Simultaneous Adverse Selection and Moral Hazard*

Daniel Gottlieb and Humberto Moreira[†]

First Version: August, 2011. This Version: April, 2013.

Abstract

We study a principal-agent model with moral hazard and adverse selection. Agents have private information about the distribution of outcomes conditional on each effort. We prove existence, characterize the solution, and establish several general properties of the resulting multidimensional screening problem. A positive mass of types with low conditional probabilities of success gets a constant payment and zero rents. Exclusion is desirable if and only if it is first-best efficient. When agents are risk neutral, an intermediate mass of types is also pooled, although they are offered contracts with variable payments and get positive rents. In addition, the region of types who exert high effort is contained in the interior of the first-best high-effort region and, unlike in pure adverse selection models, there is distortion everywhere. Under additional conditions, the optimal mechanism offers only finitely many contracts. We apply our framework to multidimensional generalizations of canonical models in insurance, regulation, and optimal taxation and show that it generates novel results.

1 Introduction

Most contracting situations combine elements of both adverse selection and moral hazard. Managers, for example, take actions that affect the firm's profitability. At the same time, they usually have better knowledge about the efficacy of each action. As another example, insurance consumers are often better informed about their riskiness than insurers. Concurrently, they may influence their riskiness by engaging in preventive effort.[1] Still, the agency literature has focused on models in which only one of these features is present. Hence, the consequences of the interaction between adverse selection and moral hazard are still poorly understood.

*Preliminary version; comments are especially welcome. We thank Vinicius Carrasco, Alex Edmans, Faruk Gul, Lucas Maestri, Stephen Morris, Luca Rigotti, Yuliy Sannikov, Jean Tirole, Glen Weyl, and seminar audiences at HEC Montreal, Johns Hopkins, Pittsburg/Carnegie Mellon, Princeton, FGV, PUC-Rio, and Wharton for comments and suggestions.

[†]Gottlieb: The Wharton School, The University of Pennsylvania, dgott@wharton.upenn.edu. Moreira: EPGE, Getulio Vargas Foundation, humberto@fgv.br.

[1] Adverse selection and moral hazard are jointly present in many other environments. For example, borrowers may have more precise information about their ability to repay a loan but may also be able to influence this probability; doctors are better informed about the adequacy of each medical treatment, but they also generally have some ability to substitute between treatments; taxpayers are often better informed about their earning abilities and can choose between activities with different distribution of earnings; and regulated firms have more precise information about their technologies but can also engage in cost-reducing actions.

In this paper, we introduce adverse selection in a standard moral hazard model. Agents choose between two costly actions ("efforts"). They have private information about the distribution of outcomes conditional on each action. There are two possible outcomes. Thus, types are two-dimensional vectors.[2] The principal has a continuous prior over the set of conditional probability distributions. We characterize the optimal mechanism and establish several properties that arise under joint adverse selection and moral hazard.

If the principal were able to observe the agents' efforts but not their outcome distributions ("pure adverse selection"), she would be able to implement the efficient allocation by compensating agents for their full effort cost. This would keep agents indifferent between each effort and, therefore, ensure that they would have no incentives to deviate. The unobservability of effort requires the principal to leave rents in order to induce high effort and to prevent each type from pretending to be another type with a slightly less favorable distribution. This generates a standard adverse selection trade-off between rent extraction and effort distortion through the local incentive-compatibility constraints. However, moral hazard also allows agents to pretend to be "distant" types by deviating in the effort dimension. For example, any type can always pretend to be someone who has the same distribution conditional on a different effort. Consequently, moral hazard generates *binding global incentive constraints*, which introduces new features in the model.

Because some agent types can pretend to be less productive and shirk, they receive variable payments but still exert low effort. When reservation utilities are type independent, a positive mass of types with low conditional probabilities of success always gets a constant payment and zero rents; all other types get variable payments and positive rents. Moreover, exclusion of some types is desirable if and only if exclusion is first-best efficient.

We establish several additional properties in the special case of risk-neutral agents. Some intermediate types immediately around the ones with zero rents are also all pooled, although their contract offers variable payments. Moreover, the region of types who exert high effort under asymmetric information is generically contained in the interior of the first-best high-effort region. Therefore, unlike adverse selection models with both one- and multi-dimensional types, the solution involves *distortion at all points* (including the top).

In some cases, the informational rents required to prevent an agent from deviating are so high that the optimal mechanism offers a very limited number of contracts to the agent. For example, when the distribution of types satisfies an increasing rents condition and the incremental output does not exceed twice the incremental cost of effort, the optimal mechanism involves offering *at most three contracts,* despite the presence of a two-dimensional continuum of types. When the probability of a high outcome is bounded away from zero and the incremental output is "not much larger" than the incremental cost of effort, the optimal mechanism involves offering *at most two contracts.* Hence, our model provides a rationale for the fact that large menus of contracts are rarely offered in practice: In the presence of simultaneous adverse selection and moral hazard, offering large menus of contracts gives too many incentives for gaming. Whenever the value from inducing high effort is "not too large," it may be optimal to offer a small number of contracts instead.

Our framework builds on the principal-agent model of Grossman and Hart (1983), which has a natural interpretation in terms of the employment relationships. However, we illustrate

[2] Grossman and Hart (1983) characterize the solution of the pure moral hazard model when there are two outcomes. However, apart from existence, they show that very little can be said about the optimal incentive scheme when there are more than two outcomes. Accordingly, we focus on the two-outcome model but allow the agent to have general private information about the distribution of outcomes.

its applicability beyond the canonical principal-agent model by considering models of insurance, procurement and regulation, and optimal taxation featuring both adverse selection and moral hazard.

Although the consequences of either adverse selection or moral hazard on insurance are well understood, only a few papers have studied the theoretical implications of their joint presence. Still, the empirical literature has shown that simultaneous moral hazard and adverse selection is a key feature of many insurance markets.[3] We show that, because the reservation utility is type-dependent in the insurance model, *exclusion is always optimal.* The optimality of exclusion is a consequence of the multidimensionality of types; it contrasts with one-dimensional type models where exclusion is not optimal if there are "enough low types" in the population (Stiglitz, 1977; Chade and Schlee, 2012). Moreover, because of moral hazard, the second-best high-effort region is strictly contained in the high-effort region in the absence of insurance. Therefore, policyholders *under-provide effort.*

We then consider an extension of the canonical regulation model of Laffont and Tirole (1986, 1993), where we allow effort to affect the regulated firm's costs stochastically. As a result, the regulator's incentive problem cannot be reduced to a pure adverse selection model. We characterize the optimal regulatory mechanism and show that it has the following features: firms with low conditional probabilities of reducing costs are pooled into a cost-plus contract; firms with intermediate conditional probabilities of reducing costs are also pooled, although they receive a contract with positive power; the high-effort region is generically contained in the interior of the first-best high-effort region; and exclusion is optimal if and only if it is first-best optimal. Under certain conditions, the optimal mechanism can be implemented with only two or three contracts.[4]

Our last application consists of an optimal taxation model, with the new feature that the mapping between effort and income is stochastic. Thus, the model can no longer be reduced to a pure adverse selection model. Individuals differ in their *conditional probabilities* of generating high or low incomes given each effort. We study the optimal nonlinear income tax for a Rawlsian planner. Tax rates are decreasing and there is always bunching at the bottom of the distribution, where all types face 100% tax rates. If utilities are quasi-linear, there is also bunching in an intermediate region, and the high-effort region is generically contained in the interior of the first-best high-effort region. We also establish conditions under which the optimal tax system can be implemented with two or three number of tax brackets.

Related Literature

Adding private information to conditional probability distributions naturally leads to a multidimensional screening environment. It is often challenging to characterize the solutions of such problems since one cannot determine from the outset the direction in which incentive constraints bind. While most of the multidimensional screening literature has focused on generalizations of the non-linear pricing model, we study a different class of models. Our framework includes, for example, generalizations of the principal-agent model common in corporate finance and labor economics, as well as models of insurance provision by a monopolist, optimal taxation, and regulation.

[3] See, for example, Karlan and Zinman (2009), Bajari, Hong, and Khwaja (2012), and Einav, Finkelstein, Ryan, Schrimpf, and Cullen (2013).

[4] Using the Laffont-Tirole framework, Rogerson (2003) and Chu and Sappington (2007) show that a pair of simple contracts can achieve a large fraction of the surplus under a certain range of parametric settings – 75 or 73 percent when costs follow either uniform or power distributions, respectively – when effort costs are quadratic.

There are some key differences between our framework and the non-linear pricing framework. In our framework, only one dimension of the type vector matters *conditional on effort.* Therefore, payoffs conditional on effort are not strictly monotone in all dimensions. However, since effort is not observable, the optimal mechanism has to provide incentives for the agent to pick the appropriate effort. As a result, local incentive compatibility is no longer sufficient to ensure global incentive compatibility: types can also deviate in the effort dimension, thereby pooling with "distant" types. In fact, all types who exert high effort in any feasible mechanism have binding global incentive-compatibility constraints. The principal's program, therefore, has to take into account a continuum of binding global constraints. Intuitively, this program corresponds to a non-standard optimal control problem with a continuum of intermediate constraints (on top of the local first- and second-order conditions). Although no general method for this class of problems exists, we are able to obtain optimality conditions using a calculus of variations approach.

Despite these differences, versions of classic results from the multidimensional screening literature also hold in our framework. For example, Armstrong (1996) established that it is generically optimal to exclude a positive mass of buyers with low valuations. Rochet and Choné (1998) showed that Armstrong's result can be generalized but, instead of exclusion, the principal would typically extract all the surplus from a positive mass of types. While it is not optimal to exclude types in our framework (as long as exclusion is not first-best optimal and participation constraints are type independent), it is also the case that the principal extracts the full surplus from a region of types with low conditional probabilities of success. In contrast, exclusion is always optimal in the insurance application of our model because reservation utilities are type dependent. Rochet and Choné also established that bunching was a generic property of multidimensional screening models. In our framework, the solution always entails "bunching at the bottom." In fact, bunching can be so extreme that, in some cases, the optimal mechanism features only a finite number of contracts.

We obtain several new results that are not present in the non-linear pricing model. For example, because all types who exert high effort have binding global constraints, the optimal allocation typically features a distortion at all points. This result contrasts with the "no distortion at the top" property from one-dimensional models, as well as Rochet and Choné's (1998, pp. 811) generalization of it ("no distortion at the boundary").[5]

In addition to the multidimensional screening literature, our paper also relates to and extends several other lines of work. The first one is the literature on insurance markets with both adverse selection and moral hazard. Stewart (1994) argued that adverse selection and moral hazard may partially offset the welfare loss associated with each other. Since low risk types are offered incomplete coverage because of adverse selection, they may exert more effort than if they were fully insured. Chassagnon and Chiappori (1997) introduced preventive effort in the seminal model of Rothschild and Stiglitz (1976) and characterized the set of separating equilibria. De Meza and Webb (2001) and Jullien, Salanie, and Salanie (2007) considered models where consumers have private information about their risk aversion and may engage in preventive effort and showed that the correlation between risk and coverage may be negative.[6] Similarly, Chiu and Karni (1998)

[5] Laffont, Maskin, and Rochet (1987) considered a natural departure from the nonlinear pricing models of Mussa and Rosen (1978) or Maskin and Riley (1984), where agents have quadratic utility functions (linear demands) and types are two-dimensional. Rochet and Stole (2002) introduced independently distributed reservation utilities in the standard nonlinear pricing model. In the monopolistic case, they show that there is no distortion at the top, and either no distortion or bunching at the bottom. For a survey of the multidimensional screening literature, see Rochet and Stole (2003).

[6] In De Meza and Webb (2001), one of the two types is risk-neutral, whereas the other type is risk averse, and

presented an explanation for the lack of private unemployment insurance based on the interaction between preferences for leisure and unobservable job effort, whereas Bond and Crocker (1991) studied a model where policyholders consume products that affect their loss probabilities and insurers do not observe their tastes for such products.

While these papers studied models with two types of consumers, we consider continuous type distributions. Therefore, our paper extends the literature by characterizing optimal insurance contracts when consumer's private information about riskiness is unrestricted. The continuous-type model allows us to determine which are the relevant binding constraints and provides a clearer representation of the richness of the incentive problem.[7]

The second line of related work concerns optimal taxation models with multidimensional taxpayer types. The seminal model of Mirrlees (1971), and most of the literature that followed, assumed that taxpayers differ only through a one-dimensional productivity parameter. In reality, however, taxpayer heterogeneity is multidimensional. Nonetheless, the theoretical difficulty of characterizing the solution of such screening programs has been a substantial barrier for the analysis of optimal taxes when taxpayer types are multidimensional. Accordingly, most of the literature has either assumed a discrete number of types, or used simulations in order to obtain properties of the optimal tax system.[8] A few recent notable exceptions are Kleven, Kreiner, and Saez (2009), Choné and Laroque (2010), Rothschild and Scheuer (2012), and Rothschild and Scheuer (2013), who studied continuous-type two-dimensional screening problems resulting from the design of taxes for couples, heterogeneity in the opportunity cost of work, self-selection into different sectors, and rent seeking, respectively.

Our paper also contributes to the literature on optimal regulation and procurement. The classic model of Laffont and Tirole (1986, 1993) features both adverse selection (the regulated firm has private information about its technology) and moral hazard (the regulator cannot observe the firm's cost-reducing effort). However, because the link between effort, types, and output is deterministic, the model can be reduced to a pure adverse selection model.[9] Caillaud, Guesnerie, and Rey (1992) and Picard (1987) introduced noise in the relationship between output and effort and showed that, under certain conditions, the principal can achieve the same utility as in the absence of noise.[10] In our model, pure adverse selection does not entail any welfare

insurance firms have positive administrative costs. Jullien, Salanie, and Salanie (2007) studied consumers with CARA utilities and showed that the power of incentives always decreases with risk aversion.

[7] As in our model, most of the insurance literature – including all the papers above – focus on two states (loss and no loss). Furthermore, with the exception of Jullien, Salanie, and Salanie (2007), these papers also assume two effort levels. However, they study competitive equilibria whereas we study the monopolist case and the constrained-efficient allocations.

[8] Tarkiainen and Tuomala (1999) and Judd and Su (2006) discuss the theoretical difficulties of characterizing optimal taxes with multidimensional types and present simulations showing that optimal taxes when types are multidimensional can be substantially different from the ones when types are one-dimensional. Several papers have analyzed models with two types in each of two dimensions, which can be suitably mapped into one-dimensional models consisting of four types. For example, Boadway, Marchand, Pestieau, and del Mar Racionero (2002) study optimal income taxes and Cremer, Pestieau, and Rochet (2001) show that the uniform commodity tax result fails to hold when types are multidimensional. Diamond (2005) and Diamond and Spinnewijn (2011) study the optimal taxation of individuals with heterogeneous skills and discount factors using a model with two types in each dimension, while Tenhunen and Tuomala (2010) consider three types in each dimension.

[9] These environments, which also include the Mirrleesian optimal taxation model, are often labeled 'false moral hazard' models (c.f. Laffont and Martimort, 2002).

[10] Caillaud, Guesnerie, and Rey (1992) describe these as 'noisy adverse selection models' rather than models of joint adverse selection and moral hazard since they "restrict attention to risk-neutral agents, which eliminates the insurance question that characterizes moral hazard problems."

losses compared to the first best, whereas pure moral hazard does. Moreover, welfare under joint moral hazard and adverse selection is lower than in the cases of both pure moral hazard and pure adverse selection. The reason for the contrasting welfare results is that agents in our model have private information about the conditional distribution of outcomes given efforts, whereas agents in Caillaud, Guesnerie, and Rey and Picard have private information about the cost of effort. Another difference between our models is that we characterize the solution under both risk neutrality and risk aversion, whereas they only consider risk-neutral agents.

The robustness of bunching indicates a relationship between the "complexity" of the environment and the number of contracts offered to the agents. When the distribution of outcomes given efforts is observable (pure moral hazard), the principal is able to perfectly design the contract for each type. Consequently, each type who exerts high effort is offered a different contract. Moreover, all types who exert low effort obtain a constant payment. When the conditional distributions of outcomes are unobservable, offering a large number of contracts introduces too many possible deviations by the agents, which requires the principal to leave large informational rents. Offering a smaller number of contracts can be an efficient way to prevent gaming by the agents. In some cases, these informational rents are so large that the optimal mechanism features only two or three contracts.

The optimality of simple contracts in "complex" environments is related to the robustness intuition of Holmstrom and Milgrom (1987). However, the notion of robustness in our model is different from the one in their seminal paper. Here, offering a limited number of contracts is robust in that it reduces the agents' incentives to misrepresent their private information about the environment. In Holmstrom and Milgrom's model, linear contracts are robust in the sense that they prevent the agent from readjusting effort over time.[11]

The structure of the paper is as follows. Section 2 presents the basic framework, and Section 3 derives some general properties of the solution. Section 4 then characterizes the solution and establishes several additional properties under the assumption of risk neutrality, whereas Section 5 generalizes the characterization for situations where agents may be risk-averse. Section 6 applies our framework to multidimensional models of insurance (6.1), regulation (6.2), and optimal taxation (6.3). Then, Section 7 concludes.[12]

2 Model

2.1 Statement of the Problem

There is a risk-neutral principal and an agent who may be either risk neutral or risk averse. The agent exerts an effort $e \in \{0,1\}$, which is unobservable by the principal. The principal does, however, observe the outcome from the partnership $x \in \{x_L, x_H\}$, which is stochastically affected by the agent's effort. We refer to $\Delta x := x_H - x_L > 0$ as the incremental output. Let p_e denote

[11] Edmans and Gabaix (2011) extend the linearity results to a model in which the realization of noise occurs before the action in each period and the principal desires to implement a fixed action in all states. Relatedly, Chassang (2013) introduces a class of calibrated contracts that are detail-free and approximate the performance of the best linear contract in dynamic environments when players are patient, while Carroll (2013) shows that the best contract for a principal who faces an agent with uncertain technology and evaluates contracts in terms of their worst-case performance is linear.

[12] Appendix A presents the benchmark cases of pure moral hazard and pure adverse selection. Proofs are in Appendix B.

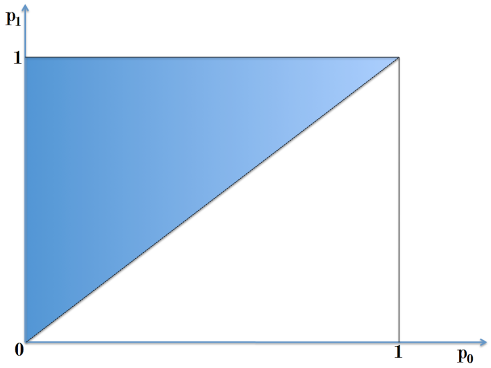

Figure 1: *Type Space (shaded area).*

the probability of outcome x_H given effort e. We refer to x_H and x_L as high and low outcomes and to $e = 1$ and $e = 0$ as high and low efforts.

The agent has private information about the conditional distribution of outcomes. Therefore, the agent's type $\mathbf{p} := (p_0, p_1)$ is a vector of conditional probabilities of a high outcome given efforts. The principal has a continuous prior distribution over types, denoted by f. Types satisfy the Monotone Likelihood Ratio Property (MLRP), which states that exerting higher effort increases the probability of the high outcome: $p_1 \geq p_0$. Under MLRP, the type space is contained in the area above the 45° line in Figure 1. Let $\bar{\Delta} := \{(p_0, p_1) \in \mathbb{R}^2 : p_1 \geq p_0\}$ denote the space of types satisfying MLRP. We assume that the distribution of types f has full support on $\bar{\Delta}$.[13]

The agent's preferences over money W and effort e are represented by an additively separable von-Neumann Morgenstern utility function, $u(W) - c_e$, where $c_0 < c_1$, u is continuously differentiable, increasing, and weakly concave, and the marginal utility function $\dot{u}$ is bounded.

There is no loss of generality in focusing on direct mechanisms in which the agent follows 'honest and obedient' strategies (Myerson, 1982). Accordingly, we can restrict mechanisms to be a fixed payment function $W : \bar{\Delta} \to \mathbb{R}$, a bonus function $B : \bar{\Delta} \to \mathbb{R}$, and an effort recommendation function $e : \bar{\Delta} \to \{0, 1\}$. We refer to the pair of payments $W(\mathbf{p})$ and $B(\mathbf{p})$ as a *contract.* An agent who reports type $\mathbf{p}$ agrees to exert effort $e(\mathbf{p})$ and receives $W(\mathbf{p})$ in case of low output and $W(\mathbf{p}) + B(\mathbf{p})$ in case of high output.

As in Grossman and Hart (1983), it is convenient to express these mechanisms in terms of the agent's utility. Let $w \equiv u(W)$ denote the utility from the fixed payment W, and let $b \equiv u(W + B) - u(W)$ denote the 'power' of the contract – the utility gain from a high output relative to a low output. With a slight abuse of notation, we will also refer to a mechanism as a function $(w, b, e) : \bar{\Delta} \to \mathbb{R}^2 \times \{0, 1\}$, and we will refer to the pair $w(\mathbf{p})$ and $b(\mathbf{p})$ as a contract.

Given a mechanism (w, b, e), a type-$\mathbf{p}$ agent obtains expected utility

$$U(\mathbf{p}) \equiv w(\mathbf{p}) + p_{e(\mathbf{p})} b(\mathbf{p}) - c_{e(\mathbf{p})}. \tag{1}$$

We refer to U as the *informational rent* function. The agent follows honest and obedient strategies if the following *incentive-compatibility* constraint is satisfied:

$$U(\mathbf{p}) \geq w(\hat{\mathbf{p}}) + p_e b(\hat{\mathbf{p}}) - c_e, \quad \forall \mathbf{p}, \hat{\mathbf{p}} \in \bar{\Delta}, \ \forall e \in \{0, 1\}. \tag{IC}$$

[13] It is immediate to generalize our results for distributions that do not satisfy MLRP as long as their support contains $\bar{\Delta}$, by projecting types outside $\bar{\Delta}$ onto the 45° line.

The mechanism satisfies *individual rationality* if the following participation constraint holds:[14]

$$U(\mathbf{p}) \geq 0, \ \ \forall \mathbf{p} \in \bar{\Delta}. \tag{IR}$$

We assume that the agent can costlessly reduce output (*free disposal*).[15] Therefore, payments have to be nondecreasing in the output:

$$b(\mathbf{p}) \geq 0, \ \ \forall \mathbf{p} \in \bar{\Delta}. \tag{FD}$$

A mechanism is *feasible* if it satisfies incentive compatibility, individual rationality, and free disposal.

The principal's expected utility is:

$$\int_{\bar{\Delta}} \left\{ \ p_{e(\mathbf{p})} \left[x_H - u^{-1} \left(w(\mathbf{p}) + b(\mathbf{p}) \right) \right] + \left(1 - p_{e(\mathbf{p})} \right) \left[x_L - u^{-1} \left(w(\mathbf{p}) \right) \right] \ \right\} f(\mathbf{p}) d\mathbf{p}. \tag{2}$$

Two mechanisms are *equivalent* if they induce the same expected utility to the principal and all agent types. A mechanism is *optimal* if it maximizes the principal's expected utility within the class of feasible mechanisms.

2.2 Feasible Mechanisms

In this subsection, we obtain necessary and sufficient conditions for a mechanism to be feasible. The first result establishes that there is no loss of generality in considering mechanisms for which there exists a continuous and non-decreasing function separating the sets of types who exert high and low efforts:[16]

Lemma 1. *For any feasible mechanism, there exists an equivalent mechanism* (w, b, e) *such that* $e(p_0, p_1) = 1$ *if and only if* $p_1 > \varphi(p_0)$ *for a continuous and non-decreasing function* $\varphi : [0, 1] \rightarrow [0, 1]$.

The intuition behind Lemma 1 is the following. Suppose a feasible mechanism recommends that type $\mathbf{p} = (p_0, p_1)$ exerts high effort, and consider a type $\hat{\mathbf{p}} = (p_0, p_1 + \varepsilon)$ for some $\varepsilon > 0$. Type $\hat{\mathbf{p}}$ has the same distribution of outcomes conditional on low effort as $\mathbf{p}$, but has a higher probability of high outcome conditional on high effort. Therefore, $\hat{\mathbf{p}}$ has an even higher incentive to exert high effort.

Next, suppose that the mechanism recommends that type $\mathbf{p} = (p_0, p_1)$ exerts low effort, and consider some type $\hat{\mathbf{p}} = (p_0 + \varepsilon, p_1)$ for some $\varepsilon > 0$. Incentive compatibility implies that $\hat{\mathbf{p}}$ will have a (weakly) higher incentive to exert low effort than type $\mathbf{p}$ has. If type $\hat{\mathbf{p}}$ is indifferent, the principal can improve by asking it to exert low effort.

The continuity of φ follows from the indirect utility function U being continuous, strictly increasing in p_1 in the region of high effort, and constant in p_1 in the region of low effort. Figure

[14] This formulation of the participation constraint with type-independent reservation utilities is standard in principal-agent models. In Section 6, we allow for type-dependent reservation utilities in order to study optimal insurance contracts.

[15] Free disposal is assumed in many principal-agent models, including Innes (1990), Acemoglu (1998), and Poblete and Spulber (2012).

[16] We will adopt the convention that indifferent types choose low effort. This will not affect our results since these types must have measure zero.

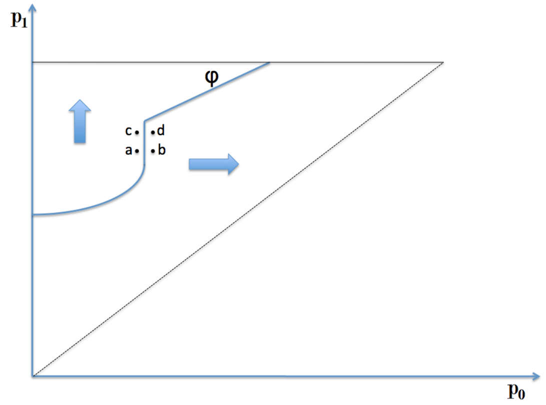

Figure 2: *Intuition behind Lemma 1 (continuity of φ).*

2 illustrates the argument. The arrows indicate the direction of growth of the informational rent function U. Consider points $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, and $\boldsymbol{d}$. Since U is continuous, if the distances between $\boldsymbol{a}$ and $\boldsymbol{b}$ and, $\boldsymbol{c}$ and $\boldsymbol{d}$ are small enough, we must have $U(\mathbf{a}) \approx U(\mathbf{b})$ and $U(\mathbf{c}) \approx U(\mathbf{d})$. Moreover, because the informational rent increases in p_1 in the region above φ, we must have $U(\mathbf{c}) > U(\mathbf{a})$, and because the informational rent is constant in p_1 in the region below φ, we must have $U(\mathbf{b}) = U(\mathbf{d})$. Therefore, we must have

$$U(\mathbf{c}) > U(\mathbf{a}) \approx U(\mathbf{b}) = U(\mathbf{d}) \approx U(\mathbf{c}),$$

which is a contradiction.

For a given feasible mechanism (w, b, e), we refer to the function φ as the *effort frontier* associated with it.[17] The effort frontier partitions the type space into types who exert low and high efforts:

$$e(p_0, p_1) = 1 \iff p_1 > \varphi(p_0). \tag{3}$$

Using the local first- and second-order conditions for incentive compatibility, we obtain the following necessary conditions:

Lemma 2. *Let (w, b, e) be a feasible mechanism and let φ and U be the effort frontier and informational rent functions associated with it. Then:*

a. *$U(p_0, p_1)$ is convex, differentiable a. e., and has gradient*

$$\nabla U(p_0, p_1) = \begin{cases} (b(p_0, p_1), 0), & \text{if } p_1 < \varphi(p_0) \\ (0, b(p_0, p_1)), & \text{if } p_1 > \varphi(p_0) \end{cases};$$

b. *$b(p_0, p_1)$ is constant in p_1 for $p_1 < \varphi(p_0)$ and constant in p_0 for $p_1 > \varphi(p_0)$;*

c. *$U(0, 0) \geq 0$ and $b(0, 0) \geq 0$.*

[17] Due to the equivalence result of Lemma 1, we focus on mechanisms for which an effort frontier function φ exists. Any other feasible mechanism will give the same payoff to the principal and all types of agents and will differ only in a set of zero measure (see the proof of the lemma). Moreover, such a mechanism exists whenever a feasible mechanism exists.

The incentive-compatibility constraints from adverse selection state that reporting each type truthfully while following the principal's effort recommendation must maximize the agent's payoff. Properties (a) and (b) are the local first- and second-order conditions of this maximization program. Property (c) is a direct consequence of the participation and free disposal constraints.

While the conditions from Lemma 2 are implied by adverse selection alone, moral hazard introduces additional incentive-compatibility constraints. In particular, under moral hazard, satisfying the local constraints is not enough to prevent global deviations from being profitable, since a type may choose a different effort level in order to pretend to be another "distant" type. The following lemma presents necessary conditions to avoid global deviations:

Lemma 3. *Let (w, b, e) be a feasible mechanism and let φ and U be the effort frontier and informational rent functions associated with it. Then:*

d. $U(p_1, p_1) = U(p_0, p_1) + \Delta c$ *for* $p_1 > \varphi(p_0)$*;*

e. $b(p_1, p_1) = b(p_0, p_1)$ *for almost all* (p_0, p_1) *such that* $p_1 > \varphi(p_0)$*.*

Because effort is costly and types on the 45° line have the same conditional distribution over outcomes under high and low efforts, these types will never exert high effort. Thus, type (p_1, p_1) exerts low effort and has the same probability of success as any type (p_0, p_1) who exerts high effort (i.e., $p_1 > \varphi(p_0)$). Therefore, they must get the same utility net of their different effort costs in any incentive-compatible mechanism – as Property (d) states. Property (e) is a consequence of the envelope theorem applied to deviations along the 45° line.

In models of pure adverse selection, conditions (a)-(c) are also sufficient for feasibility. We have seen that moral hazard introduces additional necessary conditions (d) and (e). We now establish that these necessary conditions are also sufficient.[18]

Lemma 4 (**Feasibility**)**.** *Fix a mechanism (w, b, e), and let U denote the associated informational rent function defined according to equation (1). The mechanism is feasible if and only if it satisfies conditions (a)-(e) for an effort frontier function φ satisfying condition (3).*

From Lemma 4, a mechanism is optimal if and only if it solves the following program:

$$\max_{U,b,e,\varphi} \int_{\bar{\Delta}} \left\{ \begin{array}{c} x_L - u^{-1}\left(U(\mathbf{p}) - p_{e(\mathbf{p})} b(\mathbf{p}) + c_{e(\mathbf{p})}\right) + p_{e(\mathbf{p})} \Delta x \\ -p_{e(\mathbf{p})} \left[u^{-1}\left(U(\mathbf{p}) + \left(1 - p_{e(\mathbf{p})}\right) b(\mathbf{p}) + c_{e(\mathbf{p})}\right) - u^{-1}\left(U(\mathbf{p}) - p_{e(\mathbf{p})} b(\mathbf{p}) + c_{e(\mathbf{p})}\right)\right] \end{array} \right\} f(\mathbf{p}) d\mathbf{p} \tag{P}$$

subject to conditions (a)-(e) and (3).

The direct usefulness of the characterization from Lemma 4 is limited by the fact that Program (P) is not very tractable as stated. In the next section, we will rewrite (P) as a one dimensional program, which will be key to our study of the properties of optimal mechanisms.

2.3 One-Dimensional Conditions

Let (w, b, e) be a feasible mechanism and let φ and U denote the effort frontier and informational rent functions associated with it. Let the *rent projection* associated with this mechanism be the function $\mathcal{U} : [0, 1] \to \mathbb{R}$ defined as $\mathcal{U}(t) := U(t, t)$. We say that a mechanism is *trivial* if it

[18] Modulo the conventions from footnotes 16 and 17.

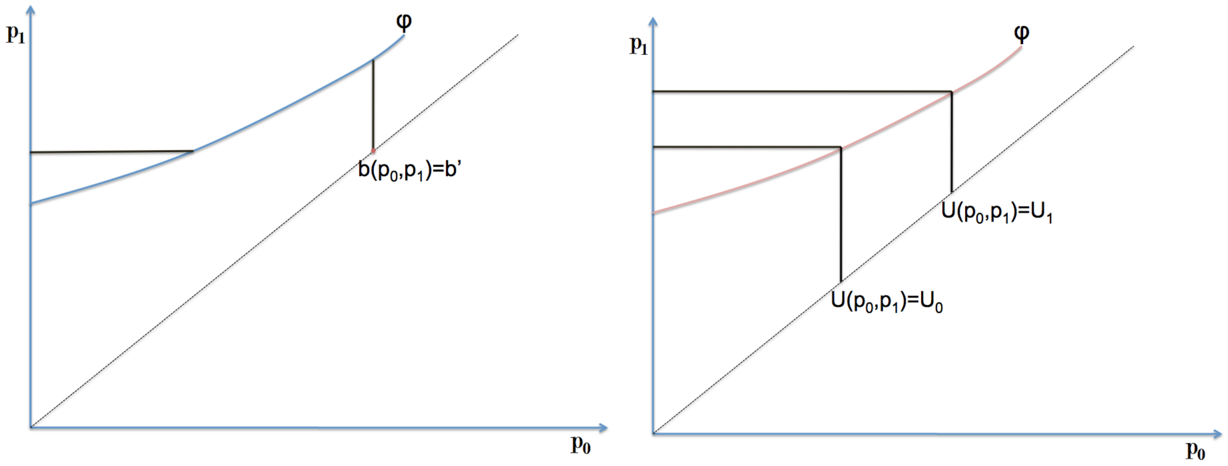

Figure 3: *Iso-power function and Iso-rent functions. The iso-power (left) depicts all types who are offered the same contract power b'. The iso-rent (right) depicts all types with the same informational rents (U_0 and U_1). Note that $U_1 > U_0$ because informational rents are increasing.*

recommends low effort to almost all types.[19] The following lemma establishes that any nontrivial feasible mechanism is characterized by the one-dimensional functions $\mathcal{U}$ and φ:[20]

Lemma 5. *Let (w, b, e) be a nontrivial feasible mechanism and let φ and $\mathcal{U}$ denote the effort frontier and rent projection functions associated with it. Then:*

$$b\left(p_0, p_1\right) = \begin{cases} \dot{\mathcal{U}}\left(p_0\right) & \text{if } p_1 \leq \varphi\left(p_0\right) \\ \dot{\mathcal{U}}\left(p_1\right) & \text{if } p_1 > \varphi\left(p_0\right) \end{cases} \quad (a.e.)\,, \tag{4}$$

$$w\left(p_0, p_1\right) = \begin{cases} \mathcal{U}\left(p_0\right) - p_0\dot{\mathcal{U}}\left(p_0\right) + c_0 & \text{if } p_1 \leq \varphi\left(p_0\right) \\ \mathcal{U}\left(p_1\right) - p_1\dot{\mathcal{U}}\left(p_1\right) + c_0 & \text{if } p_1 > \varphi\left(p_0\right) \end{cases} \quad (a.e.)\,, \quad \textit{and} \tag{5}$$

$$\mathcal{U}\left(\varphi\left(p_0\right)\right) = \min\left\{\mathcal{U}\left(p_0\right) + \Delta c; \mathcal{U}\left(1\right)\right\}. \tag{6}$$

Lemma 5 establishes that any nontrivial mechanism is (a.e.) characterized by the rent projection function $\mathcal{U}$, representing the information rent along the 45° line. Given such a rent projection function $\mathcal{U}$, the effort recommendation (characterized by the effort frontier function) is obtained by equation (6), which depicts types $(p_0, \varphi(p_0))$ who are indifferent between exerting high and low efforts.

Using the necessary and sufficient properties established previously, we can recover w, b, and φ from the rent projection function. First note that Properties (a) and (b) imply that the derivative of the rent projection function, $\dot{\mathcal{U}}$, equals the power of the contract along the 45° line, $b\left(p_0, p_0\right)$. Because the power of the contract is constant in the region of low effort for types with the same probability of success given low effort, the derivative of the rent projection function determines b in the low effort region. Moreover, since the power of the contract along the 45° line equals the one in the high effort region for each fixed probability of success given low effort (Property e), it

[19] The best trivial mechanism for the principal offers the same payments $w = c_0$ and $b = 0$ and recommends low effort $e = 0$ to (almost) all types. Similarly to Grossman and Hart (1983), it is convenient to solve for the best nontrivial mechanism and verify whether it generates an expected profit greater than the best trivial mechanism.

[20] Without loss of generality we can assume that $\dot{\mathcal{U}}(t)$ is a càdlàg function (i.e., right continuous with left limits at every point).

also determines the power of the contract for types who are recommended high effort (see Figure 3). Finally, equation (6), which is the counterpart of Property (e), allows us to recover the effort frontier φ.[21] Finally, note that Property (a) establishes that iso-rent functions have an inverse L shape with the kink at the effort frontier (Figure 3). Therefore, the informational rent U is determined by the rent along the 45° line $\mathcal{U}$ and the effort frontier φ. Then, using the definition of the informational rent (equation 1), we can recover the fixed component of the mechanism w.

We will, therefore, say that $\mathcal{U}$ is an *optimal rent projection* if the mechanism (w, b, e) associated with it is an optimal mechanism. It is more convenient to work with the one-dimensional functions $\mathcal{U}$ and φ rather than the original two-dimensional mechanism (w, b, e).

Definition 1. The function $\mathcal{U} : [0, 1] \to \mathbb{R}$ is a *feasible rent projection* if $\mathcal{U}$ is non-decreasing and convex, and $\mathcal{U}(0) \geq 0$.

The following lemma characterizes rent projection functions associated with feasible mechanisms:

Lemma 6 (One-Dimensional Characterization of Feasibility). *Let (w, b, e) be a feasible mechanism, and let $\mathcal{U}$ and φ be the rent projection and effort frontier functions associated with it. Then, $\mathcal{U}$ is a feasible rent projection and $(\mathcal{U}, \varphi)$ solve equation (6). Conversely, let $\mathcal{U}$ be a feasible rent projection, let φ be defined by the solution of equation (6), and let (w, b, e) be given by equations (3), (4) and (5). Then, (w, b, e) is a feasible mechanism.*

Let G denote the cost of providing expected utility $\mathcal{U}$ and power $\dot{\mathcal{U}}$ to an agent with probability of success t:

$$G(\mathcal{U},\dot{\mathcal{U}}, t) \equiv t u^{-1}(\mathcal{U} + (1 - t)\dot{\mathcal{U}} + c_0) + (1 - t)u^{-1}(\mathcal{U} - t\dot{\mathcal{U}} + c_0). \tag{7}$$

Rewriting the principal's expected utility in terms of $\mathcal{U}$ and φ yields

$$x_L + \int_0^1 \int_t^{\varphi(t)} (t\Delta x - G(\mathcal{U}(t),\dot{\mathcal{U}}(t), t))f(t, s)dsdt + \int_0^1 \int_{\varphi(t)}^1 (s\Delta x - G(\mathcal{U}(s),\dot{\mathcal{U}}(s), s))f(t, s)dsdt.$$

Applying Fubini's theorem, the principal's payoff (8) becomes

$$x_L + \int_0^1 \int_t^{\varphi} \left(t\Delta x - G(\mathcal{U},\dot{\mathcal{U}}, t)\right) f(t, s)\, dsdt + \int_{\varphi(0)}^1 \int_0^{\varphi^{-1}} \left(t\Delta x - G(\mathcal{U},\dot{\mathcal{U}}, t)\right) f(s, t)\, dsdt$$

$$= x_L + \int_0^1 \left(t\Delta x - G(\mathcal{U},\dot{\mathcal{U}}, t)\right) F_0(t, \varphi)\, dt + \int_{\varphi(0)}^1 \left(t\Delta x - G(\mathcal{U},\dot{\mathcal{U}}, t)\right) F_1\left(\varphi^{-1}, t\right) dt, \tag{8}$$

where $F_0(t, s) \equiv \int_t^s f(t, z)dz$ and $F_1(s, t) \equiv \int_0^s f(z, t)\, dz$, and we are omitting the dependence of the functions $\mathcal{U}$, φ and φ^{-1} on t for notational simplicity. Therefore, Program (P) can be rewritten in terms of of the one-dimensional functions $\mathcal{U}$ and φ:

$$\max_{\mathcal{U},\varphi} x_L + \int_0^1 \left(t\Delta x - G(\mathcal{U},\dot{\mathcal{U}}, t)\right) F_0(t, \varphi)\, dt + \int_{\varphi(0)}^1 \left(t\Delta x - G(\mathcal{U},\dot{\mathcal{U}}, t)\right) F_1\left(\varphi^{-1}, t\right) dt \tag{P'}$$

[21] Our projection method resembles the technique that Laffont, Maskin, and Rochet (1987) use to determine the boundary condition of the partial differential equation that characterizes incentive-compatible mechanisms in their model.

subject to (6), $\mathcal{U}$ nondecreasing and convex, and $\mathcal{U}(0) \geq 0$.

Lemma 6 simplifies the search for optimal mechanisms by restating the original maximization over the set of two-dimensional functions (P) as a maximization over the set of one-dimensional rent projection and effort frontier functions (P'). Although one-dimensional, (P') differs from standard programs from one-dimensional screening models in two important ways. First, there is no standard probability distribution or utility function that ensures the concavity of the objective function. Second, equation (6) corresponds to a non-standard constraint connecting a each type t to its projection along the effort frontier $\varphi(t)$. Mathematically, this corresponds to a continuum of intermediate value constraints. Economically, this means that, in addition to the local incentive compatibility constraints, there is also a continuum of binding global incentive-compatibility constraints.[22]

3 General Properties

This section presents general properties of optimal mechanisms. Our first proposition establishes that a positive mass of agents do not receive any informational rents $(U(\mathbf{p}) = 0)$:

Proposition 1 (**Zero Rents at the Bottom**). *No mechanism that gives strictly positive informational rents for almost all types is optimal.*

Because the rent projection function is nondecreasing, there exists $t^* > 0$ such that $\mathcal{U}(t) = 0$ if and only if $t \leq t^*$. Since $\dot{\mathcal{U}}(t) = b(t,t)$ and $\mathcal{U}$ is convex, types in the interior of the zero-rent region get the zero-power contract: $w = c_0$, $b = 0$. Then, equation (6) implies that the effort frontier $\varphi(t)$ is flat in the interval $[0, t^*]$. Figure 4 depicts these results graphically.

Recall the original description of the mechanism in monetary units (W, B, e), where $W(\mathbf{p})$ is the fixed payment and $B(\mathbf{p})$ is the bonus. Our next result establishes that any mechanism involving a bonus greater than the incremental output Δx is not optimal:

Lemma 7 (**Bounded Bonus**). *Let (W, B, e) be a feasible mechanism, and suppose that $B(\boldsymbol{p}) > \Delta x$ in a set with positive measure. Then, (W, B, e) is not optimal.*

The principal gets $x_L - W(\mathbf{p})$ if type $\mathbf{p}$ produces a low output and $x_L + \Delta x - W(\mathbf{p}) - B(\mathbf{p})$ if he produces a high output. Therefore, Lemma 7 states that it cannot be optimal for the principal to offer a contract in which she loses money in case of a high output relative to a low output. Intuitively, if the principal were offering such a large bonus, reducing it would have two effects. First, it would reduce the region of effort. In general, this would be detrimental to the principal. However, because she is losing money from a high output, this effect would actually be positive in this case. Second, it would reduce the informational rents of all types above this one (according to the projection on the 45^o line), which would, again, raise the principal's payoff. Thus, she would be able to unambiguously improve her expected payoff by reducing the bonus.

[22] The idea of working with a dual approach, which treats the informational rent as the instrument, is justified by Rochet (1987). In their classic analysis, Rochet and Choné (1998) follow this approach in a multidimensional-type model. Our approach is different from theirs in three aspects: (i) local constraints are necessary and sufficient in their model, whereas moral hazard introduces binding global constraints here; (ii) the input variable in their optimization program is the entire (multidimensional) informational rent function, whereas the domain of the input variable here is a one-dimensional subspace of the type space; and (iii) their number of instruments is equal to the dimension of the type space. In our model, the global moral hazard constraint reduces the dimensionality of the instrument from two (the dimension of the type space) to one through the one-dimensional projection method.

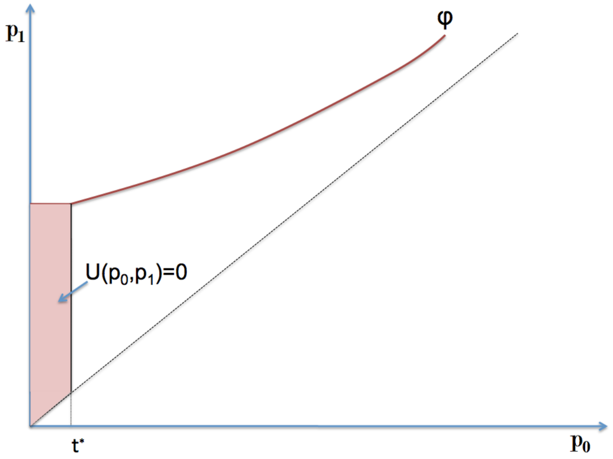

Figure 4: *Types who are offered the zero-power contract and get zero rents.*

Using the non-optimality of mechanisms with unbounded bonus functions, we can establish existence. This is an important issue since we are focusing on pure strategy mechanisms and, therefore, cannot use available existence theorems.[23]

Proposition 2 (Existence). *There exists an optimal mechanism.*

Our next result concerns the slope of the effort frontier φ. As a benchmark, consider a situation where the principal can observe both the agent's type and his effort choice. The principal would then offer an expected payment $u^{-1}(c_e)$ and require the effort level that maximizes expected surplus.[24] The expected surplus from high effort is $x_L + p_1 \Delta x - u^{-1}(c_1)$, whereas the expected surplus from low effort equals $x_L + p_1 \Delta x - u^{-1}(c_0)$. Therefore, the principal would require high effort from types for which

$$p_1 \geq p_0 + \frac{u^{-1}(c_1) - u^{-1}(c_0)}{\Delta x}. \tag{9}$$

This inequality determines the *first-best effort frontier*, which has slope 1.

Recall that, by equation (6), the effort frontier in any feasible mechanism satisfies

$$\mathcal{U}(\varphi(t)) = \mathcal{U}(t) + \Delta c$$

for all points in which $\varphi(t) < 1$. Thus, $\dot{\varphi}(t) = \frac{\dot{\mathcal{U}}(t)}{\dot{\mathcal{U}}(\varphi(t))}$ (a.e.). Since $\varphi(t) > t$ and $\mathcal{U}$ is convex, it then follows that $\dot{\varphi} \leq 1$ (a.e.). Thus, the effort frontier function in any feasible mechanism is flatter than the first-best effort frontier. Moreover, by Proposition 1, in any optimal mechanism there exists $t^* > 0$ such that $\varphi = \varphi^*$ for all $t \in [0, t^*)$. We formally state this result in the following lemma:

Lemma 8 (Slope of Effort Frontier). *Let (w, b, e) be an optimal mechanism and let φ be the effort frontier function associated with it. Then, φ is continuous, differentiable (a.e.), and $\dot{\varphi} \leq 1$ at all points of differentiability. Moreover, there exists $t^* > 0$ such that $\varphi(t) = \varphi^*$ for all $t < t^*$.*

Next, we turn to the optimality of exclusion of types.

[23] See, for example, Kadan, Reny, and Swinkels (2011) and references therein.

[24] If the agent is risk neutral, the principal can offer any random payment with expected value equal to $u^{-1}(c_e)$.

Exclusion

Our individual rationality constraint (IR) required all types to participate in the mechanism. In many situations, however, the principal can exclude some types by not offering any contracts that dominate their reservation utility. In this subsection, we consider the desirability of exclusion.

Let $\pi(\mathbf{p}) \in \{0,1\}$ denote the agent's participation: when $\pi(\mathbf{p}) = 0$, type $\mathbf{p}$ does not participate in the mechanism and gets zero utility; when $\pi(\mathbf{p}) = 1$, he participates and gets the utility specified in equation (1). A *mechanism in the model with exclusion of types* specifies, for each type $\mathbf{p}$, a utility in case of failure $w(\mathbf{p})$, a contract power $b(\mathbf{p})$, a recommended effort $e(\mathbf{p})$, and a participation decision $\pi(\mathbf{p})$. Given a mechanism (w, b, e, π), a type-$\mathbf{p}$ agent obtains expected utility:

$$U(\mathbf{p}) \equiv \pi(\mathbf{p}) \left[w(\mathbf{p}) + p_{e(\mathbf{p})} b(\mathbf{p}) - c_{e(\mathbf{p})} \right], \tag{10}$$

and the principal gets expected utility:

$$\int_{\bar{\Delta}} \left\{ \begin{array}{c} x_L - u^{-1}(w(\mathbf{p})) + \\ p_{e(\mathbf{p})} \left\{ \Delta x - [u^{-1}(w(\mathbf{p}) + b(\mathbf{p})) - u^{-1}(w(\mathbf{p}))] \right\} \end{array} \right\} \pi(p) f(\mathbf{p}) d\mathbf{p}.$$

The individual-rationality and incentive-compatibility constraints are analogous to the ones in the no-exclusion model, with the appropriate substitution of the utility function (1) by (10). All previous results can be adjusted to model with exclusion of types by restricting attention to the set of types who participate. The principal must ensure that a type gets at most zero expected utility from participating in order to exclude him.

As a benchmark, let us first consider the exclusion rule under perfect information. From the first-best effort region – condition (9) –, the principal's expected utility when contracting with type (p_0, p_1) is

$$\max \left\{ x_L + p_0 \Delta x - u^{-1}(c_0) ;\ x_L + p_1 \Delta x - u^{-1}(c_1) \right\}.$$

It is optimal to exclude a type if the principal's expected utility from that type is negative. Because the expression above is increasing in p_0 and p_1, exclusion is optimal if and only if it is optimal to exclude type $(0,0)$. Substituting $p_0 = p_1 = 0$ in the previous expression, we obtain the condition for exclusion to be first-best optimal:

$$x_L < u^{-1}(c_0). \tag{11}$$

The following proposition establishes that the same condition holds when types and efforts are not observable. Therefore, exclusion is second-best optimal if and only if it is first-best optimal.

Proposition 3 (**Exclusion**). *It is optimal to exclude a strictly positive mass of types if and only if exclusion of types is first-best optimal.*

The result from Proposition 3 contrasts with the celebrated exclusion result from Armstrong (1998) for multidimensional screening in the context of a multi-product monopolist. We return to this issue on Subsection 6.1, which considers an application to insurance, where the participation constraint is type-dependent. Note that Proposition 3 only refers to the "extensive margin," by showing that there is no exclusion if and only if the first-best features no exclusion. It does *not* imply that the exclusion regions in these two environments must coincide. In fact, it can be shown that when exclusion is optimal, the region of excluded types may either contain or be contained in the first-best exclusion region.

4 Risk Neutrality

This section characterizes optimal mechanisms when the agent is risk neutral: $u(X) = X$. In this case, the first-best region of effort – inequality (9) – is determined by $(p_1 - p_0)\Delta x \geq \Delta c$, where $\Delta c := c_H - c_L > 0$ denotes the incremental cost of effort. Therefore, a mechanism implements the first-best effort if its effort frontier solves $[\varphi(t) - t]\Delta x = \Delta c$ whenever $\varphi(t) < 1$.

Let $\mathcal{U}$ be a feasible rent projection and let φ be the effort frontier associated with it. The *effort distortion of projected type* t is

$$[\varphi(t) - t]\Delta x - \Delta c. \tag{12}$$

The effort distortion is zero if the mechanism implements the first-best effort frontier at t; it is positive if there is less effort than under first best and negative if there is more effort than under first best at t.

Let $\varphi^* := \mathcal{U}^{-1}(\Delta c)$ denote the lowest projected type for which there is high effort, let $t^* := inf\{t : \mathcal{U}(t) = 0\}$ denote the lowest projected type that gets positive rents, and let $\xi^* := \sup\{t : \varphi(t) = 1\}$ denote the projected type for which the effort frontier hits $p_1 = 1$. In the spirit of Myerson (1981), we define the *expected virtual surplus* as

$$\int_0^1 S_0(t,\mathcal{U})\mathcal{U}(t)\, f(t,\varphi)\, dt + \int_0^1 S_1(t,\mathcal{U})\mathcal{U}(t)\, f(\varphi^{-1},t)dt + S^*(\mathcal{U})\,\mathcal{U}(\varphi^*), \tag{13}$$

where

$$S_0(t,\mathcal{U}) := \begin{cases} -\frac{(\varphi-t)\Delta x-\Delta c}{\dot{\mathcal{U}}(\varphi)} - \frac{F_0(t,\varphi)}{f(t,\varphi)} \text{ if } t < \xi^* \\ -\frac{F_0(t,1)}{f(t,1)} \text{ if } t \geq \xi^* \end{cases}, \tag{14}$$

$$S_1(t,\mathcal{U}) := \begin{cases} 0 \text{ if } t \leq \varphi^* \\ \frac{(t-\varphi^{-1})\Delta x-\Delta c}{\dot{\mathcal{U}}(\varphi^{-1})} - \frac{F_1(\varphi^{-1},t)}{f(\varphi^{-1},t)} \text{ if } t > \varphi^* \end{cases}, \tag{15}$$

$$S^*(\mathcal{U}) := \frac{\{\varphi^* - E[t|t \leq t^*, \varphi^*]\}\Delta x - \Delta c}{\dot{\mathcal{U}}(\varphi^*)} F_1(t^*, \varphi^*),$$

and $E[t|t \leq t^*, \varphi^*] := \frac{\int_0^{t^*} tf(t,\varphi^*)dt}{F_1(t^*,\varphi^*)}$. Our expected virtual surplus (13) differs from Myerson's classic formula – and multidimensional generalizations of it – in one important way. Because global incentive constraints are now binding, the virtual surplus also takes into account informational rents that are left to nonadjacent types with binding incentive constraints.

S_0 and S_1 are the *marginal virtual surpluses* in the regions of low and high effort: they capture the net benefit from leaving informational rents in each of these regions. Recall that types above (t,t) in the low effort region and types to the left of (t,t) in the high effort region get the same contract (see Figure 3). Consider a small increase in the rent projection function $\mathcal{U}(t)$ for t in the region where some types exert high effort ($t > \varphi^*$). This perturbation affects all types who get the same contract as (t,t). Therefore, there is a "vertical effect" on types above (t,t) in the low effort region and a "horizontal effect" on types to the left of (t,t) in the high effort region.

Let us consider the vertical effect first. There is a marginal effect through the effort frontier and an inframarginal effect through the informational rents left to all types that keep exerting low effort but receive a higher rent. For the marginal effect, recall that type (t,φ) is indifferent between exerting high or low effort, where we omit the term t from $\varphi(t)$ for notational simplicity. Exerting high effort yields expected payoff

$$w(\varphi,\varphi) + \varphi b(\varphi,\varphi) - c_1 = \mathcal{U}(\varphi) - \Delta c.$$

Exerting low effort yields

$$w(t,t) + tb(t,t) - c_0 = \mathcal{U}(t).$$

If we increase $\mathcal{U}(t)$ while leaving $\mathcal{U}(\varphi)$ constant, type (t,φ) will strictly prefer to exert low effort (that is, condition (6) will no longer hold). The type who will now be indifferent between high and low efforts $(t,\hat{\varphi})$ will be above the original one: $\hat{\varphi} > \varphi$. Therefore, increasing the rent projection at t increases the effort frontier φ, thereby reducing the effort region. Recall that the distortion associated to the effort frontier is $(\varphi - t)\Delta x - \Delta c$, for $\varphi < 1$. The cost of increasing the effort frontier – and, thereby, increasing the distortion – is captured by the distortion per unit of bonus paid to the marginal type (t,φ):

$$\frac{(\varphi - t)\Delta x - \Delta c}{\dot{\mathcal{U}}(\varphi)}, \text{ for } t < \xi^*.$$

For the intramarginal effect, note that all types in the vertical line above (t,φ) are now receiving a higher rent. The total mass of those types is $F_0(t,\varphi)$. Since the marginal type (t,φ) has mass $f(t,\varphi)$, the cost of leaving higher rents relative to the marginal type is captured by the hazard rate:

$$\frac{F_0(t,\varphi)}{f(t,\varphi)}, \text{ for } t < \xi^*.$$

Combining both terms yields the vertical effect $S_0(t)$ (with negative signs because these are costs).

Next, consider the horizontal effect. Again, there is a marginal effect through the effort frontier and an inframarginal effect through the informational rents left to all those that keep exerting high effort but receive a higher rent. Recall that type (φ^{-1},t) is indifferent between high and low efforts. His expected payoff from high effort is

$$w(t,t) + tb(t,t) - c_1 = \mathcal{U}(t) - \Delta c.$$

His expected payoff from exerting low effort is

$$w\left(\varphi^{-1},\varphi^{-1}\right) + \varphi^{-1}b\left(\varphi^{-1},\varphi^{-1}\right) - c_0 = \mathcal{U}\left(\varphi^{-1}\right).$$

Raising $\mathcal{U}(t)$ while keeping $\mathcal{U}(\varphi^{-1})$ unchanged makes type (φ^{-1},t) strictly prefer to exert high effort. Thus, the effort frontier shifts to the right (the type who will now be indifferent between both effort levels is $(\hat{\varphi}^{-1},t)$ with $\hat{\varphi}^{-1} > \varphi^{-1}$), increasing the region of high effort. The benefit from shifting the effort frontier is captured by the distortion per unit of bonus paid to the marginal type (φ^{-1},t):

$$\frac{(t-\varphi^{-1})\Delta x - \Delta c}{\dot{\mathcal{U}}(\varphi^{-1})}, \quad \text{for } t > \varphi^*.$$

The inframarginal effect arises from the rents left to all types to the left of (φ^{-1},t), who still exert high but now obtain higher informational rents. The cost of leaving these rents is given by the mass of those types relative to the marginal type:

$$\frac{F_1(\varphi^{-1},t)}{f(\varphi^{-1},t)}, \quad \text{for } t > \varphi^*.$$

Adding the marginal and inframarginal effects yields the horizontal effect $S_1(t)$.

S^* is the *inframarginal virtual surplus*, which is the discrete counterpart of S_0 at $\varphi^* = \varphi(0)$. In order to interpret it, consider an increase in the informational rent left in a small neighborhood of φ^*. Recall that types (t,φ^*) with $t \leq t^*$ get the same contract as (φ^*,φ^*) and are indifferent between exerting high and low efforts. Thus, the expected payoff from high effort is

$$w(\varphi^*,\varphi^*) + \varphi^* b(\varphi^*,\varphi^*) - c_1 = \mathcal{U}(\varphi^*) - \Delta c = 0.$$

The payoff from low effort is zero – since, by Proposition 1, types (t,t) with $t \leq t^*$ get zero rents. Therefore, an increase in $\mathcal{U}(\varphi^*)$ makes all those types strictly prefer to exert high effort so that the effort frontier shifts down. Incorporating each of these types reduces the distortion $(\varphi^* - t)\,\Delta x - \Delta c$. Since there exists a mass $F_1(t^*,\varphi^*)$ of such types, the total gain from incorporating them equals:

$$\frac{\{\varphi^* - E[t|t \leq t^*,\varphi^*]\}\,\Delta x - \Delta c}{\dot{\mathcal{U}}(\varphi^*)} F_1(t^*,\varphi^*).$$

Moreover, because all of these types get zero payoffs, no informational rents have to be left. Hence, the hazard rate that appears in the expressions of S_0 and S_1 vanishes from S^*.

For notational clarity, let $\mathcal{S}(t,\mathcal{U}) := S_0(t,\mathcal{U})\,f(t,\varphi) + S_1(t,\mathcal{U})\,f(\varphi^{-1},t)$ denote the marginal virtual surplus weighted by its probability density. The following lemma establishes that any optimal mechanism must maximize the expected virtual surplus among the class of feasible mechanisms.

Lemma 9. *Let $\mathcal{U}$ be an optimal rent projection. Then, for any feasible rent projection $\mathcal{V} : [0,1] \to \mathbb{R}$,*

$$\int_0^1 [\mathcal{U}(t) - \mathcal{V}(t)]\,\mathcal{S}(t,\mathcal{U})\,dt + [\mathcal{U}(\varphi^*) - \mathcal{V}(\varphi^*)]\,S^*(\mathcal{U}) \geq 0.$$

In our characterization result, we will use the following notions:

Definition 2. Let $f : [0,1] \to \mathbb{R}$ be a function with a càdlàg derivative $\dot{f} : [0,1] \to \mathbb{R}$.

- f is *strongly convex* in an interval $[a,b] \subset [0,1]$ if there exists $m > 0$ such that $\dot{f}(y) - \dot{f}(x) \geq m(y-x)$ for all $x,y \in [a,b]$;
- We say f has a *kink* at $x_0 \in (0,1]$ if $\underset{x \nearrow x_0}{lim}\ \dot{f}(x) \neq \dot{f}(x_0)$; and
- An interval $[a,b] \subset [0,1]$ is called a *maximal interval where f is affine* if there exists $m \in \mathbb{R}$ such that $\dot{f}(x) = m$, for all $x \in [a,b]$, and there is no open interval containing $[a,b]$ such that $\dot{f}(x) = m$ for all x in that interval.

The following theorem gives the necessary optimality conditions:

Theorem 1 (Optimal Mechanisms under Risk Neutrality). *Let $\mathcal{U}$ be an optimal rent projection. Then:*

1. (**pointwise condition**) *If $\mathcal{U}$ is strongly convex in a non-degenerate interval $[a,b] \subset [0,1]$, then $\mathcal{S}(t,\mathcal{U}) = 0$ for almost all $t \in [a,b]$.*
2. **(bunching conditions)** *Let $[a,b] \subset [0,1]$ be a maximal interval where $\mathcal{U}$ is affine.*

- *If $\varphi^* \notin [a,b]$, then*

$$0 \geq a \int_a^b \mathcal{S}(t,\mathcal{U})\,dt \geq \int_a^b t\mathcal{S}(t,\mathcal{U})\,dt \geq b \int_a^b \mathcal{S}(t,\mathcal{U})\,dt.$$

Moreover, if $\mathcal{U}$ has kink at a (at b), then $\int_a^b (t-a)\mathcal{S}(t,\mathcal{U})dt = 0$ ($\int_a^b (t-b)\mathcal{S}(t,\mathcal{U})dt = 0$).

- *If $a = t^*$ and $b \geq \varphi^*$, then*

$$\int_{t^*}^b \mathcal{S}(t,\mathcal{U})dt + S^*(\mathcal{U})\,F_1(t^*,\varphi^*) \leq 0 \text{ and } \int_{t^*}^b (t-\varphi^*)\mathcal{S}(t,\mathcal{U})dt \leq 0.$$

Moreover, if $\mathcal{U}$ has kink at b, then

$$\int_{t^*}^b \mathcal{S}(t,\mathcal{U})dt + S^*(\mathcal{U})\,F_1(t^*,\varphi^*) = 0 \text{ and } \int_{t^*}^b (t-\varphi^*)\mathcal{S}(t,\mathcal{U})dt = 0.$$

Recall that $\mathcal{S}(t,\mathcal{U})$ is the marginal gain from increasing the rent projection $\mathcal{U}$ at t. Whenever it differs from zero in an interval where $\mathcal{U}$ is strongly convex, we can find a small perturbation that preserves convexity and raises the principal's payoff. Therefore, $\mathcal{S}(t,\mathcal{U})$ has to equal zero in any strongly convex interval.

Part 2 are the bunching conditions. In one-dimensional models, the bunching condition is determined by the ironing principle, which can be obtained by considering perturbations to the region of pooled types. Because our model has two-dimensional types, there are two admissible perturbation directions that retain the convexity of $\mathcal{U}$: translations and rotations. The two bunching conditions state that perturbing the rent projection in either of these directions does not increase the principal's payoff.

By Proposition 1, types with low probabilities of success given both high and low efforts get a constant payment equal to the cost of low effort. The next proposition shows that there exists an adjacent region where types also get a uniform contract:

Proposition 4 (Two Contracts at the Bottom). *Let (w,b,e) be an optimal nontrivial mechanism. There exists $\varphi^* > 0$ and $t^* \in (0,\varphi^*)$ such that*

- *All types $\mathbf{p} \in [0,t^*) \times [0,\varphi^*) \cap \overline{\Delta}$ get the same contract $w = c_0$ and $b = 0$, and*
- *All types $\mathbf{p} \in [t^*,\varphi^*] \times [0,1] \cap \Delta_0$ get the same contract $w < c_0$ and $b \in (\Delta c, \Delta x]$.*

Moreover, types in both regions exert low effort.

Figure 5 illustrates the result from Proposition 4. Types with sufficiently low probability of success given low and high efforts, $p_0 \leq t^*$ and $p_1 < \varphi^*$, receive a fixed payment equal to the cost of low effort c_0 and exert low effort. Region B comprises types with intermediate probabilities of success given low efforts: $t^* < p_0 \leq \varphi^*$ and $p_1 < \varphi(p_0)$. All types in this region are offered the same contract, involving a payment with a lower fixed part $w < c_0$ and a bonus between the incremental cost of effort Δc and the incremental output Δx.

The intuition for this result is the following. All types projected into diagonal types $t < \varphi^*$ exert low effort. Therefore, slightly increasing their informational rents does not affect the effort region associated with them. However, it generates an incentive for types above them to reduce

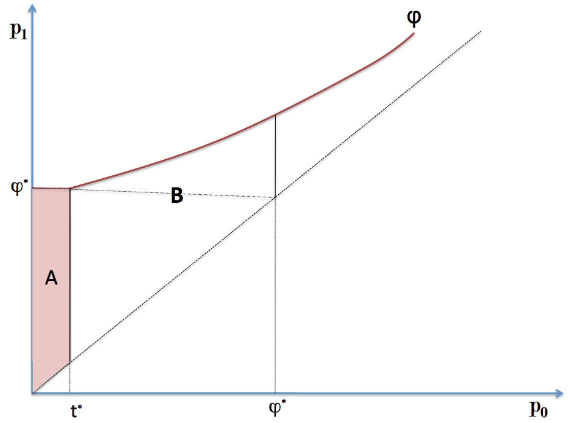

Figure 5: *Two Contracts at the Bottom: Types in Region A receive the same constant payment ($w = c_0$, $b = 0$); types in Region B receive the same contract ($w < c_0$, $b > \Delta c$).*

their effort (thereby reducing the effort region associated with types projected above φ^*) and requires more rents to be left for types that exert high effort and get the same contracts as them: (p_0, p_1) with $p_1 > \varphi(p_0)$ and $p_0 \in [\varphi^*, \varphi(\varphi^*)]$. Because both effects reduce the principal's payoff, she will want to leave as little informational rents as possible while preserving the condition that the effort frontier starts at φ^*. This is obtained by paying the zero bonus for all diagonal types that are not associated with anyone who exerts high effort (region A). For diagonal type t^*, the principal needs to pay a bonus greater than the incremental cost of effort in order to incentivize types $\{(t, \varphi^*) : t \leq t^*\}$ to exert high effort. The principal then reduces the informational rents left in this region by paying the same bonus to all those types.

We now examine the effort distortion relative to the first best. Recall that the first-best effort region under risk neutrality is determined by $(p_1 - p_0)\Delta x \geq \Delta c$. That is, a type should exert high effort if the incremental benefit from effort (i.e., the incremental effect on the probability of a high output $p_1 - p_0$ times the incremental output Δx) exceeds the incremental cost Δc. The first-best effort is implemented by making the agent a residual claimant: $b = \Delta x$.

Our first result establishes that, because the principal will never pay a bonus greater than the incremental output (Lemma 7), the effort region in any optimal mechanism is contained in the first-best effort region.

Lemma 10. *Let φ be the effort frontier function associated with it an optimal mechanism. Then, $\varphi(t) \geq t + \frac{\Delta c}{\Delta x}$ whenever $\varphi(t) < 1$.*

Lemma 10 does not rule out the possibility of implementing the first-best effort frontier for some projected types t, which would be the case if the optimal mechanism left some types as residual claimants by offering them a bonus equal to the incremental output. These contracts would eliminate the distortion but leave large informational rents to the agents.

Next, we establish that the optimal mechanism generically features "strict distortion at all points" in the sense that the low-effort region in the first-best benchmark is contained in the interior of the low-effort region of any optimal mechanism (see Figure 6):

Definition 3. Let (w, b, e) be a feasible mechanism and let φ be the associated effort frontier. We say that there is *strict distortion at all points* if $\varphi(t) > t + \frac{\Delta c}{\Delta x}$ for all t such that $\varphi(t) < 1$.

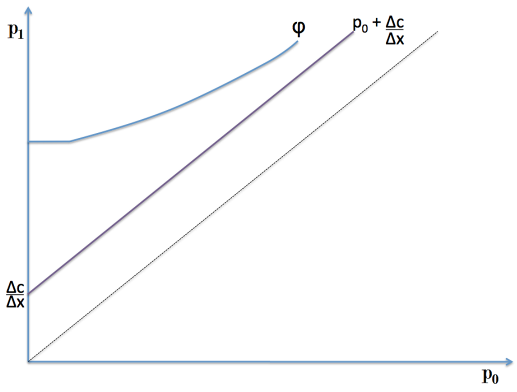

Figure 6: *Strict distortion at all points.*

We say that the a mechanism *partially sells the firm* if all types pick one of the following two contracts: $(c_0, 0)$ and $(w, \Delta x)$, for some $w \leq c_0$. Under a mechanism that partially sells the firm, agents self-select into two categories: "employees" who work for a fixed wage contract and get zero rents, and "entrepreneurs" who buy the firm for the price w and become residual claimants.

Lemma 11. *Let (w, b, e) be an optimal mechanism. Then, either there is strict distortion at all points, or the principal partially sells the firm.*

The intuition behind Lemma 11 is the following. Because distortions close to the optimum have second-order costs, it can only be desirable to implement zero-distortion for some type if there are no other types with positive distortions and rents (otherwise, the principal can improve by rebalancing the amounts of distortion between these two types). Furthermore, if the optimal mechanism sells the firm to a certain projected type $\hat{t}$, it must also sell the firm to all types with projections above $\hat{t}$. Then, all types with projections above $\hat{t}$ are also undistorted.

The result from Lemma 11 contrasts starkly with standard one-dimensional models, where all but the highest type obtain distorted allocations. Here, either the allocations of all projected types are distorted, or only projected types who get zero rents ($t \leq t^*$) obtain distorted allocations. Strict distortion at all points is a consequence of the global incentive constraint, which induces the principal to distort even the allocation of the highest types.

The lemma leaves the possibility that the optimal mechanism may involve partially selling the firm and, therefore, not distort the effort decision of all types with positive rents. Proposition 5 shows that this is generically not the case. Let $(\mathcal{D}, || \cdot ||_\infty)$ be the space of continuous density functions $f : \bar{\Delta} \to \mathbb{R}_+$ endowed with the norm of uniform convergence. A property is *generic* if the set of density functions for which it holds is open and dense in $\mathcal{D}$.

Proposition 5 (Strict Distortion). *Generically, there is strict distortion at all points.*

Because only local incentive constraints bind in standard one- and multi-dimensional screening models, there is "no distortion at the boundary." In our model, however, because all types in the high-effort region have binding global incentive-compatibility constraints, the optimal mechanism generically features distortion at all points.

Finite Mechanisms

A central message from nonlinear pricing models of multidimensional screening is the generality of bunching (Rochet and Choné, 1998). Obviously, since types are two-dimensional while instruments are one-dimensional, there has to be some bunching in our model. The interesting issue here is whether a *positive mass* of types get the same contract.

Under 'pure moral hazard' (i.e., when types are observable but effort is not), each contract is taken by at most two types. If the solution of the principal's program (P') consisted of a strictly convex rent projection $\mathcal{U}$, each contract would be taken by the vertical and horizontal projections from Figure 3, which also has zero mass. However, Proposition 1 shows that the convexity constraint binds. As a result, regions of types with positive mass are offered the same contract. The intuition is reminiscent of Rochet and Choné: type multidimensionality makes it hard to satisfy the local second-order condition from incentive compatibility (non-decreasing allocations) so that the solution involves bunching.

In this subsection, we show that, under certain conditions, the optimal mechanism can be implemented with a menu of *finitely many contracts.* Therefore, the force towards bunching is so strong that the principal will, in some cases, prefer to offer a very reduced number of contracts.

Let us define the *generalized hazard rate* function

$$R(p_0, p_1) := \frac{F_0(p_1, 1) + F_1(p_0, p_1)}{f(p_0, p_1)}.$$

The first term, $\frac{F_0(p_1,1)}{f(p_0,p_1)}$, is the ratio between the mass of types above the diagonal point (p_1, p_1) and the mass at (p_0, p_1). The second term, $\frac{F_1(p_0,p_1)}{f(p_0,p_1)}$, is the ratio between the mass of types to the left of (p_0, p_1) and the mass at (p_0, p_1). We say that the generalized hazard rate satisfies the *increasing rents condition* if

$$\frac{\partial R}{\partial p_0}(p_0, p_1) > 0 \text{ and } \frac{\partial R}{\partial p_0}(p_0, p_1) + \frac{\partial R}{\partial p_1}(p_0, p_1) \geq 0.$$

Note that the uniform distribution satisfies increasing rents.[25]

An implication of the increasing rents condition is that the informational rents associated with types along the diagonal are non-decreasing. The following lemma establishes that, under increasing rents, any optimal mechanism (w, b, e) can be implemented by offering at most two contracts to all types (p_0, p_1) with $\varphi(p_0) = 1$:

Lemma 12. *Suppose that the distribution of types satisfies increasing rents. The optimal rent projection is a piecewise linear function with at most two pieces on* $[\xi^*, 1]$.

The intuition behind Lemma 12 is the following. When the generalized hazard rate has increasing rents, the marginal virtual surplus is a strictly decreasing function of bonus. Thus, the marginal virtual surplus will be different from zero in every interval where the bonus is strictly increasing (except for at most one point). There are three possible cases: the virtual surplus may be always positive, always negative, or initially positive and then negative. In all of these cases, it is possible to increase the virtual surplus by replacing the original strictly increasing bonus by a piecewise linear one that preserves incentive compatibility. For example, if the marginal virtual surplus is negative in the entire interval $[\xi^*, 1]$, replacing the rent projection by the piecewise linear function consisting of the maximum of the tangents of the original rent projection at ξ^* and 1 preserves incentive compatibility and strictly increases the virtual surplus.

[25] Increasing rents is weaker than strict monotonicity, since R may decrease in p_1 as long as it is sufficiently increasing in p_0.

Recall that all contracts are identified by the contracts offered to types on the 45^o line. By Proposition 4, diagonal types (t,t) with $t \leq \varphi^*$ are offered exactly two contracts. Moreover, by Lemma 12, when the generalized hazard rate has increasing rents, all diagonal types (t,t) with $t \geq \xi^*$ are offered at most two contracts. In general, diagonal types (t,t) with $t \in (\varphi^*, \xi^*)$ can be offered any number of contracts. The next proposition establishes that when the incremental output Δx is "not too large" relative to the incremental cost Δc of effort, this region is empty and, therefore, the optimal mechanism features at most three contracts:

Proposition 6 (**Three Contracts**). *Suppose that the distribution of types satisfies increasing rents and let* $\frac{\Delta x}{\Delta c} \leq 2$*. Then:*
(i) *the optimal rent projection is piecewise linear with at most three pieces; and*
(ii) *the optimal mechanism can be implemented with at most three contracts.*

In particular, when the distribution is uniform the finiteness of contracts under the optimal mechanism holds for a slightly larger set of parameter values:

Corollary 1 (**Uniform Distribution**). *Suppose that types are uniformly distributed on* $\overline{\Delta}$ *and let* $\frac{\Delta x}{\Delta c} \leq 3$*. Then:*
(i) *the optimal rent projection is piecewise linear; and*
(ii) *the optimal mechanism can be implemented with a finite number of contracts.*

Finite optimal mechanisms also arise under different supports for the type distribution. For our next proposition, we depart from the full support assumption and assume that the conditional probability of a high outcome is bounded away from zero. Formally, for this proposition we consider following modified type space:

$$\bar{\Delta}\left(\underline{\mathrm{p}}\right) = \left\{(p_0, p_1) : \ \underline{\mathrm{p}} \leq p_0 \leq p_1\right\},$$

where $\underline{\mathrm{p}} > 0$. It is straightforward to adapt the characterization results previously derived for the type space $\bar{\Delta}$ to this case. Then, we obtain the following result:

Proposition 7 (**Two Contracts**). *Suppose the density* $f(p_0, p_1)$ *on* $\bar{\Delta}(\underline{p})$ *is non-increasing in* p_0*, and let* $\underline{p} \geq \frac{\Delta x - \Delta c}{\Delta x + \Delta c}$*. Then,*
(i) *the optimal rent projection is piecewise linear with at most two pieces; and*
(ii) *the optimal mechanism can be implemented with at most two contracts.*

The results above highlight the trade-off between the incentives for effort provision and rent extraction. When the incremental output is "not too large" relative to the incremental cost of effort and the distribution either satisfies increasing rents (Proposition 6) or is "sufficiently bounded away from zero" (Proposition and 7), the principal can benefit from offering a limited number of contracts, which reduces the informational rents that have to be left to the agent.

5 Risk Aversion

In this section, we generalize the characterization of optimal mechanisms obtained in the risk-neutral case (Theorem 1) for weakly concave utility functions. The generalizations of the marginal and inframarginal virtual surpluses when the utility function is weakly concave are:[26]

$$S_0(t,\mathcal{U}) := \begin{cases} -\frac{(\varphi - t)\Delta x - (G(\varphi) - G)}{\dot{\mathcal{U}}(\varphi)} - \frac{\partial G}{\partial \mathcal{U}} \frac{F_0(t,\varphi)}{f(t,\varphi)} \text{ if } t < \xi^* \\ -\frac{\partial G}{\partial \mathcal{U}} \frac{F_0(t,1)}{f(t,1)} \text{ if } t \geq \xi^* \end{cases},$$

[26] To simplify the notation, the dependence of the derivatives $\partial G/\partial \mathcal{U}$ and $\partial G/\partial \dot{\mathcal{U}}$ on $(\mathcal{U}, \dot{\mathcal{U}}, t)$ is omitted.

$$S_1(t,\mathcal{U}) := \begin{cases} 0 \text{ if } t \leq \varphi^* \\ \frac{(t-\varphi^{-1})\Delta x-(G-G(\varphi^{-1}))}{\dot{\mathcal{U}}(\varphi^{-1})} - \frac{\partial G}{\partial \mathcal{U}} \frac{F_1(\varphi^{-1},t)}{f(\varphi^{-1},t)} \text{ if } t > \varphi^* \end{cases}, \text{ and}$$

$$S^*(\mathcal{U}) := \frac{(\varphi^* - E[t|t \leq t^*, \varphi^*])\,\Delta x - G(\varphi^*)}{\dot{\mathcal{U}}(\varphi^*)} F_1(t^*, \varphi^*),$$

where we are using the following notation $G = G(\mathcal{U},\dot{\mathcal{U}},t)$, $G(\varphi) = G(\mathcal{U}(\varphi),\dot{\mathcal{U}}(\varphi),\varphi)$ and $G(\varphi^{-1}) = G(\mathcal{U}(\varphi^{-1}),\dot{\mathcal{U}}(\varphi^{-1}),\varphi^{-1})$.

The marginal virtual surpluses S_0 and S_1 differ from their risk-neutral counterparts (14) and (15) in that now the hazard rates are multiplied by the partial derivative $\partial G/\partial \mathcal{U}$. In the risk neutral case, each unit of utility left to the agent costed one dollar to the principal. Therefore, the informational rents were determined solely by the hazard rates that specified the mass of types who received these rents relative to the type on the effort frontier. Under risk aversion, each unit of utility left to the agent costs $\partial G/\partial \mathcal{U}$ to the principal. Since the principal cares about informational rents in monetary rather than in utility units, the hazard rate has to be multiplied by the "exchange rate" between utils and dollars $\partial G/\partial \mathcal{U}$. Since inframarginal types have no informational rents, S^* remains unchanged relative to the risk neutral case. As in the risk neutral case, let $\mathcal{S}(t,\mathcal{U}) \equiv S_0(t,\mathcal{U}) f(t,\varphi) + S_1(t,\mathcal{U}) f(\varphi^{-1},t)$ denote the marginal virtual surplus weighted by its probability density.

When the agent is risk averse, the cost of providing utility $\mathcal{U}$ is also a function of the power of the contract $\dot{\mathcal{U}}$. Thus, the relative cost of increasing the power of a diagonal type t equals the cost of providing power $\partial G/\partial \dot{\mathcal{U}}$ times the hazard rate of types who get the same contract on the high-effort region (horizontal effect) and the hazard rate of types who get the contract on the low-effort region (vertical effect). It is, therefore, useful to define each of these marginal costs as

$$C_0(t,\mathcal{U}) := \begin{cases} \frac{\partial G}{\partial \dot{\mathcal{U}}} \frac{F_0(t,\varphi)}{f(t,\varphi)} \text{ if } t < \xi^* \\ \frac{\partial G}{\partial \dot{\mathcal{U}}} \frac{F_0(t,1)}{f(t,1)} \text{ if } t \geq \xi^* \end{cases},$$

$$C_1(t,\mathcal{U}) := \begin{cases} 0 \text{ if } t \leq \varphi^* \\ \frac{\partial G}{\partial \dot{\mathcal{U}}} \frac{F_1(\varphi^{-1},t)}{f(\varphi^{-1},t)} \text{ if } t > \varphi^* \end{cases}.$$

As in the risk-neutral case, it is convenient to define the marginal cost of providing power weighted by its probability density:

$$\mathcal{C}(t,\mathcal{U}) := C_0(t,\mathcal{U}) f(t,\varphi) + C_1(t,\mathcal{U}) f\left(\varphi^{-1},t\right).$$

The following theorem gives the optimality conditions:

Theorem 2 (**Optimal Mechanisms under Risk Aversion**). *Let $\mathcal{U}$ be an optimal rent projection. Then:*

1. **(pointwise condition)** *If $\mathcal{U}$ is strongly convex in a non-degenerate interval $[a,b] \subset [0,1]$ such that $\varphi^* \notin [a,b]$, then*

   $$\mathcal{S}(t,\mathcal{U}) + \frac{d}{dt}\{\mathcal{C}(t,\mathcal{U})\} = 0,$$

   for almost all $t \in [a,b]$.

2. **(bunching conditions)** *Let $[a,b] \subset [0,1]$ be a maximal interval where $\mathcal{U}$ is affine.*

- *If $\varphi^* \notin [a,b]$, then*

$$0 \geq a \int_a^b \mathcal{S}\left(t, \mathcal{U}\right) dt \geq \int_a^b t\mathcal{S}\left(t, \mathcal{U}\right) dt \geq b \int_a^b \mathcal{S}\left(t, \mathcal{U}\right) dt.$$

  Moreover, if $\mathcal{U}$ has kink at a (at b), then $\int_a^b (t-a)\mathcal{S}(t,\mathcal{U})dt = 0$ ($\int_a^b (t-b)\mathcal{S}(t,\mathcal{U})dt = 0$).

- *If $a = t^*$ and $b \geq \varphi^*$, then*

$$\int_{t^*}^b \mathcal{S}(t,\mathcal{U})dt + S^*\left(\mathcal{U}\right) F_1(t^*, \varphi^*) \leq 0, \text{ and } \int_{t^*}^b (t-\varphi^*)\mathcal{S}(t,\mathcal{U})dt \leq 0.$$

  Moreover, if $\mathcal{U}$ has kink at b, then

$$\int_{t^*}^b \mathcal{S}(t,\mathcal{U})dt + S^*\left(\mathcal{U}\right) F_1(t^*, \varphi^*) = 0, \text{ and } \int_{t^*}^b (t-\varphi^*)\mathcal{S}(t,\mathcal{U})dt = 0.$$

As in the risk-neutral case, if the pointwise condition fails in an interval where $\mathcal{U}$ is strongly convex, there exists a small perturbation that preserves the convexity of the rent projection and raises the principal's payoff. The bunching conditions are obtained by applying translations and rotations to the rent projection, which also preserve convexity.

6 Other Applications

The principal-agent framework considered previously has a natural interpretation in terms of employment relationships and, therefore, is commonly used in corporate finance and labor economics. In this section, we modify our basic framework to cover models of insurance provision by a monopolist, procurement and regulation, and optimal taxation.

6.1 Insurance

The main difference between the model considered in Section 5 and the standard insurance framework is the presence of type-dependent participation constraints (c.f., Stiglitz, 1977; Chade and Schlee, 2012), since riskier types have a lower opportunity cost of remaining uninsured.

The model features a monopolistic insurance firm (principal) offering insurance to risk averse consumers (agents), with a strictly concave utility function u. Consumers have initial wealth $I > 0$ and face a potential loss $L \in (0, I)$. They exert a preventive effort $e \in \{0, 1\}$, which affects the loss probability but is unobservable by the firm. We let p_i denote the probability of *not* suffering the loss L conditional on effort e_i, $i = 0, 1$.

Consumers have private information about the loss probabilities conditional on each effort level. Therefore, their types are identified by a vector $(p_0, p_1) \in \bar{\Delta}$, satisfying MLRP. The insurance firm has a continuous prior distribution f over types with full support on $\bar{\Delta}$. A type-(p_0, p_1) consumer who does not purchase insurance gets expected utility

$$V\left(p_0, p_1\right) := \max_{i \in \{L,H\}} p_i u\left(I\right) + \left(1 - p_i\right) u\left(I - L\right) - c_i.$$

As in Section 2, we assume that consumers have access to a *free disposal* technology. In the insurance context, free disposal states that consumers can costlessly generate a loss. As a result, the insurer will not offer policies in which the indemnity exceeds the loss L.

Writing mechanisms in terms of the consumer's utility as in Section 2 (equation 10), we obtain the following individual-rationality constraint for the insurance model:

$$U(p_0, p_1) \geq V(p_0, p_1), \quad \text{for all } (p_0, p_1) \in \bar{\Delta}. \tag{IR INS}$$

Thus, a mechanism is *feasible* if it satisfies incentive-compatibility (IC), individual-rationality (IR INS), and free disposal (FD). The firm's problem is to pick a feasible mechanism that maximizes its expected profits (2).

Note that any mechanism in which some types are excluded is equivalent to a mechanism in which the principal offers the *zero coverage contract*: $W = I - L$, $B = L$. In this contract, the agent pays zero in both states. Therefore, we say that a mechanism excludes a certain type if that type is offered the zero coverage contract. Our first result establishes that it is always optimal to exclude a non-degenerate region of safer types:

Proposition 8 (**Exclusion in Insurance**). *There exists $\bar{p}_0 < 1$ such that it is optimal to exclude type (p_0, p_1) if and only if $p_0 \geq \bar{p}_0$ or $p_1 \geq \bar{p}_0 + \frac{\Delta c}{u(I)-u(I-L)}$.*

The optimality of exclusion is a consequence of the interaction between multidimensional types and type-dependent participation constraints. With pure adverse selection and one-dimensional types, Chade and Schlee (2012, Proposition 2) showed that no type is excluded if there are enough low types in the population or if agents are sufficiently risk averse. Moreover, we have shown in Section 3 that when reservation utilities are not type-dependent, exclusion is not optimal (as long as there is no exclusion in the first best). Proposition 8 contrasts with both of these results in establishing that that exclusion is always optimal in this multidimensional model. In insurance, exclusion happens "at the top" – the safest types are the ones who do not purchase any coverage.

The intuition for our "exclusion at the top" result is the following. Starting from a situation in which all risk types participate, a reduction in informational rents excludes the types with the highest outside options. When the reduction is small enough, this set only includes the highest possible types (i.e., those with p_0 close enough to 1), who never find it beneficial to exert effort. Therefore, excluding those types reduces the informational rents left to all other types and does not affect the effort region.

Next, we establish that the presence of moral hazard shrinks the effort region among types who participate relative to a situation in which insurance is not available. In the absence of insurance, type (p_0, p_1) chooses to exert high effort if

$$p_1 \geq p_0 + \frac{\Delta c}{u(I) - u(I-L)}. \tag{16}$$

By construction, the effort frontier for excluded types coincides with the uninsured effort frontier (16). The next proposition establishes that the effort frontier for types that participate lies strictly above the uninsured effort frontier. Therefore, types who participate exert "less effort" than if they were uninsured:

Proposition 9 (**Strict Distortion Relative to No Insurance**). *Let φ be the effort frontier associated with an optimal mechanism, and let $\bar{p}_0$ be the first projected type to be excluded as defined in Proposition 8. Then, $\varphi(p_0) > p_0 + \frac{\Delta c}{u(I)-u(I-L)}$ for all $p_0 < \bar{p}_0$.*

Remark 1. Because utility is non-transferable, principal and agent generally disagree over the first-best effort level. As seen above, high effort is efficient *from the agent's perspective* if condition (16) holds. On the other hand, high effort is efficient *from the principal's perspective* if $p_1 \geq p_0 + \frac{\Delta c}{L}$. The later corresponds to the first-best frontier in our model, since we are assuming that the principal has all the bargaining power.[27]

6.2 Regulation

In this subsection, we adapt our basic framework to a model of procurement and regulation. We follow the general setup from Laffont and Tirole (1986, 1993), except that we allow the firm's cost-reducing effort to affect firm costs stochastically. This modification implies that the model cannot be reduced to a pure adverse selection model anymore.

A regulated firm produces an indivisible project at a random monetary cost, which can be either low C_L or high C_H, $C_H > C_L$. The firm's manager exerts a cost-reducing effort which can be either high ($e = 1$) or low ($e = 0$), and is not observed by the regulator. The cost-reducing effort stochastically affects the firm's monetary cost. The firm faces a low cost C_L with probability p_e, and a high cost C_H with probability $1 - p_e$. Exerting effort increases the likelihood of a low cost realization: $p_1 \geq p_0$. Therefore, conditional probabilities satisfy MLRP: $(p_0, p_1) \in \bar{\Delta}$. The firm's manager has cost c_e from exerting effort e, where $c_1 > c_0$ and $\Delta c := c_1 - c_0$.

The project generates a consumer surplus of $S > 0$. The regulator observes the monetary cost incurred by the firm but not the cost-reducing effort. As an accounting convention, we assume that the regulator reimburses the firm's monetary costs in addition to paying the firm w in case of high cost and $w + b$ in case of low cost. Thus, b denotes the power of the regulated firm's contract. The expected utility of the firm's manager is then

$$U = w + p_e b - c_e, \tag{17}$$

where $e = 0, 1$. We assume that the manager has access to a free disposal technology and, therefore, can freely inflate costs. As a result, the regulator will not offer contracts with negative power. Moreover, the manager has an outside option with payoff normalized to zero.

Conditional on effort e, the regulator pays the firm an expected amount $w + p_e b + C_H - p_e (C_H - C_L)$. As in Laffont and Tirole (1986, 1993), we assume that the government has to revert to distortionary taxation in order to raise funds and, therefore, the regulator faces a shadow cost of public funds $\lambda > 0$. As a result, the net surplus of consumers/taxpayers is

$$S - (1 + \lambda) \left[w + p_e b + C_H - p_e (C_H - C_L) \right].$$

A utilitarian regulator maximizes the sum of the consumers' net surplus and the expected utility of the firm's manager (17):

$$S - (1 + \lambda) \left[w + p_e b + C_H - p_e (C_H - C_L) \right] + U. \tag{18}$$

In order to rewrite this model in terms of our basic framework, let us introduce the variables x_H and x_L, which denote the taxpayers' surplus net of the utility left to the firm's manager:

$$x_H := S - (1 + \lambda) C_L, \quad x_L := S - (1 + \lambda) C_H.$$

[27] When the agent's incremental utility from the loss is lower than the principal's incremental utility from the loss – i.e., $u(I) - u(I - L) \leq L$ – the agent picks a (weakly) lower effort than the principal would demand if effort were observable. Combining with Proposition 11, this implies that the second-best effort frontier lies above the first-best effort frontier. Note, however, that the second-best effort frontier is *not* above the first-best frontier when the opposite is true: $u(I) - u(I - L) > L$. In that case, agents who are excluded from the mechanism, for example, will choose effort according to the frontier (16), which would lie below the first-best frontier.

Note that a high output x_H corresponds to a low cost realization C_L and vice versa. Moreover, we let $\Delta x := x_H - x_L > 0$ denote the net gain from a low cost relative to a high cost realization. Rearranging expression (18), we can rewrite the regulator's objective function as

$$x_L + p_e \Delta x - (1+\lambda)c_e - \lambda U.$$

Because the shadow cost of public funds λ is positive, the regulator would like to avoid leaving rents to the firm's manager.

In the first-best benchmark where both effort and the firm's type (p_0, p_1) are observable, the regulator solves

$$\max_{(U,e)} x_L + p_e \Delta x - (1+\lambda)c_e - \lambda U$$

subject to $U \geq 0$. Therefore, the first-best mechanism leaves zero rents to the firm's manager and requires a high effort whenever $p_1 \geq p_0 + (1+\lambda)\frac{\Delta c}{\Delta x}$.[28]

We now consider the situation where the regulator does not observe either the firm manager's cost-reducing effort e or the firm's effectiveness in reducing costs (p_0, p_1). The regulator has a prior distribution about the firm's type (p_0, p_1) with full support on the set of conditional distributions that satisfy MLRP $\bar{\Delta}$ described by the continuous density f.

Following the same steps as in Section 4, we can establish the existence of an effort frontier function φ. The informational rents associated to a feasible mechanism are determined by

$$U(\boldsymbol{p}) = \begin{cases} w(\boldsymbol{p}) + p_0 b(\boldsymbol{p}) - c_0, & \text{if } p_1 \leq \varphi(p_0) \\ w(\boldsymbol{p}) + p_1 b(\boldsymbol{p}) - c_1, & \text{if } p_1 > \varphi(p_0) \end{cases}.$$

Let $\mathcal{U}(t) := w(t,t) + tb(t,t) - c_0$ denote the rent projection function. By Lemma 5, we have

$$U(t,s) = \begin{cases} \mathcal{U}(t), & \text{if } s \leq \varphi(t) \\ \mathcal{U}(s) - \Delta c & \text{if } s > \varphi(t) \end{cases}.$$

Therefore, the regulator's problem is:

$$\max_{\varphi, \mathcal{U}} \quad x_L - \lambda c_0 + \int_0^1 \int_t^{\varphi(t)} (t\Delta x - \lambda \mathcal{U}(t) - c_0) f(t,s)dsdt + \int_0^1 \int_{\varphi(t)}^1 (s\Delta x - \lambda \mathcal{U}(s) - c_1) f(t,s)dsdt,$$

subject to (6), $\mathcal{U}$ nondecreasing and convex, and $\mathcal{U}(0) \geq 0$. Adapting the results derived in Section 4, we obtain the following proposition:

Proposition 10 (**Optimal Regulation**). *There exists an optimal mechanism. Any optimal mechanism has the following properties:*

- *There exists $\varphi^* > 0$ and $t^* \in (0, \varphi^*)$ such that*
  - *All types $\mathbf{p} \in [0, t^*) \times [0, \varphi^*) \cap \overline{\Delta}$ get a cost-plus contract ($w = c_0$, $b = 0$), exert zero effort, and get zero rents;*

[28] There are two differences between this model and the framework from Section 4. First, each dollar left to the agent costs λ rather than 1. Because the regulator's payoff consists of the sum between the manager's and the taxpayers' utility, and each dollar left to the manager costs $1+\lambda$ to taxpayers, the total effect on the regulator's payoff is the shadow cost λ. Second, the regulator takes into account the additional effect of compensating the manager's disutility of effort through the requirement of raising public funds. Therefore, instead of subtracting the total surplus by c_e as in the standard model, the principal subtracts $(1+\lambda)c_e$.

  - *All types* $\mathbf{p} \in [t^*, \varphi^*] \times [0,1] \cap \Delta_0$ *get a uniform contract with positive power* ($w < c_0$, $b \geq \Delta c$)*, exert zero effort, and get positive rents.*

- *The power of the contract does not exceed the cost reduction* ($b \leq C_H - C_L$)*, and the effort region is contained in the first-best effort region;*
- *Generically, there is strict distortion at all points; and*
- *Exclusion is optimal if and only if exclusion is first-best optimal.*

The characterization of the optimal mechanism (Theorem 1) and the results on finite mechanisms can also be easily adapted for the regulation model.

6.3 Optimal Taxation

In this subsection, we consider an optimal taxation modelof a Rawlsian tax agency (principal) who wishes to design a tax system for a population of taxpayers (agents). Taxpayers generate an output that can be either high, x_H, or low, x_L. They choose effort $e \in \{0,1\}$, which is not observed by the tax agency and stochastically affects their output. Taxpayers are also privately informed about the effectiveness their effort. Thus, each taxpayer is represented by a type vector (p_0, p_1) representing the probability of a high output given each effort. Types have full support on the set of probabilities that satisfy MLRP. Taxpayers have access to a free disposal technology and, therefore, cannot be charged incremental taxes that exceed 100%.[29]

Our model can be interpreted as the optimal design of unemployment insurance. In this interpretation, unemployed workers (taxpayers) may or may not find a job. The high output x_H corresponds to the income of a worker who finds a job and the low output x_L corresponds to the income of a worker who does not find a job (possibly zero).

Our model can also be interpreted more generally as the optimal design of an income tax in the spirit of Mirrlees (1971), although the assumption of two possible outcomes may be harder to justify in this case. In the Mirrleesian framework, taxpayers also have an unobservable productivity type and choose an effort level. However, because the mapping from types and effort to income is deterministic, the model can be reduced to a screening problem with adverse selection only.[30] Here, because effort affects income stochastically, the model does not reduce to a pure adverse selection setting. Moreover, because taxpayers have private information about the probabilities of outputs given each effort level, their types are multidimensional.

We follow Piketty (1997) and Saez (2001) in assuming that the tax agency is Rawlsian and, therefore, maximizes the utility of the least favored individual.[31] By Property (a) from Lemma 2, incentive compatibility implies that taxpayers' utilities are increasing in their types. As a result, the least favored individual is the lowest type: $(0,0)$. As defined in Section 2, a mechanism $(w, b, e) : \bar{\Delta} \to \mathbb{R}^2 \times \{0,1\}$ specifies the agent's utility in case of low output w, the power of the contract b, and an effort recommendation e. The tax agency must design a mechanism that

[29] There is a large literature on optimal taxation that assumes free disposal, starting with Diamond and Mirrlees (1971) and Mirrlees (1972).

[30] Mirrlees (1990) studied optimal taxation in a model where incomes are uncertain, although he restricted the analysis to linear taxes.

[31] Saez (2001) considers both Rawlsian and utilitarianist tax agencies. Our approach can be extended to the utilitarianist case, although it requires considering an ex-ante participation constraint in our general framework.

maximizes the utility of the lowest type, $w(0,0) - c_0$, among mechanisms that satisfy incentive compatibility (IC), free disposal (FD), and the resource constraint

$$\int_{\bar{\Delta}} \left\{ \; x_L - u^{-1}(w(\mathbf{p})) + p_{e(\mathbf{p})} \left\{ \Delta x - \left[u^{-1}(w(\mathbf{p}) + b(\mathbf{p})) - u^{-1}(w(\mathbf{p})) \right] \right\} \; \right\} f(\mathbf{p}) d\mathbf{p} \geq R,$$

where the parameter $R \in \mathbb{R}$ denotes the total resources (possibly negative) that need to be financed by the tax program.

In the principal-agent framework described in Section 2, the principal wanted to extract the largest amount of expected resources from agents subject to the lowest possible type obtaining a utility above a certain reservation utility (normalized to zero). Here, the tax agency wants to maximize the utility of the lowest possible type subject to expected resources left to agents not exceeding a certain level. Hence, the tax agency's problem is the dual of the principal's problem from our general framework. It is then straightforward to adapt the analysis from the previous sections to obtain several new results for optimal taxation in the presence of joint moral hazard and adverse selection. Theorem 2 derives the optimality conditions.

Adapting Proposition 1, it follows that types in a non-degenerate region at the bottom of the distribution $\mathbf{p} \in [0, t^*) \times [0, \varphi^*) \cap \bar{\Delta}$ are all offered the same after-tax income and exert low effort. Therefore, the tax agency guarantees a constant after-tax income to these workers, regardless of their outputs (100% tax rate)[32] Moreover, the difference between the after-tax income in case of high and a low earnings, B, is a non-decreasing function of types.

Following Piketty (1997) and Diamond (1998, 2005), suppose that taxpayers have a quasi-linear utility function: $W - c_e$.[33] We can then adapt the results from Section 4. Proposition 4 establishes that types in the intermediate region, $\mathbf{p} \in [t^*, \varphi^*) \times [0, 1] \cap \bar{\Delta}$, also face a uniform tax rate (although their tax rate is lower than 100%).

Proposition 5 shows that strict distortion at all points is a generic property. Strict distortion at all points, which contrasts with the famous efficiency-at-the-top result from models with one-dimensional types, is caused by the global incentive constraints that are binding due to moral hazard. Additionally, Propositions 6 and 7 and Corollary 1 determine conditions under which optimal tax system can be implemented using a finite number of tax brackets.

7 Conclusion

Contracting situations typically combine elements of both adverse selection and moral hazard. Most of the literature, however, has focused on models in which only one of them is present. In this paper, we showed that adverse selection and moral hazard are not separable issues, and the

[32] Formally, there exists $\bar{p}_0 > 0$ and $\bar{p}_1 > 0$ such that $b(p_0, p_1) = 0$ for all $(p_0, p_1) \leq (\bar{p}_0, \bar{p}_1)$. This conclusion resembles results from the one-dimensional type model. Under a utilitarianist welfare function, the tax rate at the bottom of the earnings distribution is *zero* if and only if earnings are bounded away from zero (Seade, 1977; Ebert, 1992). Under a Rawlsian welfare function, the optimal tax rate at the bottom should be strictly lower than 100% if earnings are bounded away from zero and 100% if they are not. Since, in practice, the most disadvantaged individuals have zero earnings, the optimal income taxes at the bottom should be strictly positive under a utilitarian welfare function and 100% under a Rawlsian welfare function (c.f. Saez, 2001, Piketty and Saez, 2012). Note, however, that the optimality of the 100% tax rate in our model does not rely on the expected earnings of lowest types.

[33] Quasi-linearity is often justified empirically by the fact that income elasticities of primary earners is close to zero (although income effects are important for secondary earners). Theoretically, optimal income taxes in the Mirrleesian framework are much simpler under quasi-linear utilities.

interaction between them can generate contracts that are fundamentally different from environments featuring only one of them.

In our model, the principal always extracts all agents' surpluses when there is either pure moral hazard or pure adverse selection. Moreover, she implements the first best in the case of pure adverse selection by offering a payment equal to the agent's effort cost. Under pure moral hazard, the principal offers a fixed wage to types who exert low effort, and a positive bonus to those that exert high effort. Agents do not get positive rents, although the outcome is no longer efficient if agents are risk averse.

Optimal mechanisms are quite different when both adverse selection and moral hazard are simultaneously present. The principal has to leave rents to some agents. As a result, she faces a trade-off between rent extraction and effort distortion (via local incentive-compatibility constraints). Moral hazard introduces new features through binding global incentive compatibility constraints. Some agents who exert low effort get positive bonuses because of their ability to mimic types who exert high effort. Moreover, because even some types at the boundary have binding global incentive compatibility constraints, the optimal mechanism generically features distortion at all points. This result contrasts with the "no distortion at the boundary" result from multidimensional screening when local incentive constraints are sufficient.

Our analysis can be extended in two ways. First, the dual approach used on the optimal taxation model naturally leads to a Rawlsian planner. In order to work with a utilitarianist planner, one needs to consider an ex-ante participation constraint. Second, since the principal's program is not concave and involves a continuum of intermediate constraints, it is unlikely that a solution will in general be attainable without applying a numerical method. We believe that developing such a method could provide additional insights into the properties of optimal mechanisms.

Appendix

A Pure Moral Hazard and Pure Adverse Selection

In this appendix, we study the mechanisms when either effort or conditional probabilities are observable. We refer to the first situation as the *pure adverse selection model*, and to the second one as the *pure moral hazard model.* The main result is that the first best can be implemented under pure adverse selection but not under pure moral hazard (unless having all types exert the lowest effort is first-best efficient or agents are risk neutral). Moreover, the principal's payoff under joint adverse selection and moral hazard is strictly lower than under pure moral hazard. Therefore, adverse selection alone does not entail any payoff loss for the principal, although combining it with moral hazard further reduces the principal's payoff.[34]

A.1 Pure Moral Hazard

There is a continuum of agents in the population with different productivities: $\mathbf{p} \in \bar{\Delta}$ is distributed according to the probability distribution function f with full support. Unlike the model from

[34] Our results contrast with the ones from Caillaud, Guesnerie, and Rey (1992) and Picard (1987), who study a model in which risk-neutral agents have (one-dimensional) private information about their cost of effort. In their setting, the principal can achieve the same utility as in the absence of noise (pure adverse selection). Therefore, the moral hazard dimension does not entail any additional loss for the principal in their model, whereas pure adverse selection does.

Section 2, the principal observes the agents' productivities but still cannot monitor their efforts.

Assume that if the principal could monitor the agents' types, it would be optimal to have a non-empty set of agents exerting high effort:[35]

$$\Delta x > u^{-1}(c_1) - u^{-1}(c_0). \tag{19}$$

Following Grossman and Hart (1983), it is straightforward to characterize the optimal mechanism. In the optimal mechanism, types who exert high effort and have a different conditional probability of success p_1 get different contracts (since the principal extracts the full surplus). All types who exert low effort get the same contract which gives them utility $u^{-1}(c_0)$. Because the principal recommends high effort from types in a neighborhood of $\mathbf{p} = (0,1)$, the high-effort region is non-empty under condition (19).

Since the optimal mechanism in the case of simultaneous moral hazard and adverse selection is also feasible under pure moral hazard (but it is not optimal), the principal obtains a strictly higher profit under pure moral hazard than under simultaneous moral hazard and adverse selection (as long as the high effort region is non-empty – i.e., condition (19) holds). Moreover, as long as the agent is risk averse, the principal's expected payoff is strictly lower in the pure moral hazard model than in the first best model.

A.2 Pure Adverse Selection

This subsection considers the case of pure adverse selection. We assume that the principal is able to monitor the agent's effort but cannot observe his conditional probability of each outcome given effort. In order to stress that the implementability of the first-best under pure adverse selection does not rely on the assumptions of two effort levels or two outcomes, we will consider a framework that generalizes of the model from Section 2.

A risk-neutral principal faces an agent who may be either risk-neutral or risk-averse. The agent exerts effort $e \in \mathbf{E}$, which is *observable* by the principal. The principal also observes output $x \in \mathbf{X}$. The effort and output spaces $\mathbf{E}$ and $\mathbf{X}$ are compact and non-empty subsets of the Euclidean spaces $\mathbb{R}^N$ and $\mathbb{R}^M$. Let $c(e)$ denote the agent's cost of effort e.

Each agent's type is a set of conditional distributions of outcomes given efforts $\{p(.|e) : \mathbf{X} \to \mathbb{R} | e \in \mathbf{E}\}$. This formulation allows for infinite-dimensional types. However, when there are two outcomes and two effort levels, the framework becomes the two-dimensional model of Section 2. More generally, when $\mathbf{E}$ and $\mathbf{X}$ are both finite, a type can be represented by a matrix of conditional probabilities. In this case, types have dimension $(m-1) \times n$, where m is the number of outcomes and n is the number of effort levels. Let $\mathbf{P}$ denote the space of possible types. The principal's beliefs about the agent's private information are represented by the cumulative distribution function F on $\mathbf{P}$.[36]

A direct mechanism $\{w_{\mathbf{p}}(x), e(\mathbf{p}) : \mathbf{p} \in \mathbf{P}, x \in \mathbf{X}\}$ specifies a payment function $w_{\mathbf{p}}(.) : \mathbf{X} \to \mathbb{R}$ and a recommended effort $e(\mathbf{p})$ for each type $\mathbf{p}$. The participation and free disposal constraints (IR) and (FD) are analogous to the ones from Section 2:

$$\int_{\mathbf{X}} u(w_{\boldsymbol{p}}(x))\, p(dx|e) - c(e(\mathbf{p})) \geq 0, \tag{IR}$$

[35] If this condition does not hold, the first-best and the second-best solutions coincide and all agents exert low effort. Moreover, if agents are risk averse, the unique solution would involve paying a constant salary in both states of the world.

[36] Note that we are not imposing MLRP or full support, although the results are still true under these assumptions.

$$x \geq \hat{x} \implies w_{\boldsymbol{p}}(x) \geq w_{\boldsymbol{p}}(\hat{x}), \tag{FD}$$

for all $\mathbf{p}, \hat{\mathbf{p}} \in \mathbf{P}$ and $x, \hat{x} \in \mathbf{X}$, where the first inequality in (FD) represents vector inequality.

The incentive-compatibility constraints require each agent type to take his own contract. However, since effort is observable, the agent cannot exert a different effort than the one recommended by the principal for the type for which the contract is designed. Thus, the incentive-compatibility constraints in the pure adverse selection model are:

$$\int_{\mathbf{X}} u\left(w_{\boldsymbol{p}}(x)\right) p\left(dx|e\right) - c\left(e\left(\mathbf{p}\right)\right) \geq \int_{\mathbf{X}} u\left(w_{\hat{\boldsymbol{p}}}(x)\right) \hat{p}\left(dx|e\right) - c\left(e\left(\hat{\mathbf{p}}\right)\right), \tag{IC AS}$$

for all $\mathbf{p}, \hat{\mathbf{p}} \in \mathbf{P}$.

The principal's expected utility equals expected output minus payments:

$$\int_{\mathbf{P}} \int_{X} \left[x - w_{\mathbf{p}}(x)\right] p\left(dx|e\right) dF\left(\mathbf{p}\right).$$

A mechanism satisfying (IC AS), (IR), and (FD) is called a *feasible mechanism for the pure adverse selection model.* A mechanism is *first-best optimal* if it maximizes the principal's expected utility subject to (IR). A mechanism is *optimal for the pure adverse selection model* if it maximizes the principal's expected utility within the class of feasible mechanisms for the pure adverse selection model. The following proposition establishes that the principal is able to obtain the first-best payoff when effort is observable:

Proposition 11. *Any optimal mechanism for the pure adverse selection model is equivalent to a first-best optimal mechanism.*

Proof. In any first-best optimal mechanism, the participation constraint must bind for almost every type. Therefore, for any first-best optimal mechanism there exists an equivalent mechanism in which the participation constraint binds for all types. Fix one such mechanism and let $e(\mathbf{p})$ denote the effort exerted by type $\mathbf{p}$ in this mechanism.

Consider the mechanism $(\tilde{w}, e)$ where $\tilde{w}_{\boldsymbol{p}}(x) = c\left(e\left(\mathbf{p}\right)\right)$ for all $\mathbf{p}$. This mechanism satisfies (IC AS) and satisfies (IR) with equality. Moreover, since the payments are constant in outcomes, it also satisfies (FD). Therefore, it implements the first best. □

Therefore, we can rank the principal's and agent's payoffs in the models of the pure adverse selection, pure moral hazard and simultaneous moral hazard and adverse selection considered in the text. The principal attains the first-best payoff under pure adverse selection, which is the highest attainable profit. She attains a strictly lower payoff in the case of pure moral hazard as long as the first-best contract does not implement low effort for all types (condition 19) and agents are risk averse, and an even lower payoff in the case of joint moral hazard and adverse selection.

The agent obtains the same payoff under both pure adverse selection and moral hazard (his reservation utility). However, in the model of joint adverse selection and moral hazard, all types with projections above t^* obtain payoffs strictly above their reservation utilities (see Figure 4).

B Proofs

For notational simplicity, we will use the following notation throughout the proofs. Given a mechanism, let Δ_0 and Δ_1 denote the set of types for which the low and high efforts are recommended.

The long but straightforward proofs of Lemmata 1, 2, and 4 can be found in the online appendix.

Proof of Lemma 3

(d) From the incentive-compatibility constraints of types (p_0, p_1) and (p_1, p_1) , we have:

$$w(p_0,p_1) + p_1 b(p_0,p_1) - c_1 \geq w(p_1,p_1) + p_1 b(p_1,p_1) - c_1, \text{ and}$$

$$w(p_1,p_1) + p_1 b(p_1,p_1) - c_0 \geq w(p_0,p_1) + p_1 b(p_0,p_1) - c_0.$$

Combining these conditions yields

$$w(p_1,p_1) + p_1 b(p_1,p_1) = w(p_0,p_1) + p_1 b(p_0,p_1).$$

Therefore,

$$\begin{aligned} U(p_1,p_1) &= w(p_1,p_1) + p_1 b(p_1,p_1) - c_0 \\ &= w(p_0,p_1) + p_1 b(p_0,p_1) - c_1 + \Delta c \\ &= U(p_0,p_1) + \Delta c. \end{aligned}$$

(e) Let $\xi(p_1) \equiv \max_{\hat{p}_1} \{w(\hat{p}_1,\hat{p}_1) + p_1 b(\hat{p}_1,\hat{p}_1) - c_1\}$. From the envelope theorem, we have $\xi'(p_1) = b(p_1,p_1)$ (a. e.). Since $U(p_1,p_1) = \xi(p_1)$, it follows that $U(p_0,p_1) = \xi(p_1) - \Delta c$. From the envelope theorem, $\frac{\partial U}{\partial p_1}(p_0,p_1) = b(p_0,p_1)$ (a. e.). Thus,

$$\frac{\partial U}{\partial p_1}(p_0,p_1) = b(p_0,p_1) = b(p_1,p_1),$$

for all p_0 and almost all p_1.

Proof of Lemma 5

This Lemma will use Properties (a) and (b) from Lemma 2 and Property (d) from Lemma 3. Property (a) implies that $\mathcal{U}$ is differentiable a. e. and $\dot{\mathcal{U}}(p_0) = b(p_0,p_0)$ a.e.. Moreover, property (b) implies that $b(p_0,p_1) = b(p_0,p_0) = \dot{\mathcal{U}}(p_0)$, for almost all (p_0,p_1) such that $p_1 \leq \varphi(p_0)$, whereas property (e) implies that $b(p_0,p_1) = b(p_1,p_1) = \dot{\mathcal{U}}(p_1)$ for almost all (p_0,p_1) such that $p_1 > \varphi(p_0)$. Thus,

$$b(p_0,p_1) = \begin{cases} \dot{\mathcal{U}}(p_0), & \text{if } p_1 \leq \varphi(p_0) \\ \dot{\mathcal{U}}(p_1), & \text{if } p_1 > \varphi(p_0) \end{cases}$$

for almost all $(p_0,p_1) \in \bar{\Delta}$.

Properties (a) and (d) imply that $U(p_0,p_1) = U(p_0,p_0) = \mathcal{U}(p_0)$ if $p_1 \leq \varphi(p_0)$ and $U(p_0,p_1) = U(p_1,p_1) - \Delta c = \mathcal{U}(p_1) - \Delta c$ if $p_1 > \varphi(p_0)$. Thus,

$$U(p_0,p_1) = \begin{cases} \mathcal{U}(p_0), & \text{if } p_1 \leq \varphi(p_0) \\ \mathcal{U}(p_1) - \Delta c, & \text{if } p_1 > \varphi(p_0) \end{cases}$$

for almost all $(p_0,p_1) \in \bar{\Delta}$. Using the definition of U, we obtain:

$$w(p_0,p_1) = \begin{cases} \mathcal{U}(p_0) - p_0 \dot{\mathcal{U}}(p_0) + c_0 \text{ if } p_1 \leq \varphi(p_0) \\ \mathcal{U}(p_1) - p_1 \dot{\mathcal{U}}(p_1) + c_0 \text{ if } p_1 > \varphi(p_0) \end{cases} \quad a.e.$$

Property (d) and the continuity of U imply that

$$\mathcal{U}\left(\varphi\left(p_0\right)\right) = \mathcal{U}\left(p_0\right) + \Delta c \tag{20}$$

for $\varphi(p_0) < 1$ (a.e.). Because the mechanism is nontrivial, we have $\varphi\left(0\right) < 1$. Let $\xi \equiv \min\left\{p_0 : \varphi\left(p_0\right) = 1\right\} > 0$ (which exists because φ is a continuous function). Then, by the continuity of U, we must have

$$\begin{aligned} U\left(\xi, \varphi\left(\xi\right)\right) &= U\left(\varphi\left(\xi\right), \varphi\left(\xi\right)\right) - \Delta c \\ &= U\left(1, 1\right) - \Delta c \\ &= \mathcal{U}\left(1\right) - \Delta c. \end{aligned}$$

Moreover, because $U\left(\xi, \varphi\left(\xi\right)\right) = U\left(\xi, \xi\right) = \mathcal{U}\left(\xi\right)$ (Properties (a) and (d)) and $\mathcal{U}$ is increasing (property (a)), it follows that $\mathcal{U}\left(p_0\right) \geq \mathcal{U}\left(1\right) - \Delta c$ for all $p_0 \geq \xi$. Combining this inequality with (20) yields $\mathcal{U}\left(\varphi\left(p_0\right)\right) = \min\left\{\mathcal{U}\left(p_0\right) + \Delta c; \mathcal{U}\left(1\right)\right\}$.

Proof of Lemma 6

Lemma 5 establishes the mapping between $(\varphi, \mathcal{U})$ and (w, b, e). Given a nontrivial mechanism (w, b, e), one can recover the effort frontier φ and rent projection function $\mathcal{U}$ associated with it (a.e.). Conversely, given an effort frontier and rent projection functions $(\varphi, \mathcal{U})$, one can recover the nontrivial mechanism (w, b, e). Using this mapping, it is straightforward to check that Properties (a)-(e) from Lemmas 2 and 3 are satisfied if and only if the pair $(\varphi, \mathcal{U})$ is such that $\mathcal{U}$ is nondecreasing and convex, $\mathcal{U}\left(0\right) \geq 0$, and equation (6) is satisfied.

Proof of Proposition 1

Let $\mathcal{U}$ and φ denote the rent projection and effort frontier functions associated with a feasible mechanism. Suppose that $\mathcal{U}\left(t\right) > 0$ for all $t > 0$. For each $\epsilon > 0$ sufficiently small, consider the perturbation

$$\mathcal{U}_\epsilon(t) = max\left\{\mathcal{U}(t) - \epsilon, 0\right\}.$$

The mechanism induced by the rent function $\mathcal{U}_\epsilon$ uniformly reduces the rent of all types by ϵ and types in $[0, t_\epsilon^*] \times [0, \varphi_\epsilon^*] \cap \overline{\Delta}$ have zero rent, where t_ϵ^* and φ_ϵ^* are defined as

$$\mathcal{U}(t_\epsilon^*) = \epsilon \text{ and } \mathcal{U}(\varphi_\epsilon^*) - \epsilon = \Delta c.$$

It is immediate that $\mathcal{U}_\epsilon$ satisfies the constraints of program (P') and, therefore, the mechanism it induces is feasible.

Taking the implicit derivative of the last expression with respect to ϵ, we get

$$\frac{d\varphi_\epsilon^*}{d\epsilon} = \frac{1}{\dot{\mathcal{U}}(\varphi_\epsilon^*)} \geq 0.$$

The principal's cost from type t on each perturbed mechanism is

$$G_\epsilon(t) = \left\{ \begin{array}{l} G(\mathcal{U}(t) - \epsilon, \dot{\mathcal{U}}(t), t), \text{ if } t > t_\epsilon^* \\ u^{-1}(c_0), \text{ if } t \leq t_\epsilon^* \end{array} \right. .$$

Therefore, the principal's payoff from each perturbed mechanism is:

$$\Pi_\epsilon := \int_0^1 \left(t\Delta x - G_\epsilon(t)\right) F_0(t, \varphi_\epsilon) dt + \int_{\varphi_\epsilon^*}^1 \left(t\Delta x - G_\epsilon(t)\right) F_1(\varphi^{-1}, t) dt,$$

where we are using the fact that neither the effort frontier changes for all $t \geq t^*_\epsilon$ nor its inverse φ^{-1} for all $t \geq \varphi^*_\epsilon$.

Take the derivative of Π_ϵ with respect to ϵ and evaluate at 0:

$$\left.\frac{d\Pi_\epsilon}{d\epsilon}\right|_{\epsilon=0} = \int_0^1 \frac{\partial G}{\partial \mathcal{U}} F_0(t,\varphi)dt + \int_0^{t^*_0} \left(t\Delta x - G_0\right) f(t,\varphi) \left.\frac{d\varphi^*_\epsilon}{d\epsilon}\right|_{\epsilon=0} dt$$

$$+ \int_{\varphi^*_0}^1 \frac{\partial G}{\partial \mathcal{U}} F_1(\varphi^{-1},t)dt - (\varphi^*_0 \Delta x - G_0(\varphi^*_0))F_1(0,\varphi^*_0) \left.\frac{d\varphi^*_\epsilon}{d\epsilon}\right|_{\epsilon=0},$$

where we are omitting the arguments of G and its derivative. Notice that the first and third terms are strictly positive, the second is zero because $t^*_0 = 0$ and the fourth is zero since $F_1(0,\varphi^*_0) = 0$. Therefore, the derivative of Π_ϵ is positive at 0 which implies that principal strictly prefers the mechanism induced by $\mathcal{U}_\epsilon$ than the one induced by $\mathcal{U}$ for sufficiently small $\epsilon > 0$.

Proof of Lemma 7

Let a feasible mechanism feature $B\left(p_0,p_1\right) > \Delta x$ in a set of types with positive measure. Let $\hat{t} \equiv \inf\left\{t \in [0,1);\ B(t,t) > \Delta x\right\}$. From Proposition 1, $\hat{t} > 0$. Moreover, because $B\left(p_0,p_1\right) > \Delta x$ is true in a set of positive measure, $\hat{t} < 1$.

Let $\mathcal{U}(t)$ be the rent projection function associated with this mechanism, and let $\beta \equiv \lim_{t \nearrow \hat{t}} \dot{\mathcal{U}}(t)$ be the left derivative of $\mathcal{U}$ at $\hat{t}$. Define the following tangent affine continuation of $\mathcal{U}$ from $\hat{t}$ on:

$$\tilde{\mathcal{U}}(t) = \begin{cases} \mathcal{U}(t), \text{ if } t < \hat{t} \\ \mathcal{U}(\hat{t}) + \beta(t - \hat{t}), \text{ if } t \geqslant \hat{t} \end{cases}.$$

From the definition of $\hat{t}$, $\beta < \dot{\mathcal{U}}(1)$ (since $\underset{t\nearrow\hat{t}}{lim} B(t,t) \leq \Delta x$ and$B(t,t) > \Delta x$ for $t > \hat{t}$).

For each $\alpha \in [0,1]$, consider the following family of positive and convex perturbations $\mathcal{U}_\alpha = \mathcal{U} + \alpha(\widetilde{\mathcal{U}} - \mathcal{U})$. Since $\mathcal{U}_\alpha$ is positive, increasing, and convex, it is feasible. Let φ_α be the effort frontier associated with $\mathcal{U}_\alpha$, i.e.,

$$\mathcal{U}_\alpha(\varphi_\alpha) = \mathcal{U}_\alpha(t) + \Delta c, \tag{21}$$

for all $t \leq \xi_\alpha$, where $\xi_\alpha = \mathcal{U}_\alpha^{-1}(\mathcal{U}_\alpha(1) - \Delta c)$. Let us compute the calculus of variation of φ_α and $\mathcal{U}_\alpha$ in α. Fix $\alpha \in (0,1)$. First, let us compute the derivative of $\varphi_\alpha(t)$ with respect to α. We have the following cases:

(i) if $\xi_\alpha < \hat{t}$ or $t < \varphi_\alpha(t) < \hat{t} \leq \xi_\alpha$, then $\varphi_\alpha(t) = \varphi(t)$ which implies that $\frac{\partial \varphi_\alpha(t)}{\partial \alpha} = 0$.

(ii) if $\hat{t} < t \leq \xi_\alpha$, then using the definition of $\mathcal{U}_\alpha$ and the functional equation (21)

$$\mathcal{U}(\varphi_\alpha(t)) + \alpha\left[\mathcal{U}(t) - \mathcal{U}(\varphi_\alpha(t)) + \beta\left(\varphi_\alpha(t) - \hat{t}\right)\right] = \mathcal{U}(t) + \alpha\left[\mathcal{U}(\hat{t}) - \mathcal{U}(t) + \beta(t - \hat{t})\right] + \Delta c.$$

Taking the total derivative with respect to α we get

$$\frac{\partial \varphi_\alpha(t)}{\partial \alpha} = \frac{-(\varphi_\alpha(t) - t)\beta + \mathcal{U}(\varphi_\alpha(t)) - \mathcal{U}(t)}{(1-\alpha)\dot{\mathcal{U}}(\varphi_\alpha(t)) + \alpha\beta} \geq 0,$$

since the convexity of $\mathcal{U}$ implies that $\frac{\mathcal{U}(\varphi_\alpha(t)) - \mathcal{U}(t)}{\varphi_\alpha(t) - t} \geq \dot{\mathcal{U}}(t) \geq \beta$.

(iii) if $t < \hat{t} \leq \varphi_\alpha(t) < \xi_\alpha$, then using the definition of $\mathcal{U}_\alpha$ and the functional equation (21) we get

$$\mathcal{U}(\varphi_\alpha(t)) + \alpha\left[\mathcal{U}(\hat{t}) - \mathcal{U}(\varphi_\alpha(t)) + \beta(\varphi_\alpha(t) - \hat{t})\right] = \mathcal{U}(t) + \Delta c.$$

Taking the total derivative with respect to α we get

$$\frac{\partial \varphi_\alpha(t)}{\partial \alpha} = \frac{-(\varphi_\alpha(t) - \hat{t})\beta + \mathcal{U}(\varphi_\alpha(t)) - \mathcal{U}(\hat{t})}{(1-\alpha)\dot{\mathcal{U}}(\varphi_\alpha(t)) + \alpha\beta} \geq 0$$

by the convexity of $\mathcal{U}$.

(iv) if $\xi_\alpha \leq \varphi_\alpha(t)$, then $\varphi_\alpha(t) = 1$ and the derivative $\frac{\partial \varphi_\alpha(t)}{\partial \alpha} = 0$ when $\xi_\alpha < \varphi_\alpha(t)$. Notice that if $\xi_\alpha = \varphi(t)$, we have a kink such that the right hand derivative is non-negative by the previous case and left hand derivative is zero.

Finally, we compute the derivatives of $\mathcal{U}_\alpha(t)$ and $\dot{\mathcal{U}}_\alpha(t)$ with respect to α:

$$\frac{\partial \mathcal{U}_\alpha(t)}{\partial \alpha} = \begin{cases} 0, \text{ if } t < \hat{t} \\ \mathcal{U}(\hat{t}) - \mathcal{U}(t) + \beta(t - \hat{t}), \text{ if } t > \hat{t} \end{cases}$$

and

$$\frac{\partial \dot{\mathcal{U}}(t,t)}{\partial \alpha} = \begin{cases} 0, \text{ if } t < \hat{t} \\ \beta - \dot{\mathcal{U}}(t), \text{ if } t > \hat{t} \end{cases}.$$

Notice that the convexity of $\mathcal{U}$ implies that $\frac{\mathcal{U}(t) - \mathcal{U}(\hat{t})}{t - \hat{t}} \geq \beta$ and, hence, both derivatives above are non-positive. Let us define the principal's objective function on the family of perturbations by:

$$\Pi_\alpha = \int_0^1 \int_t^{\varphi_\alpha(t)} (t\Delta x - G(\mathcal{U}_\alpha(t), \dot{\mathcal{U}}_\alpha(t), t)) f(t,s) ds dt + \int_0^1 \int_{\varphi_\alpha(t)}^1 (s\Delta x - G(\mathcal{U}_\alpha(s), \dot{\mathcal{U}}_\alpha(s), s)) f(t,s) ds dt.$$

Notice that

$$\frac{\partial G}{\partial \mathcal{U}} = \frac{t}{u'(u^{-1}(\mathcal{U} + (1-t)\dot{\mathcal{U}} + c_0))} + \frac{1-t}{u'(u^{-1}(\mathcal{U} - t\dot{\mathcal{U}} + c_0))} > 0$$

$$\frac{\partial G}{\partial \dot{\mathcal{U}}} = t(1-t)\left(\frac{1}{u'(u^{-1}(\mathcal{U} + (1-t)\dot{\mathcal{U}} + c_0))} - \frac{1}{u'(u^{-1}(\mathcal{U} - t\dot{\mathcal{U}} + c_0))}\right) \geq 0$$

$$\frac{\partial G}{\partial t} = u^{-1}(\mathcal{U} + (1-t)\dot{\mathcal{U}} + c_0) - u^{-1}(\mathcal{U} - t\dot{\mathcal{U}} + c_0) - \dot{\mathcal{U}}\frac{\partial G}{\partial \mathcal{U}}.$$

Hence, for almost all points we have

$$\begin{aligned} \tfrac{d}{dt}\left(t\Delta x - G(\mathcal{U}(t), \dot{\mathcal{U}}(t), t)\right) &= \Delta x - \tfrac{\partial G}{\partial \mathcal{U}}(\mathcal{U}(t), \mathcal{U}\dot{(}t), t)\dot{\mathcal{U}}(t) - \tfrac{\partial G}{\partial \dot{\mathcal{U}}}(\mathcal{U}(t), \dot{\mathcal{U}}(t), t)\tfrac{d}{dt}\dot{\mathcal{U}}(t) \\ &- \tfrac{\partial G}{\partial t}(\mathcal{U}(t), \dot{\mathcal{U}}(t), t) = \Delta x - B(t,t) - \tfrac{\partial G}{\partial \dot{\mathcal{U}}}(\mathcal{U}(t), \dot{\mathcal{U}}(t), t)\tfrac{d}{dt}\dot{\mathcal{U}}(t). \end{aligned}$$

Now let us compute the total derivative of Π_α with respect to α:

$$\frac{\partial \Pi_\alpha}{\partial \alpha} = \begin{array}{c} -\int_0^1 \int_t^{\varphi_\alpha(t)} \left(\frac{\partial G}{\partial \mathcal{U}}(\mathcal{U}_\alpha(t), \dot{\mathcal{U}}_\alpha(t), t)\frac{\partial \mathcal{U}_\alpha(t)}{\partial \alpha} + \frac{\partial G}{\partial \dot{\mathcal{U}}}(\mathcal{U}_\alpha(t), \dot{\mathcal{U}}_\alpha(t), t)\frac{\partial \dot{\mathcal{U}}_\alpha(t)}{\partial \alpha}\right) f(t,s) ds dt \\ -\int_0^1 \int_{\varphi_\alpha(t)}^1 \left(\frac{\partial G}{\partial \mathcal{U}}(\mathcal{U}_\alpha(s), \dot{\mathcal{U}}_\alpha(s,s), s)\frac{\partial \mathcal{U}_\alpha(s)}{\partial \alpha} + \frac{\partial G}{\partial \dot{\mathcal{U}}}(\mathcal{U}_\alpha(s), \dot{\mathcal{U}}_\alpha(s), s)\frac{\partial \dot{\mathcal{U}}_\alpha(s)}{\partial \alpha}\right) f(t,s) ds dt \\ +\int_0^1 (t\Delta x - G(\mathcal{U}_\alpha(t), \dot{\mathcal{U}}_\alpha(t), t))\frac{\partial \varphi_\alpha(t)}{\partial \alpha} f(t, \varphi_\alpha(t)) dt \\ -\int_0^1 (\varphi_\alpha(t)\Delta x - G(\mathcal{U}_\alpha(\varphi_\alpha(t)), \dot{\mathcal{U}}_\alpha(\varphi_\alpha(t)), t))\frac{\partial \varphi_\alpha(t)}{\partial \alpha} f(t, \varphi_\alpha(t)) dt. \end{array}$$

Notice that, from the signs of the partial derivatives, the first two lines are non-negative. In the third and fourth lines both integrals for t between 0 and $\hat{t}$ are zero since $\frac{\partial \varphi_\alpha(t)}{\partial \alpha} = 0$ on $[0, \hat{t}]$. The difference between the integrals of the third and forth in the interval $[\hat{t}, 1]$ is non-negative since $\frac{d}{dt}\left(t\Delta x - G(\mathcal{U}(t),\dot{\mathcal{U}}(t),t)\right) < 0$ (because of the above expression of this total derivative, the signs of the partial derivatives and the hypothesis that $\Delta x < B(t,t)$, for all $t > \hat{t}$), $\frac{\partial \varphi_\alpha(t)}{\partial \alpha} \geq 0$ and $t \leq \varphi_\alpha(t)$. This shows that Π_α is increasing in α, i.e., $\tilde{\mathcal{U}}$ gives a weakly higher payoff for the principal than $\mathcal{U}$. We have to show that the dominance is strict to complete the proof. For this, notice that $\frac{\partial \mathcal{U}_\alpha(t)}{\partial \alpha}$ is strictly negative for all $t \geq \hat{t}$, since $\beta < \dot{\mathcal{U}}(1)$ and $\frac{\partial G}{\partial \mathcal{U}} > 0$. These imply that the integrand of the double integral in the first line is strictly negative on $[\hat{t}, 1]$, which concludes the proof.

Proof of Proposition 2

In this proof, we will denote transfers in monetary (W, B) rather than utility (w, p) units. From Lemma 7, there is no loss of generality in focusing on the space of feasible mechanisms (W, B, e) satisfying $B(\mathbf{p}) \leq \Delta x$ for all $\boldsymbol{p} \in \overline{\Delta}$. For each feasible mechanism in this space, let $\mathcal{U}$ and φ denote the rent projection and effort frontier functions associated with it.

Recall the definition of the contract power:

$$b(t,t) = u\left(W(t,t) + B(t,t)\right) - u\left(W(t,t)\right). \tag{22}$$

Because the marginal utility function is bounded, there exists $K \in \mathbb{R}$ such that $u'(x) \leq K$, for all $x \in \mathbb{R}$. Concavity of the utility function then gives

$$\dot{\mathcal{U}}(t) = u\left(W(t,t) + B(t,t)\right) - u\left(W(t,t)\right) \leq u'\left(W(t,t)\right) B(t,t) \leq K\Delta x.$$

Define the space of admissible contract powers:

$$\dot{\boldsymbol{\mathcal{U}}} \equiv \left\{\dot{\mathcal{U}} : [0,1] \to [0, K\Delta x] \text{ càdlàg and non-decreasing function}\right\},$$

which is non-empty and compact with respect to the weak topology (i.e., this is the weakest topology such that a sequence $(\dot{\mathcal{U}}_n)$ converges to $\dot{\mathcal{U}}$ if and only if $(\dot{\mathcal{U}}_n(t))$ converges to $\dot{\mathcal{U}}(t)$ in all points in which $\dot{\mathcal{U}}$ is continuous). For each $\dot{\mathcal{U}} \in \dot{\boldsymbol{\mathcal{U}}}$, define the increasing and convex function $\mathcal{U}(t) \equiv \int_0^t \dot{\mathcal{U}}(s)ds$. Let $(\dot{\mathcal{U}}_n)$ be a sequence in $\dot{\boldsymbol{\mathcal{U}}}$ weakly converging to $\dot{\mathcal{U}} \in \dot{\boldsymbol{\mathcal{U}}}$. Since $(\mathcal{U}_n, \varphi_n)$ and $(\mathcal{U}, \varphi)$ are continuous functions defined on the compact interval $[0,1]$ then, by Lesbegue's Dominated Convergence Theorem (see Rudin, 1986, pp. 26), the sequence $(\mathcal{U}_n, \varphi_n)$ pointwise (and therefore uniformly) converges to $(\mathcal{U}, \varphi)$. Hence, again by Lesbegue's Dominated Convergence Theorem, the limit of principal's objective function (8) evaluated at $(\dot{\mathcal{U}}_n)$ converges to its value at $\dot{\mathcal{U}}$.

The principal's objective function is uniformly bounded on the space of feasible mechanisms (for example, by the first-best payoff). Consider the supremum of the principal's payoff on the space of feasible mechanisms. Let $(\dot{\mathcal{U}}_n)$ be a sequence in $\dot{\boldsymbol{\mathcal{U}}}$ such that the sequence of the principal's payoff evaluated at each $\dot{\mathcal{U}}_n$ converges to its supremum. Construct the sequence of bonuses $B_n(t,t)$ according to Lemma 5 and equation (22). By Lemma 7, we can restrict to sequences for which the associated sequence of bonuses $(B_n(t,t))$ is uniformly bounded by Δx. By Helly's Selection Theorem (see Billingsley, 1995, pp. 359), there exists a subsequence $(\dot{\mathcal{U}}_{n_k})$ that converges to $\dot{\mathcal{U}} \in \dot{\boldsymbol{\mathcal{U}}}$. By the previous argument, the principal's objective function evaluated at the subsequence converges to the value at $\dot{\mathcal{U}}$. Therefore, $\dot{\mathcal{U}}$ attains the supremum value. Moreover, the associated sequence of bonuses $(B_{n_k}(t,t))$ weakly converges to the limit bonus $B(t,t)$ which must be uniformly bounded by Δx.

Proof of Proposition 3

Suppose $x_L \geq u^{-1}(c_0)$ and suppose there exists an optimal mechanism that excludes set of types with positive measure. Then, the highest payoff these types can obtain by participating in the mechanism is 0. Consider the alternative mechanism that offers a subset of these types the trivial contract: $w = u^{-1}(c_0)$, $b = 0$. For any other type, the payoff from this contract is 0 under low effort and $-\Delta c$ under high effort. Thus, no type can benefit by deviating to this contract. For each of these types, the principal gets $x_L + p_0\Delta x - u^{-1}(c_0)$ (instead of zero) by offering this contract. This is positive for all types (except for types with $p_0 = 0$, which have zero measure) if $x_L \geq u^{-1}(c_0)$. Thus, this new mechanism is also feasible and yields a higher expected payoff, contradicting the optimality of the original mechanism. Thus, whenever participation is first best optimal, there is no exclusion in the second best mechanism.

Reciprocally, suppose $x_L < u^{-1}(c_0)$ and suppose there exists an optimal mechanism with no exclusion a.e.. By Proposition 1, there exist $t^* > 0$ and $\varphi^* > t^*$ such that all types $(p_0, p_1) \leq (t^*, \varphi^*)$ are offered the trivial contract: $w = u^{-1}(c_0)$, $b = 0$. Consider the alternative mechanism that recommends non-participation to all types a set $(p_0, p_1) \leq (\epsilon, \epsilon)$ for

$$\epsilon \equiv \min\left\{t^*;\ \frac{u^{-1}(c_0) - x_L}{\Delta x}\right\} > 0. \tag{23}$$

We claim that this new mechanism is feasible. (FD) and (IR) are immediate. In order to verify (IC), note that because all types in this set are obtaining zero informational rents under the old mechanism, this recommendation is incentive-compatible. Moreover, because any other type that announces a type in this set gets zero utility it is not in their interest to do so. Thus, the new mechanism is (IC). Furthermore, the principal now gets 0 from all types in this set rather than

$$x_L + p_0\Delta x - u^{-1}(c_0) < x_L + \epsilon\Delta x - u^{-1}(c_0) \leq 0,$$

where the last inequality follows from (23). Thus, the principal obtains a strictly higher payoff under this new mechanism, which contradicts the optimality of the original one.

Proofs of Lemma 9 and Theorem 1

The lemma is an immediate consequence of Lemma 13 (presented in the proof of Theorem 2), whereas the theorem follows from Theorem 2 for the risk-neutral case.

Proof of Proposition 4

Let $(\mathcal{U}, \varphi)$ be the rent projection and effort frontier functions associated with a feasible non-trivial mechanism. Let $\mathcal{V}$ be defined as

$$\mathcal{V}(t) = \begin{cases} \max\left\{\mathcal{U}(\varphi^*) + \dot{\mathcal{U}}(\varphi^*)(t - \varphi^*),\ 0\right\}, \text{ if } t < \varphi^* \\ \mathcal{U}(t), \text{ if } t \geq \varphi^* \end{cases}.$$

Note that $\mathcal{U}(t) = \mathcal{V}(t)$ for all $t \geq \varphi^*$ and $\mathcal{U}(\varphi^*) = \Delta c$. Since the rent projection function $\mathcal{V}$ is also feasible, Lemma 9 gives

$$\int_0^{\varphi^*} \left[\frac{(\varphi(t) - t)\Delta x - \Delta c}{\dot{\mathcal{U}}(\varphi(t))} f(t, \varphi(t)) + F_0(t, \varphi(t))\right] (\mathcal{U}(t) - \mathcal{V}(t))dt \leq 0. \tag{24}$$

By Lemma 10, $\frac{(\varphi(t)-t)\Delta x-\Delta c}{\dot{\mathcal{U}}(\varphi(t))}f\left(t,\varphi\left(t\right)\right)\geq 0$, so that the term inside the first bracket is positive. Moreover, the convexity of $\mathcal{U}$ implies that, by construction, $\mathcal{U}(t)\geq\mathcal{V}(t)$, for all $t\in[0,\varphi^{*}]$. Hence, the continuity of $\mathcal{U}$ and $\mathcal{V}$ and condition (24) yield that $\mathcal{U}(t)=\mathcal{V}(t)$, for all $t\in[0,\varphi^{*}]$.

Recall that $\mathcal{U}\left(t\right)=0$ for all $t\leq t^{*}$. Therefore, the power of the contract for all types who get projected to a diagonal type $t<t^{*}$ is $b(t,t)=\dot{\mathcal{U}}\left(t\right)=0$, and, by (IR), they get $w=c_{0}$. Types who get projected to a diagonal type $t\in(t^{*},\varphi^{*})$ get the constant power $b(\varphi^{*},\varphi^{*})=\dot{\mathcal{U}}\left(\varphi^{*}\right)$. From equation (6), we have $\mathcal{U}\left(\varphi^{*}\right)=\Delta c$. Moreover,

$$\mathcal{U}\left(\varphi^{*}\right)=\int_{0}^{\varphi^{*}}\dot{\mathcal{U}}\left(\varphi^{*}\right)dt=\left(\varphi^{*}-t^{*}\right)\dot{\mathcal{U}}\left(\varphi^{*}\right).$$

Combining these two conditions yields

$$\dot{\mathcal{U}}\left(\varphi^{*}\right)=\frac{\Delta c}{\varphi^{*}-t^{*}}\leq\Delta c,$$

where the inequality uses the fact that $\varphi^{*}-t^{*}\leq 1$ (since t^{*} and φ^{*} are both between 0 and 1). Incentive compatibility then requires that the fixed payment for these types, w, be smaller than c_{0} (otherwise types projected to $t<t^{*}$ would prefer to deviate to this contract).

Proof of Lemma 10

Let (w,b,e) be an optimal mechanism and consider a type (p_{0},p_{1}) in the high effort region. By incentive compatibility, deviating to a low effort while reporting the same (true) type must yield a lower payoff:

$$w(p_{0},p_{1})+p_{1}b(p_{0},p_{1})-c_{1}\geq w(p_{0},p_{1})+p_{0}b(p_{0},p_{1})-c_{0}.$$

Subtracting $w(p_{0},p_{1})$ from both sides and rearranging, yields

$$p_{1}\geq p_{0}+\frac{\Delta c}{b(p_{0},p_{1})}\geq p_{0}+\frac{\Delta c}{\Delta x},$$

where the last inequality follows from the fact that bonuses are bounded above by Δx (Lemma 7). Therefore, the type must also exert high effort in the first-best benchmark. Noting that all types (p_{0},p_{1}) with $\varphi\left(p_{0}\right)=1$ exert low effort concludes the proof.

Proof of Lemma 11

The result will be established through a series of claims.

Let (w,b,e) be a feasible mechanism and consider a type (p_{0},p_{1}) in the high effort region: $p_{1}>\varphi\left(p_{0}\right)$. By incentive compatibility, exerting high effort must yield a higher payoff than exerting a low effort while reporting the same type:

$$w(p_{0},p_{1})+p_{1}b(p_{0},p_{1})-c_{1}\geq w(p_{0},p_{1})+p_{0}b(p_{0},p_{1})-c_{0}.$$

Subtracting $w(p_{0},p_{1})$ from both sides and rearranging yields

$$p_{1}\geq p_{0}+\frac{\Delta c}{b(p_{0},p_{1})}\geq p_{0}+\frac{\Delta c}{\Delta x},$$

where the last inequality follows from the fact that bonuses are bounded above by Δx (Lemma 7). In particular, because $(t, \varphi(t) + \epsilon)$ belongs to the high effort region for $\epsilon > 0$, it follows that, after taking the limit when $\epsilon \to 0$,

$$\varphi(t) \geq t + \frac{\Delta c}{\Delta x} \tag{25}$$

for all $t \leq 1 - \frac{\Delta c}{\Delta x}$. This already establishes that the optimal effort region is contained in the first-best effort region. We will show that it is contained in the *interior* of the first-best effort region through the two following lemmata. Since $\varphi(t)$ is strictly increasing in the region where $\varphi(t) < 1$ and constant when $\varphi(t) = 1$, its inverse is always well defined for $t < \inf\left\{\hat{t} : \varphi\left(\hat{t}\right) = 1\right\}$. We adopt the following convention: $\varphi^{-1}(t) \equiv \inf\left\{\hat{t} : \varphi\left(\hat{t}\right) \geq t\right\}$. Thus, $\varphi^{-1} : [\varphi^*, 1] \to [0, 1]$ is a strictly increasing function.

Claim 1. Suppose that $\varphi^{-1}(\bar{t}) = \bar{t} - \frac{\Delta c}{\Delta x}$ for some $\bar{t} \in [0, 1)$. Then, $\varphi^{-1}(t) = t - \frac{\Delta c}{\Delta x}$ and $\dot{\mathcal{U}}(t) = \Delta x$, for all $t \geq \bar{t}$.

Proof. Applying equation (25) to $\hat{t} = \varphi^{-1}(t)$, yields

$$\varphi^{-1}(t) \leq t - \frac{\Delta c}{\Delta x} \tag{26}$$

for all t. For notational simplicity, denote the first-best separating curve as $\varphi_f(t) \equiv t + \frac{\Delta c}{\Delta x}$ for $t \leq 1 - \frac{\Delta c}{\Delta x}$ and note that $\dot{\varphi_f^{-1}}(t) = 1$ for all such t. Then, the inequality above can be written as $\varphi^{-1}(t) \leq \varphi_f^{-1}(t)$.

Since, by Lemma 8,

$$\dot{\varphi^{-1}}(t) = \frac{\dot{\mathcal{U}}(t)}{\dot{\mathcal{U}}\left(\varphi^{-1}(t)\right)} \quad \text{a.e.,} \tag{27}$$

the convexity of $\mathcal{U}$ implies that $\dot{\varphi^{-1}}(t) \geq 1$ a.e. Therefore, $\varphi^{-1}(\bar{t}) = \varphi_f^{-1}(\bar{t})$ and $\dot{\varphi^{-1}}(t) \geq \dot{\varphi_f^{-1}}(t)$ a.e. It then follows that

$$\varphi^{-1}(t) \geq \varphi_f^{-1}(t) = t - \frac{\Delta c}{\Delta x}, \quad \text{for all } t \geq \bar{t}.$$

Combining with inequality (26), yields $\varphi^{-1}(t) = t - \frac{\Delta c}{\Delta x}$ for all $t \geq \bar{t}$.

From equation (6), $\mathcal{U}\left(t - \frac{\Delta c}{\Delta x}\right) = \mathcal{U}(t) - \Delta c$ for all $t \geq \bar{t}$. Moreover, from equation (27), we must have $\dot{\mathcal{U}}(t) = \dot{\mathcal{U}}(\varphi^{-1}(t))$ a.e., which implies that there exist constants constants $\beta > 0$ and $\alpha \in \mathbb{R}$ such that $\mathcal{U}(t) = \beta t + \alpha$ for almost all $t \geq \bar{t}$. Combining these two statements, yields

$$\alpha + \beta\left(t - \frac{\Delta c}{\Delta x}\right) = \alpha + \beta t - \Delta c,$$

for all $t \geq \bar{t}$, which gives $\beta = \Delta x$. □

Claim 2. Suppose that there exists $\bar{t} \in [0, 1]$ such that $\dot{\mathcal{U}}(t)$ is a constant function for all $t \geq \bar{t}$. Then, $\varphi(t) = \min\{\varphi(\bar{t}) - \bar{t} + t,\ 1\}$, for all $t \geq \bar{t}$.

Proof. The result is immediate if $\varphi(\bar{t}) = 1$. Let $\varphi(\bar{t}) < 1$. By Lemma 8, $\dot{\varphi}(t) = \frac{\dot{\mathcal{U}}(t)}{\dot{\mathcal{U}}(\varphi(t))}$ for almost all t such that $\varphi(t) < 1$. Because $\dot{\mathcal{U}}(t) = \dot{\mathcal{U}}(\varphi(t))$ for $t \geq \bar{t}$, it follows that $\dot{\varphi}(t) = 1$ for almost all $t \geq \bar{t}$ such that $\varphi(t) < 1$. By continuity of φ (Lemma 1), $\varphi(t) = \varphi(\bar{t}) - \bar{t} + t$ whenever $\varphi(t) < 1$. For $\varphi(t) = 1$, the result is immediate. □

We are now ready to present the proof of the main result:

Proof of the Lemma. Suppose, by contradiction, that the claim that for is false. Recall that the domain of φ^{-1} is $[\varphi^*, 1]$. Then, by condition (25), there must exist a type $t \in [\varphi^*, 1)$ for which $\varphi^{-1}(t) = t - \frac{\Delta c}{\Delta x}$. Denote the infimum of such types by

$$\bar{t} \equiv \inf\left\{t \in [0,1] : \, \varphi^{-1}(t) = t - \frac{\Delta c}{\Delta x}\right\} \in [\varphi^*, 1).$$

By Claim 1, $\varphi^{-1}(t) = t - \frac{\Delta c}{\Delta x}$ and $\dot{\mathcal{U}}(t) = \Delta x$ for all $t \geq \bar{t}$. There are two cases: $\bar{t} = \varphi^*$ and $\bar{t} > \varphi^*$.

Let $\bar{t} = \varphi^*$. It follows from the arguments in the proof of Proposition 4 that $\mathcal{U}$ cannot have a kink at φ^*. Therefore, it must be the case that $\dot{\mathcal{U}}(t) = \Delta x$ for all $t > t^*$.

Let $\bar{t} > \varphi^*$. We claim that $\mathcal{U}$ must have kink at $\bar{t}$. Otherwise, let $\delta > 0$ be small enough such that $\bar{t} - \delta > \varphi^*$ and $S_1(t, \mathcal{U}) f(\varphi^{-1}, t) = \frac{(t-\varphi^{-1})\Delta x - \Delta c}{\dot{\mathcal{U}}(\varphi^{-1})} f(\varphi^{-1}, t) - F_1(\varphi^{-1}, t) < F_0(t, \varphi)$, for all $t \geq \bar{t} - \delta$. Such $\delta > 0$ exists because φ^{-1} is a continuous function, $F_0(t, \varphi) + F_1(\varphi^{-1}, t)$ is a positive function bounded away from zero, $\dot{\mathcal{U}}(\varphi^{-1}) \geq \Delta c$ and $(t - \varphi^{-1})\Delta x - \Delta c = 0$, for all $t \geq \bar{t}$. In particular, this implies that $\mathcal{S}(t, \mathcal{U}) < 0$, for all $t \geq \bar{t} - \delta$. Define the following feasible rent projection function

$$\mathcal{V}(t) = \begin{cases} max\left\{\mathcal{U}(\bar{t} - \delta) + \dot{\mathcal{U}}(\bar{t} - \delta)(t - \bar{t} + \delta), \mathcal{U}(\bar{t}) + \Delta x(t - \bar{t})\right\}, \text{ if } t \in [\bar{t} - \delta, \bar{t}] \\ \mathcal{U}(t), \text{ if otherwise} \end{cases},$$

which is the substitution of $\mathcal{U}$ by the envelope of tangent lines at points $\bar{t} - \delta$ and $\bar{t}$ of the function $\mathcal{U}$ on the interval $[\bar{t} - \delta, \bar{t}]$. By the definition of $\bar{t}$, $\dot{\mathcal{U}}(t) < \Delta x$,[37] convexity of $\mathcal{U}$ and the hypothesis that $\mathcal{U}$ does not have kink at $\bar{t}$, $\mathcal{V}(t) < \mathcal{U}(t)$ for all $t \in (\bar{t} - \delta, \bar{t})$. Hence,

$$\int_{\bar{t}-\delta}^{\bar{t}} [\mathcal{U}(t) - \mathcal{V}(t)]\, \mathcal{S}(t, \mathcal{U}) dt < 0$$

which contradicts the optimality condition of Lemma 9. Hence, there is kink at $\bar{t}$. On the other hand, Theorem 1 would imply that $\int_{\bar{t}}^{1} \mathcal{S}(t, \mathcal{U}) dt = 0$, which is a contradiction because $\mathcal{S}(t, \mathcal{U}) < 0$ on $[\bar{t}, 1]$.

Proof of Proposition 5

Let $r = \frac{\Delta c}{\Delta x}$. Fix a density $f \in \mathcal{D}$. Since polynomial functions are dense in the space of continuous functions with respect to supremum norm, we can assume without loss of generality that f is a polynomial function. Suppose that the second-best effort frontier is not strictly above the first best-effort frontier. By Lemma 11, partially selling the firm must be optimal. This optimal mechanism is then characterized by the following rent projection function $\mathcal{U}(t) = max\{\Delta x(t - t^*), 0\}$, for some $t^* \in (0, 1)$. We also have that $\varphi^* = t^* + r < 1$. From Theorem 1, the necessary optimality bunching conditions are then given by:

$$\int_{t^*}^{1} F_0(t, t + r) dt + \int_{t^*+r}^{1} F_1(t - r, t) dt = t^* F_1(t^*, t^* + r) - \int_0^{t^*} t f(t, t^* + r) dt, \text{ and}$$
$$\int_{t^*}^{1} (t - t^*) F_0(t, t + r) dt + \int_{t^*+r}^{1} (t - t^*) F_1(t - r, t) dt = r\left(t^* F_1(t^*, t^* + r) - \int_0^{t^*} t f(t, t^* + r) dt\right).$$

[37] Notice that if $\bar{t} = 1$, then $\dot{\mathcal{U}}(t) < \Delta x = \dot{\mathcal{U}}(1)$ for all $t < 1$ and, by our convention, $\underset{t\to 1}{lim}\, \dot{\mathcal{U}}(t) = \dot{\mathcal{U}}(1)$.

Integrating by parts and reorganizing terms, we can rewrite the above equations as

$$\begin{aligned} H_1(t^*, f) &:= \textstyle\int_{t^*}^1 F_0(t, t+r)dt + \int_{t^*+r}^1 F_1(t-r, t)dt - \int_0^{t^*} F_1(t, t^*+r)dt = 0, \text{ and} \\ H_2(t^*, f) &:= \textstyle\int_{t^*}^1 (t-t^*)F_0(t, t+r)dt + \int_{t^*+r}^1 (t-t^*)F_1(t-r, t)dt - r\int_0^{t^*} F_1(t, t^*+r)dt = 0. \end{aligned}$$

Let $H \equiv (H_1, H_2) : [0,1] \times \mathcal{D} \to \mathbb{R}^2$. Then, if partially selling the firm is optimal for f, there must exist $t^* \in (0,1)$ such that $H(t^*, f) = 0$ (i.e., t^* must solve this pair of equations for the density f). In what follows, we will show that this is not possible for generic f. The following claims establish the result:

Claim 1. *The Gateaux differential of the functional $H(t^*, \cdot) : \mathcal{D} \to \mathbb{R}^2$ exists and is onto.*

Notice that $H(t^*, \cdot)$ is a linear mapping from $L_\infty(\bar{\Delta})$ into $\mathbb{R}^2$ and consequently coincides with its differential. Hence, to show that it is onto, it suffices to show that there exist f_1 and f_2 in $L_\infty(\bar{\Delta})$ such that the vectors $\{H(t^*, f_1), H(t^*, f_2)\} \subset \mathbb{R}^2$ are linearly independent. Consider $\alpha > 0$ sufficiently small and define $h_\alpha(t,s) = \mathbf{1}_{[t \le t^* - \alpha]}(t,s)$. Then,

$$F_0^\alpha(t,s) = \int_t^s h_\alpha(t,x)dx = \left\{ \begin{array}{l} s-t, \text{ if } t \le t^* - \alpha \\ 0, \text{ otherwise} \end{array} \right. , \text{ and}$$

$$F_1^\alpha(t,s) = \int_0^t h_\alpha(x,s)dx = \left\{ \begin{array}{l} t, \text{ if } t \le t^* - \alpha \\ t^* - \alpha, \text{ otherwise} \end{array} \right. .$$

Now we can compute:

$$\begin{aligned} H_1(t^*, h_\alpha) &= \textstyle\int_{t^*+r}^1 (t^* - \alpha)dt - \int_0^{t^*-\alpha} tdt - \int_{t^*-\alpha}^{t^*} (t^*-\alpha)dt \\ H_2(t^*, h_\alpha) &= \textstyle\int_{t^*+r}^1 t(t^*-\alpha)dt - (t^*+r)\left(\int_0^{t^*-\alpha} tdt + \int_{t^*-\alpha}^{t^*}(t^*-\alpha)dt\right) \end{aligned} .$$

$H(t^*, h_\alpha)$ as a function of parameter α defines a path in $\mathbb{R}^2$. Taking the derivative, we obtain its tangent field:

$$\frac{d}{d\alpha}H(t^*, h_\alpha) = \begin{pmatrix} t^* + r + \alpha - 1 \\ (t^*+r)(r+2\alpha) - 1 \end{pmatrix} = -\begin{pmatrix} 1 \\ 1 \end{pmatrix} + (t^*+r)\left(\begin{pmatrix} 1 \\ r \end{pmatrix} + \alpha \begin{pmatrix} (t^*+r)^{-1} \\ 2 \end{pmatrix}\right),$$

and the second derivative gives its curvature:

$$\frac{d^2}{d\alpha^2}H(t^*, h_\alpha) = \begin{pmatrix} 1 \\ 2(t^*+r) \end{pmatrix}.$$

Since $H(t^*, h_0) \neq 0$ and $\left\{\frac{d}{d\alpha}H(t^*, h_\alpha), \frac{d^2}{d\alpha^2}H(t^*, h_\alpha)\right\}$ are linearly independent vectors, we have that $\{H(t^*, h_0), H(t^*, h_\alpha)\}$ are also linearly independent independent, for $\alpha > 0$ sufficiently small. Considering a C^∞ function such that

$$h_\alpha(t,s) = \left\{ \begin{array}{l} 1, \text{ if } t \le t^* - \alpha \\ 0, \text{ if } t \ge t^*, \end{array} \right.$$

we that the same properties are true when $\alpha > 0$ is sufficiently small. Therefore, let us consider this smooth function instead.

Claim 2. *For every $\epsilon > 0$ there exists $\tilde{f} \in \mathcal{D}$ such that $||f - \tilde{f}||_\infty < \epsilon$ and the system of equations $H(\cdot, \tilde{f}) = (0,0)$ has no solution. In other words, for every neighborhood of f there might exist a density in the neighborhood for which partially selling the firm is not optimal.*

Since f is a polynomial function, there is only a finite number of solutions of the equation $H(t^*, f) = (0,0)$. Suppose first that there exists only one solution for this equation. From claim 1, let h_1, h_2 smooth functions such that the function $A(t,x,y) = H(t, f + x_1h_1 + x_2h_2)$ has Jacobean with respect to variables (x_1, x_2) at the point $(t^*, 0, 0)$ given by

$$\left[\begin{array}{cc} H(t^*, h_1) & H(t^*, h_2) \end{array}\right] = \left[\begin{array}{cc} e_1' & e_2' \end{array}\right],$$

where $\{e_1, e_2\}$ is the canonical basis of $\mathbb{R}^2$. In particular, it has determinant different from zero. Applying the implicit function theorem, there are small $\delta > 0$ and $\epsilon > 0$ such that $A(t, f + x_1h_1 + x_2h_2) = (a_1, a_2)$ if and only if $x_i = \varphi_i(t, a_1, a_2)$ where $\varphi_i : [t^* - \delta, t^* + \delta] \times [-\epsilon, \epsilon]^2 \to \mathbb{R}^2$ are smooth functions. Notice that $H(t, f) \neq (0,0)$, for all $t \in K := [0,1] - (t^* - \delta, t^* + \delta)$. By continuity of H and the compactness of K, we can find $(x_1, x_2) \notin \{(\varphi_1(t,0,0), \varphi_2(t,0,0)); t \in [t^* - \delta, t^* + \delta]\}$ with a sufficiently small norm such that $H(t, f + x_1h_1 + x_2h_2) \neq (0,0)$, for all $t \in [0,1]$.

Define $\tilde{f} = f + h$, where $h = x_1h_1 + x_2h_2$. Notice that, since h is a bounded function we can choose $|\epsilon| > 0$ sufficiently small such that $f + \epsilon h$ is strictly positive function. Finally, normalizing $\tilde{f}$ we have a density and get the result.

If the number of solutions of the equation $H(t^*, f) = (0,0)$ is greater than one, we proceed as before for every solution. The function A will then be defined on $2n + 1$ variables, where n is the number of solutions.

Claim 3. *The subset of $\mathcal{D}$ for which partially selling the firm is optimal is (relatively) closed. Therefore, the subset of $\mathcal{D}$ for which the second best effort frontier is strictly above the first best effort frontier is open.*

Indeed, take a sequence of densities (f_n) converging to f such that partially selling the firm is the optimal mechanism for f_n for all n. Such a mechanism is completely characterized by a cutoff $t_n^* \in (0,1)$. Take a subsequence such that $(t_{n_k}^*)$ converges to $t^* \in [0,1]$. It is easy to see that $\Pi(\mathcal{U}_{n_k}, f_{n_k})$ converges to $\Pi(\mathcal{U}, f)$, where $\mathcal{U}_n(t) = max\{\Delta x(t - t_n^*), 0\}$ and $\mathcal{U}(t) = max\{\Delta x(t - t^*), 0\}$, where we extend the notation of Π to make explicit the dependence on f. Therefore, $\mathcal{U}$ is the optimal rent projection for f.

Proof of Lemma 12

First note that for $t \geq \xi^*$,

$$\mathcal{S}(t, \mathcal{U}) = \frac{(t - \varphi^{-1})\Delta x - \Delta c}{\dot{\mathcal{U}}(\varphi^{-1})} - R(\varphi^{-1}, t).$$

By the signs of the partial derivative of R, the convexity of $\mathcal{U}$, the fact that the effort distortion is non-negative, and $\dot{\varphi^{-1}} \geq 1$ (a.s.), we have

$$\frac{d}{dt}\mathcal{S}(t, \mathcal{U}) = \frac{d}{dt}\left(\frac{(t - \varphi^{-1})\Delta x - \Delta c}{\dot{\mathcal{U}}(\varphi^{-1})} - R(\varphi^{-1}, t)\right)$$

$$= -\frac{\left(\dot{\varphi^{-1}} - 1\right)\Delta x}{\dot{\mathcal{U}}(\varphi^{-1})} - \left[\frac{\left(t - \varphi^{-1}\right)\Delta x - \Delta c}{\dot{\mathcal{U}}(\varphi^{-1})}\right]\frac{\ddot{\mathcal{U}}(\varphi^{-1})}{\dot{\mathcal{U}}(\varphi^{-1})}\dot{\varphi^{-1}} - R_1(\varphi^{-1}, t)\dot{\varphi^{-1}} - R_2(\varphi^{-1}, t) < 0$$

for almost all $t \geq \xi^*$ (where $R_1(t,s) \equiv \frac{\partial R}{\partial t}(t,s)$ and $R_2(t,s) \equiv \frac{\partial R}{\partial s}(t,s)$). Therefore, $\mathcal{S}(t,\mathcal{U})$ is a strictly increasing function of t.

Since $\mathcal{S}(t,\mathcal{U})$ is strictly decreasing, there are three possible cases: (i) $\mathcal{S}(t,\mathcal{U}) < 0$ for all $t \in [\xi^*, 1]$; (ii) $\mathcal{S}(t,\mathcal{U}) > 0$ for all $t \in [\xi^*, 1]$; and (iii) there exists $\bar{t} \in [\xi^*, 1]$ such that $\mathcal{S}(t,\mathcal{U}) \lesseqgtr 0$ if and only if $t \gtreqless \bar{t}$. Let us first suppose that $\varphi^* \leq \xi^*$. Consequently, we have the following possible cases:

(i) $\mathcal{S}(t,\mathcal{U}) < 0$ for all $t \in [\xi^*, 1]$.

Consider the following convex and piecewise linear function:

$$\mathcal{V}(t) = \begin{cases} \mathcal{U}(t), \text{ if } t \leq \xi^* \\ max\left\{\mathcal{U}(\xi^*) + \dot{\mathcal{U}}(\xi^*)(t-\xi^*), \mathcal{U}(1) + \dot{\mathcal{U}}(1)(t-1)\right\}, \text{ if } t > \xi^* \end{cases}$$

which is feasible. Notice that $\mathcal{U}(t) = \mathcal{V}(t)$ for $t \leq \xi^*$. Since $\mathcal{U}$ is optimal, by Lemma 9,

$$\int_{\xi^*}^{1} [\mathcal{U}(t) - \mathcal{V}(t)]\, \mathcal{S}(t,\mathcal{U}) dt \geq 0.$$

Because $\mathcal{S}(.,\mathcal{U})$, $\mathcal{U}$, and $\mathcal{V}$ are continuous functions and $\mathcal{U}(t) \geq \mathcal{V}(t)$ for all $t \in [\xi^*, 1]$, we must have that $\mathcal{U}(t) = \mathcal{V}(t)$, for all $t \in [\varphi^*, 1]$.

(ii) $\mathcal{S}(t,\mathcal{U}) > 0$ for all $t \in [\xi^*, 1]$.

Consider the following convex and piecewise linear function:

$$\mathcal{V}(t) = \begin{cases} \mathcal{U}(t), \text{ if } t \leq \xi^* \\ max\left\{\mathcal{U}(\varphi^*) + \dot{\mathcal{U}}(\varphi^*)(t-\varphi^*);\ \mathcal{U}(1) + \frac{\mathcal{U}(1)-\mathcal{U}(\varphi^*)}{1-\varphi^*}(t-1)\right\}, \text{ if } t > \xi^* \end{cases},$$

which is feasible. As in case (i), $\mathcal{V}$ coincides with $\mathcal{U}$ for $t \leq \xi^*$. Using Lemma 9, we obtain

$$\int_{\xi^*}^{1} [\mathcal{U}(t) - \mathcal{V}(t)]\, \mathcal{S}(t,\mathcal{U}) dt \geq 0.$$

Again, because $\mathcal{S}(.,\mathcal{U})$, $\mathcal{U}$, and $\mathcal{V}$ are continuous functions and $\mathcal{U}(t) \leq \mathcal{V}(t)$ for all $t \in [\xi^*, 1]$, we must have that $\mathcal{U}(t) = \mathcal{V}(t)$, for all $t \in [\xi^*, 1]$.

(iii) there exists $\bar{t} \in [\xi^*, 1]$ such that $\mathcal{S}(t,\mathcal{U}) \lesseqgtr 0$ if and only if $t \gtreqless \bar{t}$.

Consider the following convex and piecewise linear function:

$$\mathcal{V}(t) = \begin{cases} \mathcal{U}(t), \text{ if } t \leq \xi^* \\ \max\left\{\mathcal{U}(\xi^*) + \frac{\mathcal{U}(\bar{t})-\mathcal{U}(\xi^*)}{\bar{t}-\xi^*}(t-\xi^*);\ \mathcal{U}(1) + \dot{\mathcal{U}}(1)(t-1)\right\}, \text{ if } t > \xi^* \end{cases},$$

which is feasible. Since $\mathcal{U}(t) = \mathcal{V}(t)$ on $t \leq \xi^*$, Lemma 9 implies

$$\int_{\xi^*}^{1} [\mathcal{U}(t) - \mathcal{V}(t)]\, \mathcal{S}(t,\mathcal{U}) dt \geq 0.$$

Because $\mathcal{U}(t) \leq \mathcal{V}(t)$ on $[\xi^*, \bar{t}]$ and $\mathcal{U}(t) \geq \mathcal{V}(t)$ on $[\bar{t}, 1]$, and $\mathcal{S}(t,\mathcal{U})$, $\mathcal{U}$ and $\mathcal{V}$ are continuous functions, it follows that $\mathcal{U}(t) = \mathcal{V}(t)$, for all $t \in [\xi^*, 1]$. We conclude that $\mathcal{U}$ must have at most two pieces on the interval $[\xi^*, 1]$.

Now suppose that $\varphi^* > \xi^*$. By Proposition 4, $\mathcal{U}$ is an affine function on the interval $[\xi^*, \varphi^*]$ and $\mathcal{U}$ does not have a kink at φ^*. Proceeding as in the case where $\varphi^* \leq \xi^*$, but only substituting ξ^* for φ^* in the expression above, we also conclude that $\mathcal{U}$ must be piecewise linear with at most two pieces on the interval $[\xi^*, 1]$.

Proof of Proposition 6

(i) We claim that $\frac{\Delta x}{\Delta c} \leq 2$ implies that $\varphi^* \geq \xi^*$. Because $\mathcal{U}$ is increasing, it is enough to show that $\mathcal{U}(\varphi^*) \geq \mathcal{U}(\xi^*)$. By condition (6), $\mathcal{U}(\varphi^*) = \Delta c$ and $\mathcal{U}(\xi^*) = \mathcal{U}(1) - \Delta c$, so that

$$\mathcal{U}(\varphi^*) \geq \mathcal{U}(\xi^*) \iff \mathcal{U}(1) \leq 2\Delta c.$$

Because in any optimal mechanism we have $\mathcal{U}(0) = 0$ and, by Lemma 7, $\dot{\mathcal{U}}(t) \in [0, \Delta x]$ for all t, we have

$$\mathcal{U}(1) \leq \Delta x \leq 2\Delta c,$$

where the last inequality follows from the assumption that $\Delta x \leq 2\Delta c$. Part (ii) is an immediate consequence of (i) and equation (4).

Proof of Corollary 1

By Proposition 4, $\mathcal{U}$ is piecewise linear in $[0, \varphi^*]$. Since the uniform distribution satisfies increasing rents, it is also piecewise linear in $[\xi^*, 1]$ (Lemma 12). It remains to be shown that $\mathcal{U}$ is piecewise linear on (φ^*, ξ^*).

We claim that $\varphi(\varphi^*) \geq \xi^*$. Because $\mathcal{U}$ is increasing, it suffices to show that $\mathcal{U}(\xi^*) \leq \mathcal{U}(\varphi(\varphi^*))$. By equation (6),

$$\mathcal{U}(\varphi(\varphi^*)) = \mathcal{U}(\varphi^*) + \Delta c = 2\Delta c.$$

Since $\mathcal{U}(\xi^*) = \mathcal{U}(1) - \Delta c$, we need to show that $\mathcal{U}(1) \leq 3\Delta c$. Because $\mathcal{U}(0) = 0$, $\dot{\mathcal{U}}(t) \in [0, \Delta x]$, we have $\mathcal{U}(1) \leq \Delta x$. Then, the result follows from $\Delta x \leq 3\Delta c$.

Since $\mathcal{U}$ is piecewise linear on $[0, \varphi^*] \cup [\xi^*, 1]$ and the image of $[\varphi^*, \xi^*]$ by φ^{-1} and by φ are contained in $[0, \varphi^*]$ and $[\xi^*, 1]$, respectively, we can define a partition of the interval $[\varphi^*, \xi^*]$ such that the functions $\dot{\mathcal{U}}(\varphi^{-1})$ and $\dot{\mathcal{U}}(\varphi)$ are constant in each interval of the partition. Let $[a, b] \subset [\varphi^*, \xi^*]$ be an element of the partition and let $\dot{\mathcal{U}}(\varphi^{-1}(t)) = \beta_0$ and $\dot{\mathcal{U}}(\varphi(t)) = \beta_1$ for all $t \in [a, b]$. Then,

$$\mathcal{S}(t, \mathcal{U}) = -\frac{(\varphi - t)\Delta x - \Delta c}{\beta_1} + \frac{(t - \varphi^{-1})\Delta x - \Delta c}{\beta_0} - \varphi + t - \varphi^{-1},$$

where we have substituted the expressions for F_0 and F_1 under the uniform distribution. Differentiating with respect to t, yields:

$$-\frac{(\dot{\varphi} - 1)\Delta x}{\beta_1} + \frac{(1 - \dot{\varphi^{-1}})\Delta x}{\beta_0} - \dot{\varphi} + 1 - \dot{\varphi^{-1}}.$$

Substituting $\dot{\varphi^{-1}}(t) = \frac{\dot{\mathcal{U}}(t)}{\beta_0}$ and $\dot{\varphi}(t) = \frac{\dot{\mathcal{U}}(t)}{\beta_1}$, yields

$$-\left(\frac{\dot{\mathcal{U}}(t) - \beta_1}{\beta_1^2} + \frac{\dot{\mathcal{U}}(t) - \beta_0}{\beta_0}\right)\Delta x - \frac{\dot{\mathcal{U}}(t)}{\beta_1} + 1 - \frac{\dot{\mathcal{U}}(t)}{\beta_0}.$$

Since $\dot{\mathcal{U}}$ is a non-decreasing function, this expression is a non-increasing function on $[a, b]$. Thus, $\mathcal{S}(t, \mathcal{U})$ is an increasing function of t on $[a, b]$. Then, by the same procedure as in the proof of Lemma 12, it follows that $\mathcal{U}$ is piecewise linear on $[a, b]$. Since the partition is finite, we have that $\mathcal{U}$ is piecewise linear on $[0, 1]$.

Proof of Proposition 7

We have that

$$F_1(t,s) = \frac{\int_0^t f(x,s)dx}{f(t,s)} \geq t$$

since, by hypothesis, $f(x,s) \geq f(t,s)$, for all $x \in [0,t]$. We already know that the vertical effect is always non-positive, i.e., $S_0(t,\mathcal{U}) \leq 0$. Let us investigate the horizontal effect. For any $t > \varphi^*$, we have

$$S_1(t,\mathcal{U}) = \frac{(t-\varphi^{-1}) - \Delta c}{\dot{\mathcal{U}}(\varphi^{-1})} - \frac{F_1(\varphi^{-1},t)}{f(\varphi^{-1},t)} \leq (t-\varphi^{-1})\frac{\Delta x}{\Delta c} - 1 - \varphi^{-1}$$

since $\dot{\mathcal{U}}(\varphi^{-1}) \geq \Delta c$. The right hand side is less than or equal to zero if and only if

$$\frac{\Delta x}{\Delta c}t - 1 \leq \left(1 + \frac{\Delta x}{\Delta c}\right)\varphi^{-1}.$$

This condition is implied by the following inequality

$$\frac{\Delta x}{\Delta c} - 1 \leq \left(1 + \frac{\Delta x}{\Delta c}\right)\underline{p},$$

which is equivalent to the condition in the statement of the proposition. Therefore, given the optimal rent projection function $\mathcal{U}$, let $\mathcal{V}(t) = max\{0, \dot{\mathcal{U}}(\varphi^*)(t-\varphi^*) + \Delta c\}$, where $\mathcal{U}(\varphi^*) = \Delta c$. By Lemma 9, we must have that

$$\int_0^1 [\mathcal{U}(t) - \mathcal{V}(t)]\mathcal{S}(t,\mathcal{U})dt \geq 0.$$

Since $\mathcal{V}(t) \leq \mathcal{U}(t)$, we must have that $\mathcal{U}(t) = \mathcal{V}(t)$, for all $t \in [0,1]$. Then, the result immediately follows.

Proof of Theorem 2

The following lemma establishes that any optimal mechanism must maximize the expected virtual surplus among the class of feasible mechanisms:

Lemma 13. *Let $\mathcal{U}$ be the rent projection associated with an optimal mechanism. Then, for any feasible $\mathcal{V} : [0,1] \to \mathbb{R}$,*

$$\int_0^1 [\mathcal{U}(t) - \mathcal{V}(t)]\,\mathcal{S}(t,\mathcal{U})\,dt - \int_0^1 \left[\dot{\mathcal{U}}(t) - \dot{\mathcal{V}}(t)\right]\mathcal{C}(t,\mathcal{U})dt + [\mathcal{U}(\varphi^*) - \mathcal{V}(\varphi^*)]\,S^*(\mathcal{U}) \geq 0.$$

Proof of the lemma.

Let $h(t) \equiv \mathcal{V}(t) - \mathcal{U}(t)$ and consider the perturbation $\mathcal{U}_\epsilon \equiv \mathcal{U} + \epsilon h$. For each $\epsilon \in (0,1)$, we have that

$$\mathcal{U}(t) + \epsilon h(t) = (1-\epsilon)\mathcal{U}(t) + \epsilon\mathcal{V}(t)$$

is also feasible. Let Π denote the principal's payoff from the rent projection function $\mathcal{U}$:

$$\Pi(\mathcal{U}) = \int_0^1 (t\Delta x - G(\mathcal{U},\dot{\mathcal{U}},t))F_0(t,\varphi)dt + \int_{\varphi^*}^1 (t\Delta x - G(\mathcal{U},\dot{\mathcal{U}},t))F_1(\varphi^{-1},t)dt,$$

where φ is obtained from equation (6). Because $\mathcal{U}$ is optimal and $\mathcal{U}+\epsilon h$ is feasible, we must have

$$\Pi\left(\mathcal{U}+\epsilon h\right) \leq \Pi\left(\mathcal{U}\right)$$

for all $\epsilon \in (0,1)$. Dividing by ϵ and taking the limit, we obtain the one-sided Gâteaux derivative of Π in the direction h:

$$\lim_{\epsilon \searrow 0} \frac{\Pi\left(\mathcal{U}+\epsilon h\right) - \Pi\left(\mathcal{U}\right)}{\epsilon} \leq 0.$$

By equation (6), the effort frontier associated with $\mathcal{U}+\epsilon h$, φ_ϵ, is defined as the solution to the following functional equation:

$$\mathcal{U}\left(\varphi_\epsilon(t)\right) + \epsilon h(\varphi_\epsilon(t)) = \mathcal{U}\left(t\right) + \epsilon h\left(t\right) + \Delta c$$

for all $t \in [0, \xi^*_\epsilon]$, where ξ^*_ϵ solves $\mathcal{U}(\xi^*_\epsilon) + \epsilon h(\xi^*_\epsilon) = \mathcal{U}(1) + \epsilon h(1) - \Delta c$. Taking the total derivative of this expression with respect to ϵ and evaluating at 0, we obtain

$$\left.\frac{\partial \varphi_\epsilon}{\partial \epsilon}\right|_{\epsilon=0} = \frac{h\left(t\right) - h(\varphi)}{\dot{\mathcal{U}}\left(\varphi\right)},$$

for all $t \leq \xi^*$.

Analogously, its inverse, φ_ϵ^{-1}, satisfies an analogous functional equation:

$$\mathcal{U}\left(\varphi_\epsilon^{-1}(t)\right) + \epsilon h(\varphi_\epsilon^{-1}(t)) = \mathcal{U}\left(t\right) + \epsilon h\left(t\right) - \Delta c,$$

for all $t \in [\varphi^*_\epsilon, 1]$, where $\mathcal{U}(\varphi^*_\epsilon) + \epsilon h(\varphi^*_\epsilon) = \Delta c$. Again, taking the total derivative of this expression with respect to ϵ and evaluating at 0, we get:

$$\left.\frac{\partial \varphi_\epsilon^{-1}}{\partial \epsilon}\right|_{\epsilon=0} = \frac{h\left(t\right) - h(\varphi^{-1})}{\dot{\mathcal{U}}\left(\varphi^{-1}\right)},$$

for all $t \geq \varphi^*$. Applying the same procedure with respect to φ^*_ϵ yields

$$\left.\frac{\partial \varphi^*_\epsilon}{\partial \epsilon}\right|_{\epsilon=0} = -\frac{h(\varphi^*)}{\dot{\mathcal{U}}(\varphi^*)}.$$

Then,

$$\left.\frac{\partial \varphi_\epsilon}{\partial \epsilon}\right|_{\epsilon=0} = \frac{h\left(t\right) - h(\varphi)}{\dot{\mathcal{U}}\left(\varphi\right)}, \quad \left.\frac{\partial \varphi_\epsilon^{-1}}{\partial \epsilon}\right|_{\epsilon=0} = \frac{h\left(t\right) - h(\varphi^{-1})}{\dot{\mathcal{U}}\left(\varphi^{-1}\right)}, \quad \left.\frac{\partial \varphi^*_\epsilon}{\partial \epsilon}\right|_{\epsilon=0} = -\frac{h(\varphi^*)}{\dot{\mathcal{U}}(\varphi^*)}, \text{ and} \tag{28}$$

$$\dot{\mathcal{U}}_\epsilon = \dot{\mathcal{U}}\left(t\right) + \epsilon \dot{h}\left(t\right). \tag{29}$$

With some abuse of notation, we let $\Pi_\epsilon \equiv \Pi(\mathcal{U}+\epsilon h)$ denote the principal's profit under $\mathcal{U}_\epsilon$. Therefore,

$$\left.\frac{d\Pi_\epsilon}{d\epsilon}\right|_{\epsilon=0} = \lim_{\epsilon \searrow 0} \frac{\Pi(\mathcal{U}+\epsilon h) - \Pi(\mathcal{U})}{\epsilon}.$$

Using conditions (28), we obtain

$$\begin{aligned}
\left.\tfrac{d\Pi_\epsilon}{d\epsilon}\right|_{\epsilon=0} = \quad & -\int_0^1 \left\{\tfrac{\partial G}{\partial \mathcal{U}} h(t) + \tfrac{\partial G}{\partial \dot{\mathcal{U}}} \dot{h}(t)\right\} F_0(t,\varphi)dt \\
& -\int_{\varphi^*}^1 \left\{\tfrac{\partial G}{\partial \mathcal{U}} h(t) + \tfrac{\partial G}{\partial \dot{\mathcal{U}}} \dot{h}(t)\right\} F_1(\varphi^{-1},t)dt \\
& +\int_0^{\xi^*} (t\Delta x - G) \tfrac{h(t)-h(\varphi)}{\dot{\mathcal{U}}(\varphi)} f(t,\varphi)dt \\
& +\int_{\varphi^*}^1 (t\Delta x - G) \tfrac{h(t)-h(\varphi^{-1})}{\dot{\mathcal{U}}(\varphi^{-1})} f(\varphi^{-1},t)dt \\
& +(\varphi^*\Delta x - G(\varphi^*) \tfrac{h(\varphi^*)}{\dot{\mathcal{U}}(\varphi^*)} F_1(t^*,\varphi^*).
\end{aligned}$$

Performing a change of variables on the integrals on lines two and three, we obtain:

$$\begin{gathered}\textstyle\int_0^{\xi^*}(t\Delta x-G)\frac{h(\varphi)}{\dot{\mathcal{U}}(\varphi)}f(t,\varphi)dt=\int_0^{t^*}t\Delta x\frac{h(\varphi^*)}{\dot{\mathcal{U}}(\varphi^*)}f(t,\varphi^*)dt+\int_{\varphi^*}^1\left(\varphi^{-1}\Delta x-G\left(\varphi^{-1}\right)\right)\frac{h(t)}{\dot{\mathcal{U}}(t)}f(\varphi^{-1},t)\dot{\varphi^{-1}}(t)dt\\ \textstyle\int_{\varphi^*}^1(t\Delta x-G)\frac{h(\varphi^{-1})}{\dot{\mathcal{U}}(\varphi^{-1})}f(\varphi^{-1},t)dt=\int_0^{\xi^*}\left(\varphi\Delta x-G\left(\varphi\right)\right)\frac{h(t)}{\dot{\mathcal{U}}(t)}f(t,\varphi)\dot{\varphi}(t)dt.\end{gathered}$$

Using condition (29) yields:

$$\begin{aligned}\tfrac{d}{d\epsilon}\Pi_\epsilon\Big|_{\epsilon=0}= \quad & -\textstyle\int_0^1\left(\frac{\partial G}{\partial\mathcal{U}}F_0(t,\varphi)h(t)+\frac{\partial G}{\partial\dot{\mathcal{U}}}F_0(t,\varphi)\dot{h}(t)\right)dt\\ & -\textstyle\int_{\varphi^*}^1\left(\frac{\partial G}{\partial\mathcal{U}}F_1(\varphi^{-1},t)h(t)+\frac{\partial G}{\partial\dot{\mathcal{U}}}F_1(\varphi^{-1},t)\dot{h}(t)\right)dt\\ & -\textstyle\int_0^{\xi^*}\frac{(\varphi-t)\Delta x-(G(\varphi)-G)}{\dot{\mathcal{U}}(\varphi)}f(t,\varphi)h(t)dt\\ & +\textstyle\int_{\varphi^*}^1\frac{(t-\varphi^{-1})\Delta x-(G-G(\varphi^{-1}))}{\dot{\mathcal{U}}(\varphi^{-1})}f(\varphi^{-1},t)h(t)dt\\ & +\left(-\textstyle\int_0^{t^*}t\Delta x f(t,\varphi^*)dt+(\varphi^*\Delta x-G(\varphi^*))F_1(t^*,\varphi^*)\right)\frac{h(\varphi^*)}{\dot{\mathcal{U}}(\varphi^*)}.\end{aligned}$$

This establishes the result. Notice that, in the case of Lemma 9, substituting $\mathcal{U}-\mathcal{U}(\varphi^{-1})=\Delta c$ and $\mathcal{U}(\varphi)-\mathcal{U}=\Delta c$ into the equation above, gives the result claimed in the statement of the lemma.

Proof of the theorem.

(1) Notice that $\mathcal{S}(t,\mathcal{U})$ is an integrable function on $[a,b]$ (in the Lesbegue sense). Let $h:[0,1]\to\mathbb{R}$ be any function twice continuously differentiable function such that $h(t)=0$ for all $t\notin(a,b)$. Since $\mathcal{U}$ is strongly convex on $[a,b]$, $\mathcal{U}+\epsilon h$ is a strongly convex function if $|\epsilon|$ is sufficiently small. Performing the variational calculus (given by the previous theorem) for such feasible direction, we get

$$\int_a^b\mathcal{S}(t,\mathcal{U})h(t)dt-\int_a^b\mathcal{C}(t,\mathcal{U})\dot{h}(t)dt=0.$$

Notice that we are implicitly taking positive and negative value of ϵ to conclude that this integral is both positive and negative. Integrating by parts, we get

$$\int_a^b\left[\int_0^t\mathcal{S}(x,\mathcal{U})dx+\mathcal{C}(t,\mathcal{U})\right]\dot{h}(t)dt=0.$$

Since the function inside the brackets of the above integral is càdlàg, $\dot{h}$ is a generic continuous function. The auxiliary lemma below implies that

$$\int_0^t\mathcal{S}(x,\mathcal{U})dx+\mathcal{C}(t,\mathcal{U})$$

is constant on $[a,b]$. Since this function is a.e. differentiable (since $\dot{\mathcal{U}}$ is a.e. differentiable), we have that

$$\mathcal{S}(t,\mathcal{U})+\frac{d}{dt}\left\{\mathcal{C}(t,\mathcal{U})\right\}=0,$$

a.e. on $[a,b]$.

(2) We have two possible feasible perturbations that we can do with the rent projection function on the interval $[a, b]$: translations and rotations. Let us start with the translations and consider the case $\varphi^* \notin [a, b]$. We have that there exist $\beta > 0$ and $\alpha \in \mathbb{R}$ such that $\mathcal{U}(t) = \beta t + \alpha$, for all $t \in [a, b]$. Given $\delta > 0$ sufficiently small, define the following rent projection function:

$$\mathcal{V}_\delta(t) = max\{\mathcal{U}(t), \beta t + \alpha + \delta\}$$

which is obviously feasible. Applying Lemma 13, we get

$$\int_{a_\delta}^{b_\delta} \mathcal{S}(t,\mathcal{U}) h_\delta(t) dt - \int_{a_\delta}^{b_\delta} \mathcal{C}(t,\mathcal{U}) \dot{h}_\delta(t) dt \geq 0,$$

where $h_\delta = \mathcal{U} - \mathcal{V}_\delta$, a_δ and b_δ are the only two solutions of the equation $\mathcal{V}_\delta(t) - \mathcal{U}(t) = 0$ (which follows from the convexity of $\mathcal{U}$ and the maximality property of $[a, b]$ for sufficiently small $\delta > 0$). Let $a'_\delta \geq a_\delta$ and $b'_\delta \leq b_\delta$ be the only two solutions of the equation $\mathcal{V}_\delta(t) - \mathcal{U}(t) = \delta$ (again from convexity of $\mathcal{U}$ and the maximality of $[a, b]$ for sufficiently small $\delta > 0$). It is easy to check that $\underset{\delta\to 0}{lim}\, a_\delta = \underset{\delta\to 0}{lim}\, a'_\delta = a$ and $\underset{\delta\to 0}{lim}\, b_\delta = \underset{\delta\to 0}{lim}\, b'_\delta = b$. Therefore, since $h_\delta(t) = -\delta$ for all $t \in [a_\delta, b_\delta]$,

$$\begin{array}{l} \frac{1}{\delta}\int_{a_\delta}^{a'_\delta} \mathcal{S}(t,\mathcal{U}) h_\delta(t) dt - \frac{1}{\delta}\int_{a_\delta}^{a'_\delta} \mathcal{C}(t,\mathcal{U}) \dot{\mathcal{U}}(t) dt + \\ \frac{1}{\delta}\int_{b'_\delta}^{b_\delta} \mathcal{S}(t,\mathcal{U}) h_\delta(t) dt - \frac{1}{\delta}\int_{b'_\delta}^{b_\delta} \mathcal{C}(t,\mathcal{U}) \dot{\mathcal{U}}(t) dt + \int_{a'_\delta}^{b'_\delta} \mathcal{S}(t,\mathcal{U}) dt \geq 0. \end{array}$$

Notice that

$$\left| \frac{1}{\delta} \int_{a_\delta}^{a'_\delta} \mathcal{S}(t,\mathcal{U}) h_\delta(t) dt \right| \leq \frac{a'_\delta - a_\delta}{\delta} sup\left\{ \left| \mathcal{S}(t,\mathcal{U}) h_\delta(t) \right| ; t \in [a_\delta, a'_\delta] \right\} \leq (a'_\delta - a_\delta)\, sup\left\{ \left| \mathcal{S}(t,\mathcal{U}) \right| ; t \in [a_\delta, a'_\delta] \right\}$$

since $|h_\delta(t)| \leq \delta$, for all t. Hence, when $\delta \to 0$, the value on left hand side of the above inequality goes to 0. An analogous proof shows that the third term in the above expression goes to 0 when $\delta \to 0$.

Hence, we have that

$$\int_a^b \mathcal{S}(t,\mathcal{U}) dt = \underset{\delta\to 0}{lim} \int_{a'_\delta}^{b'_\delta} \mathcal{S}(t,\mathcal{U}) dt \geq \underset{\delta\to 0}{lim\,inf}\, \frac{1}{\delta} \left(\int_{a_\delta}^{a'_\delta} \mathcal{C}(t,\mathcal{U}) \dot{\mathcal{U}}(t) dt + \int_{b'_\delta}^{b_\delta} \mathcal{C}(t,\mathcal{U}) \dot{\mathcal{U}}(t) dt \right) \geq 0.$$

Therefore, the first result holds.

Suppose that $\mathcal{U}$ has kink at a and at b. Given $\delta > 0$ sufficiently small, define the following rent projection function:

$$\mathcal{V}_\delta(t) = \begin{cases} max\{(\beta - \delta)(t - \mathcal{U}(a)) + \mathcal{U}(a), \beta t + \alpha - \delta, (\beta + \delta)(t - \mathcal{U}(b)) + \mathcal{U}(b), & \text{if } t \in [a, b] \\ \mathcal{U}(t), \text{ if otherwise} \end{cases}$$

which is obviously feasible for δ sufficiently small. Define a_δ and b_δ the solutions of $(\beta - \delta)(t - \mathcal{U}(a)) + \mathcal{U}(a) = \beta t + \alpha - \delta$ and $\beta t + \alpha - \delta = (\beta + \delta)(t - \mathcal{U}(b)) + \mathcal{U}(b)$, respectively. It is easy to see that $\underset{\delta\to 0}{lim}\, a_\delta = a$ and $\underset{\delta\to 0}{lim}\, b_\delta = b$. Therefore, since $h_\delta(t) = \delta$ for all $t \in [a_\delta, b_\delta]$,

$$\begin{array}{l} \frac{1}{\delta}\int_{a}^{a_\delta} \mathcal{S}(t,\mathcal{U}) h_\delta(t) dt - \int_{a}^{a_\delta} \mathcal{C}(t,\mathcal{U}) dt + \\ \frac{1}{\delta}\int_{b_\delta}^{b} \mathcal{S}(t,\mathcal{U}) h_\delta(t) dt + \int_{b_\delta}^{b} \mathcal{C}(t,\mathcal{U}) dt - \int_{a_\delta}^{b_\delta} \mathcal{S}(t,\mathcal{U}) dt \geq 0. \end{array}$$

Arguing in the same we did above, we can show that the first and the third integrals converge to zero. The second and fourth integrals have bounded integrands and their integration limits converge to the same point. Hence, we have that $\int_a^b \mathcal{S}(t,\mathcal{U})dt \leq 0$. Putting the two inequalities together we get our result.

Next, consider rotations and $\varphi^* \notin [a,b]$. Given $\delta > 0$ sufficiently small, define the following rent projection function:

$$\mathcal{V}_\delta(t) = \max\left\{\mathcal{U}(t),\ (\beta+\delta)(t-a)+\mathcal{U}(a)\right\},$$

which represents a small anti-clockwise rotation of the affine function $\mathcal{U}$ on $[a,b]$ at point $(a,\mathcal{U}(a))$ in the plane type versus informational rent. This perturbation is feasible. Applying Lemma 13, we obtain

$$\int_a^{b_\delta} \mathcal{S}(t,\mathcal{U})h(t)dt \geq 0,$$

where $h_\delta = \mathcal{U} - \mathcal{V}_\delta$ and b_δ is the only solution of the equation $\mathcal{V}_\delta(t) - \mathcal{U}(t) = 0$. Proceeding in the same way as above, we conclude that

$$\int_a^b \mathcal{S}(t,\mathcal{U})(t-a)dt \geq 0.$$

Analogously, we can make a small clockwise rotation of $\mathcal{U}$ on $[a,b]$ at point $(b,\mathcal{U}(b))$ and conclude that

$$\int_a^b \mathcal{S}(t,\mathcal{U})(t-b)dt \leq 0.$$

If $\mathcal{U}$ has kink at a (at b), then we can do also a small anti-clockwise (clockwise) rotation at b (at a) and get the equality. If $\mathcal{U}$ has kink at both at a and b, using that $\int_a^b \mathcal{S}(t,\mathcal{U})dt = 0$, we conclude the last equality for this case.

The case where $a = t^*$ and $b \geq \varphi^*$ is analogous. The only difference is that we have to consider the rotation at the point $(\varphi^*, \Delta c)$ to eliminate the point effect from φ^* in the condition of Lemma 13.

Lemma 14. *Let $f \in L_\infty[a,b]$ satisfying $\int_a^b f(t)g(t)dt = 0$, for all $g \in C[a,b]$ such that $\int_a^b g(t)dt = 0$. Then, f is a constant function a.e.*

Proof. By Theorem 3.14 of Rudin (1986, pp. 69), we know that the space of real continuous functions $C([a,b])$ is dense in the space of integral functions $L_1([a,b])$ and, by the Stone-Weierstrass Theorem, every function in $C([a,b])$ is the uniform limit of a sequence of polynomial functions. Therefore, the hypothesis of the lemma implies that $\int_a^b f(t)g(t)dt = 0$, for all $g \in L_1[a,b]$ such that $\int_a^b g(t)dt = 0$.

Notice that $L_2[a,b] \subset L_1[a,b]$. Consider the closed subspace $H = \left\{g \in L_2[a,b]; \int_a^b g(t)dt = 0\right\}$ of $L_2[a,b]$. Notice that the orthogonal subspace of H in $L_2[a,b]$, $H^\perp$, is the space of constant functions.[38] Indeed, the constant functions are obviously contained in $H^\perp$ and, for each $g \in L_2[a,b]$, we have that

$$g = \left(g - \frac{1}{b-a}\int_a^b g(t)dt\right) + \frac{1}{b-a}\int_a^b g(t)dt,$$

where $g - \frac{1}{b-a}\int_a^b g(t)dt \in H$, which implies that $H^\perp$ is generated by the constant functions. Therefore, $f \in H^\perp$. □

[38] As usual for $L_p[a,b]$ space, a function g is constant when $g(t) = k$ a.e., for some $k \in \mathbb{R}$.

Proof of Proposition 8

Let $\varphi_U(t) := t + \frac{\Delta c}{u(I)-u(I-L)}$ denote the separating curve of uninsured types. Then, an uninsured type (p_0, p_1) picks high effort if $p_1 > \varphi_U(p_0)$. Proceeding as in Subsection 2.3, we can express the reservation utility of all types in terms of the separating curve φ_U and the reservation utility of diagonal types:

$$\mathcal{V}(t) := tu(I) + (1-t)u(I-L) - c_0. \tag{30}$$

Namely,

$$\max_{i\in\{0,1\}} p_i u(I) + (1-p_i)u(I-L) - c_i = \left\{ \begin{array}{c} \mathcal{V}\left(\varphi_U^{-1}(p_1)\right) \\ \mathcal{V}(p_0) \end{array} \right. .$$

As in Subsection 2.3, let $\mathcal{U}$ denote the rent projection associated with an optimal mechanism. Using these diagonal projections, the participation constraint of diagonal types becomes:

$$\mathcal{U}(t) \geq \mathcal{V}(t). \tag{31}$$

The following lemmata will be useful in the proof of the proposition:

Lemma 15. *Suppose diagonal type $t = 1$ is not excluded: $\mathcal{U}(1) \geq \mathcal{V}(1)$. Then, $b(1,1) \leq u(I) - u(I-L)$.*

Proof. Substituting the expressions for $\mathcal{U}$ and $\mathcal{V}$ at $t = 1$ yields

$$u(W+B) - c_0 \geq u(I) - c_0 \therefore W + B \geq I.$$

From Lemma 7 (and the fact that the net output Δx corresponds to the loss L in this version of the model), $B \leq L$. Hence,

$$W \geq I - L.$$

Because $B \leq L$, $W + B \geq I$, and $W \geq I - L$, concavity of u gives

$$\frac{u(W+B) - u(W)}{B} \leq \frac{u(I) - u(I-L)}{L}.$$

Substituting $B \leq L$, we obtain

$$\frac{u(W+B) - u(W)}{B} \leq \frac{u(I) - u(I-L)}{B} \therefore \underbrace{u(W+B) - u(W)}_{b(1,1)} \leq u(I) - u(I-L),$$

concluding the proof. □

Lemma 16. *In any optimal mechanism, the set of diagonal types that do not participate is an interval of the form $(\bar{t}, 1]$ for some $\bar{t} \in [0,1]$.*

Proof. First, we note that $\mathcal{U}$ is convex while $\mathcal{V}$ is affine – it has slope $\dot{\mathcal{V}}(t) = u(I) - u(I-L)$. Moreover, as established in Subsection 2.3, $\dot{\mathcal{U}}(t) = b(t,t)$ which, by convexity, is a non-decreasing function of t. There are two possible cases:

i. Suppose that type $t = 1$ is not excluded: $\mathcal{U}(1) \geq \mathcal{V}(1)$. Then, the previous lemma implies that

$$b(t,t) \leq u(I) - u(I-L),$$

for all t. As a result, $\mathcal{U}(t) \geq \mathcal{V}(t)$ for all t. Thus, all types participate if diagonal type $t = 1$ participates.

ii. Now suppose that $t = 1$ is excluded: $\mathcal{U}(1) < \mathcal{V}(1)$. Because $\mathcal{U}$ is convex and $\mathcal{V}$ is affine, there must exist $\bar{t} \in [0,1)$ such that $\mathcal{U}(t) \geq \mathcal{V}(t)$ if and only if $t \leq \bar{t}$. □

Expressing the utility of off-the-diagonal types using the projection into the diagonal, Lemma 16 implies that types will prefer not to participate if $p_0 \geq \bar{t}$, or $p_1 \geq \varphi_U(\bar{t})$.

Lemma 17. *Suppose the optimal mechanism is such that all types participate: $\mathcal{U}(t) \geq \mathcal{V}(t)$ for all t. Then, the participation constraint binds at the top: $\mathcal{U}(1) = \mathcal{V}(1)$.*

Proof. The participation constraint cannot be slack for all types. If this were the case, the principal could strictly improve by reducing $\mathcal{U}$ uniformly. Therefore, there must exist t such that $\mathcal{U}(t) = \mathcal{V}(t)$. As argued in Lemma 16, $\dot{\mathcal{V}}(t) = u(I) - u(I-L)$, and $\dot{\mathcal{U}}(t) = b(t,t)$ is a non-decreasing function of t. Moreover, by Lemma 15, $\dot{\mathcal{U}}(t) \leq \dot{\mathcal{V}}(t)$. Because there must exist some t for which $\mathcal{U}(t) = \mathcal{V}(t)$, it follows that $\mathcal{U}(1) = \mathcal{V}(1)$. □

We are now ready to establish the main result. Suppose there exists an optimal mechanism with associated projected rent function $\mathcal{U}$. By Lemma 15 $b(t,t) \leq u(I) - u(I-L)$ for all t. Because b is non-decreasing, there are two possible cases:

- There exists $\epsilon > 0$ such that $b(t,t) = u(I) - u(I-L)$ for all $t > 1-\epsilon$, and
- $b(t,t) < u(I) - u(I-L)$ for all $t < 1$.

First, suppose that $b(t,t) = u(I) - u(I-L)$ for all $t > 1-\epsilon$, where $\epsilon > 0$. By Lemma 17, we must have

$$w(1,1) + \underbrace{u(I) - u(I-L)}_{b(1,1)} - c_0 = u(I) - c_0 \therefore w(1,1) = u(I-L).$$

Moreover, since all those types t get the same power b, they must also get the same w as well (otherwise, the mechanism would not be incentive compatible). Thus, all types associated with diagonal types $t > 1-\epsilon$ are uninsured:

$$W(t,t) = I - L, \text{ and } B(t,t) = L.$$

Now, suppose that $b(t,t) < u(I) - u(I-L)$ for all $t < 1$. In order to obtain a contradiction, suppose the solution is such that all types participate. To keep the notation consistent with the rest of the paper, we write $x_H := I$, $x_L := I - L$, and $\Delta x := L$. The principal's expected utility is then

$$\Pi(\mathcal{U}) = \int_0^{\bar{t}} \left(t\Delta x - G(\mathcal{U},\dot{\mathcal{U}},t)\right) F_0(t,\varphi)dt + \int_{\varphi^*}^{\min\left\{1;\bar{t}+\frac{\Delta c}{u(x_H)-u(x_L)}\right\}} \left(t\Delta x - G(\mathcal{U},\dot{\mathcal{U}},t)\right) F_1(\varphi^{-1},t)dt,$$

where $\bar{t}$ is the last type who wants to participate.

Consider a perturbation that uniformly reduces the rents of all types by $\epsilon > 0$:

$$\mathcal{U}_\epsilon(t) \equiv \mathcal{U}(t) - \epsilon.$$

Note that the perturbation preserves $\dot{\mathcal{U}}$ and φ. Let $\bar{t}_\epsilon$ denote the highest diagonal type who chooses to participate:

$$\mathcal{U}(\bar{t}_\epsilon) - \epsilon = \mathcal{V}(\bar{t}_\epsilon).$$

(Note that, by Lemma 17, $\bar{t}_0 = 1$). Substituting the expression for $\mathcal{V}$, yields

$$\mathcal{U}(\bar{t}_\epsilon) - \epsilon = u(x_L) + \bar{t}_\epsilon\left[u(x_H) - u(x_L)\right] - c_0.$$

Total differentiation gives:

$$\frac{\partial \bar{t}_\epsilon}{\partial \epsilon} = -\frac{1}{u\left(x_H\right) - u\left(x_L\right) - \dot{\mathcal{U}}\left(\bar{t}_\epsilon\right)} = \frac{1}{b\left(\bar{t}_\epsilon, \bar{t}_\epsilon\right) - \left[u\left(I\right) - u\left(I-L\right)\right]} < 0.$$

Therefore, this perturbation excludes a positive mass of types. We will show that, for small ϵ, this perturbation raises the principal's profit, which contradicts our assumption that the original mechanism was optimal.

The principal's expected utility under the perturbation is

$$\Pi_\epsilon = \int_0^{\bar{t}_\epsilon} \left(t\Delta x - G(\mathcal{U}-\epsilon, \dot{\mathcal{U}}, t)\right) F_0(t,\varphi)dt + \int_{\varphi^*}^{\min\left\{1; \bar{t}_\epsilon + \frac{\Delta c}{u(x_H)-u(x_L)}\right\}} (t\Delta x - G(\mathcal{U}-\epsilon, \dot{\mathcal{U}}, t))F_1(\varphi^{-1}, t)dt.$$

Since $\bar{t}_0 = 1$, it follows that $1 < \bar{t}_\epsilon + \frac{\Delta c}{u(x_H)-u(x_L)}$ for ϵ small enough. Differentiating with respect to ϵ, yields

$$\frac{\partial \Pi_\epsilon}{\partial \epsilon} = \left(t\Delta x - G(\mathcal{U}-\epsilon, \dot{\mathcal{U}}, t)\right) F_0(t,\varphi)\Big|_{t=\bar{t}_\epsilon} \frac{\partial \bar{t}_\epsilon}{\partial \epsilon}$$
$$+ \int_0^{\bar{t}_\epsilon} \frac{\partial G}{\partial \mathcal{U}}(\mathcal{U}-\epsilon, \dot{\mathcal{U}}, t)F_0(t,\varphi)dt + \int_{\varphi^*}^{1} \frac{\partial G}{\partial \mathcal{U}}(\mathcal{U}-\epsilon, \dot{\mathcal{U}}, t))F_1(\varphi^{-1}, t)dt.$$

Note that $\frac{\partial G}{\partial \mathcal{U}} = \frac{t}{u'(u^{-1}(\mathcal{U}+(1-t)\mathcal{U}+c_0))} + \frac{1-t}{u'(u^{-1}(\mathcal{U}-t\mathcal{U}+c_0))} > 0$. Therefore, the terms on the second line are both strictly positive.

Moreover, $\lim_{\epsilon \searrow 0} \bar{t}_\epsilon = 1$ and

$$t\Delta x - G(\mathcal{U}-\epsilon, \dot{\mathcal{U}}, t)\Big|_{t=1} = \Delta x - u^{-1}\left(\mathcal{U}\left(1\right) + c_0\right).$$

By Lemma 17, $\mathcal{U}\left(1\right) = u\left(x_H\right) - c_0$. Therefore,

$$u^{-1}\left(u\left(x_H\right)\right) = x_H > x_H - x_L = \Delta x.$$

As a result, $\left(t\Delta x - G(\mathcal{U}-\epsilon, \dot{\mathcal{U}}, t)\right) F_0(t,\varphi)\Big|_{t=\bar{t}_\epsilon} < 0$ for small ϵ. Since $\frac{\partial \bar{t}_\epsilon}{\partial \epsilon} < 0$, it follows that the first line is also strictly positive for ϵ close to zero. Hence, $\frac{\partial \Pi_\epsilon}{\partial \epsilon} > 0$ for ϵ small enough, establishing contradicting the optimality of $\mathcal{U}$.

Proof of Proposition 9

The following lemma will be useful in the proof of the main result:

Lemma 18. *Let $\bar{t}$ be the first diagonal type to be excluded: $\mathcal{U}\left(t\right) > \mathcal{V}\left(t\right)$ for $t < \bar{t}$ and $\mathcal{U}\left(\bar{t}\right) = \mathcal{V}\left(\bar{t}\right)$. Then, $b\left(t,t\right) < u\left(I\right) - u\left(I-L\right)$ for all $t < \bar{t}$.*

Proof. As argued in Lemma 16, $\mathcal{U}$ is convex with slope $\dot{\mathcal{U}}\left(t\right) = b\left(t,t\right)$ and $\mathcal{V}$ is affine with slope $\dot{\mathcal{V}}\left(t\right) = u\left(I\right) - u\left(I-L\right)$. In order to obtain a contradiction, suppose the solution entails $b\left(\tilde{t}, \tilde{t}\right) \geq u\left(I\right) - u\left(I-L\right)$ for some $\tilde{t} < \bar{t}$. Then, because b is non-decreasing, it follows that $\dot{\mathcal{U}}\left(t\right) = b\left(t,t\right) \geq \dot{\mathcal{V}}\left(t\right)$ for all $t \in [\tilde{t}, \bar{t}]$. However, because $\tilde{t} < \bar{t}$, we must have $\mathcal{U}\left(\tilde{t}\right) > \mathcal{V}\left(\tilde{t}\right)$. Then, $\dot{\mathcal{U}}\left(t\right) \geq \dot{\mathcal{V}}\left(t\right)$ for all $t \in [\tilde{t}, \bar{t}]$ implies that $\mathcal{U}\left(\bar{t}\right) > \mathcal{V}\left(\bar{t}\right)$, contradicting $\mathcal{U}\left(\bar{t}\right) = \mathcal{V}\left(\bar{t}\right)$. □

Let (w, b, e) be an optimal mechanism with an associated effort frontier φ, and consider a type (p_0, p_1) in the high effort region: $\varphi^{-1}(p_1) > p_0$. By incentive compatibility, exerting high effort must yield a higher payoff than exerting a low effort while reporting the same type:

$$w(p_0, p_1) + p_1 b(p_0, p_1) - c_1 \geq w(p_0, p_1) + p_0 b(p_0, p_1) - c_0.$$

Subtracting $w(p_0, p_1)$ from both sides and rearranging yields

$$p_1 \geq p_0 + \frac{\Delta c}{b(p_0, p_1)} = p_0 + \frac{\Delta c}{b(p_1, p_1)},$$

where we used the fact that $b(p_0, p_1) = b(p_1, p_1)$. Taking the limit as p_0 converges to $\varphi^{-1}(p_1)$ yields

$$p_1 \geq \varphi^{-1}(p_1) + \frac{\Delta c}{b(p_1, p_1)} > \varphi^{-1}(p_1) + \frac{\Delta c}{u(I) - u(I - L)},$$

where the last inequality used the fact that $b(p_1, p_1) < u(I) - u(I - L)$. Letting $\hat{p}_0 := \varphi^{-1}(p_1)$, we obtain

$$\varphi(\hat{p}_0) > \hat{p}_0 + \frac{\Delta c}{u(I) - u(I - L)}.$$

Since this holds for any arbitrary $\hat{p}_0$, we have established the result.

Proof of Proposition 10

(i) The proof that there are two contracts at bottom is analogous to Proposition 4.

(ii) Suppose that $(w(\boldsymbol{p}), b(\boldsymbol{p}))_{\boldsymbol{p}\in\overline{\Delta}}$ is an optimal mechanism such that $(1 + \lambda)b(\hat{\boldsymbol{p}}) > \Delta x$, for some $\hat{\boldsymbol{p}} = (\hat{p}_0, \hat{p}_1) \in \overline{\Delta}$ with $\hat{p}_0 < 1$. Define $\hat{t} = inf\,\{t \in [0, 1); (1 + \lambda)b(t, t) > \Delta x\}$. Suppose that this mechanism is optimal. Since the optimal mechanism has a region with zero bonus, $\hat{t} \in (0, 1)$. Proceeding the same steps of the proof of Lemma 7 and define the principal's objective function on the family of perturbations by:

$$\Pi_\alpha = x_L - \lambda c_0 + \int_0^1 \int_t^{\varphi_\alpha(t)} (t\Delta x - \lambda \mathcal{U}_\alpha(t) - c_0) f(t, s) ds dt + \int_0^1 \int_{\varphi_\alpha(t)}^1 (s\Delta x - \lambda \mathcal{U}_\alpha(s) - c_1) f(t, s) ds dt.$$

Now let us compute the total derivative of Π_α with respect to α:

$$\begin{aligned} \tfrac{\partial \Pi_\alpha}{\partial \alpha} = \; & -\lambda \textstyle\int_0^1 \left(\int_t^{\varphi_\alpha(t)} \frac{\partial \mathcal{U}_\alpha(t)}{\partial \alpha} f(t, s) ds + \int_{\varphi_\alpha(t)}^1 \frac{\partial \mathcal{U}_\alpha(s)}{\partial \alpha} f(t, s) ds \right) dt \\ & + \textstyle\int_0^1 (t\Delta x - \lambda \mathcal{U}_\alpha(t) - c_0) \frac{\partial \varphi_\alpha(t)}{\partial \alpha} f(t, \varphi_\alpha(t)) dt \\ & - \textstyle\int_0^1 (\varphi_\alpha(t)\Delta x - \lambda \mathcal{U}_\alpha(\varphi_\alpha(t)) - c_1) \frac{\partial \varphi_\alpha(t)}{\partial \alpha} f(t, \varphi_\alpha(t)) dt. \end{aligned}$$

Notice that, since $\frac{\partial \mathcal{U}_\alpha(t)}{\partial \alpha} \leq 0$, the first line is obviously non-negative. We can rewrite the second and third lines as

$$-\int_0^1 \left[(\varphi_\alpha(t) - t)\Delta x - (1 + \lambda)\Delta c\right] \frac{\partial \varphi_\alpha(t)}{\partial \alpha} f(t, \varphi_\alpha(t)) dt$$

where we are using that $\mathcal{U}(\varphi_\alpha(t)) = \mathcal{U}(t) + \Delta c$. Suppose that $(\varphi(t) - t)\Delta x > (1 + \lambda)\Delta c$, for some $t \in [\hat{t}, 1]$. Then, by the definition of $\hat{t}$, we have that $(\varphi(t) - t)(1 + \lambda)b(t, t) > (1 + \lambda)\Delta c$. However, incentive compatibility implies that

$$w(t, t) + tb(t, t) - c_0 \geq w(t, t) + \varphi(t)b(t, t) - c_1,$$

which leads to a contradiction. Therefore, we have that

$$\left.\frac{\partial \Pi_\alpha}{\partial \alpha}\right|_{\alpha=0} \geq 0.$$

We have to show that the inequality is strict to complete the proof. For this, notice that $\frac{\partial \mathcal{U}_\alpha(t)}{\partial \alpha}$ is strictly negative for all $t \geq \hat{t}$, since $\beta < \dot{\mathcal{U}}(1)$. These imply that the integrand of the double integral in the first line is strictly negative on $[\hat{t}, 1]$. Since $\Delta x = (1+\lambda)(C_H - C_L)$, we showed that the optimal second-best power is lower that the cost reduction.

The proof of existence of an optimal mechanism is analogous to the proof of Proposition 2.

(iii) Using item (ii), the proof is analogous to the proof of Lemmas 10 and 11.

(iv) The proof is analogous to the proof of Proposition 3.

References

Acemoglu, D. (1998): "Credit Market Imperfections and the Separation of Ownership from Control," *Journal of Economic Theory*, 78(2), 355–81.

Armstrong, M. (1996): "Multiproduct Nonlinear Pricing," *Econometrica*, 64(1), 51–75.

Bajari, P., H. Hong, and A. Khwaja (2012): "A Semiparametric Analysis of Adverse Selection and Moral Hazard in Health Insurance Contracts," Discussion paper, Working Paper.

Billingsley, P. (1995): *Probability and Measure*. John Willey and Sons, 3rd edn.

Boadway, R., M. Marchand, P. Pestieau, and M. del Mar Racionero (2002): "Optimal Redistribution with Heterogeneous Preferences for Leisure," *Journal of Public Economic Theory*, 4(4), 475–98.

Bond, E. W., and K. J. Crocker (1991): "Smoking, Skydiving, and Knitting: The Endogenous Categorization of Risks in Insurance Markets with Asymmetric Information," *Journal of Political Economy*, 99(1), 177–200.

Caillaud, B., R. Guesnerie, and P. Rey (1992): "Noisy Observation in Adverse Selection Models," *Review of Economic Studies*, 59(3), 595–615.

Carroll, G. (2013): "Robustness and Linear Contracts," Working paper, Microsoft Research and Stanford University.

Chade, H., and E. Schlee (2012): "Optimal Insurance with Adverse Selection," *Theoretical Economics*, 7(3), 571–607.

Chassagnon, A., and P.-A. Chiappori (1997): "Insurance under Moral Hazard and Adverse Selection: the Case of Pure Competition," *DELTA-CREST Working Paper*.

Chassang, S. (2013): "Calibrated Incentive Contracts," *Econometrica*, Forthcoming.

Chiu, W. H., and E. Karni (1998): "Endogenous Adverse Selection and Unemployment Insurance," *Journal of Political Economy*, 106(4), 806–27.

Choné, P., and G. Laroque (2010): "Negative Marginal Tax Rates and Heterogeneity," *American Economic Review*, 100(5), 2532–47.

Chu, L. Y., and D. Sappington (2007): "Simple Cost-Sharing Contracts," *American Economic Review*, 97(1), 419–428.

Cremer, H., P. Pestieau, and J.-C. Rochet (2001): "Direct versus Indirect Taxation: the Design of the Tax Structure Revisited," *International Economic Review*, 42(3), 781–800.

De Meza, D., and D. C. Webb (2001): "Advantageous Selection in Insurance Markets," *RAND Journal of Economics*, pp. 249–62.

Diamond, P. A. (1998): "Optimal Income Taxation: an Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 88(1), 83–95.

——— (2005): *Taxation, Incomplete Markets, and Social Security*. MIT press.

Diamond, P. A., and J. A. Mirrlees (1971): "Optimal Taxation and Public Production I: Production Efficiency," *American Economic Review*, 61(1), 8–27.

Diamond, P. A., and J. Spinnewijn (2011): "Capital Income Taxes with Heterogeneous Discount Rates," *American Economic Journal: Economic Policy*, 3(4), 52–76.

Ebert, U. (1992): "A Reexamination of the Optimal Nonlinear Income Tax," *Journal of Public Economics*, 49(1), 47–73.

Edmans, A., and X. Gabaix (2011): "Tractability in Incentive Contracting," *Review of Financial Studies*, 24(9), 2865–94.

Einav, L., A. Finkelstein, S. P. Ryan, P. Schrimpf, and M. R. Cullen (2013): "Selection on Moral Hazard in Health Insurance," *American Economic Review*, 103(178–219).

Grossman, S. J., and O. D. Hart (1983): "An Analysis of the Principal-Agent Problem," *Econometrica*, 51(1), 7–45.

Holmstrom, B., and P. Milgrom (1987): "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 55(2), 303–28.

Innes, R. D. (1990): "Limited Liability and Incentive Contracting with Ex-Ante Action Choices," *Journal of Economic Theory*, 52(1), 45–67.

Judd, K., and C.-L. Su (2006): "Optimal Income Taxation with Multidimensional Taxpayer Types," Discussion paper, Working Paper.

Jullien, B., B. Salanie, and F. Salanie (2007): "Screening Risk-Averse Agents under Moral Hazard: Single-Crossing and the CARA Case," *Economic Theory*, 30(1), 151–169.

Kadan, O., P. J. Reny, and J. M. Swinkels (2011): "Existence of Optimal Mechanisms in Principal-Agent Problems," Discussion paper, Working Paper.

Karlan, D., and J. Zinman (2009): "Observing Unobservables: Identifying Information Asymmetries with a Consumer Credit Field Experiment," *Econometrica*, 77(6), 1993–2008.

KLEVEN, H. J., C. T. KREINER, AND E. SAEZ (2009): “The Optimal Income Taxation of Couples,” *Econometrica*, 77(2), 537–60.

LAFFONT, J.-J., AND D. MARTIMORT (2002): *The Theory of Incentives - Part I.* Princeton University Press.

LAFFONT, J.-J., E. MASKIN, AND J.-C. ROCHET (1987): *Optimal Nonlinear Pricing with Two-Dimensional Characteristics*pp. 256–266. University of Minnesota Press, Minneapolis.

LAFFONT, J.-J., AND J. TIROLE (1986): “Using Cost Observation to Regulate Firms,” *Journal of Political Economy*, pp. 614–641.

——— (1993): *A Theory of Incentives in Procurement and Regulation.* MIT press.

MASKIN, E., AND J. RILEY (1984): “Monopoly with Incomplete Information,” *The RAND Journal of Economics*, 15(2), 171–96.

MIRRLEES, J. A. (1971): “An Exploration in the Theory of Optimum Income Taxation,” *Review of Economic Studies*, 38(2), 175–208.

——— (1972): “On Producer Taxation,” *Review of Economic Studies*, 39(1), 105–111.

——— (1990): “Taxing Uncertain Incomes,” *Oxford Economic Papers*, 42(1), 34–45.

MUSSA, M., AND S. ROSEN (1978): “Monopoly and Product Quality,” *Journal of Economic Theory*, 18(2), 301–17.

MYERSON, R. B. (1981): “Optimal Auction Design,” *Mathematics of Operations Research*, 6(1), 58–73.

——— (1982): “Optimal Coordination Mechanisms in Generalized Principal-Agent Problems,” *Journal of Mathematical Economics*, 10(1), 67–81.

PICARD, P. (1987): “On the Design of Incentive Schemes under Moral Hazard and Adverse Selection,” *Journal of Public Economics*, 33(3), 305–31.

PIKETTY, T. (1997): “La Redistribution Fiscale Face au Chômage,” *Revue Française d’Économie*, 12(1), 157–201.

PIKETTY, T., AND E. SAEZ (2012): “Optimal Labor Income Taxation,” in *Handbook of Public Economics*, ed. by A. Auerbach, R. Chetty, and M. S. Feldstein, vol. 5. Amsterdam: Elsevier-North Holland.

POBLETE, J., AND D. SPULBER (2012): “The Form of Incentive Contracts: Agency with Moral Hazard, Risk Neutrality, and Limited Liability,” *RAND Journal of Economics*, 43(2), 215–234.

ROCHET, J.-C. (1987): “A Necessary and Sufficient Condition for Rationalizability in a Quasi-Linear Context,” *Journal of Mathematical Economics*, 16(2), 191–200.

ROCHET, J.-C., AND P. CHONÉ (1998): “Ironing, Sweeping, and Multidimensional Screening,” *Econometrica*, 66(4), 783–826.

Rochet, J.-C., and L. A. Stole (2002): "Nonlinear Pricing with Random Participation," *Review of Economic Studies*, 69(1), 277–311.

——— (2003): *Advances in Economics and Econometrics: Theory and Applications - Volume 1* vol. 1, chap. The Economics of Multidimensional Screening. Econometric Society Monographs.

Rogerson, W. P. (2003): "Simple Menus of Contracts in Cost-Based Procurement and Regulation," *American Economic Review*, 93(3), 919–926.

Rothschild, C., and F. Scheuer (2012): "Optimal Taxation with Rent-Seeking," Working paper, Middlebury College and Stanford University.

——— (2013): "Redistributive Taxation in the Roy Model," *Quarterly Journal of Economics*, Forthcoming.

Rothschild, M., and J. Stiglitz (1976): "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics*, 90(4), 629–49.

Rudin, W. (1986): *Real and Complex Analysis*. McGraw-Hill: New York, 3rd edn.

Saez, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68(1), 205–29.

Seade, J. K. (1977): "On the Shape of Optimal Tax Schedules," *Journal of Public Economics*, 7(2), 203–35.

Stewart, J. (1994): "The Welfare Implications of Moral Hazard and Adverse Selection in Competitive Insurance Markets," *Economic Inquiry*, 32(2), 193–208.

Stiglitz, J. E. (1977): "Monopoly, Non-linear Pricing and Imperfect Information: the Insurance Market," *Review of Economic Studies*, 44(3), 407–30.

Tarkiainen, R., and M. Tuomala (1999): "Optimal Nonlinear Income Taxation with a Two-Dimensional Population: A Computational Approach," *Computational Economics*, 13(1), 1–16.

Tenhunen, S., and M. Tuomala (2010): "On Optimal Lifetime Redistribution Policy," *Journal of Public Economic Theory*, 12(1), 171–98.

Online Appendix

Proof of Lemma 1

The proof of Lemma 1 uses the single-crossing condition in each region to establish the result. It is presented in the online appendix.
The proof proceeds by a series of claims. The two first claims establish that the vertical and horizontal sections of the sets Δ_0 and Δ_1 are intervals.

Claim 3. Let (w, b, e) be an incentive-compatible mechanism. If $(p_0, p_1) \in \Delta_1$ and $\hat{p}_1 > p_1$, then $(p_0, \hat{p}_1) \in \Delta_1$.

Proof. Let $\mathbf{p} = (p_0, p_1) \in \Delta_1$ and suppose that $\hat{\mathbf{p}} = (p_0, \hat{p}_1) \in \Delta_0$. Incentive-compatibility implies that

$$\begin{aligned} w(\mathbf{p}) + p_1 b(\mathbf{p}) - c_1 &\geq w(\hat{\mathbf{p}}) + p_0 b(\hat{\mathbf{p}}) - c_0, \text{ and} \\ w(\hat{\mathbf{p}}) + p_0 b(\hat{\mathbf{p}}) - c_0 &\geq w(\mathbf{p}) + \hat{p}_1 b(\mathbf{p}) - c_1. \end{aligned}$$

Combining these inequalities, we obtain $(p_1 - \hat{p}_1)\, b(\mathbf{p}) \geq 0$. Since $\mathbf{p} \in \Delta_1$, we must have $b(\mathbf{p}) > 0$. Therefore, $p_1 \geq \hat{p}_1$, which contradicts the statement of the claim. □

Claim 4. For any feasible mechanism, there exists an equivalent mechanism with the following property: if $(p_0, p_1) \in \Delta_0$ and $\hat{p}_0 > p_0$, then $(\hat{p}_0, p_1) \in \Delta_0$.

Proof. Let $\mathbf{p} = (p_0, p_1) \in \Delta_0$ and suppose that $\hat{\mathbf{p}} = (\hat{p}_0, p_1) \in \Delta_1$. Incentive-compatibility implies that

$$\begin{aligned} w\left(\mathbf{p}\right) + p_0 b\left(\mathbf{p}\right) - c_0 &\geq w(\hat{\mathbf{p}}) + p_1 b(\hat{\mathbf{p}}) - c_1, \text{ and} \\ w(\hat{\mathbf{p}}) + p_1 b(\hat{\mathbf{p}}) - c_1 &\geq w\left(\mathbf{p}\right) + \hat{p}_0 b\left(\mathbf{p}\right) - c_0. \end{aligned} \tag{32}$$

Combining these inequalities, we obtain $\left(\hat{p}_0 - p_0\right) b\left(\mathbf{p}\right) = 0$, which, because $\hat{p}_0 > p_0$, implies that $b\left(\mathbf{p}\right) = 0$. Substituting back, yields $w\left(\mathbf{p}\right) - c_0 = w(\hat{\mathbf{p}}) + p_1 b(\hat{\mathbf{p}}) - c_1$. Therefore, types $\mathbf{p}$ and $\hat{\mathbf{p}}$ are both indifferent between each others' contracts.

Consider the alternative mechanism that coincides with the original one except that we offer the same contract of type $\mathbf{p}$ to type $\hat{\mathbf{p}}$. First, we verify that new mechanism is also feasible. Limited liability follows immediately from the feasibility of the original mechanism. Moreover, because all types get exactly the same expected payoff as in both mechanisms, the participation constraint is also satisfied. To verify incentive-compatibility, note that no type other than $\hat{\mathbf{p}}$ can profit by deviating since the original mechanism was incentive compatible and no new contract was added. Moreover, because type $\hat{\mathbf{p}}$ obtains the same payoff under the new mechanism as in the original one (which was incentive-compatible), she also cannot profit by deviating.

If the set of types for which $\mathbf{p} = (p_0, p_1) \in \Delta_0$ and $\hat{\mathbf{p}} = (\hat{p}_0, p_1) \in \Delta_1$ with $\hat{p}_0 > p_0$ has zero measure, then the principal is indifferent between the original and the new mechanism. Because all agents are indifferent between them, the mechanisms are equivalent. Suppose, in order to obtain a contradiction, that the set of such types has a strictly positive measure. That is, for a set of types with positive measure, we have

$$w\left(p_0, p_1\right) - c_0 = w(\hat{p}_0, p_1) + p_1 b(\hat{p}_0, p_1) - c_1$$

where $\hat{p}_0 > p_0$. Incentive compatibility implies that expression on the left must be constant in p_1, whereas the expression on the right must be strictly increasing in p_1 (since $b(\mathbf{p}) > 0$ for all types who exert high effort). Therefore, this condition cannot hold for a set of types with positive measure. □

It follows directly from Claims 1 and 2 that there exists a non-decreasing function $\varphi : [0,1] \to \mathbb{R}_+$ such that $(p_0, p_1) \in \Delta_1$ if and only if $p_1 \geq \varphi(p_0)$. The next claim establishes that this function is continuous.

Claim 5. For every feasible mechanism, there exists an equivalent mechanism with the following property: $(p_0, p_1) \in \Delta_1$ if and only if $p_1 \geq \varphi(p_0)$ for a non-decreasing and continuous function $\varphi : [0,1] \to [0,1]$.

Proof. The existence of such a non-decreasing function φ follows straight from Claims 1 and 2. It remains to be shown that φ is continuous. Suppose, in order to obtain a contradiction, that φ is discontinuous at a point p_0. Since φ is bounded and non-decreasing, there exist $\varphi_+ > \varphi_-$ such that

$$\varphi_+ = \lim_{p \searrow p_0} \varphi(p) \text{ and } \varphi_- = \lim_{p \nearrow p_0} \varphi(p).$$

From the definition of φ, $(p_0, p_1) \in \Delta_1$ for all $p_1 \in [\varphi_-, \varphi_+]$. Moreover, for any $\delta > 0$ and $p_1 \in [\varphi_-, \varphi_+]$, it follows that $(p_0 + \delta, p_1) \in \Delta_0$.

By the Theorem of the Maximum, $U : \bar{\Delta} \to \mathbb{R}$ is a continuous function. Therefore,

$$U(p_0, p_1) = \lim_{\delta \to 0} U(p_0, p_1 + \delta), \quad \forall p_1 \in [\varphi_-, \varphi_+].$$

Let $\bar{\varphi} = \frac{\varphi_+ + \varphi_-}{2}$. Note that types $(p_0 + \delta, \bar{\varphi})$ and $(p_0 + \delta, \bar{\varphi} + \varepsilon)$ both belong to Δ_0 for any $\delta > 0$ and $\varepsilon \in \left[0, \frac{\varphi_+ + \varphi_-}{2}\right]$. Then, using the incentive-compatibility constraint of type $(p_0 + \delta, \bar{\varphi})$, we obtain

$$\begin{aligned} U(p_0 + \delta, \bar{\varphi}) &\geq w(p_0 + \delta, \bar{\varphi} + \varepsilon) + (p_0 + \delta) b(p_0 + \delta, \bar{\varphi} + \varepsilon) - c_0 \\ &= U(p_0 + \delta, \bar{\varphi} + \varepsilon). \end{aligned}$$

Similarly, the incentive-compatibility constraint of type $(p_0 + \delta, \bar{\varphi} + \varepsilon)$ yields

$$\begin{aligned} U(p_0 + \delta, \bar{\varphi} + \varepsilon) &\geq w(p_0 + \delta, \bar{\varphi}) + (p_0 + \delta) b(p_0 + \delta, \bar{\varphi}) - c_0 \\ &= U(p_0 + \delta, \bar{\varphi}). \end{aligned}$$

Combining both inequalities, we obtain

$$U(p_0 + \delta, \bar{\varphi}) = U(p_0 + \delta, \bar{\varphi} + \varepsilon), \tag{33}$$

for any $\delta > 0$ and $\varepsilon \in [0, \bar{\varphi}]$.

Moreover, from the incentive-compatibility constraint of type $(p_0, \bar{\varphi} + \varepsilon) \in \Delta_1$, we have

$$\begin{aligned} U(p_0, \bar{\varphi} + \varepsilon) &\geq w(p_0, \bar{\varphi}) + (\bar{\varphi} + \varepsilon) b(p_0, \bar{\varphi}) - c_1 \\ &= U(p_0, \bar{\varphi}) + \varepsilon b(p_0, \bar{\varphi}), \end{aligned}$$

and because $b(\mathbf{p}) > 0$ for any $\boldsymbol{p} \in \Delta_1$

$$U(p_0, \bar{\varphi} + \varepsilon) > U(p_0, \bar{\varphi}). \tag{34}$$

Equation (33) implies that

$$\lim_{\delta\searrow 0} U\left(p_0+\delta,\bar{\varphi}\right)=\lim_{\delta\searrow 0} U\left(p_0+\delta,\bar{\varphi}+\varepsilon\right), \tag{35}$$

and, by the continuity of U,

$$\lim_{\delta\searrow 0} U\left(p_0+\delta,\bar{\varphi}\right)=U\left(p_0,\bar{\varphi}\right),\ \text{and} \tag{36}$$

$$\lim_{\delta\searrow 0} U\left(p_0+\delta,\bar{\varphi}+\varepsilon\right)=U\left(p_0,\bar{\varphi}+\varepsilon\right). \tag{37}$$

Combining equations (35)-(37), we obtain $U\left(p_0,\bar{\varphi}+\varepsilon\right)=U\left(p_0,\bar{\varphi}\right)$, which contradicts inequality (34). □

Proof of Lemma 2

(a) The informational rent function U can be written as

$$U\left(p_0,p_1\right)=\max_{\hat{\boldsymbol{p}}\in\bar{\Delta}}\left\{\max_{\hat{e}\in\{0,1\}}\left\{w\left(\hat{\boldsymbol{p}}\right)+p_{\hat{e}}b\left(\hat{\boldsymbol{p}}\right)-c_{\hat{e}}\right\}\right\},$$

which is the upper envelope of linear functionals and therefore is a convex function. Convexity implies that it is a.e. differentiable. From the envelope theorem, it follows that $\nabla U\left(p_0,p_1\right)=\begin{cases}\left(b\left(p_0,p_1\right),0\right),\ \text{if}\ p_1<\varphi\left(p_0\right)\\ \left(0,b\left(p_0,p_1\right)\right),\ \text{if}\ p_1>\varphi\left(p_0\right)\end{cases}.$

(b) Monotonicity follows from standard manipulations of the incentive-compatibility constraints. The constancy properties follow from the arguments in the proof of Lemma 1.

(c) Free disposal implies that $b\left(\mathbf{p}\right)\geq 0$ for all $\mathbf{p}$. The participation constraint (IR) implies $U\left(0,0\right)\geq 0$.

Proof of Lemma 4

Let (w,b,e) be a mechanism for which there exists a continuous and non-decreasing function φ satisfying condition (3). For such a mechanism, let $U:\bar{\Delta}\rightarrow\mathbb{R}_+$ denote the informational rent function as defined in equation (2). Lemma 4 is a direct consequence of the following result, which establishes that conditions (a)-(e) from Lemmas 2 and 3 are sufficient for the feasibility of the mechanism:

Claim. Let (w,b,e) be a mechanism satisfying condition (3) for a continuous and non-decreasing function $\varphi:[0,1]\rightarrow[0,1]$. Let U be as defined in equation (1). Suppose that conditions (a)-(e) are satisfied. Then, (w,b,e) is a feasible mechanism.

Proof of the Claim. We need to establish that a mechanism satisfying conditions (a)-(e) for a continuous and nondecreasing φ satisfies incentive-compatibility (IC), individual-rationality (IR), and free disposal (FD). Condition (b) implies that $b\left(\mathbf{p}\right)\geq b\left(0,0\right)$ for all $\mathbf{p}$. Then, by condition (c), (FD) holds. Moreover, conditions (a) and (c) imply that $U\left(\mathbf{p}\right)\geq 0$ for all $\mathbf{p}$ and, therefore, (IR) is satisfied. It remains to be shown that the mechanism is incentive-compatible.

We consider deviations by types in regions Δ_0 and Δ_1 separately. There are 4 possible deviations in each region: taking a contract designed to types in regions Δ_0 or Δ_1 and exerting efforts 0 or 1. First, let $\boldsymbol{p} = (p_0, p_1) \in \Delta_0$ (i.e. $p_1 \leq \varphi(p_0)$).

Case 1: Reporting type $\mathbf{q} \in \Delta_0$ and choosing $e = 0$.

In this case, the proof follows by standard incentive-compatibility arguments (applying the one-dimensional single-crossing condition taking effort as fixed).

Case 2: Reporting a type $\mathbf{q} \in \Delta_0$ and choosing $e = 1$.

We have to verify that the following inequality is satisfied:

$$U(\boldsymbol{p}) = w(\mathbf{p}) + p_0 b(\mathbf{p}) - c_0 \geq w(\mathbf{q}) + p_1 b(\mathbf{q}) - c_1.$$

Since type $(0, \varphi(p_0)) \in \Delta_1$ and, from condition (a), $U(\boldsymbol{p}) = U(0, \varphi(p_0))$, the previous inequality is equivalent to

$$U(0, \varphi(p_0)) = w(0, \varphi(p_0)) + \varphi(p_0) b(0, \varphi(p_0)) - c_1 \geq w(\mathbf{q}) + p_1 b(\mathbf{q}) - c_1 \tag{38}$$

for all $\boldsymbol{q} \in \Delta_0$. Note that this is the incentive-compatibility constraint preventing type $(0, \varphi(p_0)) \in \Delta_1$ from getting the contract designed for $\mathbf{q} \in \Delta_0$ and choosing effort $e = 1$. As will be established in Case 8 below, this inequality is satisfied under the assumptions of the Lemma.

Case 3: Reporting type $\mathbf{q} \in \Delta_1$ and choosing $e = 0$.

We have to show that

$$w(\mathbf{p}) + p_0 b(\mathbf{p}) - c_0 \geq w(\mathbf{q}) + p_0 b(\mathbf{q}) - c_0. \tag{39}$$

Condition (e) implies that, for almost all $\mathbf{q} \in \Delta_1$, $b(\mathbf{q}) = b(q_1, q_1)$ and $w(\mathbf{q}) = w(q_1, q_1)$. Then, for all such $\mathbf{q}$, we have

$$w(\boldsymbol{q}) + p_0 b(\boldsymbol{q}) - c_0 = w(q_1, q_1) + p_0 b(q_1, q_1) - c_0.$$

Because $(q_1, q_1) \in \Delta_0$, the result from Case 1 implies that inequality (39) holds for all such $\mathbf{q}$ (which holds a. e.).

It remains to be shown that (39) holds for $\mathbf{q}$ such that $b(\mathbf{q}) \neq b(q_1, q_1)$. Let $(q_0, \hat{q}_1)$ be a type such that $b(q_0, \hat{q}_1) \neq b(\hat{q}_1, \hat{q}_1)$ and suppose $p_0 > \hat{q}_1$ (the other case is analogous). Since $b(\mathbf{q}) = b(q_1, q_1)$ for almost all $\mathbf{q} \in \Delta_1$, there exists a decreasing sequence $(q_1^n) \to \hat{q}_1$ such that $b(q_0, q_1^n) = b(q_1^n, q_1^n)$. Then, inequality (39) implies that

$$\begin{aligned} w(\mathbf{p}) + p_0 b(\mathbf{p}) - c_0 &\geq w(q_0, q_1^n) + p_0 b(q_0, q_1^n) - c_0 \\ &= U(q_0, q_1^n) + (p_0 - q_1^n) b(q_0, q_1^n). \end{aligned}$$

Because the sequence (q_1^n) is decreasing, it follows that $b(q_0, q_1^n) \geq b(q_0, \hat{q}_1)$. Hence,

$$w(\mathbf{p}) + p_0 b(\mathbf{p}) - c_0 \geq U(q_0, q_1^n) + (p_0 - q_1^n) b(q_0, \hat{q}_1).$$

Since U is continuous, it follows that the right hand side of the inequality above converges to $U(q_0, \hat{q}_1) + (p_0 - \hat{q}_1) b(q_0, \hat{q}_1)$. Rearranging, we obtain

$$\begin{aligned} w(\mathbf{p}) + p_0 b(\mathbf{p}) - c_0 &\geq w(q_0, \hat{q}_1) + \hat{q}_1 b(q_0, \hat{q}_1) - c_0 + (p_0 - \hat{q}_1) b(q_0, \hat{q}_1) \\ &= w(q_0, \hat{q}_1) + p_0 b(q_0, \hat{q}_1) - c_0, \end{aligned}$$

which concludes the proof.

Case 4: Reporting type $\mathbf{q} \in \Delta_1$ and choosing $e = 1$.

From standard single-crossing arguments, we have:

$$w(0, \varphi(p_0)) + \varphi(p_0)\, b(0, \varphi(p_0)) - c_1 \geq w(\mathbf{q}) + \varphi(p_0)\, b(\mathbf{q}) - c_1. \tag{40}$$

From condition (a), it follows that

$$w(0, \varphi(p_0)) + p_0 b(0, \varphi(p_0)) - c_0 = w(p_0, p_1) + p_0 b(p_0, p_1) - c_0$$

for all $(p_0, p_1) \in \Delta_0$. Moreover, since U is continuous, we have

$$\begin{aligned} w(0, \varphi(p_0)) + \varphi(p_0)\, b(0, \varphi(p_0)) - c_1 &= w(0, \varphi(p_0)) + p_0 b(0, \varphi(p_0)) - c_0 \\ &= w(p_0, p_1) + p_0 b(p_0, p_1) - c_0. \end{aligned}$$

Substituting in (40), we obtain

$$\begin{aligned} w(p_0, p_1) + p_0 b(p_0, p_1) - c_0 &\geq w(\mathbf{q}) + \varphi(p_0)\, b(\mathbf{q}) - c_1 \\ &\geq w(\mathbf{q}) + p_1 b(\mathbf{q}) - c_1, \end{aligned}$$

where the last inequality uses the fact that $p_1 \leq \varphi(p_0)$ (since $(p_0, p_1) \in \Delta_0$).

This concludes the possible deviations for types in Δ_0. Now, let $\boldsymbol{p} = (p_0, p_1) \in \Delta_1$ (i.e., $p_1 > \varphi(p_0)$). Again, the possible deviations can be grouped into 4 possible cases.

Case 5: Reporting type $\mathbf{q} \in \Delta_1$ and choosing $e = 1$.

This result follows from standard single-crossing arguments taking effort as fixed.

Case 6: Reporting type $\mathbf{q} \in \Delta_1$ and choosing $e = 0$.

From Case 3, the following condition holds:

$$w(p_0, \varphi(p_0)) + p_0 b(p_0, \varphi(p_0)) - c_0 \geq w(\mathbf{q}) + p_0 b(\mathbf{q}) - c_0. \tag{41}$$

Case 5 and condition (a) implies that

$$\begin{aligned} w(\boldsymbol{p}) + p_1 b(\boldsymbol{p}) - c_1 &\geq w(0, \varphi(p_0)) + \varphi(p_0)\, b(0, \varphi(p_0)) - c_1 \\ &= w(p_0, \varphi(p_0)) + p_0 b(p_0, \varphi(p_0)) - c_0. \end{aligned}$$

Then, inequality (41) yields

$$w(\boldsymbol{p}) + p_1 b(\boldsymbol{p}) - c_1 \geq w(\mathbf{q}) + p_0 b(\mathbf{q}) - c_0,$$

which concludes the proof of this case.

Case 7: Reporting type $\mathbf{q} \in \Delta_0$ and choosing $e = 0$.

Let $\varphi^{-1}(p_1) = \sup\{p_0 : \varphi(p_0) \leq p_1\}$. From Case 1, we have

$$w\left(\varphi^{-1}(p_1), p_1\right) + \varphi^{-1}(p_1)\, b\left(\varphi^{-1}(p_1), p_1\right) - c_0 \geq w(\mathbf{q}) + \varphi^{-1}(p_1)\, b(\mathbf{q}) - c_0. \tag{42}$$

From the continuity of U, we have

$$w\left(\varphi^{-1}(p_1), p_1\right) + \varphi^{-1}(p_1)\, b\left(\varphi^{-1}(p_1), p_1\right) - c_0 = w\left(\varphi^{-1}(p_1), p_1\right) + p_1 b\left(\varphi^{-1}(p_1), p_1\right) - c_1.$$

Substituting in inequality (42), yields

$$w\left(\varphi^{-1}(p_1), p_1\right) + p_1 b\left(\varphi^{-1}(p_1), p_1\right) - c_1 \geq w(\mathbf{q}) + \varphi^{-1}(p_1)\, b(\mathbf{q}) - c_0. \tag{43}$$

However, condition (a) implies that, for all $p_0 < \varphi^{-1}(p_1)$,

$$w\left(\varphi^{-1}(p_1), p_1\right) + p_1 b\left(\varphi^{-1}(p_1), p_1\right) - c_1 = w(\boldsymbol{p}) + p_1 b(\boldsymbol{p}) - c_1, \text{ and}$$

$$w(\mathbf{q}) + \varphi^{-1}(p_1)\, b(\mathbf{q}) - c_0 \geq w(\mathbf{q}) + p_0 b(\mathbf{q}) - c_0.$$

Substituting in (43), we obtain:

$$w(\boldsymbol{p}) + p_1 b(\boldsymbol{p}) - c_1 \geq w(\mathbf{q}) + p_0 b(\mathbf{q}) - c_0,$$

which concludes the proof of this case.
Case 8: Reporting type $\mathbf{q} = (q_0, q_1) \in \Delta_0$ and choosing $e = 1$.

Since $(p_1, p_1) \in \Delta_0$, standard single-crossing arguments establish that

$$w(p_1, p_1) + p_1 b(p_1, p_1) - c_0 \geq w(\mathbf{q}) + p_1 b(\mathbf{q}) - c_0.$$

Condition (e) yields:

$$w(0, p_1) + p_1 b(0, p_1) = w(p_1, p_1) + p_1 b(p_1, p_1).$$

Substituting in the previous inequality and subtracting Δc, we obtain:

$$w(0, p_1) + p_1 b(0, p_1) - c_1 \geq w(\boldsymbol{q}) + p_1 b(\boldsymbol{q}) - c_1.$$

However, from condition (d), we have

$$w(\boldsymbol{p}) + p_1 b(\boldsymbol{p}) - c_1 = w(0, p_1) + p_1 b(0, p_1) - c_1$$

for all $p_0 < \varphi^{-1}(p_1)$. Thus,

$$w(\boldsymbol{p}) + p_1 b(\boldsymbol{p}) - c_1 \geq w(\boldsymbol{q}) + p_1 b(\boldsymbol{q}) - c_1,$$

which concludes the proof.