ADVERSE SELECTION PROBLEMS WITHOUT THE SINGLE CROSSING PROPERTY

ALOISIO ARAUJO

Instituto de Matemática Pura e Aplicada
and Fundação Getúlio Vargas of Rio de Janeiro

and

HUMBERTO MOREIRA

Catholic University of Rio de Janeiro

ABSTRACT: We relax the single crossing property (SCP) for the adverse selection problems and new conditions for the incentive compatibility are derived: the marginal rate of substitution identities. The set of implementable and the optimal contracts are characterized in a specific case where the SCP is no longer valid and the characterization of the solution depends on these new conditions. The optimal contracts present discrete or continuous pooling (reswichting) and they are discontinuous even if under the monotone hazard rate condition. The question raises naturally when we are dealing with one dimensional "path" of a multidimensional characteristic problem with countervailing incentives (given by the correlation between two sources of asymmetric information). Therefore our paper gives also a framework to study multi-characteristic adverse selection problems We give some examples of them: a combination of moral hazard and adverse selection, nonlinear pricing, the regulation problem and labor market.

Keywords: Single Crossing Property; Marginal Rate of Substitution Identity; Discrete Pooling.

JEL Classification: D82

This draft: December, 1999

Financial support from CNPq and CAPES. Preliminary versions of this paper were presented at the University of Toulouse, Pompeu Fabra (Barcelona), IMPA, PUC-Rio, FGV/RJ, University of Chicago and during the XIX Meeting of the Brazilian Econometric Society and the XVII Latin America Meeting of the Econometric Society. We wish to thank A. Mas-Colell, J. Tirole, J.C. Rochet, A. Chassagnon, P.A. Chiappori, M. Harris, J. Scheinkman and L. Stole, as well as the seminar participants, for useful discussions and suggestions on earlier draft of this paper. We would like to thank Caio Ibsen for helping us in the numerical examples of the paper.

1. INTRODUCTION

The main goal of this paper is to relax the classical single crossing property (SCP) which has extensively used in the literature to characterize the solution of the adverse selection problem.

In the one-dimensional parameter case, the SCP is by definition the monotonicity of the marginal rate of substitution between the decision taken by the agent and the money transfer given by the principal with respect to the parameter (the asymmetric information).

The SCP permits the second order approach for the problem since the feasible set is a convex set: in the presence of positive (respectively negative) SCP, a decision path is implementable if and only if it is non-decreasing (respectively non-increasing) in the parameter.

The SCP enables also a full characterization of the optimal solution: in the presence of the SCP, if the optimal decision is strictly monotone in the parameter, then it should be equal to the relaxed solution.[1] Moreover, a maximal interval where it is constant is such that the marginal virtual surplus of the principal (i.e., the social surplus minus the informational rent) should be zero. These properties are sufficient to provide an algorithm allowing the computation of optimal solutions (see Guesnerie and Laffont (1984)).

The question studied in this paper is what happens when the SCP is not valid anymore? In this case, there are at least two regions in the plane of the parameter versus the decision variable: the positive and the negative single crossing regions. An implementable decision path should preserve the monotonicity property in each region and it can cross or not the curve that separates the two regions (the frontier). If the decision path crosses the frontier, what are the necessary conditions for the incentive compatibility? First, the decision path crosses the frontier in a U-shaped form (or an bell-shaped form) because of the monotonicity. Besides, we prove that a necessary condition is: if two types have the same decision, then their marginal rate of substitution should be the same. In economic terms, if two types are pooling in a given contract, then the principal guarantees truth telling only if the marginal rate of substitution of the two types is the same. We will call this condition the *marginal rate of substitution identity.* Moreover, there exist an analogous marginal condition with respect to the type (*the marginal rent identity*). In general, these conditions are not sufficient for incentive compatibility, but they are sufficient in a particular setup that we will examine.[2]

We use the second order condition of the incentive compatibility (IC) restriction and the marginal rate of substitution identities as the constraints of the

[1] This solution is obtained by imposing the first order condition of incentive compatibility constraint to reduce the problem only to the decision variable.

[2] The more general case is not finished. The difficulty of the complete characterization is that even if a decision path does not cross the two regions, the monotonicity condition is not sufficient to characterize the implementable decisions.

adverse selection problem and derive the first order conditions for the optimal contract. The constraints will be of equality and inequality type one and our problem is not concave anymore, but we still can compute the optimal contract in some cases.

Chassagnon and Chiappori (1995) studied the insurance market competitive equilibrium with adverse selection and moral hazard where the SCP is not valid. However, they studied the two type case and the second order approach remains true in the continuous version of their model. We will use the same idea of simultaneous adverse selection and moral hazard to provide the example below without the SCP and where the second order approach is not valid.

Example. (Owner-manager relationship under moral hazard and adverse selection)

Suppose that an owner (the principal) of a firm has to hire a manager (the agent) to deliver a product for him. Assume that the manager can choose between two types of technologies and the manager is more or less productive depending on his type and on the technology chosen by him. The owner has to design the reward schedule.

Let x be the units of output, θ be the manager's productivity, y be the worker's effort and t be the salary. Each type of manager has a comparative advantage in one of the technologies. More precisely, the technologies are described by $T \in \{T_1, T_2\}$:

$$\begin{aligned} T1: &\quad x = (1-\theta)y, \quad \theta \in \Theta \\ T2: &\quad x = \theta y, \qquad\quad\;\; \theta \in \Theta \end{aligned}$$

where $\Theta = [0, 1]$ is the set of types and the distribution of the manager's type is represented by a density function $p\colon \Theta \to \mathbb{R}_{++}$.

The manager's utility function is

$$V = t - y^2$$

and the owner is risk neutral with utility function given by:

$$U = x - t.$$

Define also the manager's utility function given the technology choice and as a function of the output:

$$V(x, t, \theta \mid T_1) = t - \left(\frac{x}{1-\theta}\right)^2$$

and

$$V(x, t, \theta \mid T_2) = t - \left(\frac{x}{\theta}\right)^2.$$

It is clear that $V(x, t, \theta \mid T_1) \geq V(x, t, \theta, T_2)$ if and only if $\theta \leq 1/2$, i.e., the low (high) types have comparative advantage in technology T_1 (T_2). It follows that the manager with characteristic θ close to 0 (repectively 1) is a specialist in technology T_1 (repectively T_2). We also say that types θ close to $1/2$ are generalist (they are the bad types).

The principal's problem is to maximize his expected utility over all the contracts $\{x(\cdot), t(\cdot)\}$ given the participation constraints and, in the case of asymmetric information, the incentive compatibility constraints, i.e.,

$$\max_{\{x(\cdot),t(\cdot)\}} E_\theta[x(\theta) - t(\theta)]$$

subject to incentive compatibility and participation constraints[3].

Depending on the informational structure, we have different problems. Consider the following cases:

1) *First Best*: T and θ are observable and verifiable, i.e., the owner can put the manager's type and the technology choice as part of the contract. This is the full symmetric information case. It is very easy to compute the optimal solution (see section 4)

$$\begin{aligned}
T^{FB} &= T_1(T_2) \quad \text{iff} \quad \theta \leq (\geq)\, 1/2 \\
x^{FB}(\theta) &= \begin{cases} \dfrac{(1-\theta)^2}{2}, & \text{if} \quad \theta \in [0, 1/2] \\ \dfrac{\theta^2}{2}, & \text{otherwise} \end{cases} \\
t^{FB}(\theta) &= \frac{x^{FB}(\theta)}{2}, \quad \text{for all} \quad \theta \in [0,1].
\end{aligned}$$

This solution has the well known properties: it is Pareto efficient and the manager has zero rent.

2) *Second Best with the technology choice information*: θ is not observable, but T is observable and verifiable. In this case the optimal contract can be contingent only to the technology choice. Therefore the owner designs two types of contracts (one for each chosen technology)

$$\left\{(x_i(\theta), t_i(\theta))\right\}_{\theta\in[0,1]}^{i=1,2}$$

such that it satisfies, for $i = 1, 2$,

$$\theta \in \arg\max_{\hat\theta\in[0,1]} V(x_i(\hat\theta), t_i(\hat\theta), \theta \mid T_i), \quad \forall\, \theta \in [0,1], \tag{IC$_i$}$$

[3] Assume that the manager's reservation utility is zero.

The manager with productivity θ chooses T_1 (T_2) if and only if

$$V(x_1(\theta), t_1(\theta), \theta \mid T_1) \geq (\leq) V(x_2(\theta), t_2(\theta), \theta \mid T_2)$$

and the owner has to take this last inequality into consideration to compute his expected utility.

However, since the type θ has a comparative advantage in T_1 (T_2) if and only if type $\theta \leq (\geq) 1/2$, then, in equilibrium, the optimal contract should induce type $\theta \leq (\geq) \frac{1}{2}$ to choose T_1 (T_2) when the distribution is symmetric with respect to $1/2$. Therefore, the employer will design $\{x_1, t_1\}$ ($\{x_2, t_2\}$) shutting down each type $\theta < (>) 1/2$ in equilibrium, i.e., the IR will not hold for this type. This implies that type $1/2$ will have zero rent in equilibrium. We can say that the owner will use the technology choice as signal of the manager's type (since he can control this choice).

Conditioning in each technology, the principal's problem is going to be a standard adverse selection with the SCP. The optimal decision is going to be U-shaped and to have the same properties of separating or continuous pooling equilibrium. In section 4, we will compute explicitly the solution and introduce also lotteries in the technology choice to improve the expected profit of the owner.[4]

3) *Second Best without the technology choice information*: neither θ nor T are observable. In this case the optimal contract can not be contingent to the technology, i.e., the owner can not use the technology choice as a signal of the manager's type. Therefore, the owner will face a manager that has the utility function

$$V(x, t, \theta) = \begin{cases} t - \left(\dfrac{x}{1-\theta}\right)^2, & \text{if} \quad \theta \in [0, 1/2] \\ t - \left(\dfrac{x}{\theta}\right)^2, & \text{otherwise} \end{cases}$$

It is easy to see that the derivative of the marginal rate of substitution with respect to the type changes its sign exactly at the type $1/2$. Therefore the principal's problem is going to be an adverse selection one without the SCP.

As we explained above, we have to consider now the marginal rate of substitution identity to characterize the solution. In this case, this identity is equivalent to the symmetry with respect to $1/2$, i.e., the implementable decisions x are going to be U-shaped symmetric with respect to $1/2$ (see section 4 for the details). Therefore the optimal decision will also have this property. Since the owner can not see the technology choice, he will never know this choice in equilibrium because for each output decision, there are exactly two types of managers using different technologies and delivering it. (This is what we call discrete pooling equilibrium.)

[4] The intuition here is that, by using lotteries, the principal can threat the risk averse agent and extract more rent from him. We will show that lotteries on the level of production choice do not help.

The impossibility of observing the choice made by the agent transform the bidimensional second best problem (in (T, θ)) with information into a one dimensional problem (in θ only) without the SCP. We can use this method to generate several examples. The key point is that there is a "countervailing" effect between technology choice and productivity: the principal guarantees truth telling only if he equalizes marginal utilities of the agents that choose the same output.

Another example to be considered in this paper is a natural extension of the nonlinear pricing models studied by Mussa and Rosen (1978) and Maskin and Riley (1984) where the SCP is relaxed. Suppose that a monopolist faces different types of demand with finite elasticity. More precisely, the demands are linear and the market size is decreasing with the type and the maximum price where there exists positive demand is increasing in the type. This means that the market for the low types (in the sense of wiliness to pay for the good) is the large market. In this case the SCP fails to hold and the optimal contract is going to be non-decreasing for the low types and non-increasing for the high types (the bell-shaped curve). The reason is that the monopolist wants to extract the maximal rent as in the single crossing case, but now he has to deal with the trade off between the size of the market and the wiliness to pay of the consumer: he will not extract as much rent as before of the low type because he wants to sell more the good and at the same time he wants to extract the rent of the high type without breaking the IC constraints. Therefore the condition is that the low and high type consumers that are pooling in the same quality or quantity have to have the same marginal valuation (the competitive price) for the good, i.e., in equilibrium they are treated as the same. This leads to a discrete pooling equilibrium again. The same aspects of countervailing and one dimensional "path" of a bidimensional problem are repeated here.

We also analyze a regulation model *a la* Laffont and Tirole (1993). In their basic model the cost function depends on a non observable parameter (the efficiency) and effort of the firm's manager in cutting cost. The cost function is one dimensional in the sense that there is just one source of activity that the manager can cut cost. Suppose, however, that there are two kinds of activities that the manager can put his effort to cut cost and the regulator can only observe the aggregate cost, i.e., there are two subcosts that are not observable by the regulator and the sum of them are contractile. Moreover, these activities are substitutes in the manager's point of view (i.e., the manager's disutility in efforts has positive cross derivative) and there is a decreasing relation between the parameters that characterize the subcosts. We show that this kind of interaction will result in non single crossing and again we will have the same kind of message that we had explained in the examples above.

Finally, a labor market model where workers have a vector of two characteristics (unknown for the firm) that is mixed in a verifiable signal (schooling). The firm is profit maximizer and its technology depends on this vector of characteristics: one of the characteristics has a multiplicative effect over the worker's effort and the other one is constant. There is a conflict of interest between the firm and the worker because effort is costly (and not observable) and the abilities of the worker are not totally captured by the signal. This is a standard adverse selection, however depending on the parameters of the model the SCP does not hold. In this case, discrete pooling equilibrium may appear and indicates that different workers with respect to the profile of their characteristics may be treated as the same in equilibrium.

Page (1991) shows a general existence result when the contracts are lotteries for principal-agent problems. Athey (1997) shows the existence of a pure strategy equilibria in games with incomplete information under a generalized single crossing condition. The strategies can be monotone or have "limited complexity" form (i.e., they have a finite number of peaks). These properties have a straightforward relationship with our case, but Athey (1997) does not characterize the equilibrium. We provide an existence result of deterministic optimal contract in our particular case.

The paper is organized as follows. In section 2 we present the adverse selection model. In section 3 we characterize the solution of adverse selection problems without the SCP. Section 4 presents some examples. In section 5 we gives extensions and final conclusions.

2. THE ADVERSE SELECTION MODEL

The relationship between the principal and the agent(s) involves only two types of variables: The first type is associated with a decision (or action) variable, denoted by x which is observable. The variable of the second type, denoted by t has generally the meaning of money transfer from the principal to the agent.

The principal and the agent interact through these two variables and the asymmetry of information can be described as follows: there is an one-dimensional parameter θ which is known to the agent but unobservable to the principal. This parameter belongs to some compact interval $\Theta = [\underline{\theta}, \overline{\theta}] \subset \mathbb{R}$. The principal has some *a priori* probability distribution on Θ which is associated to a continuous density $p\colon \Theta \to \mathbb{R}_{++}$. We can interpret this function as the principal's subjective assessment of the probability of θ when there is only one agent or the objective distribution of their types when there are many agents.

The principal's utility function is $U\colon I \times \mathbb{R} \times \Theta \to \mathbb{R}$, where $I \subset \mathbb{R}$ is an interval, $U(x, t, \theta) = u(x, \theta) - t$ and u is C^3 . The agent's utility function is $V\colon I \times \mathbb{R} \times \Theta \to \mathbb{R}$ such that $V(x, t, \theta) = v(x, \theta) + t$ and v is C^3.

A mechanism (contract or allocation) is a pair of functions $(x, t)\colon \Theta \to \mathbb{R}^2$. A mechanism can be viewed as a procedure giving the decision to the principal who commits himself to a decision rule relating the choice of x and t to messages sent by the agent. By the revelation principle (see Fudenberg and Tirole (1991)), any mechanism can be mimicked by a direct truthful one in the sense that there is no loss of welfare to the principal.

A decision function $x:\Theta \to \mathbb{R}$ is implementable if there exists a money transfer function $t:\Theta \to \mathbb{R}$ such that the allocation $x(\theta) \in I$, for all $\theta \in \Theta$ and satisfies the incentive compatibility constraint if for all $(\theta, \hat{\theta}) \in \Theta^2$,

$$V(x(\theta), t(\theta), \theta) \geq V(x(\hat{\theta}), t(\hat{\theta}), \theta) \tag{IC}$$

We will say that the allocation (x, t) is implementable or truth telling or that x implements t. In other words, given an implementable allocation (x, t), the announcement of the truth is an optimal strategy for the agent whatever the truth.

We say that an allocation (x, t) satisfies the individual-rationality constraint if for all $\theta \in \Theta$,

$$V(x(\theta), t(\theta), \theta) \geq 0 \tag{IR}$$

An implementable allocation that satisfies the IR constraint is called feasible. We assume that the agent's reservation utility is independent of his type[5] and, without loss of generality, we normalize it as zero.

The principal's (or the adverse selection) problem is to choose a feasible allocation with the highest expected payoff, i.e., the principal maximizes his expected utility subject to the agent's IR and IC constraints:

$$\max_{x,t} E_\theta[U(x(\theta), t(\theta), \theta)]$$

s.t.

$$V(x(\theta), t(\theta), \theta) \geq V(x(\hat{\theta}), t(\hat{\theta}), \theta),\ \forall\, (\theta, \hat{\theta}) \in \Theta^2 \tag{IC}$$

$$V(x(\theta), t(\theta), \theta) \geq 0, \quad \forall\, \theta \in \Theta \tag{IR}$$

where E_θ is the expectation with respect to the *prior*.

Definition: Let $\mathcal{C}$ be the set of all càlàg contracts, i.e., the space of all $x:\Theta \to \mathbb{R}$ right continuous and such that $\lim\limits_{\substack{\tilde{\theta}\to\theta \\ \tilde{\theta}<\theta}} x(\tilde{\theta})$ exists for each $\theta \in \Theta$ (and, in this case, it will be denoted by $x_-(\theta)$), with the topology of pointwise limit at every continuous parameter of the limit decision function (this is the weak topology in the distributional sense).

Below we present the classical first and second order conditions of the incentive constraints extended to càdlàg contracts.

Lemma 2.1

[5] We do this for simplicity. However, we can consider the case where the agent's reservation utility depends on the type. See Maggi and Rodriguez-Clare (1995), for instance.

(i) Let x be a bounded decision such that the set of its discontinuity points has zero Lebesgue measure. If t implements x, then the agent's value (or rent) function of x is given by[6]

$$\mathcal{V}^x(\theta) = v(x(\theta), \theta) + t(\theta) = \mathcal{V}^x(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} v_\theta(x(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta}, \ \forall \theta \in \Theta.$$

(ii) If x is a bounded càdlàg implementable decision, then x is non-decreasing on the region where $v_{x\theta} > 0$ (respectively non-increasing on the region where $v_{x\theta} < 0$).

Proof: See the appendix.

Lemma 2.1 (i) shows that for each implementable càdlàg x, there exist a unique càdlàg money transfer that implements x defined by

$$t(\theta) = \mathcal{V}^x(\theta) - v(x(\theta), \theta), \ \forall \theta \in \Theta.$$

Then, we define

$$\begin{aligned} \Phi^x(\theta, \hat{\theta}) &= V(x(\theta), t(\theta), \theta) - V(x(\hat{\theta}), t(\hat{\theta}), \theta) \\ &= \int_{\theta}^{\hat{\theta}} \left[\int_{x(\tilde{\theta})}^{x(\hat{\theta})} v_{x\theta}(\tilde{x}, \tilde{\theta}) d\tilde{x} \right] d\tilde{\theta} \end{aligned}$$

and, after an integration by parts, the virtual surplus (i.e., the social surplus minus the informational rent) times the probability is

$$\begin{aligned} f(x(\theta), \theta) = \big[& u(x(\theta), \theta) + v(x(\theta), \theta) \\ & + \frac{(P(\theta) - 1)}{p(\theta)} v_\theta(x(\theta), \theta) - \mathcal{V}^x(\underline{\theta}) \big] p(\theta) \end{aligned}$$

where $P(\theta) = \int_{\underline{\theta}}^{\theta} p(\tilde{\theta}) d\tilde{\theta}$ is the cumulative distribution.

If $v_\theta \geq 0$, then the rent function $\mathcal{V}^x$ assume its minimum at $\underline{\theta}$ and since money transfer is costly for the principal, the IR constraint will be binding at the optimal contract for $\underline{\theta}$, i.e., $\mathcal{V}^x(\underline{\theta}) = 0$. If $v_\theta \leq 0$, $\mathcal{V}^x(\overline{\theta}) = 0$ and

$$f(x(\theta), \theta) = \big[u(x(\theta), \theta) + v(x(\theta), \theta) + \frac{P(\theta)}{p(\theta)} v_\theta(x(\theta), \theta) \big] p(\theta).$$

Otherwise, $\mathcal{V}^x$ can assume its minimum at some point in Θ depending on x (except in special cases such as the example in the introduction where this point is 1/2

[6] The sub index in the function represents the partial derivative of the function with respect to that sub index. Also, the superior order derivative will be represented in a multi-index notation.

for every decision x). Therefore, in the general case one would need a Lagrange multiplier for the IR constraint in the problem that follows. For the sake of simplicity, let us assume that v_θ has a constant (positive) sign (see the remark 3 after theorem 3.5).

The principal's optimization program becomes

$$\begin{aligned} &\max_{x\in\mathcal{C}} E_\theta\Big[\frac{f(x(\theta),\theta)}{p(\theta)}\Big] \\ &\text{s.t. } \Phi^x(\theta,\hat\theta)\geq 0,\ \forall\theta,\hat\theta\in\Theta \end{aligned} \tag{P}$$

If we ignore the IC constraint, then the problem is called the relaxed problem and also its solution (first order approach). The first order condition of the relaxed problem is given by

$$f_x(x(\theta),\theta)=0, \quad \text{for all} \quad \theta\in\Theta$$

when $x(\theta)$ is in the interior of I.

It is well known in the literature of adverse selection problems that sufficient conditions for implementation is the constant sign of the partial derivative of the marginal rate of substitution with respect to the parameter:

$$\partial_\theta\left(\frac{V_x}{V_t}\right)=v_{x\theta}>0 \quad \text{on } I\times\Theta, \tag{$\mathbf{CS_+}$}$$

or

$$\partial_\theta\left(\frac{V_x}{V_t}\right)=v_{x\theta}<0 \quad \text{on } I\times\Theta. \tag{$\mathbf{CS_-}$}$$

This is known as the single crossing property (SCP) or sorting condition. This property implies that the indifference curves of two different types cross only one time.

In the presence of CS_+ (respectively CS_-), it is easy to show (see the proof of lemma 2.1) that if a càdlàg decision is non-decreasing (respectively non-increasing), then it is implementable. Therefore the adverse selection problem is equivalent to

$$\begin{aligned} &\max_{x\in\mathcal{C}} E_\theta\Big[\frac{f(x(\theta),\theta)}{p(\theta)}\Big] \\ &\text{s.t.} \quad x \text{ is non-decreasing (respectively non-increasing)} \end{aligned}$$

This problem is known as the second order approach because, under the SCP, the monotonicity of the decision is equivalent to the local second order condition of the IC constraint. Using the Hamiltonian approach, as in Guesnerie and Laffont (1984), one can obtain a full characterization of the solution.

3. RELAXING THE SINGLE CROSSING ASSUMPTION

Now we will introduce a natural generalization of the SCP. Assumption A1 below separates the plane θx into two regions: CS_+ and CS_-. From the local second order condition of the IC constraint we know that an implementable decision that crosses from one region to the other has to have a U-shaped form or presents continuous pooling. This will give us this kind of large scale pooling equilibrium (even under monotone hazard rate property).[7]

Formally:

A1. $v_{x\theta}(x,\theta)=0$ defines a function x_0 of θ on Θ; $v_{x^2\theta}<0$ and $v_{x\theta^2}\geq 0$ on $I\times\Theta$.

By the Implicit Function Theorem and A1, x_0 is C^1 and increasing:

$$\dot{x}_0(\theta)=-\frac{v_{x\theta^2(x_0(\theta),\theta)}}{v_{x^2\theta(x_0(\theta),\theta)}}$$

Moreover, if $x<x_0(\theta)$, $v_{x\theta}(x,\theta)>0$ (CS_+) and if $x>x_0(\theta)$, $v_{x\theta}(x,\theta)<0$ (CS_-), for all $\theta\in\Theta$ (see figure 1 below). Therefore, the assumption A1 generalizes the SCP, because $\Theta\times I$ is separated into two parts: above (respectively below) x_0, $v_{x\theta}>$ (respectively $<$) 0 on $I\times\Theta$.

Changing the sign of x or θ, there are more three other cases: $v_{x^2\theta}>0$ and $v_{x\theta^2}<0$, with x_0 increasing and reverting the regions where $v_{x\theta}>0$ and $v_{x\theta}<0$; $v_{x^2\theta}<0$ and $v_{x\theta^2}<0$; $v_{x^2\theta}>0$ and $v_{x\theta^2}>0$, for the respective cases where x_0 is decreasing.

Figure 1

We can relax the second part of A1: instead of assuming that x_0 is increasing, we could say that x_0 has a finite number of peaks. However, the analysis would be more difficult without any substantial gain in the results.

The next lemma shows when we can extend the definition of an implementable càdlàg to the associated correspondence. In what follows, $a\wedge b=\min\{a,b\}$ and $a\vee b=\max\{a,b\}$.

Theorem 3.1 Let x be a bounded implementable càdlàg decision and define $X(\theta)=[x_-(\theta)\wedge x(\theta), x_-(\theta)\vee x(\theta)]$, for all $\theta\in\Theta$, the associated correspondence and defined below.

(i) If $\theta,\hat{\theta}\in\Theta$, $\theta\leq\hat{\theta}$, then $\Phi_X(\theta,\hat{\theta})\geq 0$, i.e.,

$$\int_{\theta}^{\hat{\theta}}\left[\int_{x(\tilde{\theta})}^{y} v_{x\theta}(\tilde{x},\tilde{\theta})d\tilde{x}\right]d\tilde{\theta}\geq 0,\ \forall y\in X(\hat{\theta}).$$

[7] Chapter 9 of Laffont and Tirole (1993) studies the repeated regulation game without commitment and the "ratchet effect". They show that in equilibrium there may exist substantial pooling in every continuation equilibrium. In particular, their definition of pooling over a large scale for a continuation equilibrium is equivalent to our discrete pooling notion.

(ii) x is in the closure of the set of all continuous implementable decision if and only if X is implementable.

Proof: See the appendix.

Observe that x is discontinuous where its inverse x^{-1} is constant and vice-versa. This represents a kind of duality between the variables x and θ.

If X is implementable, then x crosses x_0 in a continuous way one time at most. In this case x must be non-increasing or non-decreasing or U-shaped. From now on, we will consider only the decision $x \in \mathcal{C}$ such that the associated X is implementable. We do this for the following reasons:

(i) economic meaning: if x is discontinuous at $\theta \in \Theta$, then it does not matter how x is defined in θ between $x_-(\theta)$ and $x(\theta)$.

(ii) tractability: if we relax this assumption, there are implementable decisions that cross x_0 many times in a discontinuous way.

(iii) when the SCP is valid, this set is the same of the implementable decisions.

(iv) discrete approximation: Theorem 3.1 says that an implementable decision can be approximated by the continuous one. But we can also prove that an implementable decision can be approximated by the "step" decisions, i.e., implementable decisions (in sense that it coincides with its associated correspondence) that are piecewise constant. This corresponds to a situation of finite types where each type has a range of decision and there is no discontinuity between the ranges[8].

The next theorems will give the necessary and sufficient conditions for the feasibility. First, we say that x is right increasing at $\theta \in \Theta$ if $x(\theta) < x(\theta + \epsilon)$, for every sufficiently small $\epsilon > 0$. Analogously, we define left increasing and right and left decreasing.

Theorem 3.2 (*Necessary conditions for implementability*) Assume A1. If x is an implementable càdlàg decision, then

(a) If x is right (left) increasing at $\hat{\theta}$ and $\Phi^x(\theta, \hat{\theta}) = 0$, then

$$v_x(x(\hat{\theta}), \hat{\theta}) \underset{(\leq)}{\geq} v_x(x(\hat{\theta}), \theta).$$

and reverting the inequalities when x is right (left) decreasing.

(b) If $\Phi^x(\theta, \hat{\theta}) = 0$, then

$$v_\theta(x(\hat{\theta}), \theta) \leq v_\theta(x(\theta), \theta) \;\text{ and }\; v_\theta(x(\hat{\theta}), \theta) \geq v_\theta(x_-(\theta), \theta)$$

[8] This corresponds to the partition equilibrium in the of the ratchet effect in Laffont and Tirole (1993). See the last footnote.

and with equality when x is continuous at θ and $\theta > \underline{\theta}$.

(c) If x is right or left increasing at $\hat{\theta}$ and $y \in X(\theta) \cap X(\hat{\theta})$, then

$$v_x(y, \hat{\theta}) = v_x(y, \theta).$$

Moreover, x is continuous at $\hat{\theta}$ (i.e., $y = x(\hat{\theta}) = X(\hat{\theta})$).

Proof: See the appendix.

Remark 1. The item (b) can be interpreted as the dual condition of (a) when we interchange θ by x, i.e., instead of looking the direct decision (x as a function of θ), we look the inverse function (θ as a function of x) and it can be interpreted as the **marginal rent equality** of a type that is considering his designed choice and an indifferent choice. Also, we can use the duality between x and θ and $v_{x\theta^2} > 0$ to show that $\Phi^x(\cdot, \hat{\theta}) \geq 0$ need to be checked only in extremities of an interval where x is constant.

Remark 2. We have the following economic interpretation for lemma 2.1 (ii) and theorem 3.2 (c): In order to provide truth telling, the principal should offer a contract that

(1) is non-decreasing (respectively non-increasing) in θ if the marginal rate of substitution is decreasing (respectively increasing) in θ.

(2) if two agents (θ and $\hat{\theta}$) choose the same contract and the agents cannot locally misrepresent their types, then the principal should equalize the marginal rate of substitution of the two agents ($\mathbf{MRS}^{\theta} = \mathbf{MRS}^{\hat{\theta}}$).

Figure 2

Remark 3. Every implementable decision x should not decrease (respectively increase) on CS_+ (respectively on CS_-). This property is a direct consequence of the local second order condition of the IC constraint.

If x hits the curve x_0, then it should cross x_0 in a constant way or preserving the marginal utility for the types that choose the same level of x.[9] This last condition is new and, when the SCP is not valid, it can play an important role in order to characterize the optimal solution of the adverse selection problem as the examples of section 4 will show. We will call this condition the **marginal rate of substitution identity**.[10]

[9] If x was identical to x_0 in an interval, then the IC constraint would not hold locally.

[10] Observe that the marginal rate of substitution identity is equivalent to v_x be constant on every level set of a feasible decision x.

Observe that, if (x, t) is feasible and $x(\theta) = x(\hat{\theta})$, then $t(\theta) = t(\hat{\theta})$. Thus, if two types are pooling in a feasible contract, then they should have the same marginal rate of substitution or a continuum of types between them should also pool.

Remark 4. Theorem 3.2 is also valid for a more general agent's utility functions. This is why we are calling our condition the marginal rate of substitution identity and not marginal utility identity.

It is important to note that we are dealing with a non-concave problem because the set of feasible decision for the agent is not a convex set when the agent's utility function does not satisfy the SCP. (The SCP guarantees concavity: the IC constraint is substituted by its second order condition).

A natural question is whether the conditions above are sufficient for the characterization of an implementable decision. Theorem 3.3 gives a partial answer of this question.

Theorem 3.3 Assume A1. Let x be a bounded càdlàg decision that satisfies the necessary condition of lemma 2.1 (ii) and theorem 3.2 (c). If $x(\underline{\theta}) \geq x(\overline{\theta})$, then x is implementable.

Proof: See the appendix.

The sufficient conditions for implementability of a decision x such that $x(\underline{\theta}) < x(\overline{\theta})$ is more difficult and is under investigation. However, theorem 3.3 can be used to solve some problems. In what follows, we show that when the relaxed solution is non-increasing, the optimal decision satisfies the assumption of theorem 3.3 and a full characterization of the optimal decision is possible.

The Existence and The Optimality Conditions

We will investigate the necessary conditions for optimality. First, we will characterize the relaxed solution. Let x_1 be the relaxed solution for (P). By the Maximum Theorem, x_1 is a continuous function of θ. Let us assume that:

A2. x_1 has a finite number of peaks on Θ; if $x < x_1(\theta)$, $f_x(x, \theta) > 0$ and if $x > x_1(\theta)$, $f_x(x, \theta) < 0$ for all $\theta \in \Theta$.

The assumption A2 is a standard one and a sufficient condition for the first part of A2 (besides the concavity of u and v) is $v_{x^2\theta} > 0$. If $x_1(\theta)$ belongs to the interior of I, then $f_x(x_1(\theta), \theta) = 0$.

Under A2, the principle of optimality for the adverse selection problem is to find an implementable decision "as close as possible" to x_1. The finite number of peaks of x_1 is to provide an analytical treatment of the problem and it is well known in the literature (see Guesnerie and Laffont (1984)).

A natural question is the existence of an optimal contract. Page (1991) provides a general result for the existence of an optimal contract in the case where the contracts are lotteries. In the case of deterministic contracts, Athey (1997) gives the existence of a pure strategy equilibrium for games with incomplete information under a generalized single crossing property or a limited complexity condition (i.e., the strategies have a finite number of peaks as a function of the parameter). In our case, we have the following:

Theorem 3.4 Assume that A1 and A2 hold. Then, there exist a solution of (P) in the set of all decision $x \in \mathcal{C}$ such that the associated correspondence X is implementable.

Proof: See the appendix.

The next theorem gives the characterization of the optimal decision in a special case. Let $x^* \in \mathcal{C}$ be an optimal decision for (P).

Theorem 3.5 Assume that A1 and A2 hold. If $x^*(\underline{\theta}) \geq x^*(\overline{\theta})$, let θ_0 be the minimum parameter for x^* and $\theta_1 \leq \theta_0$ such that $x^*(\overline{\theta}) \in X^*(\theta_1)$. Then

(a) If x^* is right and left decreasing at $\hat{\theta}$, then

$$\frac{f_x(x,\theta)}{v_{x\theta}(x,\theta)} = \delta\frac{f_x(x,\hat{\theta})}{v_{x\theta}(x,\hat{\theta})}$$

where $x = x^*(\hat{\theta})$ and if $\hat{\theta} < \theta_1$, then $\delta = 0$ and if $\theta_1 \leq \hat{\theta} \leq \theta_0$, then $\delta = 1$, $v_x(x,\hat{\theta}) = v_x(x,\theta)$ and $x = x^*(\theta)$.

(b) If $[a,b] \subset [\underline{\theta},\theta_0]$ is a maximal interval where x^* is constant, then

$$\int_a^b f_x(x,\theta)d\theta + \delta\int_{\hat{b}}^{\hat{a}\wedge\overline{\theta}} f_x(x,\theta)d\theta = 0$$

where $x = x^*(\hat{\theta})$ and if $b < \theta_1$, then $\delta = 0$ and if $\theta_1 \leq a \leq \theta_0$, then $\delta = 1$ and, in the integral, $\hat{a}$ and $\hat{b}$ are defined by the equality $v_x(x,\hat{\theta}) = v_x(x,\theta)$, where $\theta = a$ or b, respectively.

Proof: See the appendix.

The hazard rate is by definition:

$$M(\theta) = \frac{P(\theta)-1}{p(\theta)}.$$

In the following corollary assume that x^* crosses x_0 in a continuously differentiable way. For instance, this will be the case when the hazard rate is continuous at the crossing point. However, as it is shown in the examples 4.1 and 4.3, this condition may not be true. Besides, even under monotone hazard rate condition (MHRC), i.e., $M(\cdot)$ is increasing, the relaxed solution is not monotone.

Corollary (*The geometry of the curves*) Under the same assumptions of theorem 3.5, if x^* crosses x_0, then x_0, x_1, x^* and the first best solution cross at the same point and x^* is in between x_0 and x_1. Moreover, x_1 is above (below) the first best solution when it is above (below) x_0 and, under MHRC, the relaxed solution is U-shaped.

Figure 3

Remark 1. $\frac{\hat{v_{x\theta}}}{v_{x\theta}}$ is the Lagrange multiplier of marginal rate of substitution identity and we can rewrite the condition of theorem 3.5 (a) as

$$\frac{f_x(x,\theta)}{f_x(x,\hat{\theta})} = \delta \frac{v_{x\theta}(x,\theta)}{v_{x\theta}(x,\hat{\theta})}$$

with the following meaning: the rate of the virtual surplus between type θ and $\hat{\theta}$ is equal to the rate of the marginal rent between these two types.

For an illustration of the distortion effect in the case of no SCP, suppose that the principal's utility function does not depend on the agent's type. Omitting the argument of the functions and putting a hat over the function when it is evaluated at $\hat{\theta}$ and nothing when it is evaluated at θ, the first order condition given by theorem 3.5 when discrete pooling occurs is

$$\frac{u_x + v_x}{v_{x\theta}} + M = \frac{u_x + \hat{v_x}}{\hat{v_{x\theta}}} + \hat{M}.$$

Remark 2. The part (b) is the analogous ironing principle (see Mussa and Rosen (1978)). However, in our case the ironing principle may be disconnected.

Remark 3. **Rent extraction versus distortion**

As we observed in section 2, if v_θ changes its sign, then we have to consider a Lagrange multiplier for the (IR) constraint. An alternative way to deal with this problem is to define $\tilde{v}(x,\theta) = v(x,\theta) + K\theta$ where $K > 0$ is such that $\tilde{v}_\theta > 0$ and replace v by $\tilde{v}$. In order to have an equivalent problem, we have to assume now that the reservation utility of type θ is $K\theta$. This kind of situation has been studied in the literature (see Maggi and Rodríguez-Clare (1995) or Jullien (1997) for a complete treatment) and one could apply the same method to treat this problem here.

For instance, suppose that $v_{\theta\theta} < 0$ and that the curve in the plane (θ, x) defined by $v_\theta = 0$ is in the region where $v_{x\theta} > 0$. Thus this curve is increasing and it separates the plane into two regions: above this curve $v_\theta > 0$ and below it $v_\theta < 0$. Moreover, assume that the optimal decision crosses this curve just one time at $\theta_0 \in \Theta$. Fixing θ_0 and proceeding in a similar way, we will end up with the same theorems except that the objective functional will change to

$$f(x,\theta) = u(x,\theta) + v(x,\theta) + M(\theta)v_\theta(x,\theta)$$

where

$$M(\theta) = \begin{cases} \frac{P(\theta)-1}{p(\theta)} & \text{if } \theta \in [\theta_0, \overline{\theta}] \\ \frac{P(\theta)}{p(\theta)} & \text{if } \theta \in [\underline{\theta}, \theta_0] \end{cases}$$

In this case, the type θ_0 is the only one to have zero rent. The economic intuition is the same of "countervailing incentives" of Lewis and Sappington (1989) (see also Maggi and Rodríguez-Clare (1995)). The difference is that in our case it comes from the no SCP and in their case it is based on the type dependence of the agent's outside opportunities.

At last, if the optimal decision crosses the curve $v_\theta = 0$ on an interval (θ_0, θ_1), then along this interval $f(x,\theta) = u(x,\theta) + v(x,\theta)$ and the IR constraints are binding on (θ_0, θ_1).

Let us consider a particular case of the assumption A2.

Theorem 3.6 If the relaxed solution x_1 is U-shaped and $x_1(\underline{\theta}) \geq x_1(\overline{\theta})$, then the optimal solution x^* for (P) is:

$$x^*(\theta) = \begin{cases} x_1(\theta), & \text{if} \quad \theta < \theta_1 \\ x^u(\theta), & \text{it} \quad \theta \geq \theta_1 \end{cases}$$

where x^u is characterized by theorem 3.5 (a) and θ_1 is such that $x^u(\theta_1) = x^u(\overline{\theta})$.

Proof: See the appendix.

The new feature of the solution that appears in theorem 3.5 is the possibility of discrete pooling or pooling of large scale, i.e., in the optimal solution some isolated types can choose the same level of the contract. In the literature there exist just two types of equilibria: separating or continuous pooling equilibrium. In the former the agent's type is known *ex-post* by the principal and in the last the principal knows a range of types where the agent is. When the SCP does not hold, one can have discrete pooling equilibria besides separating and continuous pooling. In this case the principal does not know the true type between two types or between two ranges of types. Therefore the optimal solution can have these three characteristics: separating and continuous or discrete pooling.

Under SCP, the pooling interval of the optimal contract is characterized by the marginal welfare of the principal to be zero along this interval. Theorem 3.5 shows that this property is no longer valid when there exists discrete pooling.

The SCP obligates the principal only to check the upstream or downstream types in the case of CS_+ or CS_-, respectively (i.e., to check the second order condition or the monotonicity). However, the no SCP case obligates the principal to check also the marginal rate of substitution identity (the cross-stream condition). Therefore the IC constraint is less restrictive in the former case than in last one and, thus, the rent extraction is less powerful when there is no SCP.

4. EXAMPLES

4.1 Corporate Finance

Returning to the example in the introduction, define:

$$u(x,\theta) = x$$

$$v(x,\theta \mid T) = \begin{cases} \dfrac{-x^2}{(1-\theta)^2}\,, & \text{if} \quad T = T_1 \\ -\dfrac{x^2}{\theta^2}\,, & \text{if} \quad T = T_2 \end{cases}$$

In what follows we characterize the first and the second best solutions (with and without the technology choice information):

1) *First Best:* $\quad (x^{FB}, t^{FB}, T^{FB})$

$$T^{FB}(\theta) = \begin{cases} T_1\,, & \text{if} \quad \theta \in [0,1/2] \\ T_2\,, & \text{otherwise} \end{cases}$$

$$u_x(x^{FB}(\theta),\theta) + v_x(x^{FB}(\theta),\theta \mid T^{FB}(\theta)) = 0$$

$$t^{FB}(\theta) = -v\big(x^{FB}(\theta),\theta \mid T^{FB}(\theta)\big)$$

which implies that

$$x^{FB}(\theta) = \begin{cases} \dfrac{(1-\theta)^2}{2}\,, & \text{if} \quad \theta \in [0,1/2] \\ \dfrac{\theta^2}{2}\,, & \text{otherwise} \end{cases}$$

2) *Second Best with information:* As we explained in the introduction, the verifiability of the technology choice allows the principal to extract more rent from the agents with less distortion such that only a middle type will have zero rent. The intuition is straightforward: a generalist (bad type) will have zero rent and the specialist will have positive rent.

More striking, if the principal can commit to use lotteries on the technology choice, he can threat a group of middle risk averse agents and extract all the rent from them and improve even more his profit.

Formally, define the expected utility function of the agent with type θ on the bundle (α, x_1, x_2, t):

$$\tilde{V}(\alpha, x_1, x_2, t, \theta) = t + \alpha v(x_1, \theta \,|\, T_1) + (1-\alpha) v(x_2, \theta \,|\, T_2)$$

where $\alpha \in [0,1]$ is the probability that the principal will recommend the use of T_1 with production x_1 and the probability $1-\alpha$ with production x_2.

Now a contract is defined by $(\alpha, x_1, x_2, t) \colon \Theta \to [0,1] \times {\mathbb{R}_+}^2 \times \mathbb{R}$ (since the agent's utility function is linear on transfer, t represents the expected transfer of the lottery).

To avoid extra difficulties, assume the standard monotone hazard rate conditions:

MHRC: $M_1(\theta) = \frac{P(\theta)}{p(\theta)}$ and $M_2(\theta) = \frac{P(\theta)-1}{p(\theta)}$ are non-decreasing on θ.

Let us introduce some notation:

$$\tilde{v}(\alpha, x_1, x_2, \theta) = -[\alpha(1-\theta)^{-2} {x_1}^2 + (1-\alpha)\theta^{-2} {x_2}^2]$$

is the agent's expected cost function on the lottery, and

$$\tilde{u}(\alpha, x_1, x_2) = \alpha x_1 + (1-\alpha) x_2$$

is the principal's expected revenue. Thus, $\tilde{V}(\alpha, x_1, x_2, t, \theta) = t + \tilde{v}(\alpha, x_1, x_2, \theta)$.

Proceeding in an analogous manner, the relaxed functional is

$$\tilde{f}(\alpha, x_1, x_2, \theta) = \tilde{u}(\alpha, x_1, x_2) + \tilde{v}(\alpha, x_1, x_2, \theta) + \mathcal{V}^{(\alpha, x_1, x_2)}(\theta)$$

where $\mathcal{V}^{(\alpha, x_1, x_2)}(\theta) = \int_0^\theta \tilde{v}_\theta(\alpha(\tilde{\theta}), x_1(\tilde{\theta}), x_2(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta} + \mathcal{V}^{(\alpha, x_1, x_2)}(0)$ is the type θ agent's rent function. Finally, consider the following special cases of the certain lotteries:

$$f^i(x_i, \theta) = \tilde{u}(2-i, x_1, x_2) + \tilde{v}(2-i, x_1, x_2, \theta) + \mathcal{V}^{(2-i, x_1, x_2)}(\theta)$$

for $i = 1, 2$. Thus, after an integration by parts,

$$f^i(x_i, \theta) = u(x_i, \theta) + v(x_i, \theta \,|\, T_i) + M_i(\theta) v_\theta(x_i, \theta \,|\, T_i).$$

First, let us treat the case where the principal is prohibit to randomize[11], i.e., α is equal to 0 or 1. The game works as follows: the principal designs a reward

[11] If there is no commitment in using lotteries, i.e., after knowing the agent's type (ex-post) the principal can bias the lottery when it is profitable and this is not verifiable, then lotteries does not help ex-ante.

schedule based on the technology choice: $T_i \to (x_i, t_i)$, $i = 1, 2$. The agent accept or reject the schedule. If he accepts, he announces the verifiable technology he will use and the truthful type. Given the technology choice T_i, the principal's objective function will be $f^i(x_i, \theta)$.

Conditioning on the technology choice, the problem is a standard adverse selection with the SCP. Therefore x_1 (respectively x_2) is implementable if and only if it is non-increasing (respectively non-decreasing).

The principal provides two contracts (one for each technology choice) and decides where to shut down in each contract and take this into consideration to determine his objective function, i.e., when the agent is going to choose T_1 or T_2. Since $v_\theta(\,\cdot\,,\,\cdot\,|\,T_i)$ has constant sign, the rent function is monotone (given T_i). Therefore there exist a unique θ_i where the contract (x_i, t_i) is shut down, i.e., if $i = 1$ (respectively $i = 2$), then the rent function for all types $\theta > \theta_1$ (respectively $\theta < \theta_2$) is negative on the contract (x_i, t_i). We have the following cases:

1. $\theta_1 < \theta_2$. The IR constraint on the interval (θ_1, θ_2) is not satisfied in both contracts what is not possible.

2. $\theta_1 > \theta_2$. Denote $\mathcal{V}^i$ the rent function on the contract (x_i, t_i). Since $\mathcal{V}^1$ is decreasing, $\mathcal{V}^2$ is decreasing, $\mathcal{V}^1(\theta_1) = 0 < \mathcal{V}^2(\theta_1)$ and $\mathcal{V}^2(\theta_2) = 0 < \mathcal{V}^1(\theta_2)$, then there exist a unique $\theta_0 \in (\theta_2, \theta_2)$ such that

$$\mathcal{V}^1(\theta_0) = \int_{\theta_0}^{\theta_1} v_\theta(x_1(\theta), \theta\,|\,T_1)d\theta = \int_{\theta_2}^{\theta_0} v_\theta(x_2(\theta), \theta\,|\,T_2)d\theta = \mathcal{V}^2(\theta_0) > 0.$$

This implies that the principal can extract this rent by subtracting it from t_1 and t_2 and raise his profit.

What we have just proven is that a pair of contracts is feasible if only if case 2 is true and they are weakly dominated by one where $\theta_1 = \theta_2$. Now it is very easy to characterize the second best contract $(x_1^{SB}, x_2^{SB}, \alpha)$. The necessary (and sufficient) first order conditions are:

$$f^1{}_x(x_1^{SB}(\theta), \theta) = 0 \quad \text{if} \quad \theta \in [0, \theta_0]$$

$$f^2{}_x(x_2^{SB}(\theta), \theta) = 0 \quad \text{if} \quad \theta \in [\theta_0, 1]$$

and

$$\alpha = \begin{cases} 1 & \text{if} \quad \theta \in [0, \theta_0] \\ 0 & \text{if} \quad \theta \in [\theta_0, 1] \end{cases}$$

This means that θ_0 is determined by the intersection of the relaxed solutions and it is the only type that has zero rent. And it is optimal to make all types $\theta < \theta_0$ (respectively $\theta > \theta_0$) use technology T_1 (respectively T_2). Observe that MHRC implies that x_1^{SB} is decreasing and x_2^{SB} is increasing. Thus, this pair of contract are implementable for the second best problem. If this was not the case we would have to consider the "ironing principle".

Let us return to the case that the principal can commit to use lotteries. Taking the derivative of $\tilde{f}$ with respect to x_i, it is easy to see that non trivial lotteries[12] will be used on the intervals where the IR constraints are binding, i.e., where the rent function is null. For the rest of the interval the optimal contract is characterized by:

$$f^i{}_x(x_i, \theta) = 0$$

when $\alpha = 2 - i$.

Therefore we have to characterize the intervals where the rent function is null. However, if the rent function is constant on (θ_1, θ_2) along an implementable contract (α, x_1, x_2), then

$$\tilde{v}_\theta(\alpha(\theta), x_1(\theta), x_2(\theta), \theta) = 0, \ \forall \theta \in (\theta_1, \theta_2).$$

This implies that (omitting the dependence of the contract on θ):

$$\alpha = \frac{\theta^{-3} x_2^2}{(1-\theta)^{-3} x_1^2 + \theta^{-3} x_2^2}$$

on the interval (θ_1, θ_2).

Plug this last equation in the objective function of the principal, on the interval (θ_1, θ_2), it is going to be

$$\Psi(x_1, x_2, \theta) = \frac{x_1}{\eta(\theta)^3 \left(\frac{x_1}{x_2}\right)^2 + 1} + \frac{x_2}{\eta(\theta)^{-3} \left(\frac{x_2}{x_1}\right)^2 + 1} - \frac{(x_1 x_2)^2}{\theta^3 x_1^2 + (1-\theta)^3 x_2^2}$$

where $\eta(\theta) = \frac{\theta}{1-\theta}$.

It is immediate to show that $\Psi(\cdot, \cdot, \theta)$ is a concave functional and that $\Psi(x_1, x_2, \theta) = \Psi(x_2, x_1, 1-\theta)$, for all (x_1, x_2, θ). Therefore if $(x_1^*(\theta), x_2^*(\theta))$ is the optimal for a given θ, then $x_1^*(\theta) = x_2^*(1-\theta)$, for all $\theta \in [0, 1]$.

The first order condition gives that $x_1^* = x_2^* = x^*$ and consequently

$$x^*(\theta) = \frac{1}{2}\left[\frac{1}{\eta(\theta)^3 + 1} + \frac{1}{\eta(\theta)^{-3} + 1}\right][\theta^3 + (1-\theta)^3]$$

and

$$\alpha^*(\theta) = \frac{1}{\eta(\theta)^3 + 1}$$

for all $\theta \in (\theta_1, \theta_2)$.

The first order condition that determines θ_i $(i = 1, 2)$ is

$$u(x_i(\theta), \theta_i) + v(x_i(\theta_i), \theta_i) + M_i(\theta_i) v_\theta(x_i(\theta_i), \theta_i) = \tilde{u}(x^*(\theta_i), \theta_i) + \tilde{v}(x^*(\theta_i), \theta_i)$$

[12] A trivial lottery is the one that α is 0 or 1.

Therefore θ_i is determined by the continuity of the principal's ex-post profit when pasting from randomization to non-randomization. Moreover, the decision where to randomize or not depends on whether the right hand side of the last equation is greater than the left hand side or not, respectively.

In general, many intervals can appear in the optimal contract and it is not easy to determine them. However, the symmetric case is very simple to deal with. If the distribution is symmetric with respect to $1/2$, then it is easy to prove that those intervals are determined by intersection of x^* and x^{SB} and randomization occurs if and only if x^* is above x^{SB}.

We have to check that this randomized relaxed contract is incentive compatible in order to conclude that it is the second best solution. Let us answer this question in the symmetric distribution case. For simplicity, assume that there is just one interval where randomization occurs: (θ_1, θ_2) (thus, $\theta_2 = 1-\theta_1$). Formally, we have to check that

$$\tilde{\Phi}_{x^{SB}}(\theta, \hat{\theta}) \geq 0, \ \forall \theta, \hat{\theta} \in [0,1]$$

on the randomized relaxed solution, where $\tilde{\Phi}_x$ corresponds to the function $\tilde{v}$. There are several cases to consider:

1. $\theta, \hat{\theta} \in [0, \theta_1]$ or $\theta, \hat{\theta} \in [\theta_2, 1]$: The incentive compatibility is an immediate consequence of monotonicity of x_1^{SB} and x_2^{SB} and the SCP.

2. $\theta, \hat{\theta} \in [\theta_1, \theta_2]$: Note that

$$\tilde{v}_\theta(x^*(\hat{\theta}), \theta) = -2[\alpha^*(\hat{\theta})(1-\theta)^{-3} - (1-\alpha^*(\hat{\theta}))\theta^{-3}]x^*(\hat{\theta})^2$$

and, by the definition of α^* and $\tilde{v}_\theta(x^*(\theta), \theta) = 0$, we have that $\theta > \hat{\theta}$ if and only if $\tilde{v}_\theta(x^*(\hat{\theta}), \theta) < 0$. Thus

$$\begin{aligned}\tilde{\Phi}_{x^{SB}}(\theta, \hat{\theta}) &= \int_\theta^{\hat{\theta}} \tilde{v}_\theta(x(\hat{\theta}), \tilde{\theta}) d\tilde{\theta} - \int_\theta^{\hat{\theta}} \tilde{v}_\theta(x(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta} \\ &= \int_\theta^{\hat{\theta}} \tilde{v}_\theta(x(\hat{\theta}), \tilde{\theta}) d\tilde{\theta} \geq 0\end{aligned}$$

This implies that the IC constraint is satisfied.

The remaining cases reduce to these two.

If $p \equiv 1$ and randomization is not possible, then $\theta_0 = \frac{1}{2}$ and

$$x^{SB}(\theta) = \begin{cases} \dfrac{(1-\theta)^3}{2(1+\theta)}, & \text{if} \quad \theta \in [0, 1/2] \\ \dfrac{\theta^3}{2(2-\theta)}, & \text{otherwise} \end{cases}$$

When randomization is possible, the solution was described above.

If $p(\theta) = 2\theta$, for all $\theta \in \Theta$, then $\theta_0^{4/3} + \theta_0 = 1$ and

$$x^{SB}(\theta) = \begin{cases} \dfrac{(1-\theta)^3}{2}, & \text{if} \quad \theta \in [0, \theta_0] \\ \dfrac{\theta^4}{2}, & \text{otherwise} \end{cases}$$

3) *Second Best without information:* (x^{TB}, t^{TB})

Define now

$$\tilde{v}(x,\theta) = \begin{cases} \dfrac{-x^2}{(1-\theta)^2}, & \text{if} \quad \theta \in [0, 1/2] \\ \dfrac{-x^2}{\theta^2}, & \text{otherwise} \end{cases}$$

Because the principal can not monitor the technology choice, he will face an agent with utility function $\tilde{v}$ (and not v). Therefore we can use theorem 3.5 in order to characterize the second best solution. Here, x_0 is defined by $\theta = 1/2$.

Since $\tilde{v}_\theta(x,\theta) \underset{>}{<}$ iff $\theta \underset{<}{>} \frac{1}{2}$, the IR constraint will be binding at $\frac{1}{2}$ on the optimal contract. Thus, define

$$f(x,\theta) = u(x,\theta) + \tilde{v}(x,\theta) + M(\theta)\tilde{v}_\theta(x,\theta)$$

where

$$M(\theta) = \begin{cases} M_1(\theta) & \text{if} \quad \theta \in [0, 1/2] \\ M_2(\theta) & \text{if} \quad \theta \in [1/2, 1] \end{cases}$$

Observe that $\tilde{v}_x(x,\theta) = \tilde{v}_x(x,\hat{\theta})$ if and only if $\hat{\theta} = 1 - \theta$ and $\tilde{v}_{x\theta}(x,\theta) = -\tilde{v}_{x\theta}(x, 1-\theta)$. Thus, theorem 3.5 gives

$$f_x(x^{TB}(\theta), \theta) + f_x(x^{TB}(\theta), 1-\theta) = 0, \quad \theta \in [0, 1/2].$$

The IC constraints define a convex set and the objective function is concave. Therefore the use of lotteries on production does not improve the principal's welfare (what depends strongly on the symmetry of this example).

If the distribution is symmetric with respect to $1/2$, then it is straightforward to check that the second best problem without information is equivalent to the second best one with information when randomization is forbidden. Therefore the profit loss caused by non verifiability of information is equal to the one caused by the lack of of commitment in using lotteries when information is available.

In particular, if $p \equiv 1$, (x^{SB}, t^{SB}) is the second best solution without information (x^{SB} is symmetric with respect to $\frac{1}{2}$).

If $p(\theta) = 2\theta$, for all $\theta \in \Theta$, then

$$x^{TB}(\theta) = \begin{cases} \dfrac{(1-\theta)^4}{2-\theta}, & \text{if} \quad \theta \in [0, 1/2] \\ \dfrac{\theta^4}{1+\theta}, & \text{otherwise} \end{cases}$$

and

$$t^{TB}(\theta) = \int_{1/2}^{\theta} \tilde{v}_\theta(x^{TB}(\tilde{\theta}), \tilde{\theta})\, d\tilde{\theta} - \tilde{v}(x^{TB}(\theta), \theta).$$

The basic intuitions behind this example are the following. First, when we moved from the second best problem with information to the one without, we moved from a bidimensional problem in (T, θ) to an one dimensional problem in θ making T a function of θ. Finally, the incapacity to commit to the announcement of the technology makes the principal to equalize the marginal utilities of the pooling types in order to guarantee truth telling. This is exactly what explains the "countervailing" incentives in our case.

One may claim that this example is very particular in the sense that it is symmetric. However, we can make a perturbation of this model and obtain the same qualitative results. One way to do that is to consider (different) sunk costs for each technology.

The second best problem without information was inspired in Chassagnon and Chiappori (1995). However, in that paper (if we consider the continuous version of their model), the cross derivative of the agent's utility function does not change the sign, it only changes its magnitude, i.e., indifference curves of the agent has a kink. This is enough to produce multiple crossing of the indifference curves of two distinct types, but not to destroy the second order approach.

Include here Figure 4

Figure 4

4.2 Nonlinear Pricing

This example follows the same setup of Maskin and Riley (1984) (see also Mussa and Rosen (1978)). A monopolist produces a single product at a cost of cx^2 for x units. A buyer of type $\theta \in [0, 1]$ has preferences represented by the utility function

$$V(x, t, \theta) = \int_0^x \pi(\tilde{x}, \theta)\, d\tilde{x} - t$$

where x is the number of units purchased from the seller and t is the price paid for x. The function $\pi(\,\cdot\,,\theta)$ is the inverse demand function of the group of consumers with taste characterized by θ. The monopolist does not observe the type, but knows $P(\,\cdot\,)$, the distribution of type, with density function $p(\,\cdot\,)$.

We assume that the inverse demand has the following form:

$$\pi(x,\theta) = \theta - 2\alpha(\theta)x$$

where, α is thrice continuously differentiable, $\alpha(0)=0$, $\dot{\alpha}>0$, $\ddot{\alpha}>0$.

The assumption $\ddot{\alpha}>0$ implies that $\dfrac{\alpha(\theta)}{\theta}$ is increasing in θ, since

$$\frac{d}{d\theta}\left(\frac{\alpha(\theta)}{\theta}\right) = \frac{\dot{\alpha}(\theta)\theta - \alpha(\theta)}{\theta^2} > 0$$

if and only if

$$\dot{\alpha}(\theta) > \alpha(\theta)/\theta\,, \quad \forall\,\theta \in (0,1]$$

and this last inequality is true because α is convex. And $\theta/2\alpha(\theta)$ is the market size of the type θ demand (i.e., the number of units bought at price zero). Therefore what we are assuming is that the market size decrease with θ and θ is the supreme of prices for which there exist a positive demand. This assumption will imply no SCP and it is in contrast with the monotonicity assumption of $\pi(x,\,\cdot\,)$ in Maskin and Riley (1984).[13]

Define

$$\begin{aligned}
v(x,\theta) &= \int_0^x \pi(\tilde{\pi},\theta)\,d\tilde{x} = (\theta - \alpha(\theta)x)x \\
u(x,\theta) &= -cx^2 \\
M(\theta) &= \frac{P(\theta)-1}{p(\theta)}\,, \quad \theta \in [0,1] \\
f(x,\theta) &= u(x,\theta) + v(x,\theta) + M(\theta)v_\theta(x,\theta).
\end{aligned}$$

Therefore,

$$v_{x\theta}(x,\theta) = 1 - 2\dot{\alpha}(\theta)x \lesseqqgtr 0 \quad \text{iff} \quad x \gtreqqless x_0(\theta) = \frac{1}{2\dot{\alpha}(\theta)}\,,$$

for all $\theta \in [0,1]$.

Since $\ddot{\alpha}>0$, x_0 is decreasing and

$$\frac{1}{2\dot{\alpha}(\theta)} < \frac{\theta}{2\alpha(\theta)} < \frac{\theta}{\alpha(\theta)}\,, \quad \forall\,\theta \in [0,1]$$

[13] This is equivalent to assume that the marginal utility of consumption increases for low types and decreases for high types.

what implies that the SCP fails to hold here. (Observe that $v_x(x,\theta) \geq 0$ iff $x \leq \frac{\theta}{\alpha(\theta)}$.) Moreover, observe that $v_x(x,\theta) = v_x(x,\hat{\theta})$ if and only if $\pi(x,\theta) = \pi(x,\hat{\theta})$, i.e., two types are pooling in the same contract if they have the same marginal valuation for the good.

The relaxed solution is given by

$$x_1(\theta) = \frac{1}{2}\left[\frac{\theta + M(\theta)}{c + \alpha(\theta) + \dot{\alpha}(\theta)M(\theta)}\right]^+$$

where $[x]^+ = \max\{x, 0\}$.

Assume that:

a1. $\dot{M}(\theta) > 0$, for all $\theta \in [0,1]$.

a2. $c + \dot{\alpha}(\theta)M(\theta) + \alpha(\theta) > 0$, for all $\theta \in [0,1]$.

Assumption **a1** is the well known monotone hazard rate condition and assumption **a2** holds for c large enough. Assumption **a2** implies that $f(\,\cdot\,,\theta)$ is a concave function, for all $\theta \in [0,1]$.

If x_1 does not cross x_0, then the second best solution is going to be x_1. Observe that

$$x_1(\theta) \lesseqqgtr \frac{1}{2\dot{\alpha}(\theta)} \quad \text{iff} \quad c \gtreqqless \theta\dot{\alpha}(\theta) - \alpha(\theta).$$

Therefore, we guarantee that x_1 crosses x_0 if we assume

a3. $c < \dot{\alpha}(1) - \alpha(1)$

because $\theta\dot{\alpha}(\theta) - \alpha(\theta)$ is increasing in θ (since $\ddot{\alpha} > 0$).

Under **a1**, **a2** and **a3**, theorem 3.5 and 3.6 can be applied in order to characterize the second best solution. Observe that x_1 will be bell-shaped and that $v_\theta(x,\theta) = (1 - \dot{\alpha}(\theta)x)x \geq 0$ iff $x \leq 1/\dot{\alpha}(\theta)$. Thus, the second best solution is bell-shaped and it is such that the lowest type 0 has zero rent.

Consider the particular case that satisfies **a1**-**a3**:

$$p \equiv 1, \quad c = \frac{1}{2}$$

$$\alpha(\theta) = \theta^2$$

In this case $v(x,\theta) = \tilde{v}(\theta x)$, where $\tilde{v}(y) = y - y^2$.

Therefore the relaxed solution is

$$x_1(\theta) = \left[\frac{2\theta - 1}{1 - 4\theta + 6\theta^2}\right]^+$$

and

$$x_0(\theta) = \frac{1}{4\theta}$$

Using theorems 3.5 and 3.6, the bell-shaped part of the solution is given by

$$x^*(\theta) = \frac{3\theta + \sqrt{3\theta^2 - 1}}{2(6\theta^2 + 1)}$$

Figure 5

4.3 Regulation Problem

We are going to present a simple model of regulation of a firm like in Laffont and Tirole (1993), chapter 1. Suppose that there exist a project with social value of S that can be implemented by a firm that has the following cost structure:

$$\begin{cases} C = C_1 + C_2 \\ C_1 = \theta_1 - e_1 \\ C_2 = \theta_2 - e_2 \end{cases}$$

where the cost C is observable (C_1 and C_2 are not), θ_1 and θ_2 are the cost parameters known only to the firm, e_1 and e_2 are unobservable actions of the firm representing the efforts to reduce the subcosts C_1 and C_2, respectively.

The non-monetary disutility of effort is given by $\psi(e_1, e_2)$. We assume that it is thrice differentiable, the first and second derivatives of ψ are positive, i.e., $D\psi = (\psi_1, \psi_2) > 0$ and

$$D^2\psi = \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{pmatrix}$$

is a positive definite matrix (where the subindex represents the partial derivatives).

The firm's problem with characteristic vector (θ_1, θ_2) and total cost C is to minimize the disutility of effort:

$$\begin{aligned} &\max_{C_1, C_2 \geq 0} \ \psi(\theta_1 - C_1, \theta_2 - C_2) \\ &\text{s.t.} \\ &C_1 + C_2 = C \end{aligned}$$

Assuming an interior solution, the first order condition is

$$\psi_1(\theta_1 - C_1, \theta_2 - C_2) = \psi_2(\theta_1 - C_1, \theta_2 - C_2).$$

Following Lewis and Sappington (1989), we can introduce countervailing incentives into the model by allowing the firm's subcost parameter of production θ_2 to be a function of the subcost parameter θ_1: $\theta_2 = \eta(\theta_1)$ twice differentiable. This countervailing incentive will be associated to no SCP in two cases:[14]

[14] In the other two cases, the SCP is valid.

(a) substitute efforts and negative correlation between the subcosts: $\psi_{12} > 0$ and $\dot{\eta} < 0$.

(b) complement efforts and positive correlation between the subcosts: $\psi_{12} < 0$ and $\dot{\eta} > 0$.

Let us consider the case (a). This means that there is countervailing incentive in the effort allocation for subcost reduction since these activities are substitute and the cost parameters move in opposite directions. For instance, a family of examples are given by η a convex decreasing function and

$$\psi(e) = A^T e + e^T B e + \xi(e)$$

where $e = [e_1, e_2]$, $A > 0$, B is a definite positive matrix and $\xi(e) = \xi_1 e_1^3 + \xi_{12} e_1^2 e_2 + \xi_{21} e_2^2 e_1 + \xi_2 e_2^3$ is the third order terms with $\xi_i \geq 0$, $\xi_{ij} \geq 0$, $i, j = 1, 2$.

The consequence is that the SCP may not hold. Indeed, define the firm's surplus as

$$V = t - \psi(\theta_1 - C_1, \theta_2 - C_2)$$

where t is the net money transfer from the regulator to the firm and $\{C_i = C_i(\theta_1, C)\}_{1,2}$ are the optimal decisions of the firm given its first subcost parameter and observable aggregate cost. If

$$v(C, \theta_1) = -\psi(\theta_1 - C_1, \theta_2 - C_2)$$

then, by the Envelope Theorem,[15]

$$v_C = \psi_1$$

and

$$v_{\theta_1} = -(\psi_1 + \psi_2 \dot{\eta}) = -\psi_1(1 + \dot{\eta})$$

Thus,

$$v_{\theta_1} \gtreqless 0 \;\Leftrightarrow\; 1 + \dot{\eta} \lesseqgtr 0$$

and, consequently, if η is a concave decreasing function, the IR constraint will be binding at the parameter θ_1^0 such that $1 + \dot{\eta} = 0$. From now on assume that this is the case.

Finally, the cross derivative of v is given by:

$$v_{C\theta_1} = \psi_{11} + \psi_{12}\dot{\eta} + (\psi_{11} - \psi_{12}) \frac{\partial C_2}{\partial \theta_1}$$

and, since $\psi_{12}\dot{\eta} < 0$, the sign of $v_{C\theta_1}$ can actually change (see the specific example below).

[15] To simplify the notation we will omit the arguments of the functions.

The social welfare function is

$$W = S - (1+\lambda)(t+C) + \mathcal{V} = S - (1+\lambda)(-v+C) - \lambda\mathcal{V}$$

where

$$\mathcal{V}(\theta_1) = \int_{\theta_1^0}^{\theta_1} v_\theta(C(\tilde{\theta},\tilde{\theta})d\tilde{\theta}$$

is the firm's rent function and $\lambda > 0$ is the "shadow price of public funds" (see Laffont and Tirole (1993) for the details).

The relaxed functional is

$$\mathcal{F} = \{S - (1+\lambda)(-v+C) + \lambda M v_{\theta_1}\}p$$

where

$$M = \begin{cases} P/p & \text{if} \quad \underline{\theta} \le \theta \le \theta_1^0 \\ (P-1)/p & \text{if} \quad \theta_1^0 < \theta \le \overline{\theta} \end{cases}$$

is the hazard rate (that depends on the sign of v_{θ_1}).

The first order condition of the relaxed problem

$$\psi_1 = 1 + \frac{\lambda}{1+\lambda} M v_{C\theta_1}$$

By theorem 3.5, if θ_1 and $\hat{\theta_1}$ are pooling on the same aggregate cost, then

$$\frac{\mathcal{F}_C}{v_{C\theta_1}} = \frac{\hat{\mathcal{F}}_C}{\hat{v}_{C\theta_1}} \qquad \text{and} \qquad v_C = \hat{v}_C$$

where the hat symbol means that the functions are calculated at the $\hat{\theta_1}$ allocation. And, after some trivial manipulations, we have

$$\psi_1 = 1 + \frac{\lambda}{1+\lambda}(M - \widehat{M})\{v_{C\theta_1}^{-1} - \widehat{v_{C\theta_1}}^{-1}\}^{-1}$$

and

$$\psi_1 = \widehat{\psi}_1$$

Comparing the first order condition of the relaxed program with the second-best one, we see that the incentive correction term depends on the marginal rent of the type (for the former) and depends on an "average" of the marginal rent of the types that are pooling (for the last).

Let us take a particular example (the symmetric case): The type set is $\Theta = [0,1]$ with the uniform distribution and

$$\psi(e_1, e_2) = (e_1 + e_2)^2$$

$$\theta_2 = \eta(\theta_1) = 1 - \theta_1^2$$

Thus the firm's minimization effort program has the following solution: $C_1 = \theta_1$ and $C_2 = C - \theta_1$ (since $C_2 \leq 1 - \theta_1^2$ for $C \leq 1$).

We have the curve that separates the regions where the sign of $v_{C\theta_1}$ changes is $\theta_1 \equiv 1/2$ and $v_C = \widehat{v_C}$ if and only if $\widehat{\theta}_1 = 1 - \theta_1$. Solving the equations of the first order conditions of the second-best program we get

$$C^*(\theta_1) = \begin{cases} \frac{1}{2} + \left(1 - \frac{\lambda}{1+\lambda}\right)\theta_1 - \left(1 - \frac{2\lambda}{1+\lambda}\right)\theta_1^2 & \text{if} \quad 0 \leq \theta_1 \leq 1/2 \\ \frac{1}{2} + \left(1 - \frac{\lambda}{1+\lambda}\right)(1-\theta_1) - \left(1 - \frac{2\lambda}{1+\lambda}\right)(1-\theta_1)^2 & \text{if} \quad 1/2 \leq \theta_1 \leq 1 \end{cases}$$

This is also the relaxed of the regulator's program when he can contract on the subcosts, i.e., in this case the non-observability of the subcosts does not distort the regulatory solution. However, if the distribution is not symmetric, then the distortion will exist. For instance, when the distribution is given by $p(\theta) = 2\theta$, the relaxed solution is:

$$C^1(\theta_1) = \begin{cases} \frac{1}{2} + \left(1 - \frac{\lambda}{2(1+\lambda)}\right)\theta_1 - \left(1 - \frac{\lambda}{1+\lambda}\right)\theta_1^2 & \text{if} \quad 0 \leq \theta_1 \leq 1/2 \\ \frac{1}{2} + \left(1 - \frac{\lambda}{2(1+\lambda)}\right)\theta_1 - \left(1 - \frac{\lambda}{1+\lambda}\right)\theta_1^2 + \frac{\lambda}{2(1+\lambda)}\frac{1-2\theta_1}{\theta_1} & \text{if} \quad 1/2 \leq \theta_1 \leq 1 \end{cases}$$

and the second best solution is:

$$C^*(\theta_1) = \begin{cases} \frac{1}{2} + \left(1 - \frac{\lambda}{2(1+\lambda)}\right)\theta_1 - \left(1 - \frac{\lambda}{1+\lambda}\right)\theta_1^2 + \frac{\lambda}{2(1+\lambda)}\frac{\theta_1}{1-\theta_1} & \text{if} \quad 0 \leq \theta_1 \leq 1/2 \\ \frac{1}{2} + \left(1 - \frac{\lambda}{2(1+\lambda)}\right)(1-\theta_1) - \left(1 - \frac{\lambda}{1+\lambda}\right)(1-\theta_1)^2 + \frac{\lambda}{2(1+\lambda)}\frac{1-\theta_1}{\theta_1} & \text{if} \quad 1/2 \leq \theta_1 \leq 1 \end{cases}$$

Figure 6

The economic interpretation of the result is immediate: the subcost reductions are substitute activities and the subcost structure presents a countervailing property: the extreme types correspond to the specialist in each activity and the middle type is a "bad" generalist in both activities. Thus, the optimal contract is such that two different specialists are choosing the same aggregate cost in equilibrium (and each one is going to cut the subcost that is more inefficient); the middle type ($\theta_1 = 1/2 = \theta_2$) is the only one that has zero rent.

4.4 Labor Market

Cavallo, Heckman and Hsee (1998) presents evidence on the GED[16] as a mixed signal of cognitive and non-cognitive abilities. Comparing the GED recipients and other dropouts, there is not wage differential between them.

[16] GED is an exam taken by American high school dropouts to certify their equivalence with high school graduates.

We present a very simple model where workers have a verifiable signal s (schooling) that is the aggregation of two unknown personality characteristics: θ ("cognitive characteristic") and η ("non-cognitive characteristic"):

$$s = \theta + \eta$$

where $\theta \in [\underline{\theta}, \overline{\theta}]$ with cumulative distribution P and density function p. Given the schooling s, we can also determine the distribution of η.

If the firm hires the worker with a profile of characteristic (θ, η), then this worker will produce an output x following the technology:

$$x = \theta e + \alpha\eta$$

where e is the effort of the worker (unknown to the firm) and α is the shadow price of the non-cognitive ability for the firm. The firm maximizes its profit

$$U = \pi x - t$$

where t is the salary paid for the worker and π is the price of x. This means that the firm uses the cognitive and non-cognitive abilities of a worker, but in a different manner they are presented in the signal: θ is productive with effort while η is constant.

The firm will use the signal, salary and output as a mechanism device to extract the worker's rent. However, since effort is costly, for a given signal there exists a conflict of interest between the firm and the worker. Moreover, the schooling s is a mixed signal of θ and η and the firm can not infer the correct abilities of the worker *ex-ante*.

The worker's disutility of effort is given by $\psi \colon \Re \to \Re$. For simplicity, we assume that ψ is quadratic, i.e., $\psi(e) = e^2$. For a given schooling s, the necessary effort for a worker with characteristics $(\theta, s - \theta)$ to produce x is

$$e = \alpha - \theta^{-1} y$$

where $y = \alpha s - x$.[17] Thus, we can express the type θ worker's quasi-linear utility function in terms of the verifiable variables (x, t) and the characteristic θ: $V = t - v(x, \theta)$, where

$$v(x, \theta) = -[\alpha - \theta^{-1} y]^2$$

If we take the cross derivative of v with respect to x and θ we get

$$v_{x\theta}(x, \theta) = 2\theta^{-2}[e - \theta^{-1} y]$$

[17] Observe that $e \geq 0$ if and only if $x \leq \alpha s$.

This means that the SCP does not hold: the curve x_0 is given by

$$x_0(\theta) = \alpha s - \frac{\alpha}{2}\theta$$

Thus the discrete pooling equilibrium may happen. The discrete pooling condition is given by $v_x(x, \theta) = v_x(x, \hat{\theta})$, where $v_x(x, \theta) = -2\theta^{-1}e$. In this case

$$\hat{\theta} = \phi(x, \theta) = (\alpha y^{-1} - \theta^{-1})^{-1}$$

Since $v_\theta = -2\theta^{-2}ey$, the marginal rent of the worker is negative if and only if $x \leq \alpha s$. Then, since $v_{x^2\theta}(x, \theta) = 4\theta^{-3}$ is positive and $v_{x\theta^2}(x, \theta) = 2\theta^{-3}[6\theta^{-1}y - 2\alpha]$ is negative, an implementable contract that cross the curve x_0 is U-shaped.

Let us consider a particular example: $\Theta = [1, 2]$ with the uniform distribution.

It is easy to see that the relaxed solution crosses x_0 at $\theta^* = \alpha/\pi$. Thus, if $\alpha/\pi \in (1, 2)$, the relaxed solution is given by:

$$x_1(\theta) = \min \{\alpha s, \alpha s + \frac{2\alpha - \pi\theta^2}{2(\theta - 2)}\theta\}$$

Using theorem 3.5 we can calculate the U-shaped part of the second best contract: it is going to be one of the roots of the following third degree polynomial:[18]

$$-\frac{\pi\alpha^2}{4} + \alpha\theta^{-1}\big(\frac{\pi}{2} + \alpha\theta^{-1}\big)y - \theta^{-2}\big(\frac{\pi}{2} + 2\alpha\theta^{-1}\big)y^2 + \theta^{-4}y^3 = 0$$

Therefore, the optimal contract presents the following properties:

Figure 7

• discrete pooling: two different workers with different profile of characteristics for a given s is choosing the same contract. This property captures the idea that a worker with high cognitive ability and low non-cognitive ability can not be distinguished from a worker with low cognitive ability and high non-cognitive ability in equilibrium.

• the firm offers the same contract for an interval of high θ workers to extract all their rent. These types provide the highest output.

• The worker θ^* (a middle type) provides the lowest output and has the highest rent.

• The rent extraction and distortion trade off takes into consideration the new conditions for feasibility.

[18] The monotonicity condition eliminates the other roots.

5. CONCLUSIONS AND EXTENSIONS

In this paper we studied a generalization of the SCP. The characterization of the IC constraint depends on a new condition: the marginal rate of substitution identity. When the SCP does not hold a new type of equilibrium appears: the discrete pooling equilibrium. Four examples illustrated the pooling equilibrium: a principal-agent problem with simultaneous adverse selection and moral hazard, the nonlinear pricing problem, subcost observation in a regulation problem and mixed signal in a labor market model. In all examples multidimensional characteristics and countervailing incentives are presented.

Some extensions about the cases that were not covered follow:

(1) More general agent utility function: Assume that the agent's utility function can be any C^2 function with the same assumptions of section 2 except the quasi-linearity one. Given a feasible contract (x,t) continuously differentiable piecewise denoted $V(x(\hat{\theta}),t(\hat{\theta}),\theta)$ by $V(\hat{\theta},\theta)$, for simplicity. Then

$$V(\theta,\theta)=\int_{\underline{\theta}}^{\theta} V_\theta(\tilde{\theta},\tilde{\theta})d\tilde{\theta}+V(\underline{\theta},\underline{\theta}) \text{ and}$$

$$V(\hat{\theta},\theta)=\int_{\theta}^{\hat{\theta}}[V_x(\tilde{\theta},\theta)\dot{x}(\tilde{\theta})+V_t(\tilde{\theta},\theta)\dot{t}(\tilde{\theta})]d\tilde{\theta}+V(\theta,\theta).$$

By the first order condition of (IC),

$$\dot{t}(\theta)=-\frac{V_x(\theta,\theta)}{V_t(\theta,\theta)}\dot{x}(\theta).$$

Then

$$V(\theta,\theta)-V(\hat{\theta},\theta)=\int_{\hat{\theta}}^{\theta}\left[V_x(\tilde{\theta},\theta)-V_t(\tilde{\theta},\theta)\frac{V_x(\tilde{\theta},\tilde{\theta})}{V_t(\tilde{\theta},\tilde{\theta})}\right]\dot{x}(\tilde{\theta})d\tilde{\theta}.$$

If θ, $\hat{\theta}\in\Theta$ is such that $V(\theta,\theta)=V(\hat{\theta},\theta)$, then the derivative of the above expression with respect to $\hat{\theta}$ should be zero, i.e.,

$$\dot{x}(\hat{\theta})\left(V_x(\hat{\theta},\theta)-V_t(\hat{\theta},\theta)\frac{V_x(\hat{\theta},\hat{\theta})}{V_t(\hat{\theta},\hat{\theta})}\right)=0.$$

If $\dot{x}(\hat{\theta})\neq 0$, then

$$-\frac{V_x(\hat{\theta},\theta)}{V_t(\hat{\theta},\theta)}=-\frac{V_x(\hat{\theta},\hat{\theta})}{V_t(\hat{\theta},\hat{\theta})}.$$

This condition means that if the agent θ is indifferent between his bundle $(x(\theta),t(\theta))$ and the $\hat{\theta}$ bundle $(x(\hat{\theta}),t(\hat{\theta}))$ and the agent $\hat{\theta}$ can not locally misrepresent his type ($\dot{x}(\hat{\theta})\neq 0$), then agents θ and $\hat{\theta}$ should have the same marginal rate of substitution at the bundle $(x(\hat{\theta}),t(\hat{\theta}))$.

(2) Multidimensional decision: Assume that x is n-dimensional vector and θ is one-dimensional. A similar argument as above shows that

$$\dot{x}(\hat{\theta}) \cdot \left(-\frac{V_x(\hat{\theta},\theta)}{V_t(\hat{\theta},\theta)} \right) = \dot{x}(\hat{\theta}) \cdot \left(-\frac{V_x(\hat{\theta},\hat{\theta})}{V_t(\hat{\theta},\hat{\theta})} \right)$$

for all $\theta, \hat{\theta} \in \Theta$ such that $V(\theta,\theta) = V(\hat{\theta},\theta)$. However, observe that the dot in the above equation is the inner product of the respective vectors. The interpretation is that if the type θ is indifferent between his bundle and the type $\hat{\theta}$ bundle, then the covariation of the marginal increasing of type $\hat{\theta}$ in x with his marginal rate of substitution at the bundle $(x(\hat{\theta}), t(\hat{\theta}))$ should be equal to the covariation of the same marginal increasing with the marginal rate of substitution of type θ at the same bundle.

Observe also that the second order condition of the IC constraint is

$$\frac{d}{d\theta}\left(\frac{V_x(\theta,\theta)}{V_t(\theta,\theta)}\right) \cdot \dot{x}(\theta) \geq 0, \ \forall \theta \in \Theta.$$

(3) Discontinuous crossing when x_0 is constant: If we allows us to discontinuous crossing, the solution can be improved by a discontinuous contract. The example is the following[19]

$$\Theta = [0,1], \ p \equiv 1, \ c = 0$$
$$u(x,\theta) = (1-\theta)x - \frac{x^2}{2}$$
$$v(x,\theta) = \frac{\theta}{2}(x - x^2)$$

where we are considering quasi-linear utility functions for the principal and for the agent.

The relaxed solution is $x_1(\theta) = 1-\theta$ and $x_0(\theta) = 1/2$, for all $\theta \in \Theta$. If we only admit continuous crossing, the optimal solution will be $x^*(\theta) = 1/2$. However, if discontinuous crossing is possible, then the optimal solution will be

$$x^*(\theta) = \begin{cases} 1-\theta & \text{it} \quad 0 \leq \theta < \frac{1}{4} \\ \frac{3}{4} & \text{if} \quad \frac{1}{4} \leq \theta < \frac{1}{2} \\ \frac{1}{4} & \text{if} \quad \frac{1}{2} \leq \theta \leq 1 \end{cases}$$

APPENDIX

[19] The parameter is observable but not verifiable.

Proof of Lemma 2.1: The IC constraint implies that for $\theta > \hat{\theta}$

$$\frac{v(x(\theta),\theta) - v(x(\theta),\hat{\theta})}{\theta - \hat{\theta}} \geq \frac{\mathcal{V}^x(\theta) - \mathcal{V}^x(\hat{\theta})}{\theta - \hat{\theta}} \geq \frac{v(x(\hat{\theta}),\theta) - v(x(\hat{\theta}),\hat{\theta})}{\theta - \hat{\theta}}$$

Since v is C^3 and x is bounded, the inequality above shows that $\mathcal{V}$ is a Lipschitz function. Moreover, if x is continuous at θ, then

$$\frac{d}{d\theta}\mathcal{V}^x(\theta) = v_\theta(x(\theta),\theta).$$

By the Fundamental Theorem of Calculus, we get (i). $\square$

(ii) From (i), $t(\theta) = \mathcal{V}^x(\theta) - v(x(\theta),\theta)$, for all $\theta, \hat{\theta} \in \Theta$. Thus, it is easy to see that

$$V(x(\theta),t(\theta),\theta) - V(x(\hat{\theta}),t(\hat{\theta}),\theta) = \int_{\hat{\theta}}^{\theta}\left[\int_{x(\hat{\theta})}^{x(\tilde{\theta})} v_{x\theta}(\tilde{x},\tilde{\theta})d\tilde{x}\right]d\tilde{\theta}$$

for all $\theta, \hat{\theta} \in \Theta$.

Let $\hat{\theta}_0 \in [\underline{\theta}, \overline{\theta})$ such that $v_{x\theta}(x(\hat{\theta}_0), \hat{\theta}_0) > 0$. By the right continuity, x restrict to a small interval $I = [\hat{\theta}_0, \theta_0)$ has its graphic on CS_+. Let $A = (a, b)$ be a maximal interval in I such that $x(\hat{\theta}_0) > x(\theta)$, for all $\theta \in A$. If $a = \hat{\theta}_0$, then the double integral above will be negative when $\hat{\theta} = a$ and $\theta = b$. If $a > \hat{\theta}_0$, then the left limit of the double integral will be also negative when $\hat{\theta} = a$ and $\theta = b$ (since $x_-(a) \geq x(\hat{\theta}_0)$). In both cases we have a contradiction with the (IC) constraint. Therefore, $x(\hat{\theta}) \leq x(\theta)$, for all $\theta \in I$. $\square$

Proof of Theorem 3.1:

(i) Define the function

$$\varphi(y) = \int_{\theta}^{\hat{\theta}}\left[\int_{x(\tilde{\theta})}^{y} v_{x\theta}(\tilde{x},\tilde{\theta})d\tilde{x}\right]d\tilde{\theta}$$

for $y \in X(\hat{\theta})$. Taking the second derivative, we have

$$\varphi''(y) = v_{xx}(y,\hat{\theta}) - v_{xx}(y,\theta).$$

Thus, $\hat{\theta} \geq \theta$ and $v_{x^2\theta} < 0$ implies that $\varphi''(y) \leq 0$, for all $y \in X(\hat{\theta})$. Since $\varphi(x_-(\hat{\theta})) \geq 0$ and $\varphi(x(\hat{\theta})) \geq 0$, we have that $\varphi(y) \geq 0$, for all $y \in I_{\hat{\theta}}$.

(ii) Let (x_n) be a sequence of continuous implementable decisions such that $x_n \to x$ in the weak topology. In particular, $x_n \to x$ almost surely. By the dominated convergence theorem (see Rudin (1974)),

$$\int_{\hat{\theta}}^{\theta}\int_{x(\tilde{\theta})}^{y} v_{x\theta}(\tilde{x},\tilde{\theta})d\tilde{x}d\tilde{\theta} \geq 0 \qquad ((*))$$

when $y = x(\hat{\theta})$ or $y = x_-(\hat{\theta})$.

Given y in the interior of $X(\hat{\theta})$, for n sufficiently large, let $\hat{\theta}_n \in \Theta$ such that $x_n(\hat{\theta}_n) = y$ (such $\hat{\theta}_n$ exists because x_n assume values close to $x_-(\hat{\theta})$ and to $x(\hat{\theta})$ for large n). Then, by the monotonicity of x_n, we can choose $\hat{\theta}_n$ such that $\hat{\theta}_n \to \hat{\theta}$. Again, by the dominated convergence theorem, $(*)$ is also true for such y.

Conversely, if x is such that the associated correspondence X is implementable, x can cross continuously from CS_- to CS_+ one time at most. Thus, lemma 2.1 (ii) implies that x is non-increasing or non-decreasing or U-shaped.

If a discontinuity occurs outside the U-shaped part, we can substitute this discontinuous part by a monotone part making the decision continuous and implementable (this is possible because we can do a preserving area modification of the the decision such that the IC constraints are satisfied). If the discontinuity occurs in the U-shaped part, it should be on the left side of the "U" because of the proof of theorem 3.2 (c) and we can substitute this discontinuity by a continuous part and the right side of the discontinuity such that the resulting contract is continuous. In both cases, the modification can be done as close as one wants to the original decision in the weak topology sense. $\square$

Proof of Theorem 3.2:

(a) Define a local inverse for x at $\hat{\theta}$. Applying Fubini's Theorem (as we did in lemma 3.1) and taking the right (left) derivative at $x(\hat{\theta})$ and observing that $x(\hat{\theta})$ is a minimum point for φ, we get our result.

(b) Observe that if we fix $\hat{\theta}$, θ is a minimum point of $\Phi^x(\hat{\theta}, \cdot)$. Then, the result is a direct consequence of the first order conditions.

(c) Observe that if $y \in X(\theta) \cap X(\hat{\theta})$, with $\theta, \hat{\theta} \in \Theta$, then

$$\begin{aligned}\Phi_X(\hat{\theta}, \theta) &= \int_{\hat{\theta}}^{\theta} \left[\int_{x(\tilde{\theta})}^{y} v_{x\theta}(\tilde{x}, \tilde{\theta}) d\tilde{x} \right] d\tilde{\theta} \\ &= -\int_{\theta}^{\hat{\theta}} \left[\int_{x(\tilde{\theta})}^{y} v_{x\theta}(\tilde{x}, \tilde{\theta}) d\tilde{x} \right] d\tilde{\theta} = -\Phi_X(\theta, \hat{\theta}).\end{aligned}$$

From (i), this means that both integrals above are zero, i.e., $\Phi_X(\theta, \hat{\theta}) = 0$, i.e.,

$$\int_{\theta}^{\hat{\theta}} \left[\int_{x(\tilde{\theta})}^{y} v_{x\theta}(\tilde{x}, \tilde{\theta}) d\tilde{x} \right] d\tilde{\theta} = 0, \ \forall y \in X(\theta) \cap X(\hat{\theta}).$$

If x is right and left increasing at $\hat{\theta}$, applying (a), we get the result. If x is just right (or left) increasing at $\hat{\theta}$, use the continuity of v_x and the previous case to conclude the proof.

I claim that x is continuous at $\hat{\theta}$. Otherwise, $y(\theta)$ defined implicitly by the equality in (c) would be increasing in CS_- for a fixed $\hat{\theta} \in \Theta$ where x is discontinuous (see also the proof of theorem 3.3). $\square$

Proof of Theorem 3.3:

Let $\theta_0 \in \Theta$ be the minimum point of x and $\theta_1 \in \Theta$ such that $x(\overline{\theta}) \in X(\theta_1)$. Consider the following cases for $\hat{\theta}, \theta \in \Theta$:

(1) $\theta_1 \leq \theta \leq \theta_0 \leq \hat{\theta} \leq \overline{\theta}$ such that $x(\hat{\theta}) \in X(\theta)$.

Using Fubini's Theorem,

$$\Phi^x(\theta, \hat{\theta}) = \int_{x_0}^{x(\hat{\theta})} \left[\int_{\varphi_1(\tilde{x})}^{\varphi_2(\tilde{x})} v_{x\theta}(\tilde{x}, \tilde{\theta}) d\tilde{\theta} \right] d\tilde{x}$$

where $x_0 = x(\theta_0)$ and φ_1, φ_2 are the two inverses of x on $[x_0, x(\underline{\theta})]$, where $\varphi_2(x) = \overline{\theta}$, for all $x \in [x(\overline{\theta}), x(\underline{\theta})]$.

From theorem 3.2 (c), $v_x(\tilde{x}, \varphi_1(\tilde{x})) = v_x(\tilde{x}, \varphi_2(\tilde{x}))$, for all $\tilde{x} \in [x_0, x(\hat{\theta})]$. Thus, $\Phi^x(\theta, \hat{\theta}) = 0$.

(2) $\theta \leq \hat{\theta}$

If $x(\theta) < x(\hat{\theta})$, then $\hat{\theta} \leq \theta_0$ or $\varphi_1(\hat{\theta}) \leq \theta \leq \theta_0 \leq \hat{\theta}$. In the first case,

$$\Phi^x(\theta, \hat{\theta}) = -\int_{\theta}^{\hat{\theta}} \left[\int_{x(\hat{\theta})}^{x(\tilde{\theta})} v_{x\theta}(\tilde{x}, \tilde{\theta}) d\tilde{x} \right] d\tilde{\theta},$$

and since the region delimited in the integral is negative, we have that $\Phi^x(\theta, \hat{\theta}) \geq 0$.

In the second case, using Fubini's Theorem again and (1) above

$$\begin{aligned} \Phi^x(\theta, \hat{\theta}) &= \int_{x(\theta)}^{x(\hat{\theta})} \left[\int_{\theta}^{\varphi_2(\tilde{x})} v_{x\theta}(\tilde{x}, \tilde{\theta}) d\tilde{\theta} \right] d\tilde{x} \\ &= \int_{x(\theta)}^{x(\hat{\theta})} [v_x(\tilde{x}, \varphi_2(\tilde{x})) - v_x(\tilde{x}, \theta)] d\tilde{x}. \end{aligned}$$

Since the function $v_x(\tilde{x}, \cdot)$ is U-shaped and $\varphi_1(\tilde{x}) \leq \theta$, $v_x(\tilde{x}, \varphi_2(\tilde{x})) \geq v_x(\tilde{x}, \theta)$, for all $\tilde{x} \in [x(\theta), x(\hat{\theta})]$. Thus, $\Phi^x(\theta, \hat{\theta}) \geq 0$.

If $x(\theta) \geq x(\hat{\theta})$, then $\hat{\theta} \leq \theta_0$ or $\theta \leq \varphi_1(\hat{\theta}) \leq \theta_0 \leq \hat{\theta}$. With an analogous proof, $\Phi^x(\theta, \hat{\theta}) \geq 0$.

(3) $\theta > \hat{\theta}$

If $x(\theta) > x(\hat{\theta})$, then $\theta_0 \leq \hat{\theta} < \theta$ or $\theta_1 \leq \hat{\theta} \leq \theta_0 < \theta$. The proof is analogous to the case (2). If $x(\theta) < x(\hat{\theta})$, then, using Fubini's Theorem and (1),

$$\begin{aligned} \Phi^x(\theta, \hat{\theta}) &= -\int_{x(\theta)}^{x(\hat{\theta})} \left[\int_{\varphi_1(\tilde{x})}^{\theta} v_{x\theta}(\tilde{x}, \tilde{\theta}) d\tilde{\theta} \right] d\tilde{x} \\ &= -\int_{x(\theta)}^{x(\hat{\theta})} [v_x(\tilde{x}, \theta)) - v_x(\tilde{x}, \varphi_1(\tilde{x}))] d\tilde{x}. \end{aligned}$$

If $\tilde{\varphi}_1$ is identical to φ_1 on $[x_0, x(\overline{\theta})]$ and $v_x(\tilde{x}, \varphi_1(\tilde{x})) = v_x(\tilde{x}, \varphi_2(\tilde{x}))$, for all $\tilde{x} \in [x(\overline{\theta}), x(\underline{\theta})]$, then $\varphi_1(\tilde{x}) \leq \tilde{\varphi}_1(\tilde{x})$, for all $\tilde{x} \in [x_0, x(\underline{\theta})]$. Since $v_x(\tilde{x}, \cdot)$ is U-shaped, $v_x(\tilde{x}, \varphi_1(\tilde{x})) \geq v_x(\tilde{x}, \tilde{\varphi}_1(\tilde{x})) \geq v_x(\tilde{x}, \theta)$. Thus, $\Phi^x(\theta, \hat{\theta}) \geq 0$. □

Proof of Theorem 3.4: Consider the topology of the pointwise convergence at the continuous points of the limit, i.e., x_n converges to x if and only if $x_n(\theta) \to x(\theta)$ for every $\theta \in \Theta$ where x is continuous. It is well known that every bounded and closed set in $\mathcal{C}$ is compact with respect to this topology (see Billingsley (1986)).

Let $\underline{x}$ and $\overline{x}$ be an inferior and superior bound for x_1. It is easy to see that if $x \in \mathcal{C}$ is implementable, then y defined by $y(\theta) = \underline{x} \vee x(\theta) \wedge \overline{x}$, for all $\theta \in \Theta$ is implementable and the principal weakly prefers y than x.

It is easy to see that if x_n is a sequence of implementable decisions converging to x such that the associated sequence X_n is also implementable, then the associated X is implementable. In particular, if x crosses x_0, it should cross in a continuous way. Therefore, the space of implementable decision such that the associated correspondence is also implementable is closed.

Finally, the objective function of (P) is continuous with respect to the considered topology. Then, by Weierstrass Theorem, there exist an optimal contract for (P) as described in the theorem. □

Proof of Theorem 3.5: From theorem 3.2 and 3.3, we can rewrite (P) as

$$\max_x \int_{\underline{\theta}}^{\theta_1} f(x(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta} + \int_{\theta_1}^{\theta_0} \Big[f(x(\tilde{\theta}), \tilde{\theta}) + f(x(\tilde{\theta}), \varphi(x(\tilde{\theta}), \tilde{\theta})) g^x(\tilde{\theta}) \Big] d\tilde{\theta}$$

where $x\colon [\underline{\theta}, \theta_0] \to \mathbb{R}$ is non-increasing and φ is implicitly defined by $v_x(x, \theta) = v_x(x, \varphi)$ as a function of x and θ, for $\theta \in [\underline{\theta}, \theta_0]$ (such that $\varphi = \overline{\theta}$ for $\theta \in [\underline{\theta}, \theta_1]$) and

$$g^x(\tilde{\theta}) = \varphi_x(x(\tilde{\theta}), \tilde{\theta}) \dot{x}(\tilde{\theta}) + \varphi_\theta(x(\tilde{\theta}), \tilde{\theta})$$

comes form the theorem of change of variable in the integral. From the Implicit Function Theorem

$$\frac{\partial \varphi}{\partial x}(x, \theta) = \frac{v_{xx}(x, \theta) - v_{xx}(x, \varphi(x, \hat{\theta}))}{v_{x\theta}(x, \varphi(x, \theta))}$$

and

$$\frac{\partial \varphi}{\partial \theta}(x, \theta) = \frac{v_{x\theta}(x, \hat{\theta})}{v_{x\theta}(x, \varphi(x, \theta))}$$

In order to get the first order conditions, we have to take the Gateaux derivative of functional defined from the maximization problem above.

Fix an interval $[\theta^1, \theta^2]$ with $\theta^2 \leq \theta_0$ and the respective (optimal) decision $x^*(\theta^i) = \overline{x}^i$, $i = 1, 2$. Define the space of the admissible decision:

$$X = \{x\colon [\theta^1, \theta^2] \to \Re;\ x(\theta^i) = \overline{x}^i,\ i = 1, 2,\ \text{and } x \text{ is decreasing}\}$$

and the space of admissible perturbations:

$$H = \{h\colon [\theta^1, \theta^2] \to \Re;\ h(\theta^i) = 0,\ i = 1, 2,\ \text{and } x^* + h \text{ is decreasing}\}$$

and the objective functional:

$$\begin{aligned}
F(x) &= \int_{\theta^1}^{\theta^2} f(x(\tilde\theta), \tilde\theta) d\tilde\theta + \int_{\varphi(\overline{x}^2, \theta^2)}^{\varphi(\overline{x}^1, \theta^1)} f(x(\tilde\theta), \tilde\theta) d\tilde\theta \\
&= \int_{\theta^1}^{\theta^2} \{f(x(\tilde\theta), \tilde\theta) - [\varphi_x(x(\tilde\theta, \tilde\theta)\dot x(\tilde\theta) + \varphi_\theta(x(\tilde\theta), \tilde\theta)] f(x(\tilde\theta), \varphi(x(\tilde\theta), \tilde\theta))\} d\tilde\theta
\end{aligned}$$

The first order condition gives (omitting the arguments of the function and putting a hat when the function is calculated at φ)

$$\delta_h F(x) = \int_{\theta^1}^{\theta^2} [f_x h - [\varphi_{xx}\dot x h + \varphi_{x\theta} h + \varphi_x \dot h]\hat f - [\varphi_x \dot x + \varphi_\theta](\hat f_x h + \hat f_\theta \varphi_x h)] d\theta = 0$$

By an integration by parts we have

$$\begin{aligned}
-\int_{\theta^1}^{\theta^2} [\hat f \varphi_x \dot h] d\theta &= \int_{\theta^1}^{\theta^2} (\dot{\overline{\hat f \varphi_x}}) h d\theta \\
&= \int_{\theta^1}^{\theta^2} \{[\hat f_x \dot x + \hat f_\theta(\varphi_x \dot x + \varphi_\theta)]\varphi_x + \hat f(\varphi_{xx}\dot x + \varphi_{x\theta})\} h d\theta
\end{aligned}$$

Plug this last equation into the first order condition, we get

$$\begin{cases} \int_{\theta^1}^{\theta^2} f_x h d\theta = 0 & \text{if } [\theta^1, \theta^2] \subset [\underline{\theta}, \theta_1) \subset [\theta_1, \theta_0] \\ \int_{\theta^1}^{\theta^2} (f_x - \hat f_x \varphi_\theta) h d\theta = 0 & \text{if } [\theta^1, \theta^2] \subset [\theta_1, \theta_0] \end{cases}$$

Thus,

$$A = f_x - \frac{v_{x\theta}}{\hat{v_{x\theta}}} \hat f_x = 0$$

what is equivalent to

$$\frac{f_x}{v_{x\theta}} = \frac{\hat f_x}{\hat{v_{x\theta}}}$$

The non-decreasing condition can be identified as $dx \leq 0$ (in a distributional sense). It is easy to see that the interiority condition for the existence of a Lagrange multiplier is satisfied. Then using the same approach of Guesnerie and Laffont (1984), we have our result. $\square$

Remark. Fix θ: if x is sufficiently small, $A > 0$ and if x is sufficiently large, $A < 0$. This prove that the first order condition of the U-shaped part is also sufficient when it is the unique critical point for each θ.

Proof of Theorem 3.6: This is an immediate consequence of the remark above and theorems 3.3-3.5. $\square$

REFERENCES

Armstrong, M. (1996): "Multiproduct Nonlinear Pricing", *Econometrica*, 64, 51-75.

Arrow, K.J. (1986): "Agency and the Market", in the *Handbook of Mathematical Economics*, Vol. III, ed. by K. J. Arrow and M. D. Intrilligator. Amsterdam: North-Holland, 1184-1195.

Athey, S. (1997): "Single Crossing Properties and the Existence of Pure Strategy Equilibria in Game of Incomplete Information", *MIT Working Paper* 97-11.

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

Billingsley, P. (1986): "Probability and Measure", *John Willey & Sons*, New York.

Cavallo, A., J. Heckman, J.J. Hsee (1998): "The GED is a Mixed Signal", working paper.

Chassagnon, A. and P.A. Chiappori (1995): "Insurance under Moral Hazard and Adverse Selection: The Case of Pure Competition", mimeo.

Fudenberg, D. and J. Tirole (1991): "Game Theory", Cambridge, MA: MIT Press.

Green, J. and J.J. Laffont (1987): "Limited Communication and Incentive Compatibility", in *Information, Incentives and Economic Mechanisms, Essays in Honor of Leonid Hurwicz*, *University of Minnesota Press*, Minneapolis, 308-329.

Guesnerie, R. (1975): "Production of Public Sector and Taxation in a Simple Second Best Model", *Journal of Economic Theory*, 10, 127-156.

Guesnerie, R. and J.J. Laffont (1978): "Taxing Price Markets", *Journal of Economic Theory*, 19, 423-455.

Guesnerie, R. and J.J. Laffont (1984): "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm", *Journal of Public Economics*, 25, 329-369.

Jullien, B. (1997): "Participation Constraints in Adverse Selection Models", CREMAQ, Toulouse.

Kahn, C.M. and J. Scheinkman (1985): "Optimal Employment Contracts with Bankruptcy Constraints", *Journal of Economic Theory*, 35, 343-365.

Kreps, M. and J. Sobel (1994): "Signaling", *Handbook of Game Theory*, volume 2, chap. 25, 850-867.

Laffont, J.J. and J. Tirole (1993): "A Theory of Incentives in Procurement and Regulation", *MIT Press*, Cambridge, Massachusetts. London, England.

Laffont, J.J., E. Maskin and J.C. Rochet (1987): "Optimal Nonlinear Pricing with Two Characteristics", in *Information, Incentives and Economic Mechanisms, Essays in Honor of Leonid Hurwicz*, *University of Minnesota Press*, Minneapolis, 256-266.

Lewis, T. and D. Sappington (1988): "Regulating a Monopolist with Unknown Demand", *American Economic Review*, 78, 986-998.

Lewis, T. and D. Sappington (1989): "Countervailing Incentives in Agency Problems", *Journal of Economic Theory*, 66, 238-263.

Luenberger, D.G. (1969): "Optimization by Vector Space Methods", *John Wiley & Sons*, Inc.

Maggi, G. and Rodríguez-Clare (1995): "On Countervailing Incentives", *Journal of Economic Theory*, 66, 238-263.

Maskin, E. and J. Riley (1984): "Monopoly with Incomplete Information", *Rand Journal of Economics*, 15, 171-196.

McAfee, R. P. and J. McMillan (1988): "Multidimensional Incentive Compatibility and Mechanism Design", *Journal of Economic Theory*, 46, 335-354.

Matthews, S. and J. Moore (1987): "Monopoly Provision of Quality and Warranties: an Exploration in the Theory of Multidimensional Screening", *Econometrica*, 55, 441-467.

Mirrlees, J.A. (1971): "An Exploration in the Theory of Optimal Income Taxation", *Review of Economics Studies*, 38, 175-208.

Mirrlees, J.A. (1976): "Optimal Tax Theory: A Synthesis", *Journal of Public Economics*, 6, 327-358.

Mirrlees, J.A. (1986): "The Theory of Optimal Taxation", in the *Handbook of Mathematical Economics*, Vol. III, ed. by K. J. Arrow and M. D. Intrilligator. Amsterdam: North-Holland, 1197-1249.

Mussa, M. and S. Rosen (1978): "Monopoly and Product Quality", *Journal of Economic Theory*, 18, 301-317.

Page Jr., F.H. (1991): "Optimal Contract Mechanisms for Principal-Agent Problems with Moral Hazard and Adverse Selection", *Economic Theory*, 1, 323-338.

Rochet, J.C. (1985): "The Taxation Principle and Multi-Time Hamilton-Jacobi Equation", *Journal of Mathematical Economics*, 14, 113-128.

Rochet, J.C. and P. Choné (1997): "Ironing, Sweeping and Multidimensional Screening", *Econometrica* 66, 783-826.

Rudin, W., 1974. Real and Complex Analysis. McGraw-Hill, Inc., 2nd ed.

Spence, M. (1974): "Competitive and Optimal Responses to Signals. An Analysis of Efficiency and Distribution", *Journal of Economic Theory*, 7, 296-332.

Stole, L. (1991): "Mechanism Design under Common Agency", Mimeo. University of Chicago, GSB.