MULTIDIMENSIONAL INCENTIVE-COMPATIBILITY: THE MULTIPLICATIVELY SEPARABLE CASE[1,2]

ALOISIO ARAUJO
Instituto Nacional de Matematica Pura e Aplicada and Fundação Getulio Vargas

DANIEL GOTTLIEB
MIT

and

HUMBERTO MOREIRA
Fundação Getulio Vargas

Abstract

This paper characterizes incentive-compatible allocations when types are multidimensional and the single-crossing condition may not hold. This characterization is used to obtain the optimal contracts in multidimensional screening as well as the equilibria in multidimensional signaling models. Then, we determine the implications of signaling and screening models when the single-crossing condition is violated. The unique robust prediction of signaling is the monotonicity of transfers in (costly) actions. Any function from the space of types to the space of actions and an increasing transfer schedule can be rationalized as an equilibrium profile of many signaling models. Apart from the monotonicity of transfers in actions, we obtain an additional necessary and sufficient condition in the case of screening. In one-dimensional models, this condition states that the principal's profit as a function of the agent's type must grow at a higher rate under asymmetric information than under symmetric information.

JEL Classification: D82, D86, J41.

Keywords: Signaling and screening games, single crossing condition, incentive compatibility, non-monotonicity.

[1]For helpful comments and discussion we are grateful to Daron Acemoglu, Juliano Assunção, Vinícius Carrasco, Mathias Dewatripont, Thomas Mariotti, and Muhamet Yildiz. We also thank conference and seminar participants at Université de Toulouse 1, MIT, ESEM (Milan, 2008), EEA (Vienna, 2006), LAMES (Mexico City, 2006), Encontro SBE (Salvador, 2006), Colegio de Mexico, IBMEC-Rio, FGV and PUC-Rio.

[2]E-mail addresses: aloisioa@fgv.br (A. Araujo), daniel_g@mit.edu (D. Gottlieb) and humberto@fgv.br (H. Moreira).

1 Introduction

In several markets, some participants have information that is relevant to other participants. Frequently, information can be inferred from actions taken by the informed parties. The uninformed may move first and induce the informed to take such actions (screening) or the informed may move first and take actions in order to signal their information (signaling).

An important result in models with asymmetric information is the Revelation Principle, which states that any allocation process can be replicated by a mechanism in which participants are asked to reveal their private information. The Revelation Principle reduces a possibly complicated problem to an easy-to-state mathematical-programing problem, where each agent prefers to reveal his private information honestly (incentive-compatibility). However, the general analysis of such mathematical-programing problem is not straightforward.

Most of the literature assumes that an individual's private information consists of a one-dimensional type parameter and that the marginal utility of taking the action can be ordered (single-crossing condition, SCC). Under this assumption, Mirrlees (1971), Spence (1974), and Rothschild and Stiglitz (1976) show that a first-order and a monotonicity condition determine the solution of the programing problem, and the characterization of the set of incentive-compatible allocations becomes straightforward. McAfee and McMillan (1988) characterize incentive-compatible allocations in a multidimensional model under a single-crossing condition and Quinzii and Rochet (1985) characterize the separating equilibria of a multidimensional signaling model under a single-crossing condition.

This paper has two main purposes. First, it studies incentive-compatible allocations under a condition that is weaker than the SCC. This allows us to characterize the solution of multidimensional screening models as well as the equilibria of multidimensional signaling models where the SCC does not hold. Second, the paper determines the implications of multidimensional signaling and screening models when the SCC does not hold.

The characterization of optimal allocations in multidimensional screening models is complementary to Rochet (1987), McAfee and McMillan (1988), and Armstrong (1996). Our condition (discussed in Section 2) has three main advantages. First, it is easy to verify and is compatible with most specifications used in applied work (e.g., utility functions multiplicatively separable between the decision variable and types).[3] Second, it does not require assumptions to be made on endogenous variables and on the distribution of types. And third, it allows for utility functions that do not satisfy the SCC. The characterization of equilibria in multidimensional signaling models also provides necessary and sufficient conditions for signals to reveal all unobservable information (full-separability).[4]

[3] Armstrong (1996) assumes homogeneity of the utility function with respect to the type parameter, separability of the agent's indirect utility, and separability of the density function. McAfee and McMillan (1988) on the other hand assume a generalized single-crossing condition that requires the shadow-price indifference curves to be hyperplanes.

Matthews and Moore (1987) analyze a two-dimensional screening model where utility functions cross twice.

[4] Since Kholleppel's (1983) example of a model where no separating equilibrium existed, the existence of a fully-separating equilibrium became an important issue. Engers and Fernandez (1987) show that the SCC is sufficient for full-separability. The present article extends previous results on conditions required for full-separability by obtaining necessary and sufficient conditions.

Recently, the SCC has been criticized both on theoretical and empirical grounds. The first type of criticisms stress that, even though this property may be natural and intuitive in one-dimensional models, it cannot be extended in a sensible manner to multidimensional settings. Models with discrete multidimensional types can be transformed into one-dimensional type models but the SCC is usually broken (see Rochet and Stole, 2003). Furthermore, utility functions that satisfy the analog of the SCC in multidimensional settings may fail to satisfy this condition in the presence of other signals (Araujo et al., 2007) or when there is correlation between the characteristics (Araujo and Moreira, 2003). Also, the presence of a moral hazard dimension may cause the SCC to be violated (Acemoglu, 1998).

In some applications, it is widely recognized that the assumption of one-dimensional types may be implausible. In the context of education, for example, Heckman (2005) argues that "abilities are multiple in nature", and that one-dimensional models cannot capture important phenomena (see Araujo et al., 2007, or Heckman et al., 2005).

Criticisms made on empirical grounds stress that several examples of interesting and intuitive phenomena have been proved to arise only when one drops this assumption. Bagwell and Bernheim (1996) investigate conditions under which consumers may be willing to pay higher prices for functionally equivalent goods as away to signal wealth ('Veblen effects'). Their main finding is that Veblen effects do not arise when the SCC holds but may arise when it fails.

Bernheim (1991) proves that firms may choose to distribute dividends even when they are taxed more highly than stock repurchases ('dividend puzzle') when the SCC fails. Bernheim and Severinov (2003) provide an explanation for the equal division of bequests based on a model where the SCC fails. Rotemberg (1988) shows that a tax on a signal may be Pareto-improving if the SCC fails to hold. Similarly, Bernheim and Redding (2001) show that taxing signals can be Pareto-improving in a model where the SCC does not hold. Bernheim (1994) studies a model of conformity in social interactions where the single-crossing condition fails.

Araujo et at. (2007) study a model where high types choose to engage in a lower amount of signaling than intermediate types (see Example 1). Araujo et al. (2004) show that the relation between dividend payments and earnings may be non-monotone when the SCC is not satisfied, and Smart (2000) and Araujo and Moreira (2003) show that the relation between risk and insurance coverage may not be monotonic.

Thus, one may question what the empirical content of the signaling and screening models is once the SCC is not assumed. In other words, which implications of these models are not a result of the particular specification of the cost of the activity? Similarly, it is important to understand which results from the standard models generalize to multidimensional environments.[5]

We show that the only robust prediction of signaling is the monotonicity of transfers in costly actions. However, this prediction is shared by most alternative (symmetric information) models. Therefore, even when one employs a selection criterion, signaling can almost never be rejected. Another negative result concerns the identifiably of signaling

[5] In a model of insurance under asymmetric information, Chiappori et al. (2006) argue that coverage and risk are positively correlated even when the SCC is not imposed or when types are multidimensional. However, although we also obtain a monotonicity condition, our results are not directly applicable to insurance since we follow the standard mechanism design model and assume quasi-linear utilities.

models. It is shown that, for each signaling model, there exists a large class of signaling models with the same observable implications. Hence, it is impossible to determine which is correct among a large class of models.

The characterization of solutions of multidimensional screening allows us to identify a new necessary and sufficient condition. Under homogeneity of the distribution of types or when types are one-dimensional, this condition implies that the principal's profit must grow with respect to types at a higher rate under asymmetric information than under symmetric information.

Overall, our results imply that special attention must be devoted to the specific characteristics of the market being studied. Only with precise knowledge about the cost of engaging in the activity, the relevant number of dimensions, and the technology, one is able to obtain testable predictions of the model.

Our paper is also related to the literature on monotone comparative statics, which studies necessary and sufficient conditions for solutions to maximization programs to be monotone. One important result in this literature is that the SCC is sufficient for the solution to be monotone (c.f., Milgrom and Shannon, 1994). Furthermore, if the choice set is sufficiently rich, this condition is also necessary. In the present paper, we show that the cross-partial derivative of the cost function determines not only whether the set of incentive-compatible allocations is increasing or decreasing but the whole shape of the solution.

The rest of the paper is organized as follows. Section 2 characterizes the set of incentive-compatible allocations and studies the restrictions imposed by incentive-compatibility. Section 3 presents the screening model and characterizes the optimal solution. Then, we study the additional implications imposed by the screening model. Section 4 considers the signaling model. Then, Section 5 concludes.

2 Characterization of Incentive Compatibility

The economy consists of informed agents and an uninformed principal. Each agent is characterized by a multidimensional parameter ('type') $\theta \in \Theta$, where $\Theta \subset \mathbb{R}^n_+$ is a compact and convex set with non-empty interior. Types are distributed according to a continuous density $p : \Theta \to \mathbb{R}_{++}$. Agents may engage in a costly activity $y \in \mathbb{R}_+$. Principals have a continuously differentiable valuation function $f : \Theta \times \mathbb{R}_+ \to \mathbb{R}_+$.

The cost of the activity y is given by a C^3 function $c : \Theta \times \mathbb{R}_+ \to \mathbb{R}_+$, where

Assumption 0 $c_y(\theta, y) > 0$ and $c_{\theta_i}(\theta, y) < 0$, for all $\theta \in \Theta$, $y \in \mathbb{R}_+$, and $i = 1, ..., n$.

The first inequality states that y is costly while the second states that higher types have a lower cost of engaging in the activity y.

The standard single-crossing condition (SCC) requires $c_{\theta_i y}$ not to change signs so that having a higher type has a monotonic effect on the marginal cost of engaging in y. The following assumption generalizes the SCC in the sense that it allows $c_{\theta_i y}$ to change sign.

Assumption 1 There exist functions $\xi : \mathbb{R}_+ \to \mathbb{R}$, $\psi : \Theta \to \mathbb{R}$, and $\varphi : \Theta \to \mathbb{R}$ such that

$$c(\theta, y) = \xi(y) + y \times \psi(\theta) + \varphi(\theta).$$

Assumption 1 implies that $c_{\theta_i yy}(\theta, y) = 0$, for all $i = 1, ..., n$. It is satisfied, for example, when costs are quadratic in y :

$$\hat{c}(\theta, y) = J \times y^2 + y \times \psi(\theta) + \varphi(\theta),$$

where J is a real number. This representation includes the standard case of costs that are linear in y ($J = \varphi(\theta) = 0$ for all θ).

The SCC holds in each dimension if either $c_{\theta_i y}(\theta, y) > 0$ or $c_{\theta_i y}(\theta, y) < 0$ for all y, θ_i, and i. Assumption 1 states that $c_{\theta_i y}(\theta, y)$ is not a function of y. Hence, if we fix the $n-1$ other dimensions and consider a graph of θ_i and y, $c_{\theta_i y}$ is constant along any vertical line. In particular, the intervals where $c_{\theta_i y} > 0$ and where $c_{\theta_i y} < 0$ are separated by vertical lines (see Figures 1 and 2).

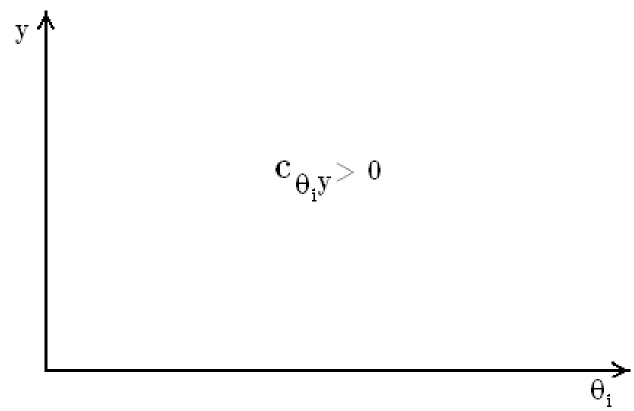

Figure 1: Cost function that satisfies SCC in each dimension.

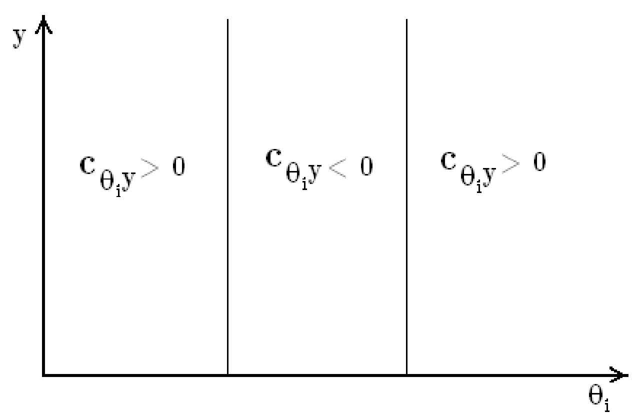

Figure 2: Cost function that satisfies Assumption 1.

In a context of education as signal, for example, the agents are workers, the principals are employers, the activity y is the amount of schooling, and the price w is the wage paid to workers. In an industrial organization context, the agents are firms, principals are potential buyers, and the activity may be the amount spent on advertisement or the duration of warranties. The following example considers the model from Araujo, Gottlieb, and Moreira (2007).

Example 1 *Consider a labor market model. Risk neutral workers engage in a costly schooling activity* $y \in \mathbb{R}_+$ *and obtain wages* $w \in \mathbb{R}_+$*. Their ability is captured by a two-dimensional vector* $\theta \equiv (\theta_1, \theta_2) \in \mathbb{R}^2$ *and the cost of schooling is*[6]

$$c(\theta, y) = \frac{y}{\theta_1 \theta_2}.$$

Employers do not observe ability but observe the amount of schooling y*. They can also interview workers, which gives them an additional measure of each worker's skills. The interview technology is represented by the following function:*

$$g(\theta, y) = \alpha\theta_1 + \theta_2 + \beta y.$$

A type-θ *worker produces a good that is worth* $f(y, \theta)$*, where* $f_{\theta_i} > 0$, $f_y \geq 0$, $f_{yy} \leq 0$*, and* $f_{y\theta_i} \geq 0$*. For a fixed* $g(\theta, y) = \bar{g}$*, we can write the cost of schooling as*

$$\hat{c}(\theta_1, y, \bar{g}) \equiv \frac{y}{\theta_1(\bar{g} - \alpha\theta_1 - \beta y)}.$$

Note that the single-crossing condition does not hold since $\hat{c}_{\theta_1 y}\left\{\substack{> \\ <}\right\}0 \iff \theta_1\left\{\substack{> \\ <}\right\}\frac{\bar{g}-\beta y}{2\alpha}$.

The single-crossing condition means that exchanging one unit of ability θ_1 *for* α *units of* θ_2 *would always be either desirable or undesirable in terms of reducing the cost of schooling. In the specification above, because the abilities are imperfect substitutes, this exchange is desirable for high levels of* θ_1 *and undesirable for low levels of* θ_1*. Therefore, the substitutability between skills breaks down the single-crossing condition. Furthermore, Assumption 1 is satisfied when* $\beta = 0$*.*

The next example describes a model of warranties:

Example 2 *Consider a model of warranties and uncertain product quality. Product quality is determined by a two-dimensional vector of characteristics* $\theta \equiv (\theta_1, \theta_2) \in \mathbb{R}^2$*. For concreteness, interpret* θ_1 *as the reliability and* θ_2 *as the complexity of the good. Consumers observe a measure of product quality given by the function* $g(\theta) = \alpha\theta_1 + \theta_2$*.*

Producers may offer a warranty that repairs any defect that may occur until time y*. Let* $w(y, g)$ *denote the price charged conditional on* $g(\theta) = g$ *when warranty* y *is provided.*

Denote by $f(\theta_1, \theta_2, y)$ *the expected value of the good to consumers given warranty* y *and* $c(\theta_1, \theta_2, y)$ *denote the expected cost of producing the good and providing warranty* y*. We assume that reliability reduces the cost of providing warranty whereas complexity increases the cost of providing warranty:*

$$c_{\theta_1 y} < 0, \ c_{\theta_2 y} > 0.$$

As in Example 1, we can rewrite the expected cost of producing the good conditional on $\bar{g}$ *as* $\hat{c}(\theta_1, y, \bar{g}) \equiv c(\theta_1, \alpha\theta_1 - \bar{g}, y)$*. Note that* $\hat{c}_{\theta_1 y} = c_{\theta_1 y} + \alpha c_{\theta_2 y}$ *may switch signs because the first term is negative while the second term is positive. Therefore, the fact that reliability decreases the cost of providing warranty whereas complexity increases this cost implies that single-crossing condition may not hold. In particular, if we assume linear costs,*

$$c(\theta_1, \theta_2, y) = A \times \theta_1 \times (K - \theta_2) \times y,$$

[6] θ_1 and θ_2 can be interpreted as cognitive and non-cognitive skills, for example.

where A and K are positive real numbers, it follows that $\hat{c}_{\theta_1 y}\left\{\substack{>\\<}\right\}0 \iff \theta_1\left\{\substack{<\\>}\right\}\frac{K+\bar{g}}{2\alpha}$ and Assumption 1 is satisfied.

The following example presents a multidimensional generalization of the standard non-linear pricing model:

Example 3 *A monopolist sells a good in different sizes (or qualities) $Q \geq 0$. Consumers have private information about their tastes, captured by a vector of types $\theta \in \Theta$. The consumer's gross surplus from the good is $V(\theta, Q)$, where $\frac{\partial V}{\partial \theta_i} > 0$ and $\frac{\partial V}{\partial Q} > 0$. A purchase of size Q is sold at price $P(Q)$. Therefore, the consumer's utility from purchasing the good is $V(\theta, Q) - P(Q)$.*

The cost of production is $FC + MC \times Q$, where $FC \geq 0$ and $MC > 0$ denote the firm's fixed and marginal costs. The firm's profit from selling the good is $P(Q) - MC \times Q - FC$.

Let $y = \bar{Q} - Q$ for $\bar{Q}$ large enough, let $w(y) = -P\left(\bar{Q} - y\right)$, and define the function $c : \Theta \times \mathbb{R}_+ \to \mathbb{R}_+$ by

$$c(\theta, y) = -V\left(\theta, \bar{Q} - y\right).$$

The consumer's utility can be written as $w(y) - c(\theta, y)$. Furthermore, $c(\theta, y)$ satisfies Assumption 0.

Let $f(y) = -MC \times \left(\bar{Q} - y\right) - FC$. The firm's profit can be written as $f(y) - w(y)$. Hence, this model is a special case of the basic framework. Moreover, Assumption 1 is satisfied if and only if there exist functions $\xi : \mathbb{R}_+ \to \mathbb{R}$, $\psi : \Theta \to \mathbb{R}$, and $\varphi : \Theta \to \mathbb{R}$ such that

$$V(\theta, Q) = \xi(Q) + Q \times \psi(\theta) + \varphi(\theta).$$

The next example describes of the random participation model of Rochet and Stole (2002):

Example 4 *Consider the same setting as in Example 3. Take $\theta = (\theta_1, \theta_2)$ and take $\xi(Q) = 0$, $\psi(\theta) = \theta_1$, $\varphi(\theta) = -\theta_2$. The parameter θ_1 denotes the consumer's taste for quality Q, and θ_2 indexes the consumer's opportunity cost.*

A mechanism $(y(\theta), w(y))$ is incentive-compatible if it satisfies the following incentive-compatibility constraint:

$$y(\theta) \in \arg\max_{\widehat{y}} w(\widehat{y}) - c(\theta, \widehat{y}) \tag{IC}$$

for all $\theta \in \Theta$. Equivalently, we say that $w(y)$ implements $y(\cdot)$. We say that a profile of activities $y(\cdot)$ is implementable if there exists a function $w(y)$ that implements it.

Before proceeding, we need to introduce some notation. We say that a type θ is regular for $y(\cdot)$ if the differential $Dy(\theta)$ has full rank; otherwise θ is critical. Finally, $\overline{y}$ is a critical value of y if there exists a critical type θ such that $y(\theta) = \overline{y}$, otherwise $\overline{y}$ is called a regular value of y. We refer to a C^1 function with measure zero of critical values as a *regular function*. In what follows we will only consider mechanisms $(y(\theta), w(y))$ such that y and w are regular.[7,8]

[7] Apart from the differentiability assumption, the only restrictions that this condition imposes to the incentive compatible profile is that there are no discontinuities in $y(\cdot)$ and its derivative. The results could be generalized with an adaptation in the proof of Theorem 1: the price profile would have kinks (which would correspond to the discontinuities of the derivative of y) and disconnected domain (which would correspond to the discontinuities in y).

[8] By Sard's theorem, this condition is automatically satisfied if $y \in C^n$, $n > 1$.

Which mechanisms $(y(\theta), w(y))$ satisfy incentive-compatibility? In the remainder of this section, we first analyze necessity and then sufficiency. The following lemma gives the usual first- and second-order conditions of the incentive compatibility constraint.

Lemma 1 *Let $(y(\cdot), w(\cdot))$ be an incentive-compatible mechanism. Then,*

$$w'(y(\theta)) = c_y(\theta, y(\theta)), \tag{1}$$

$$c_{\theta_i y}(\theta, y(\theta))\, y_{\theta_i}(\theta) \leq 0, \tag{2}$$

for all $i = 1, ..., n$ and $\theta \in \Theta$.

Proof. To simplify the proof, let us assume that $y(\cdot)$ is C^2. If $(y(\cdot), w(\cdot))$ is incentive-compatible, it must satisfy:

$$\theta \in \arg\max_{\hat{\theta}} w\left(y\left(\hat{\theta}\right)\right) - c\left(\theta, y\left(\hat{\theta}\right)\right),$$

whose local first- and second-order conditions are

$$\begin{aligned} w'(y(\theta))\, y_{\theta_i}(\theta) - c_y(\theta, y(\theta))\, y_{\theta_i}(\theta) &= 0, \\ w''(y(\theta))\left[y_{\theta_i}(\theta)\right]^2 + w'(y(\theta))\, y_{\theta_i\theta_i}(\theta) - c_{yy}(\theta, y(\theta))\left[y_{\theta_i}(\theta)\right]^2 - c_y(\theta, y(\theta))\, y_{\theta_i\theta_i}(\theta) &\leq 0. \end{aligned}$$

Equation (1) follows from the first-order condition. Differentiating the first-order condition (which must hold for every θ) yields

$$\begin{aligned} & w''(y(\theta))\left[y_{\theta_i}(\theta)\right]^2 + w'(y(\theta))\, y_{\theta_i\theta_i}(\theta) - c_{yy}(\theta, y(\theta))\left[y_{\theta_i}(\theta)\right]^2 - c_y(\theta, y(\theta))\, y_{\theta_i\theta_i}(\theta) \\ = \; & c_{\theta_i y}(\theta, y(\theta))\, y_{\theta_i}(\theta). \end{aligned}$$

Substituting in the second order condition, we obtain $c_{\theta_i y}(\theta, y(\theta))\, y_{\theta_i}(\theta) \leq 0$. ■

When the SCC holds, equation (2) reduces to a monotonicity condition. When it does not hold, equation (2) implies that incentive-compatible mechanisms may not be monotonic: y has to be increasing in the region where $c_{\theta_i y} < 0$ and decreasing in the region where $c_{\theta_i y} > 0$.

In one-dimensional models where mechanisms are monotone, if two types pool in the activity y, all intermediate types must also be pooled with them. As a consequence, pooling sets must be intervals. However, when mechanisms are not monotone, a disconnected set of types may be pooled. Araujo and Moreira (2004) have shown that a necessary condition for incentive-compatibility in this case is the so-called marginal utility identity. This condition implies that if two disconnected types are pooling in an activity y, they should have the same marginal cost. The following lemma establishes this result in our context:

Lemma 2 *If two individuals with regular types θ and $\tilde{\theta}$ choose the same signal, then their marginal cost must be the same:*

$$y(\theta) = y\left(\tilde{\theta}\right) \Rightarrow c_y(\theta, y(\theta)) = c_y\left(\tilde{\theta}, y(\theta)\right). \tag{3}$$

Proof. Suppose that $(w(y(\theta)), y(\theta))$ is incentive-compatible. Then, (1) must hold. Therefore, if θ and $\widetilde{\theta}$ are regular types such that $y(\theta) = y\left(\tilde{\theta}\right)$, equation (1) implies that

$$c_y(\theta, y(\theta)) = c_y\left(\tilde{\theta}, y\left(\tilde{\theta}\right)\right).$$

■

We have shown that the local conditions (1), (2) and the global condition (3) are necessary for incentive-compatibility even when Assumption 1 does not hold. The following example shows that they may not be sufficient when Assumption 1 does not hold.

Example 5 *Suppose that the cost of activity y is*

$$c(\theta, y) = \frac{y}{2|\theta - 1|} + \frac{1}{8}\left[y - (\theta - 1)^2\right]^2 - \frac{5\theta}{3}\left[y - (\theta - 1)^2\right]^3,$$

so that Assumption 1 is not satisfied. Consider the following mechanism:

$$\begin{aligned} y(\theta) &= (\theta - 1)^2, \\ w(y) &= \sqrt{y}, \end{aligned}$$

where $\Theta = [1.1,\ 2]$.

Note that conditions (1), (2), and (3) are satisfied. Under the proposed mechanism, type $\theta = \frac{3}{2}$ chooses $y\left(\frac{3}{2}\right) = \frac{1}{4}$ and obtains utility $w\left(\frac{1}{4}\right) - c\left(\frac{3}{2}, \frac{1}{4}\right) = \frac{1}{4}$. However, by choosing $y = 1$, he obtains $w(1) - c\left(\frac{3}{2}, 1\right) = \frac{63}{64} > \frac{1}{4}$, which is a profitable deviation. Therefore, (1), (2), and (3) may not prevent non-local deviations when Assumption 1 does not hold.

The proposition below establishes that, under Assumption 1, the necessary conditions are sufficient as well.

Proposition 1 *Suppose Assumption 1 holds. A mechanism $(y(\cdot), w(\cdot))$ is incentive-compatible if and only if the first- and second-order conditions (1) and (2), and the pooling condition (3) are satisfied.*

Proof. ($\Rightarrow$) Follows from Lemmata 1 and 2.

($\Leftarrow$) Let us define the wage schedule $w \in C^1$. By (3) we can define the derivative of w by $w'(y) = c_y(\theta, y)$ for $y = y(\theta)$ and all regular types $\theta \in \Theta$. Since the set of critical values has zero measure, we can extend continuously the definition of $w'(y)$ for critical values. Then, define the following wage schedule for $y \in y(\Theta)$:[9]

$$w(y) = \int_{y^{\min}}^{y} w'(x)dx + w^{\min},$$

where $w^{\min} \geq \max_{\theta \in \Theta} c(\theta, 0)$ (which ensures that all types will participate) and $y^{\min} = \min_{\theta \in \Theta} y(\Theta)$.

[9] For $y \notin y(\Theta)$, we extend the wage function linearly such that the derivative is continuous.

Let θ be first a regular type (for critical θ the argument is made by continuity). The first-order condition of the previous problem is $w'(y(\theta)) = c_y(\theta, y(\theta))$, which holds by the definition of $w(\cdot)$.

Given a regular $\widehat{y} = y(\widehat{\theta})$, there exists $i \in \{1, ..., n\}$ such that $y_{\theta_i}(\widehat{\theta}) \neq 0$. By the implicit function theorem there exists a function φ such that, locally, $\widehat{\theta}_i = \varphi(\widehat{y}, \widehat{\theta}_{-i})$, where $\widehat{\theta}_{-i}$ is the vector $n-1$ dimensional $\widehat{\theta}$ but coordinate i. Taking the derivative with respect to $\widehat{y}$ in the equality $w'(\widehat{y}) = c_y(\varphi(\widehat{y}, \widehat{\theta}_{-i}), \widehat{\theta}_{-i}, \widehat{y})$ we get

$$w''(\widehat{y}) = c_{yy}(\widehat{y}, \widehat{\theta}) + c_{\theta_i y}\left(\varphi\left(\hat{y}, \widehat{\theta}_{-i}\right), \widehat{\theta}_{-i}, \hat{y}\right)\varphi_y\left(\hat{y}, \widehat{\theta}_{-i}\right).$$

Therefore, the second derivative of the agent's utility at $\widehat{y}$ is

$$w''(\widehat{y}) - c_{yy}(\theta, \widehat{y}) = c_{yy}(\widehat{\theta}, \widehat{y}) - c_{yy}(\theta, \widehat{y}) + c_{\theta_i y}\left(\varphi\left(\hat{y}, \widehat{\theta}_{-i}\right), \widehat{\theta}_{-i}, \hat{y}\right)\varphi_y\left(\hat{y}, \widehat{\theta}_{-i}\right).$$

However, Assumption 1 implies that $c_{yy}(\widehat{\theta}, \widehat{y}) - c_{yy}(\theta, \widehat{y}) = 0$ because $c_{\theta_i yy} = 0$ for all i. Then, the sign of the second derivative is given by (2), i.e., it is negative. Again, for critical $\widehat{y}$ we can use continuity. This implies that $y(\theta)$ is the maximum on $y(\Theta)$. This concludes the proof. ■

Next, we use the characterization from Proposition 1 to study implications of incentive-compatibility when the SCC is not imposed. Our assumption that the activity y is costly implies that transfers must be strictly increasing in y. Therefore, $w'(y) > 0$ is a necessary condition for incentive-compatibility. The theorem below states that it is also a sufficient condition. More specifically, given any mechanism $(y(\theta), w(y))$ we can find a cost function satisfying Assumptions 0 and 1 for which such schedule is incentive-compatible.

Theorem 1 *Let $y(\cdot)$ be a regular function and let $w(\cdot)$ be a positive C^2 function. There exists a C^1 cost function satisfying Assumption 1 for which $(y(\cdot), w(\cdot))$ is incentive-compatible if and only if $w(\cdot)$ is strictly increasing. Moreover, if $w(\cdot)$ is concave, such cost function can be chosen such that Assumption 0 is also satisfied.*

Proof. ($\Rightarrow$) Follows from revealed preference.

($\Leftarrow$) Let us define the following C^1 function:

$$c(\theta, y) = A(\theta) + w'(y(\theta))y + \frac{K}{2}(y - y(\theta))^2,$$

where $K > 0$ is a constant such that $w''(y(\theta)) < K$ and $A(\theta)$ is such that $c(\theta, y) > 0$ and $c_{\theta_i}(\theta, y) < 0$. Such function $A(\theta)$ exists.

The function c is C^1 because $y(\Theta)$ is a compact set and $w(\cdot)$ is C^2. Moreover, the marginal cost $c_y(\theta, y) = w'(y(\theta)) + K(y - y(\theta))$ is always positive along $y = y(\theta)$. A sufficient condition for the existence of a K such that the marginal cost is positive for all $y \geq 0$ and $w''(y(\theta)) < K$ is $w''(y(\theta)) < \frac{w'(y(\theta))}{y(\theta)}$ for all θ. This condition is obviously satisfied when the wage function is concave since $w' > 0$ and Θ is compact.

Note that c satisfies Assumption 1. We claim that the pair $(y(\cdot), w(\cdot))$ is incentive-compatible. First, (1) is trivial and (2) holds because

$$c_{\theta_i y}(\theta, y(\theta))y_{\theta_i}(\theta) = \left[w''(y(\theta)) - K\right]\left[y_{\theta_i}(\theta)\right]^2 \leq 0.$$

Furthermore, if θ and $\widetilde{\theta}$ are regular values of $y(\cdot)$, then

$$c_y(\theta, y) = c_y(\widetilde{\theta}, y) \Longleftrightarrow w'(y(\theta)) - Ky(\theta) = w'(y(\widetilde{\theta})) - Ky(\widetilde{\theta})$$
$$\Longleftrightarrow y = y(\theta) = y(\widetilde{\theta}),$$

since $w'(y) - Ky$ is decreasing on y (by the assumption on K), i.e., (3) holds. Using Proposition 1, we conclude the proof. ■

The theorem above implies that, apart from the monotonicity of the transfer function, incentive-compatibility by itself does not lead to any additional restrictions on the space of incentive-compatible mechanisms. In the next sections, we characterize the solutions of screening and signaling models and analyze which additional restrictions arise.

3 The Screening Model

In this section, we embed the structure of the previous section into a screening model. There is one uninformed principal who makes a take-it-or-leave-it offer to an informed agent. The agent's private information is characterized by the parameter θ, which is distributed according to the density p.

The revelation principle allows us to restrict the space of contracts to direct mechanisms that satisfy the incentive-compatibility constraint (IC). Each type has an outside option that gives a constant reservation utility normalized to zero.[10] Therefore, the principal faces the following participation constraint:

$$w(y(\theta)) - c(\theta, y(\theta)) \geq 0 \qquad \forall \theta \in \Theta. \tag{IR}$$

The following definition states the principal's problem:

Definition 1 *The principal's program is*

$$\max_{(y(\theta), w(y))} E[f(\theta, y(\theta)) - w(y(\theta))] \tag{4}$$
$$\textit{s.t. (IC) and (IR).}$$

Proposition 1 states that, under Assumption 1, the space of incentive-compatible mechanisms is characterized by conditions (1), (2) and (3). Define the agent's informational rent given an incentive-compatible mechanism $(y(\theta), w(y))$ as

$$r(\theta) \equiv w(y(\theta)) - c(\theta, y(\theta)). \tag{5}$$

Applying the Envelope Theorem (see Milgrom and Segal, 2002), we obtain the following condition:

$$\nabla r(\theta) = -\nabla_\theta c(\theta, y(\theta)), \tag{6}$$

where ∇ is the gradient operator. Note that condition (6) is equivalent to (1).

[10] One could also allow for type-dependent reservation utilities. For the purposes of our results on the additional restrictions imposed by screening, we can always ajust the cost function to avoid countervailing incentives.

In order to solve the principal's program, we follow the standard approach of considering first a relaxed program which ignores some of the constraints. Then, we state conditions that ensure that the ignored constraints do not bind so that the solution of the relaxed program is the same as the solution of the principal's program.

The relaxed program is defined as the maximization of the principal's profit subject to the agent's first-order condition and the participation constraint:

$$\max_{(y(\theta),w(y))} E[f(\theta,y(\theta)) - w(y(\theta))]$$

$$s.t.\ (1) \text{ and (IR)}.$$

Note that the program above does not take into account the local second-order conditions (2) and the global conditions (3). For clarity, it is convenient to analyze the one-dimensional and the multidimensional cases separately.

3.1 The one-dimensional case

This subsection characterizes the solutions of screening models when the parameter of private information θ is one-dimensional and considers their empirical implications. We obtain a new necessary and sufficient condition for a mechanism to be a solution to a screening model. The new condition, which does not depend on the SCC, states that the principal's profit must increase in the agent's type at a higher rate under asymmetric information than in the symmetric information case. Equivalently, the condition identifies the regions where the principal's profit is increasing and decreasing in the activity y chosen by the agent.

From our assumption about the type space, it can be written as $\Theta = \left[\underline{\theta}, \overline{\theta}\right] \subset \mathbb{R}$. Note that Assumption 0 implies that the informational rent is increasing in the agent's type. Therefore, the participation constraint (IR) is satisfied if and only if $r\left(\underline{\theta}\right) \geq 0$. Furthermore, in the solution to Program (4) it must be the case that $r(\underline{\theta}) = 0$.

Integrating (6), the agent's informational rent becomes

$$r(\theta) = r(\underline{\theta}) - \int_{\underline{\theta}}^{\theta} c_\theta(\widetilde{\theta}, y(\widetilde{\theta}))d\widetilde{\theta}.$$

Applying integration by parts, we obtain

$$E[w(y(\theta))] = r(\underline{\theta}) + E\left[c(\theta,y(\theta)) - \frac{1-P(\theta)}{p(\theta)}c_\theta(\theta,y(\theta))\right],$$

for every incentive-compatible mechanism $(y(\theta), w(y))$.

Substituting into the objective function, the relaxed program becomes:

$$\max_{(y(\theta),w(y))} E\left[f\left(\theta,y\left(\theta\right)\right) - c\left(\theta,y\left(\theta\right)\right) + \frac{1-P(\theta)}{p(\theta)}c_\theta\left(\theta,y\left(\theta\right)\right)\right].$$

The pointwise first-order condition of the relaxed program is equivalent to

$$f_y\left(\theta,y\left(\theta\right)\right) - c_y\left(\theta,y\left(\theta\right)\right) + \frac{1-P(\theta)}{p(\theta)}c_{\theta y}\left(\theta,y\left(\theta\right)\right) = 0. \tag{7}$$

This condition depicts the usual trade-off between rent extraction and distortion that the principal faces. Instead of equating the marginal valuation to the marginal cost, the principal sets it equal to the marginal cost plus the marginal cost of information rents.

For simplicity, suppose that w is C^2. Taking the total derivative of equation (1) with respect to θ gives:

$$w''(y(\theta))y'(\theta) = c_{\theta y}(\theta, y(\theta)) + c_{yy}(\theta, y(\theta))y'(\theta).$$

Thus, (7) can be written as

$$f_y\left(\theta, y\left(\theta\right)\right) - w'(y(\theta)) + \frac{1-P(\theta)}{p(\theta)}[w''(y(\theta)) - c_{yy}(\theta, y(\theta))]y'(\theta) = 0. \tag{8}$$

Note that the (necessary) local second-order condition of Program (IC) is $w''(y(\theta)) - c_{yy}\left(\theta, y\left(\theta\right)\right) \leq 0$. Substituting from equation (8), it follows that

$$[f_y\left(\theta, y\left(\theta\right)\right) - w'(y(\theta))]y'(\theta) \geq 0, \text{ for all } \theta \in \Theta, \tag{9}$$

since the hazard rate is always positive.

Condition (9) has a natural interpretation in terms of the principal's profit: $\pi\left(\theta\right) \equiv f\left(\theta, y\left(\theta\right)\right) - w\left(y\left(\theta\right)\right).$ Under symmetric information, we would have $\pi'\left(\theta\right) = f_\theta\left(\theta, y\left(\theta\right)\right).$ Differentiating the profit function, we obtain:

$$\pi'\left(\theta\right) = f_\theta\left(\theta, y\left(\theta\right)\right) + \left[f_y\left(\theta, y\left(\theta\right)\right) - w'\left(y\left(\theta\right)\right)\right] y'\left(\theta\right) \geq f_\theta\left(\theta, y\left(\theta\right)\right),$$

where the inequality uses equation (9). Therefore, under asymmetric information, the principal's profit increases at a greater rate than the increase in productivity. Note that this result is quite general in that it does not depend on any assumptions on the cost function except for it being increasing in y and decreasing in θ and the assumption that types are one-dimensional. The next subsection shows that this result can be somewhat generalized to multidimensional types.

Condition (9) can also be interpreted as determining the effect of activity y on the principal's profit. From condition (2), it can be written as:

$$f_y\left(\theta, y\left(\theta\right)\right) \left\{ \begin{array}{c} \geq \\ \leq \end{array} \right\} w'(y(\theta)) \iff c_{\theta y}\left(\theta, y\left(\theta\right)\right) \left\{ \begin{array}{c} \leq \\ \geq \end{array} \right\} 0. \tag{10}$$

Thus, the profit must be increasing in y in the region where the type decreases the marginal cost of the activity (controlling for the type θ).[11] Conversely, the profit must be decreasing when types increase the marginal cost of the activity y. In the standard case where $c_{\theta y} < 0$ (the single-crossing condition is satisfied), it implies that the principal's profit is increasing in the activity y controlling for the agent's type.

The previous argument shows the necessity of condition (9). It turns out that this condition is also sufficient to rationalize any incentive-compatible mechanism as the solution of the principal's program when Assumption 1 is satisfied:

[11] By 'controlling for the type', we mean the effect on the profit if an agent chose a different amount of activity y. Of course, in equilibrium we only observe one action for each type: $y = y\left(\theta\right)$.

Theorem 2 *Suppose $\Theta \subset \mathbb{R}$. Let $y(\cdot)$ be a regular function and let $w(\cdot)$ be a positive C^2 function. There exists a C^1 cost function satisfying Assumption 1 and a distribution of types p for which $(y(\cdot), w(\cdot))$ is the optimal mechanism if and only if $w(\cdot)$ is strictly increasing and condition (9) is satisfied.*

Proof. ($\Rightarrow$) Follows from the preceding argument and Theorem 1.

($\Leftarrow$) Let $(y(\theta), w(y))$ be a mechanism satisfying the conditions of the theorem. Take the cost function of the proof of Theorem 1. Then, the mechanism satisfies incentive-compatibility and equation (7) becomes:

$$\frac{f_y(\theta, y(\theta)) - w'(y(\theta))}{[w''(y(\theta)) - K]y'(\theta)} + \frac{1 - P(\theta)}{p(\theta)} = 0. \tag{11}$$

Hence, by (9) we can then define the following function:

$$P(\theta) = 1 - A \times \exp\left[\int_\theta^{\overline{\theta}} \frac{K - w''(y(t))}{f_y(t, y(t)) - w'(y(t))} y'(t) dt\right],$$

where A is chosen such that $P(\underline{\theta}) = 0$. It is easy to see that $P(\cdot)$ is a cumulative distribution function which satisfies (11). Note that for such economy the second-order condition of the relaxed functional holds if and only if

$$f_{yy}(\theta, y(\theta)) - K \leq 0.$$

Therefore, choosing K with this property we conclude that $(y(\theta), w(y))$ is the solution of the principal's program.[12] ■

3.2 The multidimensional case

As in McAfee and McMillan (1988) and Rochet (1987), this subsection obtains necessary and sufficient conditions for implementability and optimality in a multidimensional screening model. The implementability result follows straight from Proposition 1, which characterizes the set of incentive-compatible mechanisms under Assumption 1. In this subsection, we use this result to characterize the optimal mechanism.

In general, the first difficulty in multidimensional screening is to deal with the integration of equation (6). In order to deal with this, we will follow the approach proposed by Armstrong (1996).

Assume that there exists a $\underline{\theta} \in \Theta$ such that $\underline{\theta} \leq \theta$ for all $\theta \in \Theta$. With no loss of generality, take $\underline{\theta} = 0$.[13] Under Assumption 0, the informational rent (5) is increasing in the agent's type. Thus, the participation constraint is satisfied if and only if $r(0) \geq 0$ so that, in the solution to Program (4), we must have $r(0) = 0$.

Consider the expected value of the agent's informational rent (5):

$$R = \int_\Theta r(\theta)p(\theta)d\theta.$$

[12] We can choose a cost function such that the marginal cost is always positive whenever $f_{yy}(\theta, y(\theta)) < \frac{w'(y(\theta))}{y(\theta)}$. A sufficient condition is that w is more concave than $f(\theta, \cdot)$ for all θ, i.e., $f_{yy}(\theta, y) < w''(y)$ for all $y \geq 0$ (see the proof of Theorem 1).

[13] This can be obtained by redefining the types as $\tilde{\theta} = \theta - \underline{\theta}$.

Define the function $v : [0,1] \to \mathbb{R}$ by $v(t) = \int\limits_{\Theta} r(t\theta)p(\theta)d\theta$. Then, it follows that $v'(t) = \int\limits_{\Theta} \theta \cdot \nabla r(t\theta)p(\theta)d\theta$, where $\cdot$ denotes the inner product. Because $r(0) = 0$, we have

$$v(0) = 0 \text{ and } v(1) = R.$$

Since $v(1) - v(0) = \int\limits_{\Theta} v'(t)dt$, it follows that

$$R = \int_0^1 \left(\int_{\Theta} \theta \cdot \nabla r(t\theta)p(\theta)d\theta \right) dt. \tag{12}$$

The envelope condition (6) implies that $R = -\int_0^1 \left(\int_{\Theta} \theta \cdot \nabla_{\theta} c(t\theta, y(t\theta))p(\theta)d\theta \right) dt$. For each $t > 0$, we apply the change of variables $\eta = t\theta$, which takes Θ into itself. Under this transformation, the term $\theta_i c_{\theta_i}(t\theta, y(t\theta))$ becomes

$$\frac{1}{t}\eta_i c_{\theta_i}(\eta, y(\eta)).$$

Then, the integral with respect to θ in equation (12) can be transformed according to

$$\int_{\Theta} \theta \cdot \nabla_{\theta} c(t\theta, y(t\theta))p(\theta)d\theta = t^{-n-1} \int_{\Theta} \eta \cdot \nabla_{\theta} c(\eta, y(\eta))p\left(\frac{\eta}{t}\right) d\eta,$$

so that the expected informational rent becomes

$$R = -\int_{\Theta} \eta \cdot \nabla_{\theta} c(\eta, y(\eta)) \left(\int_0^1 t^{-n-1} p\left(\frac{\eta}{t}\right) dt \right) d\eta.$$

Finally, letting $\tau = 1/t$ and defining $q(\eta) = \int_1^{\infty} \tau^{n-1} p(\tau\eta)d\tau$, we obtain

$$R = -\int_{\Theta} \eta \cdot \nabla_{\theta} c(\eta, y(\eta))q(\eta)d\eta.$$

Substituting R back into the principal's objective function, we obtain the following relaxed maximization problem

$$\max_{y(\cdot)} E\left[f(\theta, y(\theta)) - c(\theta, y(\theta)) + \theta \cdot \nabla_{\theta} c\left(\theta, y\left(\theta\right)\right) \frac{q(\theta)}{p(\theta)} \right], \tag{13}$$

where the expectation operator is taken with respect to the probability measure defined by the density $p(\theta)$.

The procedure used to derive the expression above is known as integration along rays. It only takes into account the constraints (1) along rays.[14] The term inside the expectation is the virtual surplus. It consists of the first-best social surplus $f(\theta, y(\theta)) - c(\theta, y(\theta))$ plus the distortion needed to prevent deviations along rays $\theta \cdot \nabla_{\theta} c(\theta, y(\theta)) \frac{q(\theta)}{p(\theta)}$ (which is negative because $c_{\theta_i}(\theta, y) < 0$).

[14] Note that the gradient of the rent function,

$$\nabla r(\theta) = \nabla_{\theta} c(\theta, y(\theta)),$$

is a conservative vector field since, under assumption A1,

$$\frac{\partial}{\partial \theta_j} \left(\frac{\partial c}{\partial y}(\theta, y(\theta)) \right) = \frac{\partial}{\partial \theta_i} \left(\frac{\partial c}{\partial y}(\theta, y(\theta)) \right)$$

for all $i, j = 1, ..., n$. Therefore, the integration along rays is unimportant to define procedure above.

Remark 1 *Armstrong (1996) characterizes the optimal "cost-based" tariff assuming homogeneity of the utility function with respect to the type parameter and separability of the agent's "cost-based" indirect utility. With an additional separability condition on the density function which depends directly on the endogenous separability of the indirect utility, he also shows that this tariff is optimal. McAfee and McMillan (1988) generalize the single-crossing condition for the multidimensional case. However, their condition are so restrictive that imply that the shadow price indifference curves have to be hyperplanes.*

Although we are restricted to utility functions that satisfy Assumption 1, we do not need any homogeneity and separability assumptions. Moreover, the marginal cost indifference curves may not be hyperplanes and, therefore, we do not assume the generalized single-crossing condition. Thus, our setup is not contained neither in Armstrong (1996) nor in McAfee and McMillan (1988).

The following lemma will be useful to characterize the solution of Program (13).

Lemma 3 *Suppose the cost function is convex. Let $y : \Theta \to \mathbb{R}$ be a profile of activities and define $\gamma(\theta) \equiv c_y(y(\theta), \theta)$ as the marginal cost associated with it. Let $\tilde{y} : \gamma(\Theta) \to \mathbb{R}$ be a function. Then:*

i. if $y = \tilde{y} \circ \gamma$ and $\tilde{y}$ is decreasing, then $y(\cdot)$ is implementable;

ii. if $y(\cdot)$ is implementable, then there exists a non-increasing $\tilde{y}$ such that $y = \tilde{y} \circ \gamma$.

Proof. First, note that Proposition 1 implies that $y(\cdot)$ is implementable if and only if it satisfies conditions (2) and (3).

(i) Let us verify conditions (2) and (3). We have

$$\begin{aligned} c_{\theta_i y}(\theta, y(\theta)) y_{\theta_i}(\theta) &= c_{\theta_i y}(\theta, y(\theta)) \tilde{y}'(\gamma(\theta)) \gamma_{\theta_i}(\theta) \\ &= c_{\theta_i y}(\theta, y(\theta)) \tilde{y}'(\gamma(\theta)) [c_{yy}(\theta, y(\theta)) y_{\theta_i}(\theta) + c_{\theta_i y}(\theta, y(\theta))], \end{aligned}$$

which implies that

$$[1 - c_{yy}(\theta, y(\theta)) \tilde{y}'(\gamma(\theta))] c_{\theta_i y}(\theta, y(\theta)) y_{\theta_i}(\theta) = [c_{\theta_i y}(\theta, y(\theta))]^2 \tilde{y}'(\gamma(\theta)).$$

Since $c(\theta, y)$ is convex and $\tilde{y}(\cdot)$ is decreasing, it follows that (2) holds.

Let $\theta, \hat{\theta} \in \Theta$ be such that $y(\theta) = y(\hat{\theta})$. Since $\tilde{y}(\cdot)$ is decreasing, $\gamma(\theta) = \gamma(\hat{\theta})$. Therefore, $c_y(\theta, y(\theta)) = c_y(\hat{\theta}, y(\hat{\theta}))$ or, equivalently,

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

Hence, (3) holds.

(ii) Now, suppose that conditions (2) and (3) hold. From condition (3), it follows that

$$y(\theta) = y(\hat{\theta}) \Rightarrow \gamma(\theta) = \gamma(\hat{\theta}),$$

for all $\theta, \hat{\theta} \in \Theta$. This means that the indifference curve of $y(\cdot)$ are contained in the indifference curves of $\gamma(\cdot)$. Using condition (2), we have, through radial directions, that

$$y(\theta) > y(\hat{\theta}) \Rightarrow \gamma(\theta) \leq \gamma(\hat{\theta}),$$

for all $\theta, \hat{\theta}$ in a given radius from 0. Again using (3), the last inequality holds in all Θ. Applying the representation theorem for preferences, there must exist a non-increasing $\tilde{y} : \gamma(\Theta) \to \mathbb{R}$ such that $y(\theta) = \tilde{y} \circ \gamma(\theta)$, for all $\theta \in \Theta$. ∎

Lemma 3 and Proposition 1 imply that $y(\theta)$ is implementable by $w(y)$ if (1) holds and there exists a non-increasing function $\tilde{y}(\gamma)$ such that $y(\theta) = \tilde{y}(\gamma(\theta))$, where $\gamma(\theta) = c_y(\theta, y(\theta))$. Therefore, there is no loss of generality in considering the indirect mechanism where the message space corresponds to the set of possible marginal costs of taking the action y. Each type reveals its marginal cost $\gamma(\theta)$ of taking the prescribed action, takes the action $\tilde{y}(\gamma(\theta))$, and receives a transfer of $w\left(\tilde{y}(\gamma(\theta))\right)$.

Taking the conditional expectation of Program (13) and applying the law of iterated expectations, we obtain the following first-order condition:

$$E\left[f_y(\theta, \tilde{y}(\gamma(\theta)))|\gamma(\theta) = \gamma\right] - \gamma + E\left[\theta.\nabla_\theta c_y(\theta, y(\theta)) \frac{q(\theta)}{p(\theta)}|\gamma(\theta) = \gamma\right] = 0. \quad (14)$$

Suppose that there exists an implicit decreasing and non-negative solution of equation (14), $\tilde{y}^*(\gamma)$.[15] Then, the following theorem establishes that $y^*(\theta) = \tilde{y}^*(\gamma(\theta))$ is the solution of Program (13):

Theorem 3 *Suppose the valuation function f concave in y and the cost function c is convex in y and satisfies Assumption 1. If equation (14) defines a decreasing function $\tilde{y}^* : \gamma(\Theta) \to \mathbb{R}_+$ which is integrable,[16] then $y^*(\theta) = \tilde{y}^*(\gamma(\theta))$ is an optimal profile of activities.*

Equation (14) has an intuitive interpretation in terms of projections. The unrestricted optimum of Program (13) is the pointwise maximization of the virtual surplus. However, this solution may not satisfy conditions (2) and (3). From Lemma 3, any profile of actions that satisfies these conditions can be written as an indirect profile that is a function of θ only through the marginal cost of taking the action $\gamma(\theta)$. Then, condition (15) states that *the solution of the principal's program is determined by the first-order condition of the projection of the virtual surplus on the space of marginal costs $\gamma(\Theta)$.*

Recall that, from Assumption 1, the cost function is $c(\theta, y) = \xi(y) + y \times \psi(\theta) + \varphi(\theta)$. In order to prove Theorem 3, it is useful to consider indirect mechanisms where the message space consists of $\psi(\Theta)$. Although the space of marginal costs $\gamma(\Theta)$ is more intuitive than the space $\psi(\Theta)$, it is harder to work with since the marginal cost $\gamma(\theta)$ is a function of the profile of actions $y(\cdot)$, which is endogenous. However, when costs are convex, working with both message spaces is equivalent:

Lemma 4 *Consider a convex cost function satisfying Assumption 1, let $y : \Theta \to \mathbb{R}$ be a profile of activities, and define $\gamma(\theta) \equiv c_y(\theta, y(\theta))$ as the marginal cost associated with it. There exists a strictly increasing function $\phi : \gamma(\Theta) \to \psi(\Theta)$ such that $\phi(\gamma(\theta)) = \psi(\theta)$ for all $\theta \in \Theta$.*

[15] If this relaxed solution does not satisfy the monotonicity condition one has to perform the usual "ironing principle" (see Mussa and Rosen, 1978). Notice that since $y(\gamma)$ is one-dimensional function, this is straightforward exercise.

[16] Integrability is ensured if, for example, $\lim_{y\to\infty} f(\theta, y) - c(\theta, y) = -\infty$ for all θ. This implies that the implicit solution of (14) is bounded. Because Θ is compact, it follows that it is integrable.

Proof. Since $\gamma(\theta) = \xi'(y(\theta)) + \psi(\theta)$, it follows that

$$\gamma - \xi'(z(\gamma)) = \psi.$$

The left hand side is an increasing function of γ because ξ' is a non-decreasing function (remember that $\xi'' = c_{yy} \geq 0$). Thus, γ is an increasing transformation of ψ, which concludes the proof. ■

Therefore, there is no loss of generality in considering indirect mechanisms where each type sends message $\psi(\theta)$, takes an action $\hat{y}(\psi(\theta))$ and obtains a transfer $w(\hat{y}(\gamma(\theta)))$. Proceeding as in equation (14), we obtain:

$$E\left[f_y(\theta, \hat{y}(\psi(\theta)))|\psi(\theta) = \psi\right] - \xi'(y) - \psi(\theta) + E\left[\theta.\nabla_\theta \psi(\theta) \frac{q(\theta)}{p(\theta)}|\psi(\theta) = \psi\right] = 0. \quad (15)$$

Lemma 5 *Suppose the valuation function f concave in y and the cost function c is convex in y and satisfies Assumption 1. If equation (15) defines a decreasing function $\hat{y}^* : \psi(\Theta) \to \mathbb{R}$ which is integrable, then $y^*(\theta) = \hat{y}^*(\psi(\theta))$ is an optimal profile of activities.*

Proof. Suppose $\hat{y}(\psi)$ satisfies (15) and let $\hat{z}(\psi)$ be an arbitrary implementable profile of activities. By the previous lemma we can suppose that $\hat{y}$ and z are non-increasing functions. Let $y(\theta) := \hat{y}(\psi(\theta))$ and $z(\theta) := \hat{z}(\psi(\theta))$.

Define the following functional

$$F[z] := E\left[D(\theta, z(\theta))\right],$$

where

$$D(\theta, z(\theta)) := f(\theta, z(\theta)) - c(\theta, z(\theta)) + \theta.\nabla c(\theta, z(\theta)) \frac{q(\theta)}{p(\theta)}.$$

Note that $F[z]$ consists of the objective function from Program (13) evaluated at $z(\theta)$.

Since $D(\theta, \cdot)$ is a concave function for each θ,

$$D(\theta, z(\theta)) - D(\theta, y(\theta)) \leq D_y(\theta, y(\theta))\left[z(\theta) - y(\theta)\right].$$

Taking the law of iterated expectations, yields

$$F[z] - F[y] \leq E\left\{E\left[D_y(\theta, \hat{y}(\psi))|\psi(\theta) = \psi\right]\left[\hat{z}(\psi) - \hat{y}(\psi)\right]\right\}.$$

However, (15) implies that $E[D_y(\theta, \hat{y}(\psi))|\psi(\theta) = \psi] = 0$. Thus, $F[z] - F[y] \leq 0$. ■

Theorem 3 then follows immediately from Lemmata 4 and 5. Note that, because $\nabla_\theta y(\theta) = \widehat{y}'(\psi)\nabla_\theta \psi(\theta)$,

$$\widehat{y}'(\psi) E\left[\theta \cdot \nabla_\theta \psi(\theta) \frac{q(\theta)}{p(\theta)}|\psi(\theta) = \psi\right] = E\left[\theta \cdot \nabla_\theta y(\theta) \frac{q(\theta)}{p(\theta)}|\psi(\theta) = \psi\right].$$

Substituting in equation (15), we obtain

$$\begin{aligned} & \{E[f_y(\widehat{y}(\psi), \theta)|\psi(\theta) = \psi] - \gamma\} \times E\left[\theta \cdot \nabla_\theta y(\theta) \frac{q(\theta)}{p(\theta)}|\psi(\theta) = \psi\right] \quad (16) \\ = \quad & -\widehat{y}'(\psi)\left(E\left[\theta \cdot \nabla_\theta \psi(\theta) \frac{q(\theta)}{p(\theta)}|\psi(\theta) = \psi\right]\right)^2 \geq 0, \end{aligned}$$

where the inequality uses the fact that, by Lemmata 3 and 4, $\widehat{y}$ is non-increasing. Note that equation (1) implies that $\gamma(\theta) = w'(y(\theta))$. Therefore, the inequality above generalizes condition (9) for the multidimensional case.

The following examples illustrate the usefulness of the characterization in Theorem (3):

Example 6 (One-dimensional model with SCC) *Take Θ to be an interval in $\mathbb{R}$ and let $c_{\theta y} < 0$, $c_{yy} \geq 0$. Since $n = 1$, we have*

$$\theta q(\theta) = \int_1^\infty \theta p(\tau\theta) d\tau = 1 - P(\theta).$$

Because $\gamma(\theta)$ is a decreasing function, we can apply a change of variables from γ to θ. Then, equation (14) becomes

$$f_y(\theta, y(\theta)) - c_y(\theta, y(\theta)) = -c_{\theta y}(\theta, y(\theta)) \frac{1 - P(\theta)}{p(\theta)},$$

which is the standard first-order condition of the unidimensional relaxed problem.

Example 7 (One-dimensional Labor Market without SCC) *In the model from Example 1, let $\beta = 0$.*[17] *The cost function becomes $\hat{c}(\theta_1, y, \overline{g}) = \frac{y}{\theta_1(\overline{g} - \alpha\theta_1)}$, which satisfies Assumption 1. The marginal cost of schooling is $\gamma(\theta_1) = \psi(\theta_1) = \frac{1}{\theta_1(\overline{g} - \alpha\theta_1)}$. Suppose that θ_1 conditional on $\overline{g}$ is uniformly distributed on $\left[\underline{\theta}, \overline{\theta}\right]$. In the appendix, we show that condition (14) becomes:*

$$f_y\left(\frac{\overline{g} - \sqrt{\overline{g}^2 - \frac{4\alpha}{\psi}}}{2\alpha}, \hat{y}(\psi)\right) + f_y\left(\frac{\overline{g} + \sqrt{\overline{g}^2 - \frac{4\alpha}{\psi}}}{2\alpha}, \hat{y}(\psi)\right) = \overline{g}\left(\overline{\theta} + \frac{\overline{g}}{\alpha}\right)\psi^2 - 2\psi. \quad (17)$$

This equation characterizes the solution of the signaling model presented by Araujo, Gottlieb, and Moreira (2007) in a screening environment. If f is type-independent (so that it can be written as $f(y)$), the solution is $\hat{y}(\psi) = f_y^{-1}\left(\frac{\overline{g}}{2}\left(\overline{\theta} + \frac{\overline{g}}{\alpha}\right)\psi^2 - \psi\right)$.

Example 8 (Two-dimensional Labor Market) *Let $\theta = (\theta_1, \theta_2)$ be uniformly distributed on $[0,1]^2$ and suppose the valuation function is type-independent (so that it can be written as $f(y)$). As in Example 1, take $c(\theta, y) = \frac{y}{\theta_1\theta_2}$. The marginal cost function is given by $\gamma(\theta) = \psi(\theta) = \frac{1}{\theta_1\theta_2}$. Then, condition (14) becomes*

$$f'(\widetilde{y}(\gamma)) = \gamma\left\{1 + E\left[\left.\frac{q(\theta)}{p(\theta)}\right| \gamma(\theta) = \gamma\right]\right\}.$$

In the appendix, we show that $E\left[\left.\frac{q(\theta)}{p(\theta)}\right| \gamma(\theta) = \gamma\right] = \frac{1}{2\gamma} - \frac{1}{2\gamma^2} - \frac{\ln(\gamma)}{2\gamma^2}$. Thus, the solution is characterized by[18]

$$f'(\tilde{y}(\gamma)) = \frac{2\gamma^2 + \gamma - 1 - \ln(\gamma)}{2\gamma},$$

[17] See Araujo and Moreira (2004) for the case $\beta \neq 0$. The analysis of this case is more complex since implementability is not necessary by the local and global conditions presented in this paper.

[18] It is straightforward to show that the implicit solution $\tilde{y}(\gamma)$ is decreasing in γ.

or, in terms of the agent's type,

$$f'\left(y\left(\theta_1,\theta_2\right)\right)=\frac{1}{\theta_1\theta_2}+\frac{1}{2}-\frac{\theta_1\theta_2}{2}\left[1-\ln\left(\theta_1\theta_2\right)\right].$$

Example 9 (Nonlinear Pricing) *Consider the model of Example 3. Since* $f_y = MC$ *and* $\gamma(\theta) = c_y(\theta, y(\theta)) = V_Q\left(\theta,\bar{Q}-y(\theta)\right)$, *equation (14) becomes*

$$V_Q\left(\theta,Q\left(\theta\right)\right)=MC+E\left[\theta.\nabla_\theta V_Q(\theta,Q(\theta))\frac{q(\theta)}{p(\theta)}|\gamma(\theta)=\gamma\right].$$

When types are one-dimensional and the SCC is satisfied, this equation reduces to the standard nonlinear pricing rule:

$$V_Q\left(\theta,Q\left(\theta\right)\right)=MC+V_{\theta Q}\left(\theta,Q\left(\theta\right)\right)\frac{1-F\left(\theta\right)}{f\left(\theta\right)}.$$

The example below relates condition (16) to its one-dimensional counterpart (9):

Example 10 (Multidimensional Screening) *Take* $n \geq 2$ *and suppose that the distribution function is homogeneous of degree* $\alpha < 2-n$. *By homogeneity,* $\frac{q(\theta)}{p(\theta)} > 0$ *can be factored out of the conditional expectation in condition (16). Thus, we obtain*

$$\left\{E[f_y(\theta y(\theta))|\psi(\theta)=\psi]-w'\left(y\left(\theta\right)\right)\right\}E\left[\theta\cdot\nabla_\theta y(\theta)|\psi(\theta)=\psi\right]\geq 0,$$

which is the multidimensional equivalent of condition (9).

Suppose that the valuation function is type-independent (i.e., it can be written as $f(y)$*) and the cost cost function is homogenous of degree* $\beta < 0$ *on* θ.[19] *Then, the first-order condition (14) yields*

$$f'\left(\widetilde{y}\left(\gamma\right)\right)-\gamma+a\beta\gamma=0,$$

where $a=\int_1^\infty \tau^{\alpha+n-1}d\tau<\infty$ *since* $\alpha < 2-n$. *From (1), this condition, after multiplying by* $\widetilde{y}'(\gamma)$, *becomes*

$$\left[f'(\widetilde{y}\left(\gamma\right))-w'(\widetilde{y}\left(\gamma\right))\right]\widetilde{y}'\left(\gamma\right)=a\beta\gamma\widetilde{y}'\left(\gamma\right)\leq 0.$$

Thus, in this case, condition (9) holds when we identify each type by its marginal cost of the activity γ *(as was shown in the proof of Lemma 3,* $\tilde{y}$ *is decreasing).*

Reciprocally, for each mechanism $(y(\cdot), w(y))$ we are able to find an economy such that the mechanism is the solution of the principal's program. This is formally stated in the following theorem:

Theorem 4 *Suppose* $\Theta = \mathbb{R}_+^n$. *Let* $y : \Theta \to \mathbb{R}_+$ *be a regular function and* $w(y)$ *be a positive* C^2 *function. There exist a valuation function linear in* y, *a* C^1 *cost function satisfying Assumptions 0 and 1, and a distribution of types* p *for which* $(y(\theta), w(y))$ *is the optimal mechanism if and only if* $w(y)$ *is strictly increasing and condition (16) is satisfied.*

[19] Armstrong (1996) assumes $\beta = -1$.

Proof. ($\Rightarrow$) Follows from an adaptation of Theorem 1.

($\Leftarrow$) Let $(y(\theta), w(y))$ be a mechanism satisfying the conditions of the Theorem. Applying Theorem 1, we can find a cost function satisfying Assumptions A0 and A1 such that this mechanism is incentive compatible. Notice that $\gamma(\theta) = w'(y(\theta))$. Take a density p over Θ that is homogeneous of degree $\alpha < 2 - n$. Hence, maximizing (13) pointwise is equivalent to setting $f(\theta, y) = \mu(\gamma)y$, where

$$\mu(\gamma) = \gamma - aE\left[\theta.\nabla_\theta c_y(\theta, y(\theta))|\gamma(\theta) = \gamma\right]$$

satisfies equation (15), which is necessary and sufficient for optimality, and a is defined in the previous example. ■

The participation constraint (IR) implies that all types participate in the mechanism. When $\varphi(\theta) + \xi(0) = 0$ for all θ (as in Examples 6, 7, 8) there is no loss of generality in assuming so. In general, however, it may be optimal to exclude some types. In that case, it is useful to distinguish between two types of models:

1. Certain Participation: $\varphi(\theta) = \bar{\varphi}$ for all $\theta \in \Theta$,
2. Random Participation: $\varphi(\theta)$ is non-constant in $\theta \in \Theta$.

Next, we study the exclusion of types in models with certain and random participation separately.

3.2.1 Certain Participation

Without loss of generality, we can normalize $\bar{\varphi} = 0$. It is straightforward to show that the exclusion region is defined by $\Theta_0 = \{\theta \in \Theta; \gamma(\theta) \geq \gamma_0\}$, where γ_0 is such that $\tilde{y}(\gamma_0) = 0$. Thus, it follows that, if equation (14) defines a decreasing function $\tilde{y}^* : \gamma(\Theta) \to \mathbb{R}$ which is integrable, the optimal profile of activities is given by

$$y^*(\theta) = \max\{\tilde{y}^*(\gamma(\theta)), 0\}.$$

Equivalently, if equation (15) defines a decreasing function $\hat{y}^* : \psi(\Theta) \to \mathbb{R}$, then the optimal profile of activities is $y^*(\theta) = \max\{\tilde{y}^*(\psi(\theta)), 0\}$.

3.2.2 Random Participation

Recall that, by Lemmata 3 and 4, there is no loss of generality in considering indirect mechanisms where the message space is $\psi(\Theta)$. Let $\hat{r}(\psi) = w(\hat{y}(\psi)) - \xi(\hat{y}(\psi)) - \psi\hat{y}(\psi)$. Given any fixed mechanism, types which participate are those whose outside options are lower than $\hat{r}$, i.e.,

$$\hat{r}(\psi) \geq \varphi(\theta).$$

Then, the principal's expected payoff is

$$E\left\{\left[f(\theta, \hat{y}(\psi)) - \xi(\hat{y}(\psi)) - \psi\hat{y}(\psi) - \hat{r}(\psi)\right]\mathbf{1}_{[\hat{r}(\psi) \geq \varphi(\theta)]}\right\},$$

where $\mathbf{1}$ denotes the indicator function. If the solution $\hat{y}(\psi)$ of this program is a decreasing function, then $y(\theta) = \hat{y}(\psi(\theta))$ is the solution of the principal's program.

Assume that the valuation function f is type-independent. Hence, as in models of second-degree price discrimination, the agent's parameter of private information does not enter the principal's payoff directly. Then, applying the law of iterated expectations, the principal's payoff is

$$E\left\{\left[f\left(\hat{y}\left(\psi\right)\right)-\xi\left(\hat{y}\left(\psi\right)\right)-\psi\hat{y}\left(\psi\right)-\hat{r}(\psi)\right]E\left[\mathbf{1}_{\left[\hat{r}(\psi)\geq\varphi(\theta)\right]}\middle|\,\psi\left(\theta\right)=\psi\right]\right\}.$$

Let $G^{\varphi}\left(\psi,x\right)=E\left[\mathbf{1}_{\left[x\geq\varphi(\theta)\right]}\middle|\,\psi\right]$ denote the probability that a type with rent $r\left(\theta\right)=x$ participates conditional on $\psi\left(\theta\right)=\psi$. Denote the social surplus by $S\left(\psi,y\right)\equiv f\left(y\right)-\xi\left(y\right)-\psi y$. Following the approach of Rochet and Stole (2002), the principal's program can be written as the maximization of

$$E\left\{\left[S\left(\psi,\hat{y}\left(\psi\right)\right)-\hat{r}\left(\psi\right)\right]G^{\varphi}\left(\psi,\hat{r}\left(\psi\right)\right)\right\},$$

subject to $\hat{y}\left(.\right)$ being non-increasing and there existing a $w\left(.\right)$ such that $w(\hat{y}(\psi))=\hat{r}(\psi)+\xi(\hat{y}(\psi))+\psi\hat{y}(\psi)$.

By the envelope theorem, this program is equivalent to

$$\begin{gathered}\max_{\hat{y}(.)}E\left\{\left[S\left(\psi,\hat{y}\left(\psi\right)\right)-\hat{r}\left(\psi\right)\right]G^{\varphi}\left(\psi,\hat{r}\left(\psi\right)\right)\right\}\\ \text{subject to } \hat{r}'\left(\psi\right)=-\hat{y}\left(\psi\right),\ \text{and}\\ \hat{y}\left(\psi\right)\ \text{non-increasing.}\end{gathered}$$

Ignoring the monotonicity condition, the solution is characterized by the second-order differential (Euler equation):

$$\frac{d}{d\psi}\left[G^{\varphi}\left(\psi,\hat{r}\left(\psi\right)\right)S_{y}\left(\hat{y}\left(\psi\right),\psi\right)\right]+\frac{\partial}{\partial r}\left\{\left[S\left(\psi,\hat{y}\left(\psi\right)\right)-\hat{r}\left(\psi\right)\right]G^{\varphi}\left(\psi,\hat{r}\left(\psi\right)\right)\right\}=0,$$

with boundary conditions $\hat{y}\left(\psi_{m}\right)=y^{FB}(\theta_{m})$ and $\hat{y}\left(\psi_{M}\right)=y^{FB}(\theta_{M})$, where $\psi_{m}=\min\psi\left(\theta\right),\ \theta_{m}=\arg\min_{\theta\in\Theta}\psi\left(\theta\right),\ \psi_{M}=\max\psi\left(\theta\right),\ \theta_{M}=\arg\max_{\theta\in\Theta}\psi\left(\theta\right),\ y^{FB}\left(\theta\right)$ is the first-best solution (i.e., it satisfies $S_{y}(\theta,y^{FB}(\theta))=0$), and $\hat{r}'(\psi)=-\hat{y}(\psi)$.

Simplifying and using the fact that $\hat{r}'(\psi)=-\hat{y}(\psi)$, we obtain:

$$G^{\varphi}\left(\psi,\hat{r}\left(\psi\right)\right)\left[\left(\xi''\left(\hat{y}\left(\psi\right)\right)-f''\left(\hat{y}\left(\psi\right)\right)\right)\hat{r}''\left(\psi\right)-2\right]+\frac{\partial}{\partial\psi}G^{\varphi}\left(\psi,\hat{r}\left(\psi\right)\right)\left[f'\left(\hat{y}\left(\psi\right)\right)-\xi'\left(\hat{y}\left(\psi\right)\right)-\psi\right]$$

$$+\frac{\partial}{\partial r}G^{\varphi}\left(\psi,\hat{r}\left(\psi\right)\right)\left[\left(f'\left(\hat{y}\left(\psi\right)\right)-\xi'\left(\hat{y}\left(\psi\right)\right)\right)\hat{r}\left(\psi\right)+f\left(\hat{y}\left(\psi\right)\right)-\xi\left(\hat{y}\left(\psi\right)\right)-\hat{r}\left(\psi\right)\right]=0. \quad (18)$$

Assuming that equation (18) implicitly defines a non-increasing function $\hat{y}\left(\psi\right)$, it characterizes the solution of the random participation model when valuations are type-independent. If the solution $\hat{y}\left(\psi\right)$ is not non-increasing, the solution is obtained by applying an ironing procedure. The proposition below summarizes this result:

Proposition 2 *Suppose the valuation function f is type-independent and concave and suppose that the cost function c is convex in y and satisfies Assumption 1. If equation (18) defines a decreasing function $\hat{y}^{*}:\psi\left(\Theta\right)\rightarrow\mathbb{R}$ which is integrable, then $y^{*}\left(\theta\right)=\hat{y}^{*}(\psi\left(\theta\right))$ is an optimal profile of activities.*

Equation (18) generalizes the characterization of Rochet and Stole (2002) for arbitrary distributions of types and arbitrary cost functions satisfying Assumption 1. The following example shows that their model can be obtained as a special case of our characterization:

Example 11 (Rochet and Stole, 2002) *Let $\Theta \subset \mathbb{R}^2$ be an interval and denote types by $(t, x) \in \Theta$. Let $\psi(t, x) = -t$, $\varphi(t, x) = -x$, $f(y) - \xi(y) = \frac{y^2}{2}$. Assume that types t and x are independently distributed and denote by $f(t)$ and $G(x)$ the probability distribution of t and the cumulative distribution of x, respectively. Let $M(t, u) \equiv G^{\varphi}(t, u) = f(t)G(u)$.*

Then, equation (18) becomes

$$M_u(t, u)\left(u - \frac{1}{2}\dot{u}^2\right) + M(t, u)(2 - \ddot{u}) + M_t(t, u)(t - \dot{u}) = 0,$$

which is the equation obtained by Rochet and Stole.

Therefore, this section analyzed the solution of screening models under Assumption 1. Theorem 3 characterized the solution when all types participate. In 3.2.1, it was shown that this characterization can be easily generalized to the case of certain participation (i.e., $\varphi(\theta) = \bar{\varphi}$ for all θ). In 3.2.2, we characterized the solution with the exclusion of types when participation is random but valuation functions are type-independent. This characterization generalized the one presented by Rochet and Stole (2002).

In terms of empirical implications, Theorem 4 established that the screening model imposes two restrictions on the space of mechanisms $(y(\theta), w(y))$. First, by incentive-compatibility, it requires w to be monotonic. Second, maximization of the principal's payoff imposes condition (16). When types are one-dimensional or when the distribution function and the cost function are homogeneous and the valuation function is type-independent, this condition implies that the principal's profit as a function of types must increase at a greater rate under asymmetric information than the increase in productivity (which is equal to the rate of growth under symmetric information).

4 The Signaling Game

In this section, we consider a standard signaling game where preferences are as described in Section 2. There are many identical principals ('receivers') who act competitively. For simplicity, we consider the case where the principal's valuation function does not depend on the amount of the activity ('signal') exerted by the agent ('sender'):

$$f(\theta, y) = f(\theta).$$

Therefore, the only way through which the activity affects the transfer is through its informational content. Of course, our results extend to the case where the activity also affects the valuation. Our competitive assumption implies that the transfer is equal to the expected valuation of the sender conditional on the signaling activity.

The timing of the signaling game is as follows. First, nature determines the type of each sender, θ, according to the density function p. Then, senders choose the amount of signaling y contingent on their types. Subsequently, the market offers a transfer $w(y)$ conditional on the observed signal.

Since all receivers are equal, we will study symmetric equilibria where the offered wage schedule is the same for every receiver. As usual, we adopt the perfect Bayesian equilibrium concept. In what follows $E[\cdot|\cdot]$ represents the conditional expectation operator with respect to the measure of beliefs μ.

Definition 2 *A perfect Bayesian equilibrium (PBE) for the signaling game is a profile of strategies $(y(\theta), w(y))$ and beliefs $\mu(\cdot \mid y)$ such that:*

1. *The sender's strategy is optimal given the equilibrium wage schedule, i.e., (IC) holds.*
2. *The market is competitive (i.e., receivers earn zero profits):*
$$w(y) = E[f(\theta) \mid y(\cdot) = y]. \tag{19}$$
3. *Beliefs are consistent: $\mu(\theta \mid y)$ is derived from the sender's strategy using Bayes' rule where possible.*

Substituting the zero-profit condition (19) into the first-order condition from incentive-compatibility (1), we obtain:

$$y_{\theta_i}(\theta) = \frac{\frac{\partial}{\partial \theta_i} E[f(\cdot)|y(\cdot) = y(\theta)]}{c_y(\theta, y(\theta))}. \tag{20}$$

Consider a separating equilibrium. Bayes' rule implies that $w(y(\theta)) = E[f(\cdot)|y(\cdot) = y(\theta)] = f(\theta)$ for all θ. Therefore, equation (20) becomes $y_{\theta_i}(\theta) = \frac{f_{\theta_i}(\theta)}{c_y(\theta, y(\theta))}$. Therefore, Assumption 0 implies that we must have $y_{\theta_i}(\theta) > 0$ in any separating equilibrium. However, from the second-order condition from incentive compatibility (2), we can only have $y_{\theta_i}(\theta) > 0$ if $c_{\theta_i y}(\theta, y(\theta)) \leq 0$. Hence, *a fully separating equilibrium exists only if $c_{\theta_i y}(\theta, y(\theta)) \leq 0$ for all θ: The SCC is satisfied along the equilibrium signal y.*[20]

Given an equilibrium profile $(y(\cdot), w(\cdot))$, denote the type with the lowest amount of activity y, the amount he chooses, and the wage he gets by

$$\begin{aligned} \theta^{\min} &= \arg\min_{\hat{\theta} \in \Theta} y\left(\hat{\theta}\right), \\ y^{\min} &= y(\theta^{\min}), \text{ and} \\ w^{\min} &= w\left(\theta^{\min}\right). \end{aligned} \tag{21}$$

Let $f^{\min}$ denote the lowest valuation in the economy:

$$f^{\min} = \min_{\hat{\theta} \in \Theta} f\left(\hat{\theta}\right).$$

In a PBE we need to make sure that agents have no incentive to deviate to actions off the equilibrium path. In order to obtain a sufficient condition for types not to benefit

[20] This condition is slightly less demanding than the SCC, which states that $c_{\theta_i y}(\theta, y) < 0$ for all θ, y.

from deviating to actions off the equilibrium path, let beliefs off the equilibrium be given by

$$\mu(\theta|y) = \begin{cases} 1 \text{ if } \theta = \arg\min f\left(\hat{\theta}\right) \\ 0 \text{ if } \theta \neq \arg\min f\left(\hat{\theta}\right) \end{cases},$$

for $y \notin y(\Theta)$. By deviating to any $y \notin y(\Theta)$, an agent gets transfer $f^{\min}$, which gives a payoff of at most $f^{\min} - c(\theta, 0)$. Then, type θ does not want to deviate to any action off the equilibrium path $y \notin y(\Theta)$ if this payoff is lower than $E\left[f(\cdot) \mid y^{\min}\right] - c\left(\theta, y^{\min}\right)$. Therefore, types do not benefit from deviating to actions off the equilibrium path if $y^{\min} \geq 0$ satisfies

$$E\left[f(\theta) \mid y^{\min}\right] - f^{\min} \geq c\left(\theta, y^{\min}\right) - c(\theta, y), \quad \forall\theta, \tag{22}$$

for all $y \notin y(\Theta)$.

Notice that the set of actions $y^{\min} \geq 0$ such that this inequality is satisfied is nonempty (since $y^{\min} = 0$ satisfies this condition). Moreover, when $\theta^{\min}$ is separated, the only $y^{\min}$ compatible with (22) is $y^{\min} = 0$. Inequality (22) gives boundary conditions for the differential equation (1).

Conditions (1), (2), (3), (19), and (22) are necessary for a PBE. As in Example 5, they may not be sufficient when Assumption 1 does not hold. However, the following lemma states that they are sufficient for a PBE when Assumption 1 holds.

Lemma 6 *Suppose Assumption 1 holds and let $y(\cdot)$ and $w(\cdot)$ be regular functions. There exists a set of beliefs $\mu(\cdot \mid y)$ such that $(y(\cdot), w(\cdot), \mu(\cdot))$ is a PBE if and only if the first- and second-order conditions (1) and (2), the pooling condition (3), the zero-profit condition (19), and the boundary condition (22) are satisfied.*

Proof. ($\Rightarrow$) Straight from Proposition 1.

($\Leftarrow$) Define the transfer schedule as in Proposition 1, where $w^{\min}$ is as defined in (21).

Observe that $w(y^{\min}) = w^{\min}$ and, by (20), $w'(y(\theta))y_{\theta_i}(\theta) = \frac{\partial}{\partial\theta_i}E[f(\cdot)|y(\cdot) = y(\theta)]$ for every regular type θ and every i. Therefore, by continuity, $w(y) = E[f(\theta) \mid y(\cdot) = y]$ for all $y \in y(\Theta)$, i.e., the zero profit condition holds.

Lemma 1 has shown that the agent's strategy is optimal given the equilibrium transfer schedule. This concludes the proof. ■

Inequality (22) implies that, when $\theta^{\min}$ is not separated, there may exist several boundary conditions $y^{\min}$ that satisfy the requirements of a PBE. In order to deal with the multiplicity of equilibria, we will follow part of the literature by selecting based on an efficiency criterion. Thus, we impose as a selection criterion that

$$y^{\min} = 0. \tag{23}$$

We will refer to a PBE satisfying condition (23) as a *least costly equilibrium*. The proposition below states necessary and sufficient conditions for a least costly equilibrium.

Proposition 3 *Suppose Assumption 1 holds and let $y(\cdot)$ and $w(\cdot)$ be regular functions. There exists a set of beliefs $\mu(\cdot \mid y)$ such that $(y(\cdot), w(\cdot), \mu(\cdot))$ is a least costly equilibrium if and only if the first- and second-order conditions (1) and (2), the pooling condition (3), and the boundary condition (23) are satisfied.*

Proof. Follows straight from Lemma 6. ■

Proposition 3 characterizes the least costly equilibria of the signaling model. This characterization allows us to study which restrictions follow from a signaling model when the single-crossing condition is not imposed.

We know from Section 2 that incentive-compatibility implies that the transfer schedule must be strictly increasing. Furthermore, our selection criterion implies that there must be some type θ such that $y(\theta) = 0$. The theorem below shows that these are the only implications of the signaling model.

Theorem 5 *Let $y(\cdot)$ be a regular function and let $w(\cdot)$ be a positive C^2 function. There exists a C^1 cost function satisfying Assumption 1 and a distribution of types p for which $(y(\cdot), w(\cdot))$ is a least costly equilibrium profile of signals and transfers if and only if $w(\cdot)$ is strictly increasing and there exists a $\theta \in \Theta$ such that $y(\theta) = 0$. Furthermore, this equilibrium profile is the same for all distribution of types p.*

Proof. ($\Rightarrow$) Obvious.

($\Leftarrow$) Let us define the following C^1 functions: $c(\theta, y)$ as in the proof of Theorem 1 and

$$f(\theta) = w(y(\theta)).$$

Observe that c and f are non-negative C^1 functions. We claim that the pair $(y(\cdot), w(\cdot))$ are the profiles of signals and transfers in a least costly equilibrium for the economy $\{c, f, p\}$, for any density p. First, by Proposition 1 $(y(\cdot), w(\cdot))$ is incentive compatible which implies Condition 1 of the definition of the PBE. Moreover, $y(\theta) = y(\widetilde{\theta}) \Rightarrow f(\theta) = f(\widetilde{\theta})$, which implies that $E[f(\cdot)|y(\theta) = y] = f(\theta)$ and $\frac{\partial}{\partial \theta_i}\{E[f(\cdot)|y(\theta) = y]\} = f_{\theta_i}(\theta)$, for all θ and density p. We then only need to prove the first-order conditions (20) which is equivalent to

$$\frac{f_{\theta_i}(\theta)}{c_y(\theta, y(\theta))} = \frac{w'(y(\theta))y_{\theta_i}(\theta)}{w'(y(\theta))} = y_{\theta_i}(\theta).$$

Finally, the boundary condition (23) obviously holds. Using Theorem 1, we conclude the proof. ■

Since the equilibrium constructed in Theorem 5 holds for all distributions p and for any cost function $c(\theta, y) = w'(y(\theta))y + \frac{\hat{K}}{2}(y - y(\theta))^2$ with $\hat{K} > K$, it follows that the model is not identified given data on signals and transfers. Indeed, any distribution of types is compatible with the same equilibrium profile of signals and transfers. Thus, we have the following corollary:

Corollary 1 *Signaling models are not identified given data on signals and transfers: for every profile $(y(\cdot), w(\cdot))$ satisfying the conditions of Theorem 5, there is an infinite number of signaling models that have this profile as a least costly equilibrium.*

Remark 2 *The only robust property that emerges from the equilibrium is the monotonicity of the wage schedule. If the cost of signaling function is increasing on the signal then, by revealed preference, equilibrium wages are increasing.*

Therefore, in the context of education as a signal, the only robust implication of the signaling hypothesis is the monotonicity of wages in education. However, this result is also *shared by the usual human capital (symmetric information) models. Indeed, the revealed preference argument holds regardless of the link between education and productivity.*

In the context of advertising as a signal, Theorem 5 implies that revenues must be increasing in advertisement. Because this is implied by revealed preference it holds regardless of the relationship between quality and advertising. It is also shared by Becker and Murphy's (1993) model of advertisement as a good, for example.

Remark 3 *For one-dimensional type models, Theorem 5 says that signaling models are compatible with non-monotonic signaling functions. However, from (2), this is only possible when the SCC is violated. Recent works show that such non-monotonic equilibria may emerge and have important empirical consequences (see Araujo et al., 2007, and Araujo and Moreira, 2003).*

Theorem 5 does not allow one to control for the valuation function. From an applied perspective, it assumes that the econometrists do not observe the valuation function. Next, we show that the results from Theorem 5 remain even if one controls for the valuation function. More precisely, let $y(\theta)$ be a profile of signals and $w(y)$ be the associated transfer schedule that satisfy the conditions of Theorem 2. Fix any valuation function $f(\theta)$. We say that f is consistent with the given profile of signals and transfers if it satisfies the zero-profit condition:

$$w(y) = E[f(\cdot)|y(\theta) = y], \tag{24}$$

where the expectation is taken with respect to some distribution of types p. Definition 2 implies that this consistency requirement must be satisfied in any PBE. The following corollary establishes that any profile of signals and transfers satisfying (24) can be rationalized by a cost of signaling function and a distribution of types as a part of a least costly equilibrium of this economy (characterized by the fixed valuation function and the distribution of types). Formally, we have the following:

Corollary 2 *Let $y(\cdot)$ be a regular function, let $w(\cdot)$ be a C^2 function, and let $f(\cdot)$ be a valuation function satisfying (24) with respect to a continuous density p. There exists a C^1 cost function satisfying Assumption 1 for which $(y(\cdot), w(\cdot))$ is a least costly equilibrium profile of signals and transfers if and only if $w(\cdot)$ is strictly increasing and there exists a $\theta \in \Theta$ such that $y(\theta) = 0$.*

Proof. ($\Rightarrow$) Obvious.

($\Leftarrow$) Take the cost of signaling exactly as in the proof of Theorem 1 and the density p. Then, following the same steps of the proof of Theorem 5 establishes the result. ■

Theorem 5 and Corollary 2 show that there are only three implications of signaling models. First, the zero-profit condition implies that transfers have to be in the convex hull of the valuation of agents. Second, our selection criterion based on efficiency requires some type to choose zero amount of the activity (least costly equilibrium). However, this result is not robust since there are PBEs that do not satisfy this condition. Third, the fact that signaling is costly implies that transfers must be strictly increasing in signals. Although this conclusion is extremely robust, it is also implied by most alternative models. Any profiles of activities y and transfers w satisfying these three conditions is compatible with a least costly equilibrium of a signaling model.

Remark 4 (Multidimensional Actions) *It is possible to generalize the model presented here to allow for n-dimensional actions. Let $\mathbf{y} = (y_1, ..., y_n)$ be a vector of actions. Quinzii and Rochet (1985) assume that the cost of signaling function satisfies*

$$c(\theta, \mathbf{y}) = \sum_{i=1}^{n} \frac{y_i}{\theta_i}.$$

This specification implies that costs are additively separable in actions and that it satisfies the SCC in each dimension. Following the approach in this paper, it is straightforward to generalize our characterization of incentive-compatibility to cost functions of the form:

$$c(\theta, \mathbf{y}) = \xi(\mathbf{y}) + \varphi(\theta) + \sum_{i=1}^{n} y_i \times \psi_i(\theta).$$

where $\xi : \mathbb{R}_+^n \to \mathbb{R}$, $\varphi : \Theta \to \mathbb{R}$, $\psi_i : \Theta \to \mathbb{R}$, $i = 1, ..., n$. In particular, such specification does not imply that the SCC holds. It is also immediate to generalize the results from Theorems 1 and 5 for the $n \times n$ case. However, the characterization of the solution of the screening problem is not trivial in the $n \times n$ case.[21]

5 Conclusion

This paper studied incentive-compatibility when the single-crossing condition is not satisfied. This allowed us to provide a characterization of the solution of multidimensional screening models and the equilibria of multidimensional signaling models. Then, using our characterization, we analyzed the restrictions imposed by incentive-compatibility, screening, and signaling once the single-crossing condition does not hold.

First, it was shown that the only implication of incentive-compatibility was the monotonicity of transfers in actions. Then, in the case of screening models, we obtained a new necessary and sufficient condition. In the one-dimensional models, it implies the principal's profit to grow at a higher rate under asymmetric information than the it would grow under symmetric information.

In the case of signaling models, the zero-profit condition implies that transfers must be in the convex hull of valuations in each set of pooled types. We have also imposed a selection criterion that requires some type to choose a zero amount of the signal. We have shown that any profile of actions and transfers satisfying these conditions is an equilibrium of some economy.

Therefore, apart from these mild restrictions, the implications of signaling and screening models arise from assumptions of the cross-partial derivative of the cost function. In the absence of more precise knowledge of the cost function, we cannot obtain other testable predictions from these models.

[21] Although one can extend the integration by rays to get an expression analogous to Program (13), it is not obvious that Lemma 3 will remain true in this case. Consequently, the projection method used here to derive the first-order condition (14) may not be applicable.

Appendix

One-dimensional Labor Market Model without SCC: First note that $\psi = \frac{1}{\theta_1(\overline{g}-\alpha\theta_1)}$ implies that

$$\theta_1 = \frac{\overline{g} + \sqrt{\overline{g}^2 - \frac{4\alpha}{\psi}}}{2\alpha}, \text{ or } \theta_1 = \frac{\overline{g} - \sqrt{\overline{g}^2 - \frac{4\alpha}{\psi}}}{2\alpha}.$$

Because types are uniformly distributed, both types have the same probability conditional on $\psi(\theta) = \psi$. Thus,

$$\frac{\theta_1 q(\theta_1)}{p(\theta_1)} = \frac{1 - P(\theta_1)}{p(\theta_1)} = \overline{\theta} - \theta_1,$$

and, since $\psi'(\theta_1) = -(\overline{g} - 2\alpha\theta_1)\psi^2 = -\frac{\psi}{\theta_1} + \alpha\theta_1\psi^2$,

$$\begin{aligned} \frac{\theta_1 q(\theta_1)}{p(\theta_1)}\psi'(\theta_1) &= (\overline{\theta} - \theta_1)\left(-\frac{\psi}{\theta_1} + \alpha\theta_1\psi^2\right) \\ &= \overline{\theta}\left(-\frac{\psi}{\theta_1} + \alpha\theta_1\psi^2\right) + \psi - \alpha\theta_1^2\psi^2. \end{aligned}$$

Therefore,

$$E\left[\frac{\theta_1 q(\theta_1)}{p(\theta_1)}\psi'(\theta_1)\middle|\psi(\theta_1) = \psi\right] = -\overline{\theta}\psi E\left[\frac{1}{\theta_1}\middle|\psi(\theta_1) = \psi\right] + \psi - \alpha\psi^2 E\left[\theta_1^2\middle|\psi(\theta_1) = \psi\right].$$

Note that:

$$\begin{aligned} E\left[\frac{1}{\theta_1}\middle|\psi(\theta_1) = \psi\right] &= \frac{\psi\overline{g}}{2}, \text{ and} \\ E\left[\theta_1^2\middle|\psi(\theta_1) = \psi\right] &= \frac{\overline{g}^2}{2\alpha^2} - \frac{1}{\alpha\psi}. \end{aligned}$$

Hence, we obtain:

$$\begin{aligned} E\left[\frac{\theta_1 q(\theta_1)}{p(\theta_1)}\psi'(\theta_1)\middle|\psi(\theta_1) = \psi\right] &= -\frac{\overline{\theta}\overline{g}}{2}\psi^2 + \psi - \frac{\overline{g}^2}{2\alpha}\psi^2 + \psi \\ &= 2\psi - \frac{\overline{g}}{2}\left(\overline{\theta} + \frac{\overline{g}}{\alpha}\right)\psi^2. \end{aligned}$$

Substituting in condition (14), we obtain equation (17).

Two-dimensional Labor Market Model: Let $p = 1$ on $[0,1]^2$ and $\gamma = \frac{1}{\theta_1\theta_2}$. Thus, $\theta \cdot \nabla_\theta \gamma = -\gamma$, and

$$q(\theta) = \int_1^\infty \tau^{n-1} p(\tau\theta) d\tau = \left[\int_1^{\min\{1/\theta_1, 1/\theta_2\}} \tau d\tau\right]^+ = \frac{1}{2}\left[\min\{1/\theta_1, 1/\theta_2\}^2 - 1\right]^+.$$

The following lemma states the result claimed in the text.

Lemma. For all $\gamma \in [1, \infty)$,

$$E\left[\frac{q(\theta)}{p(\theta)}\middle|\gamma(\theta) = \gamma\right] = \frac{1}{2\gamma} - \frac{1}{2\gamma^2} + \frac{\ln(\gamma)}{2\gamma^2}.$$

Proof. We have to show that

$$E\left[E\left[\frac{q(\theta)}{p(\theta)}\middle|\gamma(\theta) = \gamma\right] h(\gamma)\right] = E\left[\frac{q(\theta)}{p(\theta)} h(\gamma)\right] \tag{25}$$

for any measurable function $h : [1, \infty) \to \mathbb{R}$. In order to calculate the expression on the right hand side of equation (25), we change variables from (θ_1, θ_2) to (θ_1, γ) :

$$T : [0,1]^2 \to [0,1] \times [1, \infty]$$
$$(\theta_1, \theta_2) \to (\theta_1, 1/\theta_1\theta_2)$$

Thus,

$$T^{-1} : D \to [0,1]^2$$
$$(\theta_1, \gamma) \to (\theta_1, 1/\gamma\theta_1)$$

where $D = T\left([0,1]^2\right) = \{(\theta_1, \gamma) \in [0,1] \times [1, \infty]; \gamma \geq 1/\theta_1\}$, and

$$|DT^{-1}| = \begin{vmatrix} 1 & 0 \\ -\frac{1}{\gamma\theta_1^2} & -\frac{1}{\gamma^2\theta_1} \end{vmatrix} = -\frac{1}{\gamma^2\theta_1}.$$

Hence,

$$\begin{aligned} E\left[\frac{q(\theta)}{p(\theta)}h(\gamma)\right] &= \int_0^1 \int_0^1 q(\theta)h(\gamma)d\theta_2 d\theta_1 \\ &= \int_0^1 \int_{1/\theta_1}^1 \widetilde{q}(\theta_1, \gamma)h(\gamma) \times \left(\frac{1}{\gamma^2\theta_1}\right) d\gamma d\theta_1 \\ &= \int_1^\infty \int_{1/\gamma}^1 \widetilde{q}(\theta_1, \gamma)\frac{1}{\gamma^2\theta_1}h(\gamma)d\theta_1 d\gamma, \end{aligned} \tag{26}$$

where the last equality comes from Fubini's theorem, and $\widetilde{q}(\theta_1, \gamma) = q(\theta_1, 1/\gamma\theta_1)$. Note that

$$\begin{aligned} \frac{1}{\theta_1} &\geq \frac{1}{\theta_2} = \gamma\theta_1 \Leftrightarrow \theta_1 \leq \frac{1}{\gamma^{1/2}} \\ \theta_2 &= \frac{1}{\gamma\theta_1} \leq 1 \Leftrightarrow \frac{1}{\gamma} \leq \theta_1. \end{aligned}$$

Therefore, we have that

$$\widetilde{q}(\theta_1, \gamma) = \frac{1}{2}\begin{cases} \gamma^2\theta_1^2 - 1, & \frac{1}{\gamma} \leq \theta_1 \leq \frac{1}{\gamma^{1/2}} \\ \frac{1}{\theta_1^2} - 1, & \frac{1}{\gamma^{1/2}} \leq \theta_1 \leq 1 \end{cases}.$$

Substituting in equation (26), yields

$$E\left[\frac{q(\theta)}{p(\theta)}h(\gamma)\right] = \int_1^\infty \frac{h(\gamma)}{2\gamma^2}\left\{\int_{1/\gamma}^{1/\gamma^{1/2}} \left(\gamma^2\theta_1 - \frac{1}{\theta_1}\right) d\theta_1 + \int_{1/\gamma^{1/2}}^1 \left(\frac{1}{\theta_1^3} - \frac{1}{\theta_1}\right) d\theta_1\right\} d\gamma. \tag{27}$$

Noting that

$$\int_{1/\gamma}^{1/\gamma^{1/2}} \left(\gamma^2\theta_1 - \frac{1}{\theta_1}\right) d\theta_1 = \gamma^2 \frac{\theta_1^2}{2}\bigg|_{1/\gamma}^{1/\gamma^{1/2}} - \ln(\theta_1)|_{1/\gamma}^{1/\gamma^{1/2}} = \frac{\gamma^2}{2}\left(\frac{1}{\gamma} - \frac{1}{\gamma^2}\right) - \frac{1}{2}\ln(\gamma),$$

and

$$\begin{aligned} \int_{1/\gamma^{1/2}}^1 \left(\frac{1}{\theta_1^3} - \frac{1}{\theta_1}\right) d\theta_1 &= -\frac{1}{2\theta_1^2}\bigg|_{1/\gamma^{1/2}}^1 - \ln(\theta_1)|_{1/\gamma^{1/2}}^1 \\ &= -\frac{1}{2} + \frac{\gamma}{2} - \frac{1}{2}\ln(\gamma), \end{aligned}$$

equation (27) implies that:

$$E\left[\frac{q(\theta)}{p(\theta)}h(\gamma)\right] = \int_1^\infty \frac{h(\gamma)}{2\gamma^2}\left(\gamma - 1 - \ln(\gamma)\right)d\gamma$$

for all h, which establishes the result. ■

References

Acemoglu, Daron, 1998. Credit Market Imperfections and the Separation of Ownership from Control. Journal of Economic Theory 78, 355-381.

Araujo, A., Moreira, H., 2003. Non-monotone insurance contracts and their empirical consequences. Discussion Paper #512, Getulio Vargas Foundation.

Araujo, A., Moreira, H., 2004. Adverse Selection Problems without the Spence-Mirrlees Condition. Discussion Paper #425, Getulio Vargas Foundation.

Araujo A., Gottlieb D., Moreira H., 2007. A Model of Mixed Signals with applications to Countersignaling, RAND Journal of Economics 38, 1020-1043.

Araujo, A., Moreira, H., Tsuchida, M., 2004. Do Dividends Signal More Earnings? A theoretical analysis. Discussion Paper #524, Getulio Vargas Foundation.

Armstrong, M., 1996. Multiproduct Nonlinear Pricing, Econometrica 64, 51-75.

Bagwell, L. S., Bernheim, B. D., 1996. Veblen Effects in a Theory of Conspicuous Consumption. American Economic Review 86, 394-377.

Becker, G. S., Murphy, K. M., 1993. A Simple Theory of Advertising as a Good or Bad. Quarterly Journal of Economics 108, 941-964.

Bernheim, B. D., 1991. Tax Policy and the Dividend Puzzle. Rand Journal of Economics 22; 455-476.

Bernheim, B. D., 1994. A Theory of Conformity. Journal of Political Economy 102; 841-877.

Bernheim, B. D., Redding, L. S., 2001. Optimal Money Burning: Theory and Application to Corporate Dividend Policy, Journal of Economics & Management Strategy, 10, 463-507.

Bernheim, B.D., Severinov, S., 2003. Bequests as Signals: An Explanation for the Equal Division Puzzle, Journal of Political Economy, 111, 733-764.

Chiappori, P.-A., Jullien, B., Salanié, B., Salanié, F., 2006. Asymmetric Information in Insurance: General Testable Implications, RAND Journal of Economics, 37, 783-798.

Engers, M., Fernandez, L., 1987. Market Equilibrium with Hidden Knowledge and Self-Selection. Econometrica 55, 425-439.

Heckman, J., 2005. Lessons from the Technology of Skill Formation. NBER Working Paper #11142.

Heckman, J., J. Stixrud, Urzua S., 2005. The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. Discussion Paper, University of Chicago.

Kohllepel, L. Multidimensional 'Market Signaling', 1983. Discussion Paper, Universität Bonn.

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

McAfee, R.P., McMillan, J., 1988. Multidimensional incentive compatibility and mechanism design. Journal of Economic Theory 46, 335-354

Milgrom, P. R., Segal, I. R., 2002. Envelope Theorems for Arbitrary Choice Sets, Econometrica 70, 583-601.

Milgrom, P. R., Shannon, C., 1992. Monotone Comparative Statics, Econometrica 62, 157-180.

Mirrlees, J. A., 1971. An Exploration in the Theory of Optimum Income Taxation. Review of Economic Studies 38, 175-208.

Quinzii, M., Rochet, J.-C., 1985. Multidimensional Signaling. Journal of Mathematical Economics 14, 261-284.

Rochet, J.-C., Stole, L., 2003. The Economics of Multidimensional Screening. In: M. Dewatripont, L. P. Hansen and S. J. Turnovsky (eds.) Advances in Economics and Econometrics: Theory and Applications - Eight World Congress, 2003.

Rotemberg. J. J., 1988. Pareto Improving Distortionary Taxes in the Presence of Signaling. Alfred P. Sloan School of Management Working Paper No. 2039-88.

Rothschild, M. and Stiglitz, J. E., 1976. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. Quarterly Journal of Economics 90, 629-650.

Smart, M., 2000. Competitive Insurance Markets with Two Unobservables. International Economic Review 41, 153-169.

Spence, M., 1974. Market Signalling. Harvard University Press, Cambridge, MA.