \theta, \underline{\eta}) \beta f(\theta | \underline{\eta}) + B(\bar{x}(\theta), q_0, \theta, \bar{\eta}) (1 - \beta) f(\theta | \bar{\eta})$$

$$+ \bar{\kappa} \bar{x} + \underline{\kappa} \underline{x} + \mu (\underline{\pi} - (\underline{z} - \bar{z}))$$

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

$$\frac{\partial H}{\partial \bar{z}} = \mu = -\bar{\kappa}_\theta$$

$$\frac{\partial H}{\partial \underline{z}} = -\mu = -\underline{\kappa}_\theta;$$

differentiating with respect to the controls I get

$$\frac{\partial H}{\partial \bar{x}} = \left( V_{\bar{x}}^1(\bar{x}(\theta)) - \theta - \delta q_0 - (1 - \alpha) \frac{F(\theta | \bar{\eta})}{f(\theta | \bar{\eta})} \right) (1 - \beta) f(\theta | \bar{\eta}) + \bar{\kappa} = 0 \quad (37)$$

$$\frac{\partial H}{\partial \underline{x}} = \left( V_{\underline{x}}^1(\underline{x}(\theta)) - \theta - \delta q_n - (1 - \alpha) \frac{F(\theta | \underline{\eta})}{f(\theta | \underline{\eta})} \right) \beta f(\theta | \underline{\eta}) + \underline{\kappa} = 0$$

The Kuhn-Tucker conditions are

$$\underline{\pi} - (\underline{z} - \bar{z}) \leq 0, \quad \mu \geq 0, \quad \text{and} \quad \mu (\underline{\pi} - (\underline{z} - \bar{z})) = 0.$$

Since both  $\bar{z}(\theta)$  and  $\underline{z}(\theta)$  are free, the transversality conditions are

$$\bar{\kappa}(\underline{\theta}) = \underline{\kappa}(\underline{\theta}) = 0.$$

Finally, the costate variables are allowed to take jumps at points where the inequality state constraint switches from being active to slack.

If  $\underline{\pi} - (\underline{z}(\theta) - \bar{z}(\theta)) = 0$  then

$$\underline{\kappa}(\underline{\theta}^+) - \underline{\kappa}(\underline{\theta}) = \gamma_0$$

and

$$\bar{\kappa}(\underline{\theta}^+) - \bar{\kappa}(\underline{\theta}) = -\gamma_0.$$

for some  $\gamma_0$ . If  $\underline{\pi} - (\underline{z}(\underline{\theta}) - \bar{z}(\underline{\theta})) > 0$ , then  $\gamma_0 = 0$ .

If  $\theta_i \in (\underline{\theta}, \bar{\theta})$  is a point where the inequality state constraint switches from being binding to slack or vice-versa then

$$\underline{\kappa}(\theta_i^+) - \underline{\kappa}(\theta_i) = \gamma_i$$

and

$$\bar{\kappa}(\theta_i^+) - \bar{\kappa}(\theta_i) = -\gamma_i$$

for some  $\gamma_i$ . See Seierstad and Sydsaeter (1999) chapter 5 for a statement and proof of optimality of these conditions.

Suppose that the state inequality constraint is slack at and continues to be slack on a set of positive measure  $[\underline{\theta}, \theta']$  implying that  $\gamma_0 = 0$ ,  $\mu(\underline{\theta}) = 0$  and  $\mu(\theta) = 0$  for all  $\theta \in [\underline{\theta}, \theta']$ . From conditions (37) it is clear that  $\bar{\kappa}$  and  $\underline{\kappa}$  are continuously differentiable in  $\theta$  whenever  $\bar{x}$  and  $\underline{x}$  are continuously differentiable in  $\theta$ . Using the conditions of optimality for the state variables,  $\bar{\kappa}_\theta = -\mu$  and  $\underline{\kappa}_\theta = \mu$ , and the transversality conditions, I have for  $\theta \leq \theta'$

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

and

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

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

$$V_{\bar{x}}^1(\bar{x}(\theta)) - \theta - \delta q_0 - (1 - \alpha) \frac{F(\theta | \bar{\eta})}{f(\theta | \bar{\eta})} = 0$$

and

$$V_{\underline{x}}^1(\underline{x}(\theta)) - \theta - \delta q_n - (1 - \alpha) \frac{F(\theta | \underline{\eta})}{f(\theta | \underline{\eta})} = 0.$$

These conditions are equivalent to (13) and (14), so these conditions would imply that  $\underline{x}(\theta) = \underline{x}^\dagger(\theta)$  and  $\bar{x}(\theta) = \bar{x}^\dagger(\theta)$  for  $\theta \leq \theta'$ .

The proof that  $\underline{\pi} < \bar{\pi}^\dagger$  is made formally in the proof of Proposition 2 below. The argument in the main text shows that the above conditions lead into a contradiction. ■

**Proof of Lemma 5.** Notice that

$$\underline{\kappa}(\theta) = -\bar{\kappa}(\theta) \text{ for all } \theta.$$



To see this, notice that at points where the controls are differentiable, we have  $\bar{\kappa}_\theta = -\mu(\theta)$  and  $\underline{\kappa}_\theta = \mu(\theta)$ . At points where the state inequality constraint switches from being active to passive or vice versa we have  $\underline{\kappa}(\theta_i^+) - \underline{\kappa}(\theta_i) = -(\bar{\kappa}(\theta_i^+) - \bar{\kappa}(\theta_i))$ , or  $\underline{\kappa}(\underline{\theta}^+) - \underline{\kappa}(\underline{\theta}) = -(\bar{\kappa}(\underline{\theta}^+) - \bar{\kappa}(\underline{\theta}))$  if this point is  $\underline{\theta}$ .

Moreover, since  $\bar{\kappa}_\theta = -\mu(\theta)$  and  $\underline{\kappa}_\theta = \mu(\theta)$  whenever, the controls are differentiable, the costate variabls are constants on any interval where the inequality state constraint is slack. Suppose  $[\theta_1, \theta_2]$  is such an interval. Then, I can write

$$\underline{\kappa}(\theta) = \underline{\kappa}(\theta_1^+) \text{ for all } \theta \in (\theta_1, \theta_2)$$

and

$$\bar{\kappa}(\theta) = -\underline{\kappa}(\theta_1^+) \text{ for all } \theta \in (\theta_1, \theta_2),$$

where  $\underline{\kappa}(\theta)$  is non-negative.

Inserting these values into the conditions of optimality for the control variables, 37, I have

$$\left( V_{\underline{x}}^1(\underline{x}^*(\theta)) - \theta - \delta q_0 - (1 - \alpha) \frac{F(\theta|\bar{\eta})}{f(\theta|\bar{\eta})} \right) - \frac{\underline{\kappa}(\theta_1^+)}{(1 - \beta) f(\theta|\bar{\eta})} = 0$$

and

$$\left( V_{\underline{x}}^1(\underline{x}^*(\theta)) - \theta - \delta q_n - (1 - \alpha) \frac{F(\theta|\underline{\eta})}{f(\theta|\underline{\eta})} \right) + \frac{\underline{\kappa}(\theta_1^+)}{\beta f(\theta|\underline{\eta})} = 0.$$

In the remainder of the proof, I show that these schedules are consistent with constraint (17) if they satisfy (20). In this case we can take the state inequality constraint binding on the minimal set  $\underline{\theta}$ . Moreover, I show that, when these schedules satisfy (21), then they are only consistent with (17) if (17) binds on a single interval.

i) Suppose that the conditions of optimality satisfy (20). Distinguish two cases,  $\theta_2 < \bar{\theta}$  and  $\theta_2 = \bar{\theta}$ . If  $\theta_2 < \bar{\theta}$ , then (17) is binding at  $\theta_2$ , so  $\underline{\kappa}(\theta_1^+)$  adjusts such that.

$$\int_{\theta_1}^{\theta_2} \underline{x}^*(y) dy - \int_{\theta_1}^{\theta_2} \bar{x}^*(y) dy = 0.$$

Given condition (20), this implies that  $\bar{x}^*(\theta_1^+) > \underline{x}^*(\theta_1^+)$ ,  $\bar{x}^*(\theta_2^-) < \underline{x}^*(\theta_2^-)$ , and that the schedules cross exactly once. In turn, since

$$\rho_\theta(\theta, \underline{\pi}) = \bar{x}^*(\theta) - \underline{x}^*(\theta)$$

this implies that  $\rho(\theta, \underline{\pi})$  is increasing at  $\theta_1^+$ , reaches a maximum in the interior of  $(\theta_1, \theta_2)$  and decreases again towards  $\theta_2^-$ , where  $\rho(\theta, \underline{\pi}) = 0$  by definition. Hence, the constraint is met for all  $\theta \in (\theta_1, \theta_2)$ .

Consider now the case where  $\theta_2 = \bar{\theta}$ , where  $\rho(\bar{\theta}, \underline{\pi}) = \underline{\pi} \geq 0$ . This case differs from the former one only if  $\underline{\pi} > 0$ , so suppose this is the case. In this case,  $\underline{\kappa}(\theta_1^+)$  is reduced to allow for a deficit  $\int_{\theta_1}^{\theta_2} \underline{x}^*(y) dy - \int_{\theta_1}^{\theta_2} \bar{x}^*(y) dy = -\underline{\pi}$ . Reducing  $\underline{\kappa}(\theta_1^+)$  decreases  $\underline{x}^*(\theta)$  pointwise and increases  $\bar{x}^*(\theta)$  pointwise, so it weakly increases  $\rho_\theta(\theta, \underline{\pi})$  pointwise. Since for any  $\theta \in (\theta_1, \theta_2)$ ,

$$\rho(\theta, \underline{\pi}) = \rho(\theta_1, \underline{\pi}) + \int_{\theta_1}^{\theta} \rho_y(y, \underline{\pi}) dy,$$

this implies that  $\rho(\theta, \underline{\pi}) \geq 0$  for all  $\theta \in (\theta_1, \theta_2)$ .

ii) Suppose now that condition (21) holds, so  $\bar{x}(\theta) = \underline{x}(\theta) \implies \frac{d\bar{x}(\theta)}{d\theta} > \frac{d\underline{x}(\theta)}{d\theta}$ . Then, the schedules  $\bar{x}(\theta)$  and  $\underline{x}(\theta)$  can only satisfy condition

$$\int_{\theta_1}^{\theta_2} \underline{x}^*(y) dy - \int_{\theta_1}^{\theta_2} \bar{x}^*(y) dy = 0$$

if  $\bar{x}^*(\theta_1^+) < \underline{x}^*(\theta_1^+)$ ,  $\bar{x}^*(\theta_2^-) > \underline{x}^*(\theta_2^-)$ , and if the schedules cross exactly once. But then, I would have

$$\rho_\theta(\theta_1^+, \underline{\pi}) = \bar{x}^*(\theta_1^+) - \underline{x}^*(\theta_1^+) < 0,$$

so (17) would be violated for some  $\theta \in (\theta_1, \theta_2)$ . Hence, it is not possible that (17) is slack on the interval  $(\theta_1, \theta_2)$  and becomes binding at  $\theta_2$  again. So, under condition (21), constraint (17) is, if at all, binding on an interval. ■

**Proof of Proposition 2.** The Lagrangian of the problem is

$$L = \max_{\bar{x}(\theta), \underline{x}(\theta), \underline{\pi}} \left[ \begin{aligned} & \beta \int_{\underline{\theta}}^{\bar{\theta}} B(\underline{x}(\theta), q_n, \theta, \underline{\eta}) f(\theta | \underline{\eta}) d\theta - \beta(1 - \alpha)\underline{\pi} \\ & + (1 - \beta) \int_{\underline{\theta}}^{\bar{\theta}} B(\bar{x}(\theta), q_0, \theta, \bar{\eta}) f(\theta | \bar{\eta}) d\theta \\ & + k \left( \int_{\underline{\theta}}^{\bar{\theta}} (\underline{x}(\theta) - \bar{x}(\theta)) d\theta + \underline{\pi} \right). \end{aligned} \right] \quad (38)$$

In this formulation,  $k$  is independent of  $\theta$ . It is useful to solve this problem sequentially. In the first step, I take  $\underline{\pi}$  as given and solve for the quantity schedules that are optimal against that value of  $\underline{\pi}$ . In this step, the value of the multiplier depends on  $\underline{\pi}$ , so I write  $k = k(\underline{\pi})$ . The quantity schedules depend on  $\underline{\pi}$  through the level of the multiplier, but I will simply write  $\underline{x}^*(\theta; k)$  and  $\bar{x}^*(\theta; k)$  to keep notation compact.

Step 1: For any given  $\underline{\pi}$ , the problem is concave in  $\bar{x}(\theta)$  and  $\underline{x}(\theta)$ , so by a standard sufficiency theorem, the first-order conditions are necessary and sufficient for an optimum. The pointwise first-order conditions with respect to the controls are

$$\left( V_x^1(\underline{x}^*(\theta; k)) - \theta - \delta q_n - (1 - \alpha) \frac{F(\theta)}{f(\theta)} \right) \beta f(\theta) + k(\underline{\pi}) = 0$$

and

$$\left( V_x^1(\bar{x}^*(\theta; k)) - \theta - \delta q_0 - (1 - \alpha) \frac{F(\theta)}{f(\theta)} \right) (1 - \beta) f(\theta) - k(\underline{\pi}) = 0.$$

Step 2: Invoking the envelope theorem, the derivative of the objective with respect to  $\underline{\pi}$  is

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

the second derivative with respect to  $\underline{\pi}$  is

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

I now show that  $\frac{dk(\underline{\pi})}{d\underline{\pi}} \leq 0$ , implying that the objective is concave in  $\underline{\pi}$ .

Totally differentiating the constraint

$$\int_{\underline{\theta}}^{\bar{\theta}} (\underline{x}^*(\theta; k) - \bar{x}^*(\theta; k)) d\theta + \underline{\pi} = 0,$$

I obtain

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

Differentiating the first-order conditions for the controls totally, I obtain

$$\frac{d\underline{x}^*(\theta; k)}{dk} = -\frac{1}{V_{xx}^1(\underline{x}^*(\theta; k)) \beta f(\theta)} \quad (39)$$

and

$$\frac{d\bar{x}^*(\theta; k)}{dk(\underline{\pi})} = \frac{1}{V_{xx}^1(\bar{x}^*(\theta; k)) (1 - \beta) f(\theta)}, \quad (40)$$

so  $\underline{x}_k^*(\theta; k) > 0$  and  $\bar{x}_k^*(\theta; k) < 0$ , and thus  $\frac{dk(\underline{\pi})}{d\underline{\pi}} \leq 0$ .

Since the objective is concave in  $\underline{\pi}$ , the first-order condition is necessary and sufficient for an optimum. I have either  $\underline{\pi}^* = 0$  and

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

or  $\underline{\pi}^* > 0$  and

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

Notice that in either case we must have  $\underline{x}^* < \underline{x}^\dagger$  since  $\underline{x}^\dagger > 0$  and  $k(\underline{x}^\dagger) = 0$ . The former case arises if and only if  $k(0) \leq \beta(1 - \alpha)$ . We now determine whether or not this condition holds.

Consider the function

$$\Phi(k) \equiv \int_{\underline{\theta}}^{\bar{\theta}} (\underline{x}^*(\theta; k) - \bar{x}^*(\theta; k)) d\theta.$$

We know that  $\Phi(0) < 0$ , since otherwise the constraint would not be binding at all.  $\Phi(k)$  is increasing in  $k$ .  $k(0)$  is the value of  $k$  solving  $\Phi(k)|_{k=k(0)} = 0$ . Since  $\Phi(k)$  is increasing in  $k$ , we have  $k(0) \leq \beta(1 - \alpha)$  if and only if

$$\Phi(k)|_{k=k(0)} = 0 \leq \Phi(k)|_{k=\beta(1-\alpha)}$$

Substituting  $k = \beta(1 - \alpha)$  into the first-order conditions we get the schedules

$$V_x^1(\underline{x}^{\dagger\dagger}(\theta)) = \theta + \delta q_n - (1 - \alpha) \frac{1 - F(\theta)}{f(\theta)}$$

and

$$V_x^1(\bar{x}^{\dagger\dagger}(\theta)) = \theta + \delta q_0 + (1 - \alpha) \frac{\frac{\beta}{1-\beta} + F(\theta)}{f(\theta)}.$$

Hence, we have  $k(0) \leq \beta(1 - \alpha)$  if and only if

$$\int_{\underline{\theta}}^{\bar{\theta}} \underline{x}^{\dagger\dagger}(\theta) d\theta \geq \int_{\underline{\theta}}^{\bar{\theta}} \bar{x}^{\dagger\dagger}(\theta) d\theta.$$

Finally, we need to check incentive compatibility in the  $\theta$  and in the  $\eta$  dimension.

Consider first incentive compatibility in the  $\theta$  dimension. The schedules (23) and (24) are continuous; hence they are differentiable everywhere and if they satisfy  $\frac{d\underline{x}(\theta)}{d\theta} \leq 0$  and  $\frac{d\bar{x}(\theta)}{d\theta} \leq 0$ , they are monotonic. Consider first the schedule  $\bar{x}(\theta)$ . From a total differentiation of (23), I obtain

$$\frac{d\underline{x}^*}{d\theta} = \frac{\left(1 + (1 - \alpha) \frac{\partial}{\partial \theta} \frac{F(\theta)}{f(\theta)}\right) + \frac{k}{\beta} \frac{f_\theta(\theta)}{f(\theta)^2}}{V_{xx}^1(\underline{x}^*(\theta))}. \quad (41)$$

A total differentiation of (24) gives

$$\frac{d\bar{x}^*}{d\theta} = \frac{\left(1 + (1 - \alpha) \frac{\partial}{\partial \theta} \frac{F(\theta)}{f(\theta)}\right) - \frac{k^*}{1-\beta} \frac{f_\theta(\theta)}{f(\theta)^2}}{V_{xx}^1(\bar{x}^*(\theta))}, \quad (42)$$

Observe that  $f_\theta(\theta) \leq 0$  implies that  $\frac{\partial}{\partial \theta} \frac{F(\theta)}{f(\theta)} \geq 0$ . Hence, the numerator on the right-hand side of (42) is non-negative and thus  $\frac{d\bar{x}^*}{d\theta} \leq 0$ . The numerator on the right-hand side of (41) is minimized for  $k$  as large as possible. As shown above, the highest possible value of  $k$  is  $\beta(1 - \alpha)$ . Hence we have

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

where the first inequality follows from substituting  $k = \beta(1 - \alpha)$  and the second one follows from  $\frac{\partial}{\partial \theta} \frac{1-F(\theta)}{f(\theta)} \leq 0$ .

Consider next incentive compatibility in the  $\eta$  dimension; we need to verify that the solution schedules satisfy (20). Clearly,  $f_\theta(\theta) \leq 0$  implies that the numerator on the right-hand side of (42) is at least as large as the numerator on the right-hand side of (41). Hence,  $\underline{x}(\theta) = \bar{x}(\theta) \implies \frac{d\underline{x}(\theta)}{d\theta} \geq \frac{d\bar{x}(\theta)}{d\theta}$ . ■

**Proof of Lemma 6.** I demonstrate that affiliation implies that (21). Differentiating (27) and (28) I obtain

$$\frac{d\underline{x}^*}{d\theta} = \frac{(2 - \alpha) - \frac{f_\theta(\theta|\eta)}{f(\theta|\eta)} \left[ (1 - \alpha) \frac{F(\theta|\eta)}{f(\theta|\eta)} - \frac{\kappa}{\beta f(\theta|\eta)} \right]}{V_{xx}^1(\underline{x}^*(\theta))}$$

and

$$\frac{d\bar{x}^*}{d\theta} = \frac{2 - \alpha - \frac{f_\theta(\theta|\bar{\eta})}{f(\theta|\bar{\eta})} \left[ (1 - \alpha) \frac{F(\theta|\bar{\eta})}{f(\theta|\bar{\eta})} + \frac{\kappa}{(1-\beta)f(\theta|\bar{\eta})} \right]}{V_{xx}^1(\bar{x}^*(\theta))}.$$

Notice that at  $\underline{x}(\theta) = \bar{x}(\theta)$ , the terms in brackets are equal to  $V_x^1(\underline{x}^*(\theta)) - \theta = V_x^1(\bar{x}^*(\theta)) - \theta$ ; moreover, as the latter term in brackets is necessarily positive, the former one is positive too. Hence,  $\bar{x}(\theta) = \underline{x}(\theta) \implies \frac{d\bar{x}(\theta)}{d\theta} > \frac{d\underline{x}(\theta)}{d\theta}$  if and only if  $\frac{f_\theta(\theta|\bar{\eta})}{f(\theta|\bar{\eta})} > \frac{f_\theta(\theta|\eta)}{f(\theta|\eta)}$ , which is precisely the affiliation assumption. ■

**Proof of Lemma 7.** To ease notation in this proof, I suppress the dependence of the solution schedules on  $\underline{\pi}$ .

Consider again the control problem spelled out in the proof of Lemma 4. The proof is organized into three parts. In part i, I derive the solution for  $\theta \leq \theta'$ . In part ii, I derive the solution for  $\theta > \theta'$ . In part iii, I show that the solution displays continuity at  $\theta'$ .

**Part i:** the solution for  $\theta \leq \theta'$

I demonstrate that that for  $\theta \leq \theta'$ , the optimal schedule satisfies  $\bar{x}(\theta) = \underline{x}(\theta) = x^*(\theta)$  and  $x^*(\theta)$  solves

$$V_x^1(x^*(\theta)) - \theta - (1 - \alpha) \frac{F(\theta)}{f(\theta)} = 0. \quad (43)$$

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

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

and

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

Substituting back into (37)

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

Note that  $\bar{x} = \underline{x}$  as  $\mu > 0$  for  $\theta \leq \theta'$ . Adding the two conditions of optimality for the control variables, and dividing by  $f(\theta)$  I get

$$V_x^1(x^*(\theta)) - \theta - (1 - \alpha) \frac{F(\theta)}{f(\theta)} = \frac{-\underline{\kappa}(\underline{\theta}) - \bar{\kappa}(\underline{\theta})}{f(\theta)} \quad (44)$$

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

To complete the argument I now argue that  $-\underline{\kappa}(\underline{\theta}) - \bar{\kappa}(\underline{\theta}) = 0$ . Since  $\bar{x}(\theta) = \underline{x}(\theta) = x^*(\theta)$  for  $\theta \leq \theta'$ , any solution of (44) for given  $-\underline{\kappa}(\underline{\theta}) - \bar{\kappa}(\underline{\theta})$  satisfies constraint (17). Moreover,  $\underline{\kappa}(\underline{\theta})$  and  $\bar{\kappa}(\underline{\theta})$  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,  $\underline{\kappa}(\underline{\theta})$  and  $\bar{\kappa}(\underline{\theta})$  have no influence on the location of the switching point  $\theta'$  either. Hence, at the optimum  $\underline{\kappa}(\underline{\theta})$  and  $\bar{\kappa}(\underline{\theta})$  must be such that, conditional on  $\theta$ , the expected value of the objective is maximized. Hence,  $\bar{\kappa}(\underline{\theta}) = -\underline{\kappa}(\underline{\theta})$ , and I obtain the expression in the Proposition.

**Part ii:** the solution for  $\theta > \theta'$

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

I first show that  $\bar{k} + \underline{k} = 0$ . To see this, consider a candidate pair of schedules that give rise to a switch point  $\theta'$ . Clearly, for the subinterval  $[\theta', \bar{\theta}]$ , constraint (17) is binding only at  $\theta'$ . But then, choosing the optimal schedules  $\bar{x}(\theta)$  and  $\underline{x}(\theta)$  on the subinterval  $[\theta', \bar{\theta}]$  is equivalent to the isoperimetric problem (38) with  $\theta'$  replacing  $\underline{\theta}$ . Hence,  $k \equiv \bar{k} = -\underline{k}$ .

**Part iii:** Continuity at the switch-point

I can write the value of the objective as

$$\Gamma(\theta') = W^1(\theta') + W^2(\theta') - \beta(1 - \alpha)\underline{\pi}, \quad (45)$$

where

$$W^1(\theta') \equiv \beta \int_{\underline{\theta}}^{\theta'} B(x^*(\theta), q_n, \theta, \underline{\eta}) f(\theta|\underline{\eta}) d\theta + (1 - \beta) \int_{\underline{\theta}}^{\theta'} B(x^*(\theta), q_0, \theta, \bar{\eta}) f(\theta|\bar{\eta}) d\theta$$

and

$$W^2(\theta') \equiv \beta \int_{\theta'}^{\bar{\theta}} B(\underline{x}(\theta), q_n, \theta, \underline{\eta}) f(\theta | \underline{\eta}) d\theta + (1 - \beta) \int_{\theta'}^{\bar{\theta}} B(\bar{x}(\theta), q_0, \theta, \bar{\eta}) f(\theta | \bar{\eta}) d\theta \\ + k \left( \underline{\pi} + \int_{\theta'}^{\bar{\theta}} \underline{x}(y) dy - \int_{\theta'}^{\bar{\theta}} \bar{x}(y) dy \right).$$

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

$$\Gamma_{\theta'}(\theta') = W_{\theta'}^1(\theta') + W_{\theta'}^2(\theta')$$

where

$$W_{\theta'}^1(\theta') \equiv \beta B(x^*(\theta'), q_n, \theta', \underline{\eta}) f(\theta' | \underline{\eta}) d\theta + (1 - \beta) B(x^*(\theta'), q_0, \theta', \bar{\eta}) f(\theta' | \bar{\eta})$$

and

$$W_{\theta'}^2(\theta') = -\beta B(\underline{x}(\theta'), q_n, \theta', \underline{\eta}) f(\theta' | \underline{\eta}) - (1 - \beta) B(\bar{x}(\theta'), q_0, \theta', \bar{\eta}) f(\theta' | \bar{\eta}) \\ - k(\underline{x}(\theta') - \bar{x}(\theta')) \quad (46)$$

Clearly, at the optimum I must have  $W_{\theta'}^1(\theta') + W_{\theta'}^2(\theta') = 0$ , so the values of the objectives evaluated at the bound  $\theta'$  must match. One solution is clearly reached when  $\bar{x}(\theta') = \underline{x}(\theta') = x^*(\theta')$ . I now show this solution is unique.

To make the dependence of  $\bar{x}(\theta')$  and  $\underline{x}(\theta')$  on  $k$  explicit, I write these schedules as  $\bar{x}(\theta'; k) = \underline{x}(\theta'; k)$ , respectively. A total differentiation of the integral constraint

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

delivers

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

Recall from (39) and (40) in the proof of Proposition 2 that  $\frac{dx}{dk} > 0$  and  $\frac{d\bar{x}}{dk} < 0$ . Hence, the denominator of the expression for  $\frac{dk}{d\theta'}$  is positive. Hence, I have  $\frac{dk}{d\theta'} < 0$  for  $\underline{x}(\theta'; k) < \bar{x}(\theta'; k)$  and  $\frac{dk}{d\theta'} > 0$  for  $\underline{x}(\theta'; k) > \bar{x}(\theta'; k)$ . Thus  $k$  is minimized when  $\theta'$  is such that  $\bar{x}(\theta'; k) = \underline{x}(\theta'; k)$ . Again using (39) and (40), for any other value of  $\theta'$ , I will have  $\underline{x}(\theta'; k) > \bar{x}(\theta'; k)$ . However, it

is easy to see that the values

$$\bar{x} \equiv \arg \max_x \{ (V^1(x) - \theta) f(\theta' | \bar{\eta}) - (1 - \alpha) F(\theta' | \bar{\eta}) \}$$

and

$$\hat{x} \equiv \arg \max_x \{ (V^1(x) - \theta) f(\theta' | \underline{\eta}) - (1 - \alpha) F(\theta' | \underline{\eta}) \}$$

satisfy  $\bar{x} \geq \hat{x}$  due to affiliation. Hence, the sum of the terms in the first line in (46) decreases by an increase in  $k$ . Moreover,  $-k(\underline{x}(\theta') - \bar{x}(\theta'))$  becomes negative. Hence, there can be no other solution.

Taken together these arguments imply the structure of the solution given in the Lemma. ■

**Proof of Proposition 3.** Complementing the proof of Lemma 7, I demonstrate that the optimal value of  $\underline{\pi}$  is zero.

It is easy to see that the derivative of (45) with respect to  $\pi$  is still given by

$$\Gamma_{\underline{\pi}}(\underline{\pi}) = -\beta(1 - \alpha) + k; \quad (47)$$

and the second derivative is still given by  $\Gamma_{\underline{\pi}\underline{\pi}} = \frac{dk}{d\underline{\pi}}$  whenever this is well defined. Letting  $\bar{x}(\theta; k)$  and  $\underline{x}(\theta; k)$  denote the functions defined by (23) and (24), and using the fact that (17) is binding at  $\theta'(k)$ , I have for any  $\underline{\pi} > 0$

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

where I have used the fact that effects of  $k$  on  $\theta'(k)$  exactly cancel out as  $\underline{x}(\theta'(k); k) = \bar{x}(\theta'(k); k)$ . 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 Lemmas 5 and 6, for  $\underline{\pi} = 0$ , constraint (17) is binding at  $\bar{\theta}$ ; by convexity of the bunching region, the constraint is binding for all  $\theta$ . Hence, when evaluating the derivative  $\Gamma_{\underline{\pi}}(\underline{\pi})$  at  $\underline{\pi} = 0$ , I can use the fact that  $\bar{x}(\bar{\theta}) = \underline{x}(\bar{\theta}) = x^*(\bar{\theta})$  for  $\underline{\pi} = 0$ . Hence, for  $\theta = \bar{\theta}$  and  $\delta = 0$ , I can write (43) in explicit form as

$$(V_x^1(x^*(\bar{\theta})) - \bar{\theta}) f(\bar{\theta}) = (1 - \alpha) \quad (48)$$

From (24),  $k(0)$  satisfies

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

Substituting from (48), we have

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



and thus

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

■

## 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 Economics Research Paper Series (TWERPS) 858, University of Warwick, Department of Economics, 2008
- Jullien, B. “Participation Constraints in Adverse Selection Models,” *Journal of Economic Theory*, **93**, 2000, 1–47.
- 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
- Kleven, H. J., C. T. Kreiner, and E. Saez “The Optimal Income Taxation of Couples,” *Econometrica* Vol. 77, No. 2 March, 2009, 537–560
- Laffont, J.-J., E. Maskin, and J.-C. 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.
- Laffont, J.-J, Tirole, J, “A Theory of Incentives in Procurement and Regulation”, MIT Press, 1993
- Lewis, T. and D. Sappington, “Regulating a Monopolist with Unknown Demand and Cost Function”, *Rand Journal of Economics* **19**, 1988, 438-457.
- Lewis, T. and D. Sappington, “Countervailing Incentives in Agency Problems,” *Journal of Economic Theory*, **49** (1989), 294–313.

- Loeb, M. and Magat, W. “A decentralized method of utility regulation”, *Journal of Law and Economics* 22, 1979, 399-404
- Maggi, G., and A. Rodriguez-Clare, “On Countervailing Incentives,” *Journal of Economic Theory* 66, 1995, 238–263.
- McAfee, R.P. and J. MacMillan, “Multidimensional Incentive Compatibility and Mechanism Design”, *Journal of Economic Theory* 46, 1988, 335-354.
- 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.
- Sappington, D. Optimal Regulation of a Multiproduct Monopoly with Unknown Technological Capabilities, *The Bell Journal of Economics*, 1983, Vol. 14, No. 2 (Autumn, 1983), pp. 453-463
- Sawkins, J.W., and S. Reid, “The Measurement and Regulation of Cross Subsidy: The Case of Scottish Water Industry,” *Utilities Policy*, 15 (2007), 36–48.
- 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
- Tirole, J. “The Theory of Industrial Organization”, *MIT Press, Cambridge MA*, 1988