ADVERSE SELECTION PROBLEMS WITHOUT THE SPENCE-MIRRLEES CONDITION

By Aloisio Araujo and Humberto Moreira[1]

We relax the single crossing property or Spence-Mirrlees condition (SMC) for the adverse selection problem and derive the conditions for incentive compatibility (IC). Economically, this requires that the principal has to take care of wider varieties of strategic behavior of the agent who can now mimic the behavior of other types that are not necessarily nearby. We prove that in order to do that he has to equalize the marginal rate of substitution and the marginal rent between pooling types. This implies a change in the trade off between rent extraction and distortion leading to a lower level of total welfare. Mathematically, our necessary conditions are expressed by a generalized Lagrangian which contains terms other than the standard first and second order conditions of the IC constraints. The incentive compatible contracts can present discrete or continuous pooling and they are discontinuous even under the monotone hazard rate condition. We do not enter into the full complexity of contracts, so we just consider the case of at most one discontinuity. In the space of the limit of continuous contracts, the optimal contract is characterized. In the context of a wider set of contracts we present a class of incentive compatible contracts which approximates the optimal. The non SMC arises naturally when we are dealing with a multidimensional characteristic problem with countervailing incentives (given by the correlation between two sources of asymmetric information). Therefore, our paper also gives a framework to study multi-characteristic adverse selection problems. We give some examples of these in the following contexts: combination of moral hazard and adverse selection, nonlinear pricing, regulation and labor contract.

Keywords: Spence-Mirrlees condition; marginal rate of substitution and rent identities; discrete pooling.

1. INTRODUCTION

The main goal of this paper is to relax the classical Spence-Mirrlees condition (SMC) which has been extensively used in the literature to characterize the solution of the adverse selection problem.

[1] Financial support from CNPq and ANEEL (Agência Nacional de Energia Elétrica). Preliminary versions of this paper were presented at the University of Toulouse, Pompeu Fabra (Barcelona), IMPA, PUC-Rio, EPGE-RJ, University of Chicago, the Nineteenth Meeting of the Brazilian Econometric Society and the Seventeenth Latin American Meeting of the Econometric Society. We wish to thank A. Chassagnon, P. A. Chiappori, M. Harris, J. J. Laffont, A. Mas-Colell, J. C. Rochet, J. Scheinkman, L. Stole and J. Tirole, as well as the seminar participants, for useful discussions and suggestions on earlier drafts of this paper. We would like to thank Caio Ibsen and Daniel Chrity for helping us with the numerical examples.

In the one-dimensional parameter case, the SMC 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 (asymmetric information).

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

The SMC also enables a full characterization of the optimal solution: if the optimal decision is strictly monotone in the parameter, then it should be the relaxed solution.[2] 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 Fudenberg and Tirole (1991) or Guesnerie and Laffont (1984)).

The question studied in this paper is: What happens when the SMC 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 negative single crossing regions. An implementable decision path should preserve monotonicity in each region, and it may or may not cross 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 a bell-shaped form) because of the monotonicity. Then, 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 marginal condition with respect to the decision the *marginal rate of substitution identity.* Moreover, there exists 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.[3]

We use the second order condition of the incentive compatibility (IC) constraint and the marginal rate of substitution identity as the constraints of the adverse selection problem and derive the first order conditions for the optimal contract. The constraints will be both of the equality and inequality type and our problem is not concave anymore, but we still can compute the optimal contract in a particular setup. These constraints indicate that the principal has to take care of wider varieties of strategic behavior of the agent who can now mimic the behavior of other types that are not nearby. Mathematically, our results are expressed by

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

[3] The more general case is not finished. The difficulty of a 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.

a generalized Lagrangian which contains terms other than the first and second order conditions of the IC constraints. We do not get into the full complexity of contracts: we just consider the case of at most one discontinuity. However, many examples show that this is enough to approximate the principal's relaxed welfare.

Chassagnon and Chiappori (1995) studied the insurance market competitive equilibrium with adverse selection and moral hazard where the SMC 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 SMC and where the second order approach is not valid.

Example 1: (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 a reward schedule.

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

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

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 managers with characteristic θ close to 0 (respectively 1) are specialists in technology T_1 (respectively T_2). We also say that types θ close to $1/2$ are generalists (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 (where the manager's reservation utility is zero).

Depending on the informational structure, we have different problems:

1) *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 on 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 each $i = 1, 2$,

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

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 $\theta \leq\ (\geq)\, 1/2$, 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 a signal of the manager's type (since he can control this choice).

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

2) *Second Best without the technology choice information*: neither θ nor T are observable. In this case the optimal contract can not be contingent on the technology,

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

i.e., the owner can not use the technology choice as a signal of manager's type. Therefore, the owner will face a manager that has the following 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 type 1/2. Therefore, the principal's problem does not have the SMC.

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., implementable decisions are going to be U-shaped symmetric with respect to 1/2 (see Section 4 for the details). In particular, 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 a discrete pooling equilibrium).

The impossibility of observing the choice made by the agent transforms the bidimensional second best problem (in (T,θ)) with information into a one-dimensional problem (only in θ) without the SMC. 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 nonlinear pricing models studied by Mussa and Rosen (1978) and Maskin and Riley (1984) where the SMC is relaxed. Suppose that a monopolist faces different types of demand with finite elasticity. More precisely, the demands are linear, the market size is decreasing in 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 willingness to pay for the good) is the large one. In this case the SMC fails to hold and the optimal contract is going to be non-decreasing for the low types and non-increasing for the high types (a 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 consumer's willingness to pay: he will extract less rent from low types because he wants to sell the good more and at the same time he wants to extract the rent of high types without breaking the IC constraints. Therefore, the condition is that 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 bidimensional characteristic problem are repeated here.

There are some contributions on the literature of multi-characteristic and multi-decision problem: McAfee and McMillan (1988) gives a generalization of the

SMC which is enough to imply a strong structure on the agent's utility function. Matthews and Moore (1987) studies the case of when there are more decision variables than characteristics. Armstrong (1995) characterized the optimal solution of a specific model of multiproduct nonlinear pricing. He showed that bunching is common in this case. Our paper also addresses this question. More recently, Rochet and Choné (1997) studies the ironing principle in a multidimensional nonlinear pricing model.

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 the effort of the firm's manager in cutting costs. The cost function is one-dimensional in the sense that there is just one source of activity that allows the manager to cut cost. Suppose, however, that there are two kinds of activities that the manager employs to cut cost and the regulator can only observe the aggregate cost, i.e., there are two sub-costs 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 of effort has positive cross derivative) and there is a decreasing relation between the parameters that characterize the sub-costs. 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 explained in the examples above.

Finally, we study a labor contract where workers have a vector of two characteristics (unknown to the firm) that are mixed in a verifiable signal (schooling). This example follows, in spirit, Cavallo, Heckman and Hsee (1998). The firm is a 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 workers's abilities are not totally captured by the signal. This is a standard adverse selection problem. However, depending on the parameters of the model, the SMC does not hold. In this case, discrete pooling equilibrium may appear, and indicates that different workers with respect to the profile of characteristics may be treated as the same in equilibrium.

The paper is organized as follows. In Section 2 we present the adverse selection model. In Section 3 we give necessary conditions for the solution of adverse selection problems without the SMC. Section 4 presents some examples. Section 5 gives the final conclusions and the Appendix A contains the proofs.

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) generally has 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 a one-dimensional parameter θ which is known to the agent and 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 $x\colon \Theta \to I$ is implementable if there exists a money transfer function $t\colon \Theta \to \mathbb{R}$ that satisfies the incentive compatibility constraint: for all $(\theta, \hat{\theta}) \in \Theta^2$,

$$\text{(IC)} \qquad V(x(\theta), t(\theta), \theta) \geq V(x(\hat{\theta}), t(\hat{\theta}), \theta).$$

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

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

$$\text{(IR)} \qquad V(x(\theta), t(\theta), \theta) \geq 0.$$

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:

$$\begin{aligned}
& \max_{x,t} E_\theta[U(x(\theta), t(\theta), \theta)] \\
& \text{s.t.} \\
\text{(IC)} \quad & V(x(\theta), t(\theta), \theta) \geq V(x(\hat{\theta}), t(\hat{\theta}), \theta), \ \forall (\theta, \hat{\theta}) \in \Theta^2 \\
\text{(IR)} \quad & V(x(\theta), t(\theta), \theta) \geq 0, \quad \forall \theta \in \Theta
\end{aligned}$$

[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.

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

DEFINITION 1: Let $\mathcal{C}$ be the set of all càdlàg decisions, i.e., the space of all $x\colon\Theta \to \mathbb{R}$ 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 pointwise limit topology at every continuous parameter of the limit decision function (this is the weak topology in the distributional sense - see Rudin (1974)).

Below we present the classical first and second order conditions of the incentive constraints extended to càdlàg decisions. All the proofs are presented in Appendix A.

LEMMA 1: *(The first and second order conditions of the IC constraint)*

(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 (resp. non-increasing) on the region where $v_{x\theta}>0$ (resp. $v_{x\theta}<0$).

Lemma 1 (i) shows that for each implementable càdlàg x, there exists a unique (except by a constant $\mathcal{V}^x(\underline{\theta})$) 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

$$f(x(\theta),\theta) = \left[u(x(\theta),\theta) + v(x(\theta),\theta) + \frac{(P(\theta)-1)}{p(\theta)}v_\theta(x(\theta),\theta) - \mathcal{V}^x(\underline{\theta})\right]p(\theta)$$

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

[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.

If $v_\theta \geq 0$, then the rent function $\mathcal{V}^x$ assumes its minimum at $\underline{\theta}$, and since money transfer is costly for the principal, the IR constraint will be binding at the optimal contract in $\underline{\theta}$, 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 for every decision x). Therefore, in the general case one might 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 remark 3 after Theorem 4').

The principal's optimization program becomes

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

If we ignore the IC constraint, then the problem is called relaxed and so is 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 a sufficient condition for implementation is the constant sign of the partial derivative of the marginal rate of substitution with respect to the parameter:

$$\textbf{(CS}_+\textbf{)} \qquad \partial_\theta \left(\frac{V_x}{V_t} \right) = v_{x\theta} > 0 \quad \text{on } I \times \Theta,$$

or

$$\textbf{(CS}_-\textbf{)} \qquad \partial_\theta \left(\frac{V_x}{V_t} \right) = v_{x\theta} < 0 \quad \text{on } I \times \Theta.$$

This is known as the Spence-Mirrlees Condition (SMC) or sorting condition and it 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 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 program is known as the second order approach because, under the SMC, 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 SPENCE-MIRRLEES CONDITION

Now we will introduce a natural generalization of the SMC. Assumption A1 below separates the plane (θ, x) into two regions: CS_+ and CS_-.

ASSUMPTION 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 SMC, 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 three other equivalent cases: $v_{x^2\theta} > 0$ and $v_{x\theta^2} \leq 0$, with x_0 increasing and reverting the regions where $v_{x\theta} > 0$ and $v_{x\theta} < 0$; $v_{x^2\theta} < 0$ and $v_{x\theta^2} \leq 0$ (see Example 2); $v_{x^2\theta} > 0$ and $v_{x\theta^2} \geq 0$ (see Example 4), for the respective cases where x_0 is decreasing.

FIGURE 1.– The curve x_0.

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

The next theorem will give necessary conditions for implementability. 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 1: *(Necessary conditions for implementability) Assume A1. If x is an implementable càdlàg decision, then*

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

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

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

(ii) if $\Phi^x(\theta,\hat{\theta}) = 0$, then

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

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

(iii) If x is right and left increasing at $\hat{\theta}$ and $x = x(\theta) = x(\hat{\theta})$, then

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

Remark 1: Item (ii) is the dual condition of (i) when we interchange θ and x, i.e., instead of looking at the direct decision (x as a function of θ), we look at the inverse function (θ as a function of x), which can be interpreted as the *marginal rent equality* of a type that considers his designed choice and an indifferent choice.

Remark 2: The economic interpretation for Lemma 1 and Theorem 1: In order to provide truth telling, the principal offers a contract that

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

• if two agents (θ and $\hat{\theta}$) choose the same decision and the agents cannot locally misrepresent their types, then the principal must equalize the marginal rate of substitution of the two agents ($\mathbf{MRS}^{\theta} = \mathbf{MRS}^{\hat{\theta}}$) - see figure 2 below.

• if a type is indifferent between his designed choice and type $\hat{\theta}$ choice, then the principal must equalize the agent's rent of these choices.

Figure 2.– The U-shaped decision.

The SMC 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 SMC case obligates the principal also to check the marginal rate of substitution identity (the cross-stream condition). Therefore, the IC constraint is less restrictive in the former case than in latter one and, thus, the rent extraction is less powerful when there is no SMC.

If x hits the curve x_0, then it should cross x_0 in a constant way or preserve the marginal utility for types that choose the same level of x (if x were identical to x_0 in an interval, then the IC constraint would not hold locally). This last condition is new, and when the SMC 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 and rent identities*.[7]

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.

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

It is important to note that we are dealing with a non-concave problem because the set of feasible decisions for the agent is not convex when the agent's utility function does not satisfy the SMC.

3.1. *The Necessary Conditions for Optimality*

From the local second order condition of the IC constraint we know that an implementable decision that crosses continuously from one region to the other has to have a U-shaped form or present continuous pooling. This will give us a large scale pooling equilibrium (even under monotone hazard-rate property)[8]. We will study two cases: the continuous and the discontinuous crossing.

We will study 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.

Definition 2: The hazard rate is by definition: $M(\theta) = \frac{P(\theta)-1}{p(\theta)}$.

Even under a monotone hazard rate condition (MHRC), i.e., $M(\cdot)$ is increasing, the relaxed solution is not monotone.

Proposition 1: *(Geometry of the curves) Assume that A1 and MHRC hold, u is concave and does not depend on θ, and $v(\cdot, \theta)$ is concave. If x_1 crosses x_0, then x_1 is U-shaped and is above (below) the first best solution when it is above (below) x_0.*

The proof of this proposition is straightforward and it is not presented here.

Figure 3.– The possible cases for x_0 and x_1.

The two top cases correspond to the SMC situation. In this paper we are going to treat the third case. The other three cases represent the possibilities which although not considered specifically in this paper can be given similar treatment.

Assumption A2: *x_1 is U-shaped, crosses x_0 in a decreasing way, $x_1(\underline{\theta}) \geq x_1(\overline{\theta})$ and $x \lesseqqgtr x_1(\theta) \Leftrightarrow f_x(x, \theta) \gtreqqless 0$, for all $\theta \in \Theta$.*

A sufficient condition for the second part of A2 is the concavity of $f(\cdot, \theta)$, for each $\theta \in \Theta$. 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.

We are going to divide the analysis into two cases. The first one is the space of the limit of continuous implementable decisions. In the second one the decision

---

[8] 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.

can have at most one discontinuity (and the associated correspondence is not necessarily implementable).

The former situation corresponds to the case where the space of feasible contracts is more restricted. This diminishes the space of contract that the principal can design and enlarges the strategic space of the agents because they have more options to mimic the agents that have a set of choices. In the last situation the principal controls the choice of the agent more, increasing his surplus. In particular, we will see that the set of realized equilibrium choices is disconnected, i.e., there exists a range of decision choices between two equilibrium realizations that is not attainable. The intuition is that the principal designs a game for the agent to play and he strictly prefers to limit the agent's decision choice set, comparing the first situation with the second one.

If more discontinuities are possible, the principal's welfare increases. However, as will be shown in Examples 2 and 4, the single-discontinuity case approximates the relaxed welfare very well.

3.2. *The Case of the Closure of Continuous Decisions*

Define $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$. 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 of x. The characterization of the set of the closure of continuous implementable decisions is given by:[9]

Theorem 2: *(The closure of continuous decisions) Let x be a bounded implementable càdlàg decision and X is the associate correspondence. Thus, x is in the closure of the set of all continuous implementable decisions if and only if X is implementable, i.e.,* $\int_\theta^{\hat\theta}\left[\int_{x(\tilde\theta)}^{y} v_{x\theta}(\tilde x, \tilde\theta)d\tilde x\right]d\tilde\theta \geq 0,\ \forall y \in X(\hat\theta)$ *(or $\Phi^X(\theta, \hat\theta) \geq 0$).*

Under A1, 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. In this section, we will consider only the decision $x \in \mathcal{C}$ such that the associated X is implementable. Observe that when the SMC is valid, this set is equivalent to the continuous implementable decisions.

The next theorem gives a situation such that the necessary conditions of Lemma 1 and Theorem 1 are sufficient for the characterization of an implementable decision (the proof is analogous to the proof of Lemmas 3 and 4).

Theorem 3: *(Sufficient conditions for implementability) Assume A1. Let x be a bounded càdlàg decision that satisfies the necessary condition of Lemma 1 (ii) and Theorem 1 (iii). If $x(\underline{\theta}) \geq x(\overline{\theta})$, then x is implementable.*

[9] Indeed, as we will see in Theorem 5, the optimal decision can have discontinuities (except in cases like Examples 1 and 3). But we are assuming that the discontinuity is part of the decision (i.e., the associated correspondence is implementable). The sketch of the proof of Theorem 2 is presented in Appendix A.

The sufficient conditions for implementability of a decision x such that $x(\underline{\theta}) < x(\overline{\theta})$ is more difficult. However, Theorem 3 can be used to solve some problems.

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 Spence-Mirrlees Condition 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:

Lemma 2': *(Existence) Suppose that A1 and A2 hold. Then, there exists x^* a solution for (P) in the set of the closure of continuous implementable decisions.*

We will present an analogous version of this lemma (Lemma 2) whose proof is given in the Appendix A.[10] The next theorem gives the characterization of the optimal U-shaped part of the decision.

Theorem 4': *(Necessary conditions for optimality) Suppose that A1 and A2 hold. Let θ_0 be the continuous crossing parameter for x^* and $\theta_1 < \theta_0 < \theta_2$ such that $x^*(\theta_2) = x^*(\theta_1)$. Then:*

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

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

where $\theta_1 \leq \hat{\theta} \leq \theta_0$, $v_x(x,\hat{\theta}) = v_x(x,\theta)$ and $x = x^*(\theta) = x^*(\hat{\theta})$.

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

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

where $\theta_1 \leq a \leq \theta_0$, $x^*|_{[a,b]} \equiv \bar{x}$ and, in the integral, $\hat{a}$ and $\hat{b}$ are defined by the equality $v_x(x,\cdot) = v_x(x,\theta)$, where $\theta = a$ or b, respectively.

As we will see later, Theorem 4' is generalized by Theorem 4.

Remark 1: $\frac{v_{x\theta}(x,\hat{\theta})}{v_{x\theta}(x,\theta)}$ is the Lagrange multiplier of marginal rate of substitution identity and we can rewrite the condition of Theorem 4' (i) as

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

[10] We only have to observe that if a sequence of implementable correspondence (X_n) converges to X, then X is also implementable.

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 them.

For an illustration of the distortion effect in the case of no SMC, 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 no hat when it is evaluated at θ, the first order condition given by Theorem 4' 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 (ii) is the analogous to the 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 $\tilde{v}$. However, 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, separating 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 once 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" in Lewis and Sappington (1989) (see also Maggi and Rodríguez-Clare (1995)). The difference is that in our case it comes from the no SMC and in their case it is based on the type dependence of the agent's outside opportunity.

Finally, 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).

Remark 4: Suppose 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 Examples 1 and 3, this condition may not be true. In this case, the critical decision characterized by Theorem 4', x_0 and x_1 cross at the same point and this critical decision is strictly between x_0 and x_1.

The new feature of the solution that appears in Theorem 4' is the possibility of discrete pooling or large scale pooling, 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 latter the principal knows a range of types where the agent is. When the SMC 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, continuous pooling or discrete pooling.

Under SMC, the pooling interval of the optimal contract is characterized by the marginal welfare of the principal being zero along this interval. Theorem 4' shows that this property is no longer valid when there exists discrete pooling.

Under the assumption that follows we can characterize the optimal decision. For each $\mu \geq 0$, let $x_1^\mu(\theta)$ be the implicit solution of $f_x(\cdot,\theta) + \mu v_{x\theta}(\cdot,\theta) = 0$, for each $\theta \in \Theta$. Observe that $x_1^0 \equiv x_1$.

Assumption A3: *For each $\hat{\theta}$ and $\mu \geq 0$, the equations $v_x(x_1^\mu(\cdot),\cdot) = v_x(x_1^\mu(\cdot),\hat{\theta})$ and $v_\theta(x_1^\mu(\cdot),\cdot) = v_\theta(x_1^\mu(\hat{\theta}),\cdot)$ have at most one solution in the increasing part of x_1^μ on CS_+.*

This assumption means that the curve x_1^μ crosses the implicit solutions of $v_x(\cdot,a) - v_x(\cdot,\theta) = 0$ and $v_\theta(\cdot,\theta) - v_\theta(\bar{x},\theta) = 0$, at most once for each $a \in \Theta$ and $\bar{x} \in I$. For examples, see 2 and 4.

Theorem 5: *(Optimal decision) Assume A1, A2 and the first part of A3 for $\mu = 0$. Then the optimal solution for (P) is:*

$$x^*(\theta) = \begin{cases} \bar{x}, & \text{if} \quad \theta < \theta_1 \\ x^1(\theta), & \text{if} \quad \theta \geq \theta_1 \end{cases}$$

where $\bar{x} = x_1(\theta_1)$, $\int_{\underline{\theta}}^{\theta_1} f_x(\bar{x},\theta)d\theta = 0$ or

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

where x^u is characterized by Theorem 4' (i) and θ_1 is such that $x^u(\theta_1) = x^u(\overline{\theta})$.

The intuition of Theorem 5 is straightforward: every monotone implementable decision is dominated by a U-shaped decision or a continuous decision that is constant plus the relaxed solution (see the proof in the Appendix A). Theorem 5 displays the optimal decision in each case.

Observe that the optimal decision given in the Theorem 5 can be discontinuous at θ_1, because x^u is strictly between x_0 and x_1 (see remark 4 after Theorem 4').

3.3. *The Case of One Jump Decisions*

We are not going to characterize the space of implementable decisions because it is very large. For instance, there are implementable decisions that cross the curve x_0 many times. Our strategy instead is to derive the necessary optimal conditions. In order to do this, we deal with a relaxed program and give conditions under which this relaxed solution is the optimal second best decision. We are going to consider contracts of limited complexity: we restrict the number of crossing to at most one. The following lemma guarantees that there is a solution in this space and it is analogous to Lemma 2'.

LEMMA 2: *Assume that A1 and A2 hold and that I is bounded. There is a solution of (P) in the set of all decisions $x \in \mathcal{C}$ that crosses x_0 one time at most.*

In what follows we are going to give the necessary optimal conditions when there is just one discontinuity. First we provide a generalization of Theorem 4' that can be useful in the case where the relaxed solution is increasing in CS$_+$ (the case that is not treated here). This theorem derives the first order condition of a strictly increasing (or decreasing) part of a decision that satisfies the necessary conditions of Theorem 1. Formally:

THEOREM 4: *(Necessary conditions for optimality) Suppose that A1 and A2 hold. Let x^* be an optimal decision and $[\theta^1, \theta^2]$ an interval in Θ on which x^* is continuous and strictly monotone (in particular, the necessary conditions of Theorem 1 are satisfied under equality). If for each $\theta \in [\theta^1, \theta^2]$ there exist a unique $\hat{\theta} \neq \theta$ such that $\Phi^{x^*}(\theta, \hat{\theta}) = 0$. Then,*

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

where $\theta \in [\theta^1, \theta^2]$, $v_x(\hat{x}, \hat{\theta}) = v_x(\hat{x}, \theta)$, $v_\theta(x, \theta) = v_\theta(\hat{x}, \theta)$ and $x = x^*(\theta)$, $\hat{x} = x^*(\hat{\theta})$.

The interpretation of Theorem 4 follows the same intuitions of Theorem 4' (see the remarks after that theorem) except that besides the marginal rate of substitution identity, the principal must equalize the marginal rent identity in order to obtain truth telling. Observe that Theorem 4' is a particular case of Theorem 4 when $x = \hat{x}$.

Continuous Crossing: (but not constant)[11] Fix $\theta_0 \in \Theta$ as a parameter where x^* has a discontinuity and $\theta_1 \geq \theta_0$ is the minimal parameter that defines the U-shaped part of the optimal decision. Consider the space of càdlàg decisions x such that $x(\theta_1) = x^*(\theta_1)$, x is non-increasing on $[\underline{\theta}, \theta_1]$ and non-decreasing on $[\theta_2, \overline{\theta}]$, where $\theta_2 = \varphi(\theta_1, x^*(\theta_1))$.[12]

A relaxed program[13]

$$\begin{aligned}
&\max_x \; E_\theta[\frac{f(x(\theta), \theta)}{p(\theta)}] \\
&\text{s.t.} \\
&\dot{x}(\theta) \leq 0, \;\; \forall \theta \in [\underline{\theta}, \theta_1] \\
&\dot{x}(\theta) \geq 0, \;\; \forall \theta \in [\theta_2, \overline{\theta}] \\
&x(\theta) = x^u(\theta), \;\; \forall \theta \in [\theta_1, \theta_2] \\
&\Phi^x(\overline{\theta}, \theta_{0_}) = \int_{\theta_0}^{\overline{\theta}} \int_{x_-(\theta_0)}^{x(\tilde{\theta})} v_{x\theta}(\tilde{x}, \tilde{\theta}) d\tilde{x} d\tilde{\theta} \geq 0 \\
&x_-(\theta_0) \geq x(\overline{\theta}) \\
&x(\theta_i) = x^u(\theta_i), \;\; i = 1, 2
\end{aligned}$$

where $x^u(\cdot)$ is the implicit solution of the first order condition given in the Theorem 4' (observe that $x^u(\theta_1) = x^u(\theta_2)$). Observe that if there exists $\theta > \theta_2$ such that $x(\theta) > x_-(\theta_0)$, this would imply that there is a U-shaped part on $[\underline{\theta}, \theta_0)$, which is not possible (because this would generate another discontinuity - see the remark 4 after Theorem 4').

The Lagrangian will be:

$$\begin{aligned}
L(x, \mu, \lambda_1, \lambda_2, \Gamma) = &\int_{\underline{\theta}}^{\theta_1} [f(x(\theta), \theta) - \Gamma(\theta)\dot{x}(\theta)] d\theta + \int_{\theta_2}^{\overline{\theta}} [f(x(\theta), \theta) + \Gamma(\theta)\dot{x}(\theta)] d\theta \\
&+ \mu \int_{\theta_0}^{\overline{\theta}} \int_{x_-(\theta_0)}^{x(\tilde{\theta})} v_{x\theta}(\tilde{x}, \tilde{\theta}) d\tilde{x} d\tilde{\theta} + \lambda_0 [x_-(\theta_0) - x(\overline{\theta})] \\
&+ \lambda_1 [x^u(\theta_1) - x(\theta_1)] + \lambda_2 [x^u(\theta_2) - x(\theta_2)]
\end{aligned}$$

where $\Gamma \geq 0$ is the monotonicity multiplier (with $\dot{\Gamma} = \gamma$), $\mu \geq 0$ is the Lagrange multiplier of the IC constraint of the pair $(\overline{\theta}, \theta_{0\_})$, $\lambda_0 \geq 0$ and $\lambda_i \in \Re$ $(i = 1, 2)$ are the Lagrange multipliers associated to the other three constraints.

Taking the Gateaux derivative with respect to the space of admissible directions: $H = \{h \colon \Theta \to \Re;\ C^1 \text{ on } [\underline{\theta}, \theta_1] \cap [\theta_2, \overline{\theta}] \text{ and } h|_{(\theta_1, \theta_2)} \equiv 0\}$, the first order

[11] The constant crossing will be considered in the next case.

[12] $\varphi(\theta, x)$ is defined in the proof of Theorem 4 and the U-shaped part of the optimal decision is characterized by Theorem 4'.

[13] The relaxed program is given by a less relaxed program than the original one and it essentially consists of adding the monotonicity conditions, the IC constraint of the highest type with respect to the type where the discontinuity occurs, and the U-shaped characterization of Theorem 4'.

condition is (δ means the Gateaux derivative):

$$\begin{aligned}\delta_h L(x) = &\int_{\underline{\theta}}^{\theta_1} [f_x(x(\theta),\theta) + \gamma(\theta)]h(\theta)d\theta - \Gamma_-(\theta_0)h_-(\theta_0) - \Gamma(\theta_1)h(\theta_1) \\ &+ \int_{\theta_2}^{\overline{\theta}} [f_x(x(\theta),\theta) - \gamma(\theta)]h(\theta)d\theta - \Gamma(\theta_2)h(\theta_2) \\ &+ \mu\int_{\theta_0}^{\overline{\theta}} [v_{x\theta}(x(\theta),\theta)h(\theta) - v_{x\theta}(x_-(\theta_0),\theta)h_-(\theta_0)]d\theta \\ &+ \lambda_0[h_-(\theta_0) - h(\overline{\theta})] - \lambda_1 h(\theta_1) - \lambda_2 h(\theta_2) = 0\end{aligned}$$

and the Kuhn and Tucker slackness conditions are:

$$\begin{cases} \Gamma(\theta)\dot{x}(\theta) = 0 \\ \mu\Phi^x(\overline{\theta},\theta_{0_-}) = 0 \\ \lambda_0[x_-(\theta_0) - x(\overline{\theta})] = 0 \\ x(\theta_i) = x^u(\theta_i),\ i = 1,2 \end{cases}$$

Then,

$$\begin{cases} f_x(x,\theta) + \gamma(\theta) = 0, \quad \text{for } \theta < \theta_0 \\ f_x(x,\theta) + \gamma(\theta) + \mu v_{x\theta}(x,\theta) = 0, \quad \text{for } \theta_0 \le \theta < \theta_1 \\ f_x(x,\theta) - \gamma(\theta) + \mu v_{x\theta}(x,\theta) = 0, \quad \text{for } \theta_2 < \theta \\ -\Gamma_-(\theta_0) + \mu[v_x(x_-(\theta_0),\theta_0) - v_x(x_-(\theta_0),\overline{\theta})] + \lambda_0 = 0, \quad \text{for } \theta = \theta_{0_-} \\ \lambda_i + \Gamma(\theta_i) = 0, \quad \text{for } \theta = \theta_i,\ i = 1,2 \\ \Gamma(\overline{\theta}) - \lambda_0 = 0, \quad \text{for } \theta = \overline{\theta} \end{cases}$$

and

$$\Gamma(\theta) = \begin{cases} \int_{\theta}^{\hat{\theta}_0} f_x(\bar{x}_1,\tilde{\theta})d\tilde{\theta}, \quad \theta \in [\hat{\theta}_0,\theta_0] \\ \Gamma(\theta_1) + \int_{\theta}^{\theta_1} [f_x(\bar{x}_2,\tilde{\theta}) + \mu v_{x\theta}(\bar{x}_2,\tilde{\theta})]d\tilde{\theta}, \quad \theta \in [\theta_0,\theta_1] \\ \Gamma(\theta_2) + \int_{\theta_2}^{\theta} [f_x(\bar{x}_2,\tilde{\theta}) + \mu v_{x\theta}(\bar{x}_2,\tilde{\theta})]d\tilde{\theta}, \quad \theta \in [\theta_2,\hat{\theta}_1] \\ \int_{\hat{\theta}_2}^{\theta} [f_x(\bar{x}_1,\tilde{\theta}) + \mu v_{x\theta}(\bar{x}_1,\tilde{\theta})]d\tilde{\theta}, \quad \theta \in [\hat{\theta}_2,\overline{\theta}] \end{cases}$$

where $\gamma \equiv 0$ for $\theta \in [\underline{\theta},\hat{\theta}_0]$ or $\theta \in [\hat{\theta}_1,\hat{\theta}_2]$ and $\bar{x}_i$'s are the constant parts of the optimal decision.[14]

We can calculate the second order derivative of the Lagrangian:

$$\begin{aligned}\delta_{hh} L(x) = &\int_{\underline{\theta}}^{\theta_1} f_{xx}(x(\theta),\theta)h(\theta)^2 d\theta + \int_{\theta_2}^{\overline{\theta}} f_{xx}(x(\theta),\theta)h(\theta)^2 d\theta \\ &+ \mu\int_{\theta_0}^{\overline{\theta}} [v_{x^2\theta}(x(\theta),\theta)h(\theta)^2 - v_{x^2\theta}(x_-(\theta_0),\theta)h_-(\theta_0)^2]d\theta\end{aligned}$$

[14] Using the first order condition of the Lagrangian, one can easily verify that there exists no other constant part in the optimal decision.

Since $\mu \geq 0$ and $v_{x^2\theta} < 0$, if $f_{xx} \leq 0$ and $f_{xx}(\cdot,\theta_0)+\mu(v_{x^2}(\cdot,\theta_0)-v_{x^2}(\cdot,\overline{\theta})) \leq 0$, then the second order derivative is always negative. This means that our Lagrangian is a concave functional and the first order condition is necessary and sufficient (for each θ_0 and θ_1).

Therefore, the candidate of second best solution of our program in this case is

$$x(\theta) = \begin{cases} x_1(\theta), & \theta < \hat{\theta}_0 \\ \bar{x}_1, & \hat{\theta}_0 \leq \theta < \theta_0 \text{ or } \hat{\theta}_2 \leq \theta \leq \overline{\theta} \\ x^u(\theta) & \theta_1 \leq \theta \leq \theta_2 \\ \bar{x}_2, & \theta_0 \leq \theta \leq \theta_1 \text{ or } \theta_2 \leq \theta \leq \hat{\theta}_1 \\ x_1^{\mu}(\theta), & \hat{\theta}_1 \leq \theta \leq \hat{\theta}_2 \end{cases}$$

where x_1^{μ} is the implicit solution of $f_x(\cdot,\theta) + \mu v_{x\theta}(\cdot,\theta) = 0$.

FIGURE 4.– The critical decision in the case of continuous crossing.

LEMMA 3: *(Implementability of a critical decision) Under A1-A3, for each* θ_0 *and* θ_1*, the decision defined above is implementable.*

Finally, one must optimize with respect to θ_0. This is done in the numerical Examples 2 and 4 (observe that this program is not concave anymore).

DISCONTINUOUS CROSSING:[15] Fix $\theta_0 \in \Theta$ as a parameter where x^* may cross x_0 discontinuously. We are going to solve the following relaxed program:

$$\begin{aligned} &\max_x \, E_\theta[\frac{f(x(\theta),\theta)}{p(\theta)}] \\ &\text{s.t.} \\ &\dot{x}(\theta) \leq (\geq) 0, \ \forall \theta \leq (\geq) \theta_0 \\ &\Phi^x(\overline{\theta}, \theta_{0\_}) = \int_{\theta_0}^{\overline{\theta}} \int_{x_-(\theta_0)}^{x(\tilde{\theta})} v_{x\theta}(\tilde{x},\tilde{\theta}) d\tilde{x} d\tilde{\theta} \geq 0 \\ &x_-(\theta_0) \geq x(\overline{\theta}) \end{aligned}$$

Assume that the optimal decision is piecewise continuously differentiable with at most one discontinuity. More precisely, we assume that discontinuity occurs at θ_0. The Lagrangian will be

$$\begin{aligned} L(x,\mu,\lambda,\Gamma) = &\int_{\underline{\theta}}^{\overline{\theta}} [f(x(\theta),\theta) + \Gamma(\theta)\dot{x}(\theta)] d\theta \\ &+ \mu \int_{\theta_0}^{\overline{\theta}} \int_{x_-(\theta_0)}^{x(\tilde{\theta})} v_{x\theta}(\tilde{x},\tilde{\theta}) d\tilde{x} d\tilde{\theta} \\ &+ \lambda[x_-(\theta_0) - x(\overline{\theta})] \end{aligned}$$

[15] We can adapt this case to the continuous crossing one.

where $\mu \geq 0$ is the Lagrange multiplier of the IC constraint of the pair $(\overline{\theta}, \theta_{0\_})$,$\Gamma\colon \Theta \to \Re$ is the Lagrangian multiplier of the monotonicity of x with $\dot{\Gamma} = \gamma$ ($\Gamma \leq 0$ for $\theta < \theta_0$ and $\Gamma \geq 0$ for $\theta \geq \theta_0$) and $\lambda \geq 0$ is the multiplier of the third constraint.

Taking the Gateaux derivative with respect to the space of admissible directions: $H = \{h\colon \Theta \to \Re;\ C^1\}$, the first order condition is:

$$\begin{aligned}\delta_h L(x) = &\int_{\underline{\theta}}^{\overline{\theta}} [f_x(x(\theta),\theta) - \gamma(\theta)]h(\theta)d\theta + \Gamma_-(\theta_0)h_-(\theta_0)\\ &+ \mu \int_{\theta_0}^{\overline{\theta}} [v_{x\theta}(x(\theta),\theta)h(\theta) - v_{x\theta}(x_-(\theta_0),\theta)h_-(\theta_0)]d\theta\\ &+ \lambda[h_-(\theta_0) - h(\overline{\theta})] = 0\end{aligned}$$

and the Kuhn and Tucker slackness conditions are:

$$\begin{cases} \Gamma(\theta)\dot{x}(\theta) = 0 \\ \mu \Phi^x(\overline{\theta}, \theta_{0\_}) = 0 \\ \lambda[x_-(\theta_0) - x(\overline{\theta})] = 0 \end{cases}$$

Then,

$$\begin{cases} f_x(x,\theta) = 0, \quad \text{for} \quad \theta < \hat{\theta}_0 \\ f_x(x,\theta) - \gamma(\theta) = 0, \quad \text{for} \quad \hat{\theta}_0 \leq \theta < \theta_0 \\ f_x(x,\theta) - \gamma(\theta) + \mu v_{x\theta}(x,\theta) = 0, \quad \text{for} \quad \theta_0 \leq \theta \leq \overline{\theta} \\ \Gamma_-(\theta_0) + \mu[v_x(x_-(\theta_0),\theta_0) - v_x(x_-(\theta_0),\overline{\theta})] + \lambda = 0, \quad \text{for} \quad \theta = \theta_{0\_} \\ \Gamma(\overline{\theta}) - \lambda = 0, \quad \text{for} \quad \theta = \overline{\theta} \end{cases}$$

Observe that

$$\Gamma(\theta) = \begin{cases} \int_{\hat{\theta}_0}^{\theta} f_x(\bar{x}_1, \tilde{\theta})d\tilde{\theta}, \quad \theta \in [\hat{\theta}_0, \theta_0] \\ \int_{\theta_0}^{\theta} [f_x(\bar{x}_2, \tilde{\theta}) + \mu v_{x\theta}(\bar{x}_2, \tilde{\theta})]d\tilde{\theta}, \quad \theta \in [\theta_0, \hat{\theta}_1] \\ \int_{\hat{\theta}_2}^{\theta} [f_x(\bar{x}_1, \tilde{\theta}) + \mu v_{x\theta}(\bar{x}_1, \tilde{\theta})]d\tilde{\theta}, \quad \theta \in [\hat{\theta}_2, \overline{\theta}] \end{cases}$$

where $\gamma \equiv 0$ for $\theta \in [\hat{\theta}_1, \hat{\theta}_2]$ or $\theta \in [\underline{\theta}, \hat{\theta}_0$ and $\bar{x}_i$'s are the constant parts of the optimal decision.

We can calculate the second order derivative of the Lagrangian:

$$\delta_{hh} L(x) = \int_{\underline{\theta}}^{\overline{\theta}} f_{xx}(x(\theta),\theta)h(\theta)^2 d\theta + \mu \int_{\theta_0}^{\overline{\theta}} [v_{x^2\theta}(x(\theta),\theta)h(\theta)^2 - v_{x^2\theta}(x_-(\theta_0),\theta)h_-(\theta_0)^2]d\theta$$

and again the same remark of the continuous crossing case is valid here.

Therefore, the candidate for second best solution of our program is

$$x(\theta) = \begin{cases} x_1(\theta), & \theta < \hat{\theta}_0 \\ \bar{x}_1, & \hat{\theta}_0 \leq \theta \leq \theta_0 \text{ or } \hat{\theta}_2 \leq \theta \\ \bar{x}_2, & \theta_0 \leq \theta < \hat{\theta}_1 \\ x_1^{\mu}(\theta), & \hat{\theta}_1 \leq \theta \leq \hat{\theta}_2 \end{cases}$$

FIGURE 5.– The critical decision in the case of discontinuous crossing.

LEMMA 4: *(Implementability of a critical decision) Under A1-A3, the decision defined above is implementable, for each* θ_0.

Again, one can optimize with respect to θ_0 and obtain the optimal decision in the class of decisions that cross x_0 discontinuously.

Our procedure is to use these first order conditions to implement the critical solution in numerical examples (see Examples 2 and 4). Lemmas 3 and 4 guarantee that the decisions characterized above are implementable under assumption A3.

4. EXAMPLES

EXAMPLE 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 second best solutions with and without the technology choice information:

1) *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 a specialist will have positive rent.

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

If $p \equiv 1$ and randomization is not possible, then

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

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

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

where $\theta_0^{4/3} + \theta_0 = 1$.

2) *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 4' in order to characterize the second best solution. Here, x_0 is defined by $\theta = 1/2$.

Since $\tilde{v}_\theta(x, \theta) \lesseqqgtr 0$ if and only if $\theta \gtreqqless 1/2$, the IR constraint will be binding at 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 4' gives

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

The IC constraint defines a convex set and the objective function is concave. Therefore, the use of lotteries on production does not improve the principal's welfare (which 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. Thus, the loss of profit caused by non verifiability of information is equal to that caused by the lack of commitment in using lotteries when information is available (see Appendix B).

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

If $p(\theta)=2\theta$, for all $\theta\in\Theta$, then the second best without information is

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

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 a one-dimensional problem in θ making T a function of θ. Finally, the incapacity to commit to announcement of the technology makes the principal equalize the marginal utilities of the pooling types in order to guarantee truth telling. This is exactly what explains the "countervailing" incentives in this example.

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 this is to consider (different) sunk costs for each technology.

The second best problem without information was inspired by 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 have 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.

FIGURE 6.– The corporate finance example for $p(\theta)=2\theta$.

The top figure presents the second best with (xsb) and without (xtb) technology choice information. The other two figures present the IC constraints of xtb and xsb (i.e., the graphs of Φ^{xtb} and Φ^{xsb}), respectively. The minimal value in right bottom figure is -0.3950.

EXAMPLE 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 $\frac{c}{2}x^2$ for x units. A buyer of type $\theta\in[a,a+1]$ $(a\geq 0)$ 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 three times 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 $\dot\alpha(\theta) > \alpha(\theta)/\theta, \quad \forall\,\theta \in (a, a+1]$ and this last inequality is true because α is convex. Since $\theta/2\alpha(\theta)$ is the market size of the type θ demand (i.e., the number of units bought at price zero), what we are assuming is that the market size decreases with θ and θ is the supreme of prices for which there exists a positive demand. This assumption can imply no SMC, and it is in contrast with the monotonicity assumption of $\pi(x,\cdot)$ in Maskin and Riley (1984).[16]

Define

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

Therefore,

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

for all $\theta \in [a, a+1]$.

Since $\ddot\alpha > 0$, x_0 is decreasing and $\frac{1}{2\dot\alpha(\theta)} < \frac{\theta}{2\alpha(\theta)} < \frac{\theta}{\alpha(\theta)}, \quad \forall\,\theta \in [a, a+1]$, which implies that the SMC fails to hold here. (Observe that $v_x(x,\theta) \geq 0$ if and only if $x \leq \frac{\theta}{\alpha(\theta)}$). 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 $\theta \in [a, a+1]$.

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

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

Assumption *a1* is the 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 $\theta \in [a, a+1]$.

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

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

Hence, we guarantee that x_1 crosses x_0 if we assume

a3: $\quad c < \dot{\alpha}(a+1) - \alpha(a+1)$

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

Under *a1*, *a2* and *a3*, Lemmas 3 and 4 can be applied in order to characterize the second best solution. Observe that x_1 will be bell-shaped and that $v_\theta(x, \theta) = [1 - \dot{\alpha}(\theta)x]x \geq 0$ if and only if $x \leq 1/\dot{\alpha}(\theta)$ (thus, the lowest type a has zero rent).

Consider the particular case that satisfies *a1*-*a3*: $p \equiv 1, \quad c = 1$ and $\alpha(\theta) = \frac{b}{2}\theta^2$.

For instance, if $b = 2$, then $v(x, \theta) = \tilde{v}(\theta x)$, where $\tilde{v}(y) = y - y^2$.

Therefore, the relaxed solution is

$$x_1(\theta) = \left[\frac{2\theta - a - 1}{c + b(3\theta - a - 2)\theta}\right]^+$$

and

$$x_0(\theta) = \frac{1}{2b\theta}$$

Using Theorem 4', the bell-shaped part of the solution is given by

$$x^*(\theta) = \frac{3\theta + \sqrt{3\theta^2 - 2c/b}}{2(3b\theta^2 + c)}.$$

We present the two candidates of the second best decision for $a = 1$, $b = 1$ and $c = 1.5$:[17]

FIGURE 7.1.– The curves *xopt*0 and x_1 and their respective IC constraints.

The top figure gives the optimal decision (*xopt*0) given in Theorem 5; its expected profit is 0.19906; the relaxed and the first best profits are 0.20302 and 0.29387, respectively. The bottom left figure is the IC constraint of *xopt*0 (the graph of Φ^{xopt0} which minimal value is zero, i.e., *xopt*0 is implementable). The bottom right figure is the IC constraint of the relaxed solution (the graph of Φ^{x_1} which minimal value is -0.0096129, i.e., it is not implementable).

[17] They are calculated using a routine in MATLAB.

FIGURE 7.2.– The curves *xopt*1 and *xopt*2 and their respective IC constraints.

The top (bottom) left figure gives the optimal decision - *xopt*1 (*xopt*2) - characterized by Lemma 3 (4); its expected profit is 0.20302 (0.20265). The top (bottom) right figure is the respective IC constraint - the graph of Φ^{xopt1} (Φ^{xopt2}) - which minimal value is zero, i.e., *xopt*1 (*xopt*2) is implementable.

Where x_0 is the curve that separates CS$_-$ and CS$_+$, x_1 is the relaxed solution and x_{FB} is the first best solution. In Appendix C we present a table showing a comparative static exercise in c (the cost parameter). As we can see in the table both the first and second best profit decreases with c. Observe also that the value $c = 4.25$ corresponds to the SMC case (see the Table 1 in Appendix C).

EXAMPLE 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 is a project with social value 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 (but C_1 and C_2 are not), θ_1 and θ_2 are the cost parameters known only to the firm, and e_1 and e_2 are unobservable actions of the firm representing the efforts to reduce the sub-costs 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 three times 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:

$$\min_{C_1, C_2 \geq 0} \psi(\theta_1 - C_1, \theta_2 - C_2)$$
$$\text{s.t. } C_1 + C_2 = C$$

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 sub-cost parameter of production θ_2 to be a function of the sub-cost parameter θ_1: $\theta_2 = \eta(\theta_1)$, twice differentiable.

This countervailing incentive will be associated with the no SMC in two cases (in the other two cases, the SMC is valid).

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

(b) complementary efforts and positive correlation between the sub-costs: $\psi_{12} < 0$ and $\dot{\eta} > 0$.

Let us consider case (a). This means that there is countervailing incentive in the effort allocation for sub-cost reduction, since these activities are substitutes and the cost parameters move in opposite directions. For instance, a family of examples is 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 $\xi(e) = \xi_1 e_1^3 + \xi_{12} e_1^2 e_2 + \xi_{21} e_2^2 e_1 + \xi_2 e_2^3$ are the third order terms with $\xi_i \geq 0$, $\xi_{ij} \geq 0$, $i, j = 1, 2$.

The consequence is that the SMC can 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 sub-cost parameter and observable aggregate cost. If

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

then, by the Envelope Theorem (to simplify the notation we will omit the arguments of the functions), $v_C = \psi_1$ 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).

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

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

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.

The first order condition of the relaxed problem is

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

By Theorem 4', if θ_1 and $\hat{\theta_1}$ are pooling on the same aggregate cost, then

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

where the hat means that the functions are calculated at the $\hat{\theta}_1$ allocation. And, after some trivial manipulations, $\psi_1 = 1 + \dfrac{\lambda}{1+\lambda}(M - \widehat{M})\{v_{C\theta_1}^{-1} - \widehat{v}_{C\theta_1}^{-1}\}^{-1}$ 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 latter).

Let us take a particular example (the symmetric case): The type set is $\Theta = [0,1]$ with the uniform distribution, $\psi(e_1, e_2) = (e_1 + e_2)^2$ and $\theta_2 = \eta(\theta_1) = 1 - \theta_1^2$. Thus the firm's minimization effort program has the following solution: $C_1 = \theta_1$ and $C_2 = C - \theta_1$ (since $C_2 \le 1 - \theta_1^2$ for $C \le 1$).

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 $\widehat{\theta_1} = 1 - \theta_1$. Applying Theorem 5 we get the second best cost:

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

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

$$C1(\theta_1) = \begin{cases} \frac{1}{2} + \left(1 - \frac{\lambda}{2(1+\lambda)}\right)\theta_1 - \left(1 - \frac{\lambda}{1+\lambda}\right)\theta_1^2 & \text{if} \quad 0 \le \theta_1 \le 1/2 \\ \frac{1}{2} + \left(1 - \frac{\lambda}{2(1+\lambda)}\right)\theta_1 - \left(1 - \frac{\lambda}{1+\lambda}\right)\theta_1^2 + \frac{\lambda}{2(1+\lambda)}\frac{1-2\theta_1}{\theta_1} & \text{if} \quad 1/2 \le \theta_1 \le 1 \end{cases}$$

and the second best solution is:

$$C^*(\theta_1) = \frac{1}{2} + (1 - \frac{\lambda}{2(1+\lambda)})\theta_1 - (1 - \frac{\lambda}{1+\lambda})\theta_1^2 + \frac{\lambda}{2(1+\lambda)}\frac{\theta_1}{1-\theta_1},$$

for $\theta_1 \in [0, 1/2]$ and $C^*(\theta_1) = C^*(1-\theta_1)$, for $\theta_1 \in [1/2, 1]$.

FIGURE 8.– The regulation problem for $\lambda = 0.05$.

The top figure presents the relaxed ($C1$) and second best (C^*) solutions. The other two figures present the IC constraints of $C1$ and C^* (the graph of Φ^{C1} and Φ^{C^*}), respectively. The minimal value of the bottom right figure is -0.0012.

The economic interpretation is immediate: the sub-cost reductions are substitute activities and the sub-cost structure presents a countervailing property: the extreme types correspond to specialists 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 sub-cost that is more inefficient); the middle type ($\theta_1 = 1/2 = \theta_2$) is the only one that has zero rent (this is analogous to Example 1).

EXAMPLE 4: LABOR CONTRACT[18]

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 ability") and η ("non-cognitive ability"):

$$s = \theta + \eta$$

where $\theta \in [\underline{\theta}, \overline{\theta}]$ has cumulative distribution P and density function p. Therefore, 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 worker's effort (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 to 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

[18] This example is motivated in part by Cavallo, Heckman and Hsee (1998) who present evidence on the GED as a mixed signal of cognitive and non-cognitive abilities (GED is an exam taken by American high school dropouts to certify their equivalence with high school graduates): comparing the GED recipients and other dropouts, there is no wage differential between them.

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 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 correctly *ex-ante* the worker's abilities.

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 $y = \alpha s - x$.[19] Thus, we can express 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

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

This means that the SMC can 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

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

Since $v_\theta = -2\theta^{-2}ey$, the marginal rent of the worker is negative if and only if $x \leq \alpha s$ (thus v_θ changes its sign - see remark 3 after Theorem 4'). 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 crosses 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 4' 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:[20]

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

---

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

[20] The monotonicity condition eliminates the other roots.

We present a particular example: $\pi = 1$, $\alpha = 3.6$, $s = 1.5$, $\pi = 2$, $\underline{\theta} = 1$ and $\overline{\theta} = 2$. The candidates for the second best decision are:[21]

FIGURE 9.1.– The curves $xopt0$ and x_1 and their respective IC constraints.

The top figure is the optimal decision ($xopt0$) given by Theorem 5; its expected profit is 4.1834; the relaxed and the first best profits are 4.1957 and 5.3333, respectively. The bottom left figure is the IC constraint of $xopt0$ (the graph of Φ^{xopt0} which minimal value is zero). The bottom right figure is the IC constraint of the relaxed solution (the graph of Φ^{x_1} which minimal value is -0.096005, i.e., it is not implementable).

FIGURE 9.2.– The curves $xopt0$ and $xopt1$ and their respective IC constraints.

The top (bottom) left figure gives the optimal decision - $xopt1$ ($xopt2$) - characterized by Lemma 3 (4); its expected profit is 4.1876 (4.1942). The top (bottom) right figure is the respective IC constraint - the graph of Φ^{xopt1} (Φ^{xopt2}) - which minimal is zero, i.e., $xopt1$ ($xopt2$) is implementable.

In Appendix C we do a comparative static exercise with α (the shadow price of the non-cognitive ability). Observe that the parameter $\alpha = 2$ corresponds to the no SMC situation (see the Table 2 in Appendix C).

We summarize this example:

• Discrete pooling: two different workers with different profiles of characteristics for a given s choose 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 type workers to extract all their rent. These types provide the highest output.

• The worker θ^0 (a middle type) provides the lowest output and has the highest rent.

• The rent extraction and distortion trade-off take into consideration the new conditions for implementability.

5. CONCLUSIONS

In this paper we studied a generalization of the SMC. Although this leads to a non-convex problem, we are able to give economically meaningful solution. We first give a necessary condition for implementability which could be translated into the marginal rate of substitution and rent identities. These conditions are also sufficient for some cases. Next we derive the Lagrangian for the second best

[21] As in Example 2, we used a MATLAB program to calculate these candidates.

problem taking into consideration these necessary conditions. This leads to the equality of the marginal virtual surplus rate between two pooling types and the rate of the marginal rent between them.

Four numerical examples illustrated our method: a principal-agent problem with simultaneous adverse selection and moral hazard, the nonlinear pricing problem, sub-cost observation in a regulation problem and mixed signal in a labor contract model. In all examples multidimensional characteristics and countervailing incentives are presented.

Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina, 110, Rio de Janeiro, 22460-320, Brasil; aloisio@impa.br

Fundação Getúlio Vargas, Praia de Botafogo, 190, Rio de Janeiro, 22253-900, Brasil

and

Pontifícia Universidade Católica do Rio de Janeiro, Rua Marquês de São Vicente, 225, Rio de Janeiro, 22453-900, Brasil; humberto@econ.puc-rio.br

APPENDICES

Appendix A
Proofs

1. *Proof of Lemma 1:*

(i) 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 $\mathcal{V}^x$ 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).

(ii) From (i), $t(\theta) = \mathcal{V}^x(\theta) - v(x(\theta),\theta)$, for all $\theta \in \Theta$. 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, there exists $\theta_0 > \hat{\theta}_0$ such that the convex hull of $\{(\theta_2, x(\theta_1));\ \theta_1,\theta_2 \in [\hat{\theta}_0,\theta_0)$ and $\theta_1 \leq \theta_2\}$ is in the region $v_{x\theta} > 0$. Let (a,b) be a maximal interval in I such that $x(\hat{\theta}_0) > x(\theta)$, for all $\theta \in (a,b)$. If

$a = \hat{\theta}_0$, then the double integral above will be negative for $\hat{\theta} = a$ and $\theta = b$. If $a > \hat{\theta}_0$, then the left limit of the double integral will be also negative for $\hat{\theta} = a_-$ and $\theta = b$ (since (a, b) is maximal and $x_-(a) \geq x(\hat{\theta}_0)$). In both cases we have a contradiction with the IC constraint. Therefore, $x(\hat{\theta}) \leq x(\theta)$, for all $\theta \in I$. *Q.E.D.*

2. *Proof of Theorem 1:*

(i) Define a local inverse for x at $\hat{\theta}$. Applying Fubini's Theorem and taking the right (left) derivative at $x(\hat{\theta})$ and observing that $x(\hat{\theta})$ is a minimum point for $\Phi(\cdot, \theta)$, we get the result.

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

(iii) Observe that if $y = x(\theta) = x(\hat{\theta})$, with $\theta, \hat{\theta} \in \Theta$, then

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

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

Since x is implementable, $\Phi^x(\theta, \hat{\theta}) = 0$.

If x is right and left increasing at $\hat{\theta}$, applying (i), we get the result. *Q.E.D.*

3. *Proof of Theorem 2:*

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

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

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 $\hat{\theta}_n \longrightarrow \hat{\theta}$. 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 1 (ii) implies that x is non-increasing or non-decreasing or U-shaped. Let $[\theta_1, \theta_2] \subset \Theta$. If the IC constraint is not binding on this interval, we can approximate the discontinuities of x by a continuous monotone part. Otherwise, we can use the marginal rate of substitution and rent identities to approximate x by a monotone decision on this interval. This procedure in both cases leads to an implementable continuous decision as close as one wants to x in the weak topology sense. *Q.E.D.*

4. *Proof of Lemma 2:*

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)).

It is easy to see that if x_n is a sequence of implementable uniformly bounded càdlàg decisions converging to x, then $x \in \mathcal{C}$ is bounded. Moreover, for each $\theta, \hat{\theta} \in \Theta$ continuous parameters for x, $\Phi^{x_n}(\theta, \hat{\theta}) \to \Phi^{x}(\theta, \hat{\theta})$ because $x_n(\theta) \to x(\theta)$ and $x_n(\hat{\theta}) \to x(\hat{\theta})$. Since x is a càdàg decision, the IC constraint is satisfied at x for every θ and $\hat{\theta}$. If x_n crosses x_0 at most one time, for each n, then it is easy to see that x has the same property (otherwise, there wuold exist n such that x_n crosses x_0 more than one time). Therefore, the set of implementable decisions in $\mathcal{C}$ that cross x_0 at most one time is closed.

Finally, the objective function of (P) is continuous with respect to the considered topology. Thus, by the Weierstrass Theorem, there exists an optimal decision for (P) in the set of uniformly bounded càdlàg decisions. $Q.E.D.$

5. *Proof of Theorem 4:*

Consider the following transformation: T from $\Re^4$ to $\Re^2$, $T = (T_1, T_2)$, where

$$\begin{cases} T_1(\theta, x, \hat{\theta}, \hat{x}) = v_x(\hat{x}, \theta) - \hat{v}_x \\ T_2(\theta, x, \hat{\theta}, \hat{x}) = v_\theta - v_\theta(\hat{x}, \theta) \end{cases}$$

where hat means that the function is calculated at $(\hat{x}, \hat{\theta})$ and without hat means that it is calculated at (x, θ).

The Jacobian of T with respect to the variables $(\hat{\theta}, \hat{x})$ is:

$$J_{(\hat{\theta},\hat{x})}T = \begin{pmatrix} -\hat{v}_{x\theta} & v_{xx}(\hat{x}, \theta) - \hat{v}_{xx} \\ 0 & -v_{x\theta}(\hat{x}, \theta) \end{pmatrix}$$

with determinant $|J_{(\hat{\theta},\hat{x})}T| = \hat{v}_{x\theta} v_{x\theta}(\hat{x}, \theta) < 0$. By the Implicit Function Theorem, for each $(\theta, x, \hat{\theta}, \hat{x})$ such that $T(\theta, x, \hat{\theta}, \hat{x}) = (0, 0)$, there exists a local neighborhood of (θ, x) where we can put $(\hat{\theta}, \hat{x})$ as a continuous differentiable function of (θ, x): $(\hat{\theta}, \hat{x}) = \varphi(\theta, x)$, $\varphi = (\varphi_1, \varphi_2)$. In this case:

$$\left(J_{(\hat{\theta},\hat{x})}T\right)^{-1} = \frac{1}{\hat{v}_{x\theta} v_{x\theta}(\hat{x}, \theta)} \begin{pmatrix} -v_{x\theta}(\hat{x}, \theta) & \hat{v}_{xx} - v_{xx}(\hat{x}, \theta) \\ 0 & -\hat{v}_{x\theta} \end{pmatrix}$$

and

$$J_{(\theta,x)}T = \begin{pmatrix} \hat{v}_{x\theta}(\hat{x}, \theta) & 0 \\ v_{\theta\theta} - v_{\theta\theta}(\hat{x}, \theta) & v_{x\theta} \end{pmatrix}$$

which implies that

$$J_{(\theta,x)}\varphi = \frac{1}{\hat{v}_{x\theta} v_{x\theta}(\hat{x}, \theta)} \begin{pmatrix} v_{x\theta}(\hat{x}, \theta)^2 + [\hat{v}_{xx} - v_{xx}(\hat{x}, \theta)][v_{\theta\theta} - v_{\theta\theta}(\hat{x}, \theta)] & v_{x\theta}[\hat{v}_{xx} - v_{xx}(\hat{x}, \theta)] \\ \hat{v}_{x\theta}[v_{\theta\theta}(\hat{x}, \theta) - v_{\theta\theta}] & -v_{x\theta}\hat{v}_{x\theta} \end{pmatrix}$$

and $|J_{(\theta,x)}\varphi| = v_{x\theta}/\hat{v}_{x\theta}$.

In order to get the first order conditions, we have to take the Gateaux derivative of the functional defined from the maximization problem above. Assume that $x|_{[\theta^1,\theta^2]}$ is increasing and define the space of admissible perturbations:[22]

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

and the objective functional:

$$\begin{aligned} F(x) &= \int_{\theta^1}^{\theta^2} f(x(\theta), \theta) d\theta + \int_{\varphi(\theta^2, x^*(\theta^2))}^{\varphi(\theta^1, x^*(\theta^1))} f(x(\theta), \theta) d\theta \\ &= \int_{\theta^1}^{\theta^2} \{f(x(\theta), \theta) - [\partial_x \varphi^1(\theta, x(\theta))\dot{x}(\theta) + \partial_\theta \varphi^1(\theta, x(\theta))] f(x(\theta), \varphi(\theta, x(\theta)))\} d\theta \end{aligned}$$

where $x = x^* + h$

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

$$\begin{aligned} \delta_h F(x^*) = \int_{\theta^1}^{\theta^2} \{f_x h - [(\partial_{xx}\varphi^1 \dot{x}^* + \partial_{x\theta}\varphi^1)h + \partial_x \varphi^1 \dot{h}]\hat{f} \\ - [\partial_x \varphi^1 \dot{x}^* + \partial_\theta \varphi^1](\hat{f}_x \partial_x \varphi^2 + \hat{f}_\theta \partial_x \varphi^1)h\} d\theta = 0. \end{aligned}$$

By an integration by parts we have

$$\begin{aligned} -\int_{\theta^1}^{\theta^2} [\hat{f}\partial_x \varphi^1]\dot{h} d\theta &= \int_{\theta^1}^{\theta^2} (\dot{\overline{\hat{f}\partial_x \varphi^1}}) h d\theta \\ &= \int_{\theta^1}^{\theta^2} \{[\hat{f}_x(\partial_x \varphi^2 \dot{x}^* + \partial_\theta \varphi^2) + \hat{f}_\theta(\partial_x \varphi^1 \dot{x}^* + \partial_\theta \varphi^1)]\partial_x \varphi^1 \\ &\quad + \hat{f}(\partial_{xx}\varphi^1 \dot{x}^* + \partial_{x\theta}\varphi^1)\} h d\theta. \end{aligned}$$

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

$$\int_{\theta^1}^{\theta^2} \{f_x - \hat{f}_x[\partial_\theta \varphi^1 \partial_x \varphi^2 - \partial_\theta \varphi^2 \partial_x \varphi^1]\} h d\theta = 0$$

or

$$\int_{\theta^1}^{\theta^2} [f_x - \hat{f}_x |J\varphi_{(\theta,x)}|] h d\theta = 0.$$

Thus,

$$A = f_x - \frac{v_{x\theta}}{\hat{v}_{x\theta}} \hat{f}_x = 0, \quad \text{which is equivalent to}$$

$$\frac{f_x}{v_{x\theta}} = \frac{\hat{f}_x}{\hat{v}_{x\theta}}. \qquad Q.E.D.$$

---

[22] By the uniqueness of $\hat{\theta}$ for each $\theta \in [\theta^1, \theta^2]$ such that $\Phi^{x^*}(\theta, \hat{\theta}) = 0$, if $T \equiv 0$ along $x + h$, then $x^* + h$ is implementable for $h \in H$ sufficiently small.

6. *Proof of Theorem 5:*[23]

If x is contained in CS$_-$, then it is non-increasing and the decision $y(\cdot) = x_1(\cdot) \vee x_0(\overline{\theta})$ is implementable and dominates x (y is a degenerate U-shaped decision[24].) Suppose now that x is in CS$_+$. Thus, x crosses or does not cross x_1. In both cases, we will construct an implementable decision that crosses x_0 and dominates x.

Let θ_m be the minimum for x_1. For each $\theta_1 \in [\theta_m, \overline{\theta}]$ we can define the following decision:

$$y(\theta) = \begin{cases} x_1(\theta) & \text{for} \quad \theta \geq \theta_1 \\ x_1(\theta_1) & \text{for} \quad \varphi(\theta_1, x_1(\theta_1)) \vee \underline{\theta} < \theta \leq \theta_1 \\ x_1(\varphi(\theta, x_1(\theta))) & \text{for} \quad \varphi(\overline{\theta}, x_1(\overline{\theta})) \vee \underline{\theta} \leq \theta \leq \varphi(\theta_1, x_1(\theta_1)) \vee \underline{\theta} \\ x_1(\theta) & \text{for} \quad \underline{\theta} \leq \theta \leq \varphi(\overline{\theta}, x_1(\overline{\theta})) \vee \underline{\theta} \end{cases}$$

where φ is defined in the proof of Theorem 4 above such that $x = \hat{x}$.

We claim that A3 implies that y is U-shaped when $\varphi(\theta_1, x_1(\theta_1)) > \underline{\theta}$. Indeed, define $\gamma(\theta) = \varphi(\theta, x_1(\theta))$. We have that $\dot{\gamma} = \varphi_\theta + \varphi_x \dot{x_1}$ (omitting the arguments of the functions) and, in particular,

$$\dot{\gamma}(\theta_m) = \varphi_\theta(\theta_m, x_1(\theta_m)) = \frac{v_{x\theta}(x_1(\theta_m), \theta_m)}{v_{x\theta}(x_1(\theta_m), \varphi(\theta_m, x_1(\theta_m)))} < 0.$$

because $\dot{x_1}(\theta_m) = 0$.

If there exist $\theta_1, \theta_2 \in \Theta$, $\theta_1 \neq \theta_2$, such that $\gamma(\theta_1) = \gamma(\theta_2)$ and $\dot{x}(\theta_i) \geq 0$, then the implicit solution of $v_x(\cdot, \gamma(\cdot)) - v_x(\cdot, \theta) = 0$ crosses x_1 at θ_i, $i = 1, 2$, which contradicts A3. Thus, γ is monotone where x_1 is increasing. Since $\dot{\gamma}(\theta_m) < 0$, γ is non-increasing or, equivalently, y is non-increasing on $[\underline{\theta}, \varphi(\theta_1, x_1(\theta_1))]$.

As in the proof of Lemma 3, y is implementable. If x does not cross x_1, let $\theta_1 = \theta_m$, and if it does cross, consider θ_1 as the parameter where the crossing occurs. In both cases, it is easy to see that y dominates x.

By Lemma 2', let x^* be the optimal decision and $[\theta_1, \theta_2] \subset \Theta$ and interval. We have the following possibilities on $[\theta_1, \theta_2]$:

(i) the IC constraint is not binding: making the standard variational calculus argument, $x^* \equiv x_1$ on $[\theta_1, \theta_2]$;

(ii) the IC constraint is binding: if x^* is part of the "U" (i.e., where the marginal rate of substitution is binding), using Theorem 4', then $x^* \equiv x^u$; otherwise, x^* is constant and it is not part of a "U" on $[\theta_1, \theta_2]$. Then, consider the maximal interval containing $[\theta_1, \theta_2]$ where x^* is constant. This case will lead to $x^* \equiv \bar{x}$ on this maximal interval.

Thus, the optimal decision has one of the shapes presented in the statement of this theorem (i.e., x_1 plus U-shaped part or constant part plus x_1).

Observe that the first case is equivalent to the single crossing situation and it is optimal in the set of the decision that are non-decreasing (and it is implementable when $\varphi(\theta_1, \bar{x}) \leq \underline{\theta}$). The second

---

[23] The basic idea of the proof is domination: the definition of a set of implementable decisions that the principal is better off, i.e., given any implementable decision, there exist an implementable decision in this set that is more close to x_1. In this proof, this set is given by x_1 plus a U-shaped part or constant part plus x_1.

[24] By U-shaped part we mean an interval where the marginal rate of substitution and rent identities are "binding" and a U-shaped decision is the one that contains a U-shaped part.

case is the U-shaped one characterized by Theorem 4 (i) (and its implementability by Theorem 3). We argue that in this case the decision presented is optimal in the set of U-shaped decision. Indeed, fix $\theta \in \Theta$. Observe that in the case of Theorem 4', by assumptions A1 and A2, if x is sufficiently small, the expression $A > 0$ (defined in the proof of Theorem 4) and if x is sufficiently large, $A < 0$.[25] This proves that the first order condition of the U-shaped part is also sufficient when it is the unique critical point for each θ that defines an implementable decision. *Q.E.D.*

7. *Proof of Lemma 3:*[26]

We have to show that $\Phi^x(\theta, \hat{\theta}) \geq 0$, for all $\theta, \hat{\theta} \in \Theta$. Let us consider the following cases:

0. $\hat{\theta}, \theta \in [\theta_1, \theta_2]$. Without loss of generality, assume that $\hat{\theta} \leq \theta$. There are two cases to consider:

0.1. $\theta \leq \varphi(\hat{\theta}, x(\hat{\theta}))$. Using Fubini's Theorem,

$$\begin{aligned} \Phi^x(\theta, \hat{\theta}) &= \int_{x_m}^{x(\hat{\theta})} \left[\int_{\varphi_1(\tilde{x})}^{\varphi_2(\tilde{x})} v_{x\theta}(\tilde{x}, \tilde{\theta}) d\tilde{\theta} \right] d\tilde{x} \\ &= \int_{x_m}^{x(\hat{\theta})} [v_x(\tilde{x}, \varphi_1(\tilde{x})) - v_x(\tilde{x}, \varphi_2(\tilde{x}))] d\tilde{x} \end{aligned}$$

where x_m is the minimum of x and φ_1, φ_2 are the two inverses of x on $[x_m, x(\theta_1)]$, where $\varphi_2(x) = \theta$, for all $x \in [x(\theta), x(\hat{\theta})]$. From Theorem 1 (iii), $v_x(\tilde{x}, \varphi_1(\tilde{x})) \geq v_x(\tilde{x}, \varphi_2(\tilde{x}))$, for all $\tilde{x} \in [x_m, x(\hat{\theta})]$ (with equality on $[x_m, x(\theta)]$). Thus, $\Phi^x(\hat{\theta}, \varphi(\hat{\theta}, x(\hat{\theta}))) \geq 0$. (Observe that if $\theta = \varphi(\hat{\theta}, x(\hat{\theta}))$, then $\Phi^x(\hat{\theta}, \varphi(\hat{\theta}, x(\hat{\theta}))) = 0$.)

0.2. $\theta \geq \varphi(\hat{\theta}, x(\hat{\theta}))$. We have that

$$\begin{aligned} \Phi^x(\theta, \hat{\theta}) &= \int_{\theta}^{\varphi(\hat{\theta}, x(\hat{\theta}))} \int_{x(\tilde{\theta})}^{x(\hat{\theta})} v_{x\theta}(\tilde{x}, \tilde{\theta}) d\tilde{x} d\tilde{\theta} + \int_{\varphi(\hat{\theta}, x(\hat{\theta}))}^{\hat{\theta}} \int_{x(\tilde{\theta})}^{x(\hat{\theta})} v_{x\theta}(\tilde{x}, \tilde{\theta}) d\tilde{x} d\tilde{\theta} \\ &= \int_{\theta}^{\varphi(\hat{\theta}, x(\hat{\theta}))} \int_{x(\tilde{\theta})}^{x(\hat{\theta})} v_{x\theta}(\tilde{x}, \tilde{\theta}) d\tilde{x} d\tilde{\theta} \geq 0 \end{aligned}$$

because $\Phi^x(\hat{\theta}, \varphi(\hat{\theta}, x(\hat{\theta}))) = 0$ and by the monotonicity on $[\varphi(\hat{\theta}, x(\hat{\theta})), \theta]$.

1. $\hat{\theta} \leq \theta_0$ which generates the following sub-cases:

1.1. $\theta \leq \hat{\theta}$. The monotonicity implies that, for each $\tilde{\theta} \in [\theta, \hat{\theta}]$, $x(\tilde{\theta}) \geq x(\hat{\theta})$ and the derivative of Φ^x with respect to θ is

$$\partial_\theta \Phi^x(\tilde{\theta}, \hat{\theta}) = v_\theta(x(\tilde{\theta}), \tilde{\theta}) - v_\theta(x(\hat{\theta}), \tilde{\theta}) \leq 0.$$

(because $v_{x\theta} < 0$ on this region). Since $\Phi^x(\theta, \theta) = 0$, $\Phi^x(\hat{\theta}, \theta) \geq 0$.

---

[25] Since $v_{x\theta}$ and $\hat{v}_{x\theta}$ have opposite signs and f_x and $\hat{f}_x$ have positive signs when x is small and are negative when x is large.

[26] Observe that the main difference between this case and the one treated in Lemma 4 is on the interval $[\theta_1, \theta_2]$: in former case, the decision is U-shaped and in last one, the decision is constant.

1.2. $\hat{\theta} < \theta$. First, assume that $\hat{\theta} = \theta_{0\_}$ (the left hand side limit at θ_0). Define $y(\cdot) \equiv \max\{x(\cdot), x(\theta_1)\}$. By assumption A3, the derivative $\partial_\theta \Phi^y(\cdot, \hat{\theta})$ changes its sign at most one time: non-negative and then non-positive. From the case 0,

$$\Phi^x(\overline{\theta}, \hat{\theta}) \geq \Phi^y(\overline{\theta}, \hat{\theta})$$

and $\Phi^x(\hat{\theta}, \hat{\theta}) = 0$ imply this case.

For the general case, observe that $\Phi^x(\theta, \hat{\theta}) \geq \Phi^x(\theta, \theta_{0\_})$, because the derivative

$$\partial_{\hat{\theta}} \Phi^x(\theta, \hat{\theta}) = [v_x(x(\hat{\theta}), \hat{\theta}) - v_x(x(\hat{\theta}), \theta)]\dot{x}(\hat{\theta})$$

is non-positive.

2. $\hat{\theta} > \theta_0$

2.1 $\theta \geq \hat{\theta}$. This case is a consequence of monotonicity and case 0.

2.2. $\theta_0 \leq \theta < \hat{\theta}$. Taking the derivative of Φ^x with respect to $\hat{\theta}$, we get

$$\partial_{\hat{\theta}} \Phi^x(\theta, \hat{\theta}) = [v_x(x(\hat{\theta}), \hat{\theta}) - v_x(x(\hat{\theta}), \theta)]\dot{x}(\hat{\theta}).$$

By assumption A3, this derivative changes its signal at most one time: non-negative and then non-positive. Since $\Phi^x(\hat{\theta}, \hat{\theta}) = 0$, if $\Phi^x(\overline{\theta}, \hat{\theta}) \geq 0$, then $\Phi^x(\theta, \hat{\theta}) \geq 0$. Thus, it is enough to prove that

$$\int_{\theta}^{\overline{\theta}} \int_{x(\tilde{\theta})}^{x_-(\theta_0)} v_{x\theta}(\tilde{x}, \tilde{\theta}) d\tilde{x} d\tilde{\theta} \geq 0 \quad (*).$$

Observe that

$$\int_{\theta_0}^{\overline{\theta}} \int_{x(\tilde{\theta})}^{x_-(\theta_0)} v_{x\theta}(\tilde{x}, \tilde{\theta}) d\tilde{x} d\tilde{\theta} \geq 0$$

and $\partial_\theta \Phi^x(\theta, \hat{\theta}) = v_\theta(x(\theta), \theta) - v_\theta(x(\hat{\theta}), \theta)$ changes its signal at most one time (by assumption A3): non-negative and then non-positive. These two claims imply $(*)$.

2.3. $\theta < \theta_0$. We have that

$$\Phi^x(\theta, \hat{\theta}) = \Phi^x(\hat{\theta}, \theta_0) + \int_{\theta}^{\theta_0} \int_{x(\tilde{\theta})}^{x(\hat{\theta})} v_{x\theta}(\tilde{x}, \tilde{\theta}) d\tilde{x} d\tilde{\theta}.$$

By monotonicity and case 0, $\Phi^x(\theta_0, \hat{\theta}) \geq 0$ and $\int_{\theta}^{\theta_0} \int_{x(\tilde{\theta})}^{x(\hat{\theta})} v_{x\theta}(\tilde{x}, \tilde{\theta}) d\tilde{x} d\tilde{\theta} \geq 0$, we get the result. *Q.E.D.*

8. *Proof of Lemma 4:*

The proof is analogous to the Lemma 3 proof, but the case 2.3 which has the following modification:

2.3'. We have to prove that $\int_{\theta}^{\theta_0} \int_{x(\tilde{\theta})}^{y} v_{x\theta}(\tilde{x}, \tilde{\theta}) d\tilde{x} d\tilde{\theta} \geq 0$, for all $y \in [x(\theta_0), x_-(\theta_0)]$. First, this inequality is true for $y = x_-(\theta_0)$. Taking the derivative of this double integral with respect to y, we get $v_x(y, \theta_0) - v_x(y, \theta)$, which changes its signal at most one time. Thus, it is enough to check the inequality at $y = x(\theta_0)$. Taking the right derivative of $\Phi^x(\theta, \theta_{0\_}) \geq 0$ with respect to θ at $\theta = \theta_0$, we get

$$v_\theta(x_-(\theta_0), \theta_0) - v_\theta(x(\theta_0), \theta_0) \geq 0.$$

Since $\Phi^x(\theta_0, \theta_0) = 0$, we have the desired result. *Q.E.D.*

Appendix B
Second Best With Information: The Use of Lotteries

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 $(\alpha, x_1, x_2, t)\colon \Theta \to [0,1] \times {\mathbb{R}_+}^2 \times \mathbb{R}$ (t is the expected transfer of the lottery).

To avoid extra difficulties, assume the monotone hazard rate condition:

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 in 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 degenerate lotteries:[27]

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

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 prohibited from randomizing[28], i.e., α is equal to 0 or 1. The game works as follows: the principal designs a reward schedule based on the technology choice: $T_i \to (x_i, t_i)$, $i = 1, 2$. The agent accepts or rejects this 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 $f^i(x_i, \theta)$.

Depending on the technology choice, the problem is a standard adverse selection with the SMC. 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 takes this into consideration to determine his objective function, i.e., when

[27] A degenerate lottery is the one in which α is 0 or 1.

[28] If there is no commitment to 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 do not help ex-ante.

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 exists 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 is not satisfied on the interval (θ_1, θ_2) for both contracts, which 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 exists 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 and 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 induce 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 contracts are implementable for the second best problem. If this were not the case we would have to consider the "ironing principle".

Let us return to the case in which 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 degenerate lotteries 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).

Plugging this last equation into the objective function of the principal, on the interval (θ_1, θ_2), it will be

$$\Psi(x_1, x_2, \theta) = \frac{x_1}{\eta(\theta)^3\left(\frac{x_1}{x_2}\right)^2 + 1} + \frac{x_2}{\eta(\theta)^{-3}\left(\frac{x_2}{x_1}\right)^2 + 1} - \frac{(x_1x_2)^2}{\theta^3 x_1^2 + (1-\theta)^3 x_2^2}$$

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

It immediately follows 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,

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

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

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

Therefore, θ_i is determined by the continuity of the principal's ex-post profit when passing from randomization to non-randomization. Moreover, when 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 the 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 SMC.

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.

Appendix C
The Tables

We present the tables of Examples 2 and 4.

REFERENCES

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

ARROW, K. J. (1986): "Agency and the Market," *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 Games 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*. New York: John Willey & Sons.

CAVALLO, A., J. HECKMAN AND 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," *working paper*.

FUDENBERG, D. AND J. TIROLE (1991): *Game Theory*. Cambridge: 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. 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.

LAFFONT, J. J. AND J. TIROLE (1993): *A Theory of Incentives in Procurement and Regulation*. Cambridge: MIT Press.

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 (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," *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. (1987): "A Necessary and Sufficient Condition for Rationalizability in a Quasi-Linear Context," *Journal of Mathematical Economics*, 16, 191-200.

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.

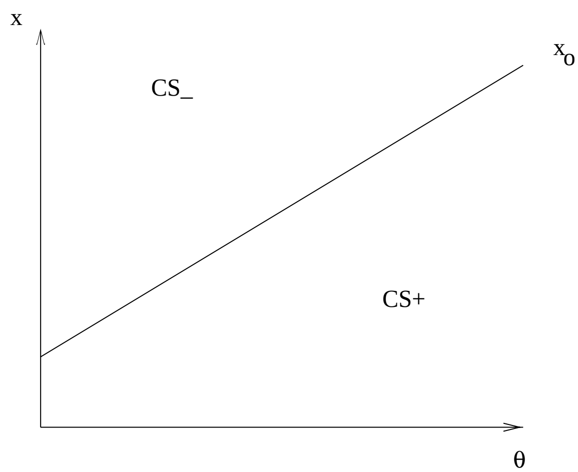

Figure 1

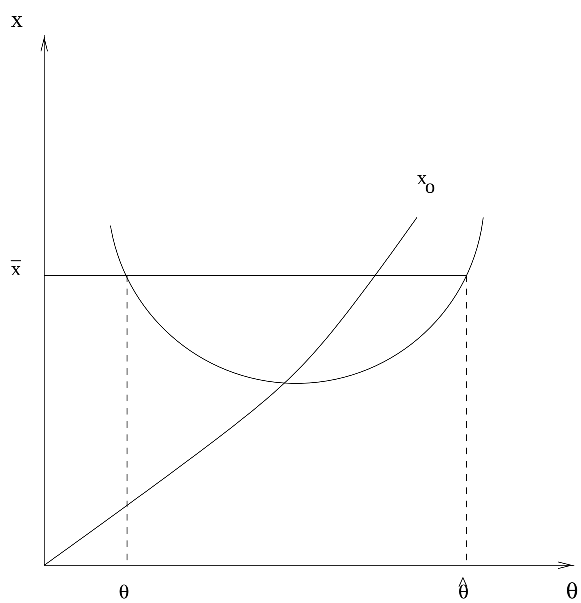

Figure 2

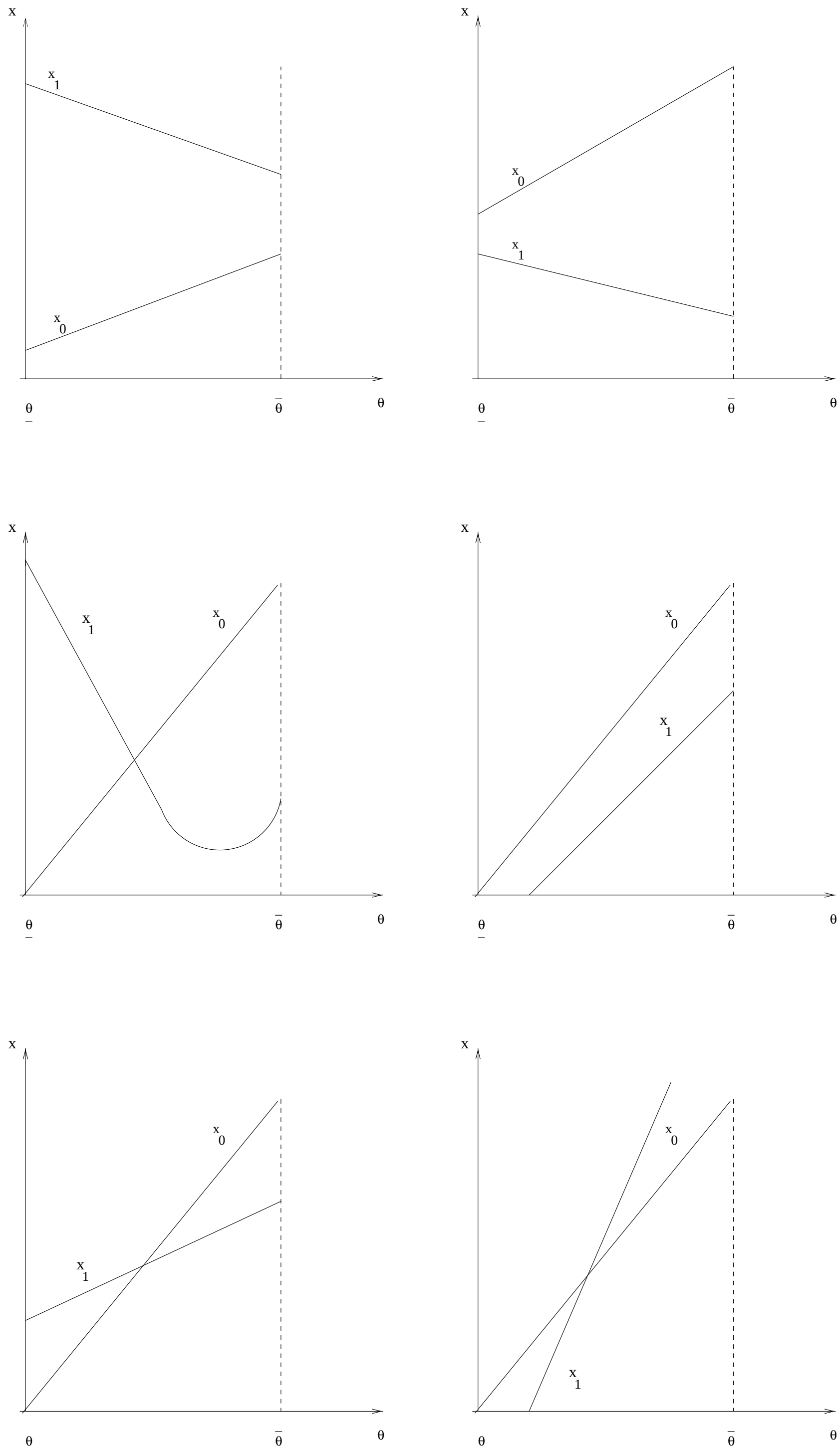
x
x_1
x_0
$\underline{\theta}$
$\bar{\theta}$
θ
x
x_0
x_1
$\underline{\theta}$
$\bar{\theta}$
θ
x
x_1
x_0
$\underline{\theta}$
$\bar{\theta}$
θ
x
x_0
x_1
$\underline{\theta}$
$\bar{\theta}$
θ
x
x_0
x_1
$\underline{\theta}$
$\bar{\theta}$
θ
x
x_0
x_1
$\underline{\theta}$
$\bar{\theta}$
θ

FIGURE 3

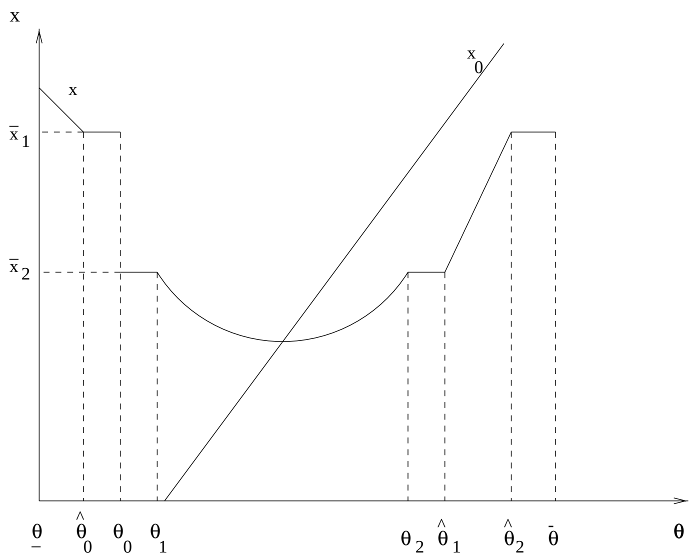

Figure 4

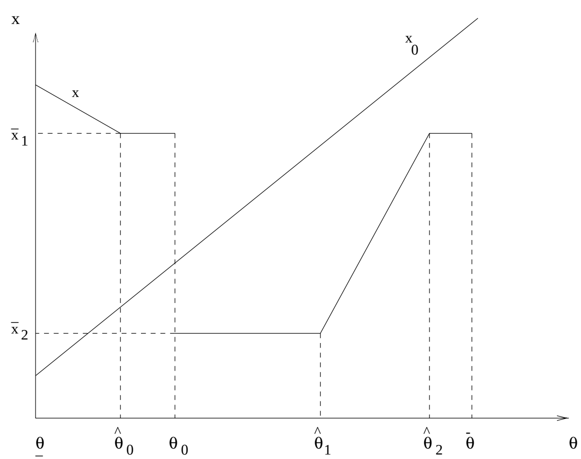

Figure 5

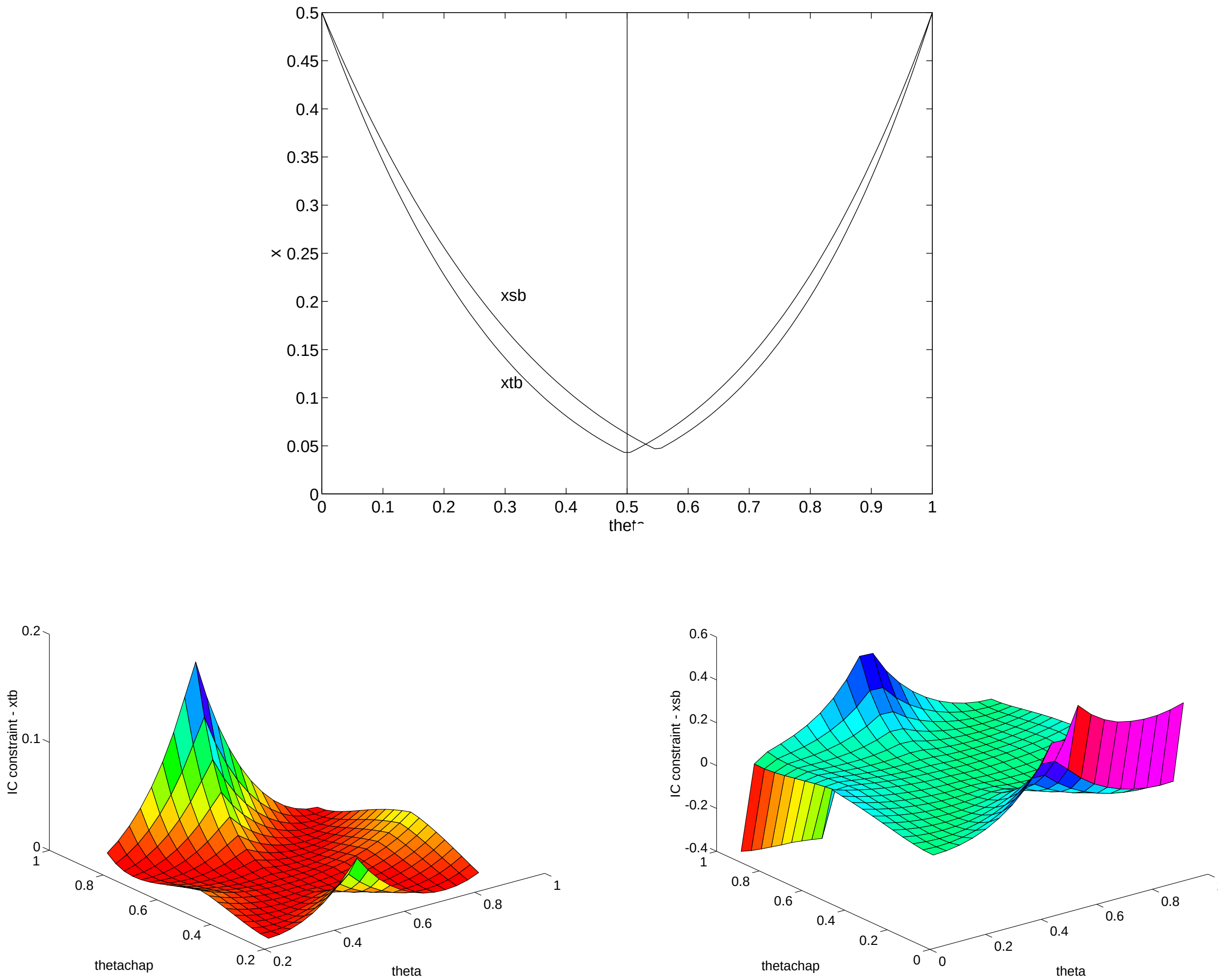
x
theta
xsb
xtb
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
IC constraint - xtb
thetachap
theta
IC constraint - xsb
thetachap
theta

Figure 6

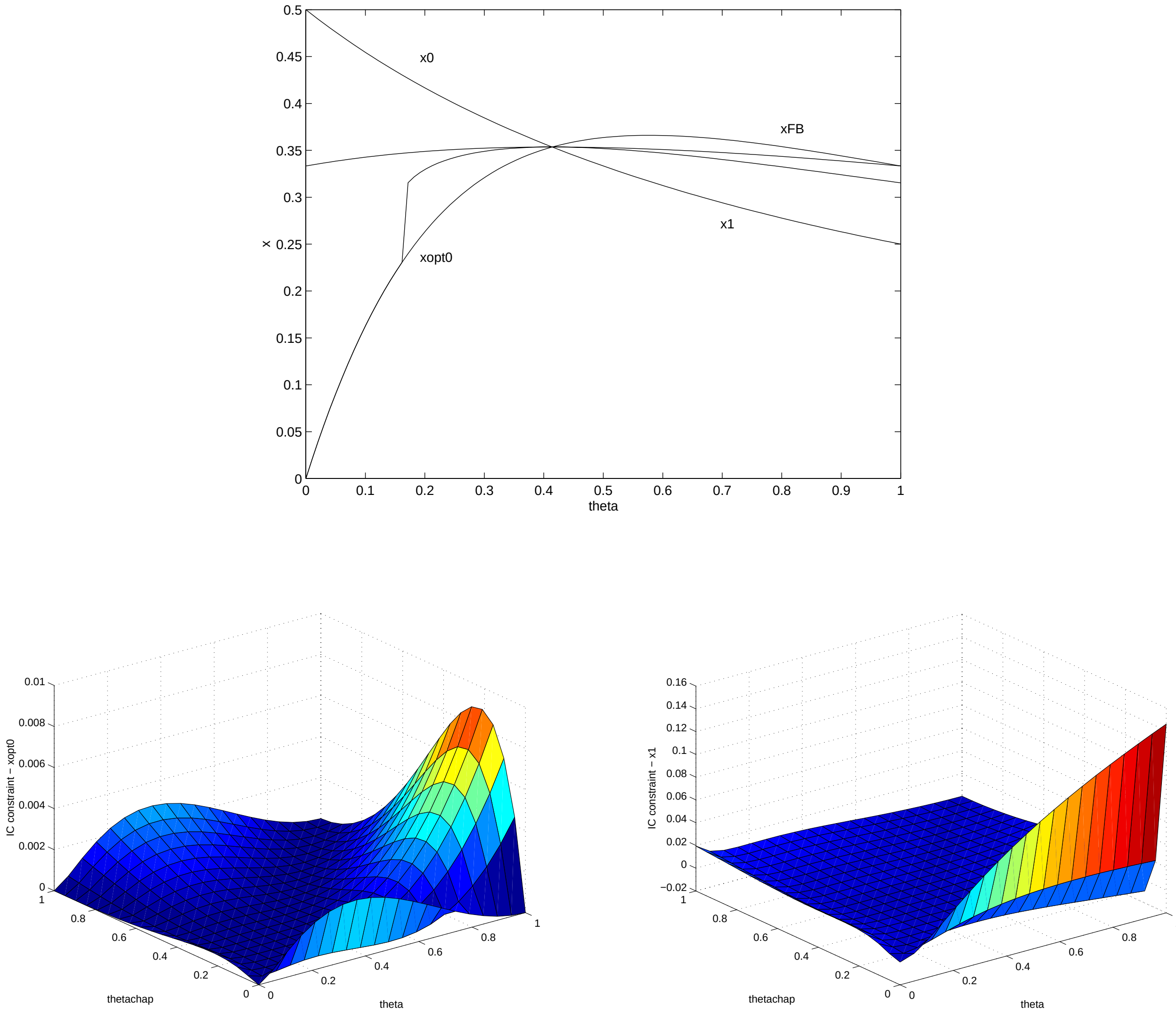
x0
xFB
x1
xopt0
x
theta
IC constraint − xopt0
thetachap
theta
IC constraint − x1
thetachap
theta

FIGURE 7.1

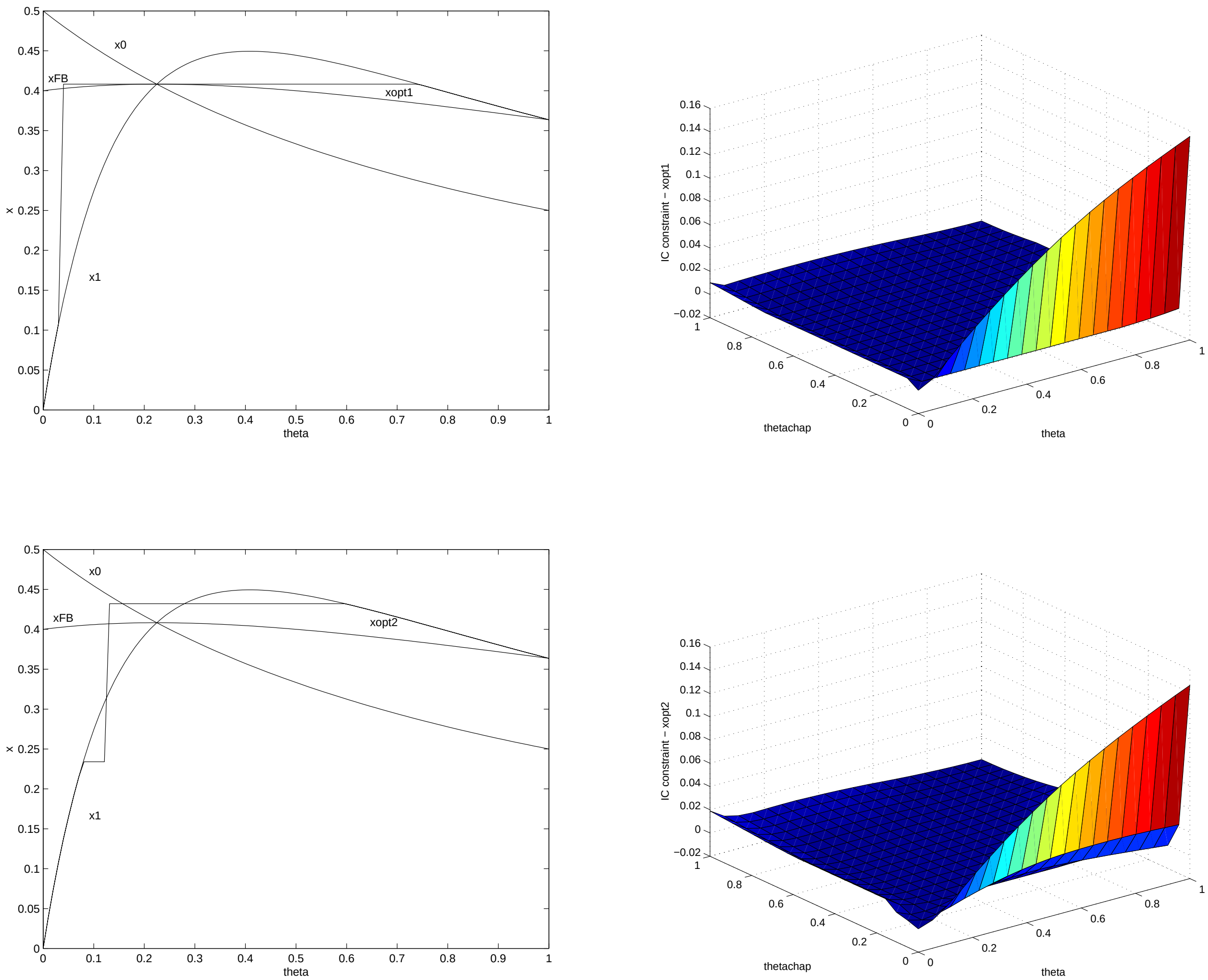
x0
xFB
xopt1
x1
x
theta
IC constraint − xopt1
thetachap
theta
x0
xFB
xopt2
x1
x
theta
IC constraint − xopt2
thetachap
theta

FIGURE 7.2

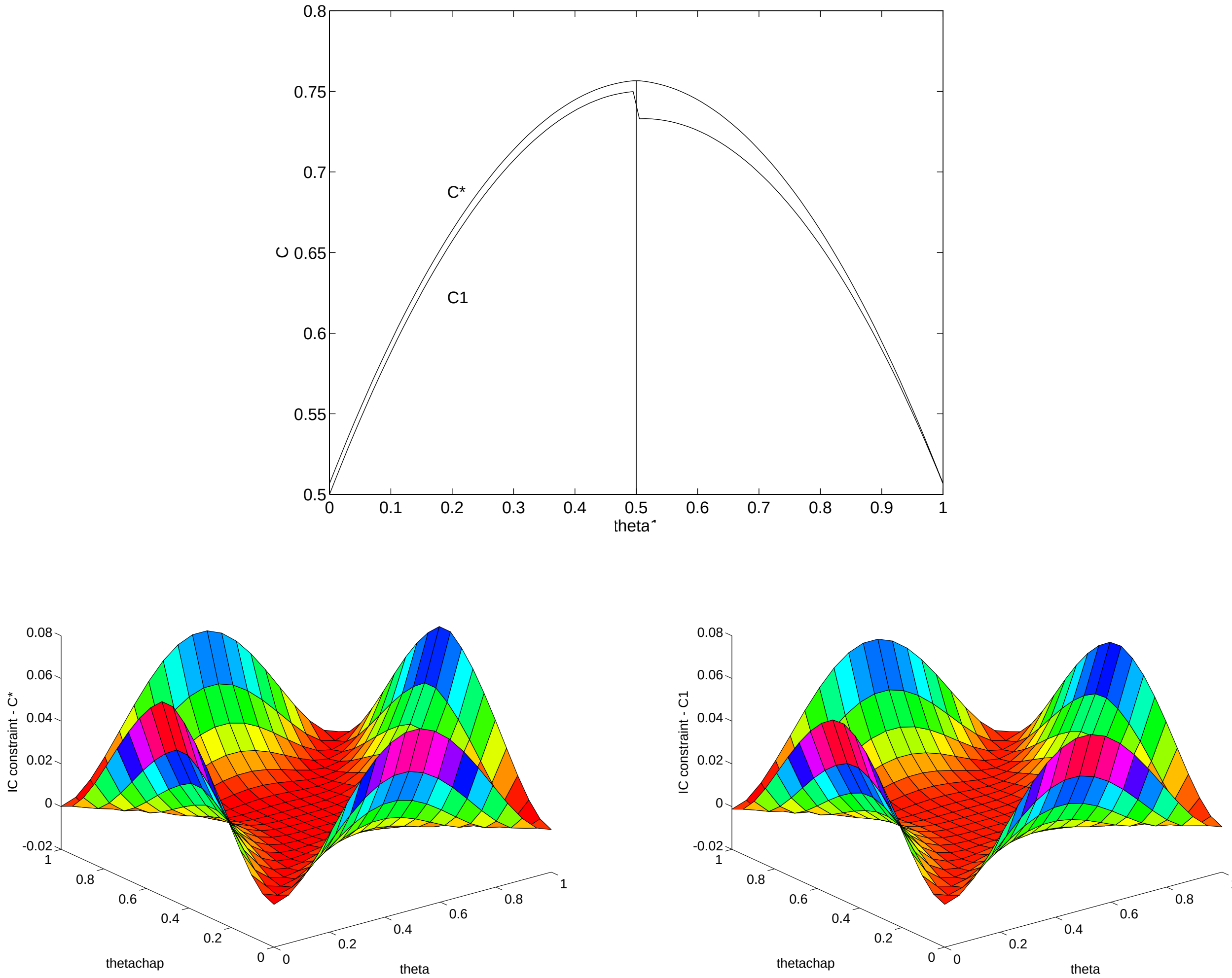
0.8
0.75
0.7
0.65
0.6
0.55
0.5
C
C*
C1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
theta*
IC constraint - C*
0.08
0.06
0.04
0.02
0
-0.02
thetachap
theta
IC constraint - C1

FIGURE 8

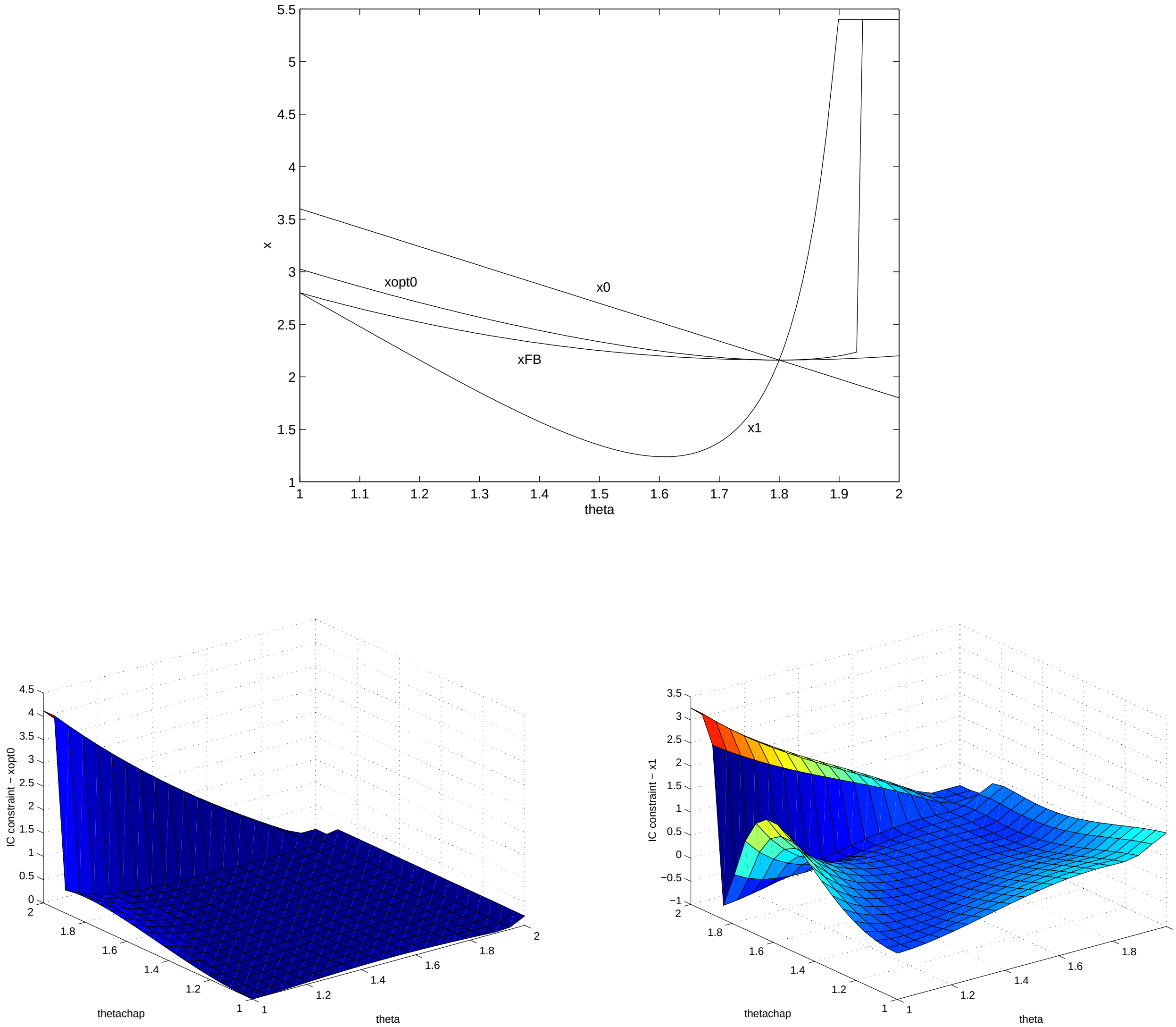
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
x
xopt0
x0
xFB
x1
1
1.1
1.2
1.3
1.4
1.5
1.6
1.7
1.8
1.9
2
theta
IC constraint − xopt0
thetachap
theta
IC constraint − x1
thetachap
theta

Figure 9.1

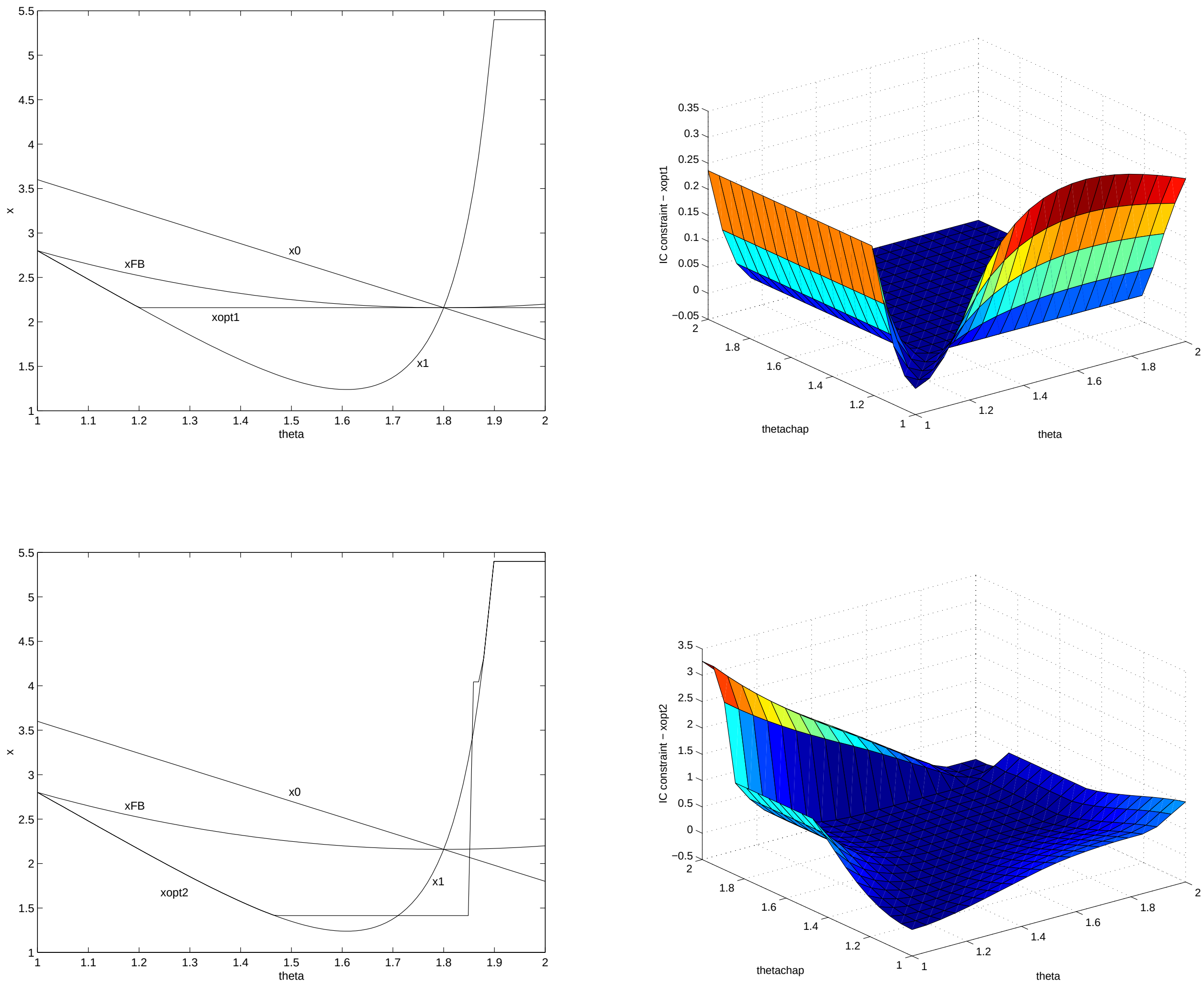
x
theta
x0
xFB
xopt1
x1
IC constraint − xopt1
thetachap
theta
x
theta
x0
xFB
xopt2
x1
IC constraint − xopt2
thetachap
theta

FIGURE 9.2

TABLE 1

NonLinear Pricing

1.1 $a = 1$ $b = 1$

c	$O.V.0$	$O.V.1$	$O.V.2$	$R.V.$	$F.B.V.$
1.25	0.21968	0.22322	0.22514	0.22547	0.31478
1.5	0.19906	0.20136	0.20265	0.20302	0.29387
1.75	0.18402	0.18504	0.18547	0.18578	0.27568
2.0	0.17096	0.17154	0.17155	0.17178	0.2597
2.25	0.15968	0.15989	0.15905	0.16007	0.24553
2.5	0.14984	0.14991	0.14992	0.15004	0.23287
2.75	0.14121	0.14125	0.14123	0.14132	0.2215
3.0	0.1336	0.1336	0.1336	0.13365	0.2112
3.25	0.12679	0.12679	0.12679	0.12683	0.20185
3.5	0.12069	0.12069	0.12069	0.12073	0.1933
3.75	0.11518	0.11518	0.11518	0.11521	0.18546
4.0	0.11017	0.11017	0.11021	0.11021	0.17825
4.25	0.10564	0.10564	0.10564	0.10564	0.17158

1.2 $a = 2$ $b = 1$

c	$O.V.0$	$O.V.1$	$O.V.2$	$R.V.$	$F.B.V.$
4.5	0.234840	0.235400	$0,235430$	0.235490	0.288560
5.0	0.222202	0.222260	$0,222260$	0.222380	0.275780
5.5	0.210520	0.210660	0.210650	0.210750	0.264100
6.0	0.200260	0.200310	$*$	0.200360	0.253390
6.5	0.190970	0.190980	$*$	0.191010	0.243520
7.0	0.182520	0.182530	0.182530	0.182530	0.234400
7.5	0.174810	0.174810	0.174810	0.174810	0.225940
8.0	0.167750	0.167750	0.167750	0.167750	0.218080
8.5	0.161250	0.161250	0.161250	0.161260	0.210750

where $*$ means that there is no feasible solution for the optimization program.

TABLE 2

Labor Contract

2.1 $s = 2$ $\pi = 2$

α	$O.V.0$	$O.V.1$	$O.V.2$	$R.V.$	$F.B.V.$
2.0	4.1533	4.1533	4.1533	4.1533	4.3333
2.2	4.2224	4.2224	4.2224	4.2224	4.5333
2.4	4.2535	4.2535	4.2537	4.2538	4.7333
2.6	4.2532	4.2534	4.2542	4.2546	4.9333
2.8	4.2277	4.2298	4.2316	4.2325	5.1333
3.0	4.1834	4.1876	4.1942	4.1957	5.3333
3.2	4.1266	4.1363	4.1517	4.1538	5.5333
3.4	4.0641	4.0953	4.1149	4.1175	5.7333
3.6	4.0087	4.064	4.0972	4.1003	5.9333
3.8	3.9605	4.0508	4.1161	4.1207	6.1333

2.2 $s = 1.5$ $\pi = 2$

α	$O.V.0$	$O.V.1$	$O.V.2$	$R.V.$	$F.B.V.$
2.0	2.1533	2.1533	2.1533	2.1533	2.3333
2.2	2.0224	2.0224	2.0224	2.0224	2.3333
2.4	1.8535	1.8536	1.8537	1.8538	2.3333
2.6	1.6532	1.6534	1.6542	1.6546	2.3333
2.8	1.4277	1.4298	1.4316	1.4325	2.3333
3.0	1.1834	1.1876	1.1942	1.1957	2.3333
3.2	0.92663	0.93627	0.95166	0.95378	2.3333
3.4	0.66409	0.6953	0.71494	0.71749	2.3333
3.6	0.40871	0.46398	0.49722	0.50032	2.3333
3.8	0.16048	0.25084	0.31606	0.32069	2.3333

where
O.V. 0: Optimal Value 0 - the expected virtual surplus at $xopt0$
O.V. 1: Optimal Value 1 - the expected virtual surplus at $xopt1$
O.V. 2: Optimal Value 2 - the expected virtual surplus at $xopt2$
R. V.: Relaxed Value - the expected virtual surplus at x_1
F.B. V.: First Best Value - the principal's first-best expected utility