Simultaneous Adverse Selection and Moral Hazard*

Daniel Gottlieb and Humberto Moreira†

First Version: August, 2011. This Version: May, 2017.

Abstract

We study a principal-agent model with moral hazard and adverse selection. Agents have private information about the distribution of outputs conditional on each effort and, possibly, the cost of 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. Unlike in pure adverse selection models, there is distortion everywhere: the region of types who exert high effort is contained in the interior of the first-best high-effort region. Under additional conditions, the optimal mechanism offers only finitely many contracts. Our model, therefore, provides a multidimensional screening rationale for the lack of rich menus of contracts observed in certain environments. We apply our framework to multidimensional generalizations of canonical models in insurance, regulation, and optimal taxation and show that it generates novel results.

*We thank Eduardo Azevedo, Vinicius Carrasco, Sylvain Chasing, Alex Edmans, Faruk Gul, Lucas Maestri, George Mailath, Roger Myerson, Stephen Morris, Luca Rigotti, Yuliy Sannikov, Jean Tirole, Rakesh Vohra, and seminar audiences at HEC Montreal, Johns Hopkins University, Princeton University, FGV, PUC-Rio, Universidad de Chile, University of Pennsylvania, University of Pittsburg/Carnegie Mellon University, the Wharton School, and the BYU Computational Public Economics, the 2013 LAMES, the 2013 SBE, the IWGTS 2014, and the 2014 ESEM meetings for comments and suggestions. Rafael Mourão provided outstanding research assistance. Gottlieb gratefully acknowledges financial support from the Dorinda and Mark Winkelman Distinguished Scholar Award. Moreira acknowledges CNPq for financial support.

†Gottlieb: Washington University in St. Louis dgottlieb@wustl.edu. Moreira: FGV/EPGE, humberto@fgv.br.

Contents

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, most of 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 not well understood.

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 outputs conditional on each action. There are two possible outputs. 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 output 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 choose the principal's preferred effort. Because effort is unobservable, the principal has to leave informational rents to prevent each type from pretending to be another type with a 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 exerting a different level of effort. Consequently, moral hazard introduces new features in the model through *binding global incentive constraints*. The optimal contracts are, therefore, remarkably different from the ones from pure adverse selection models.

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 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 when agents are risk neutral. Intermediate types 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

[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.

[2] Grossman and Hart (1983) characterize the solution of the pure moral hazard model when there are two outputs. However, apart from existence, they show that very little can be said about the optimal incentive scheme when there are more than two outputs. Accordingly, we focus on the two-output model but allow the agent to have general private information about the distribution of outputs and on the incremental cost of effort. In the Online Appendix I, we generalize our model to allow the agent to have private information about their cost of effort as well. In that case, types are three-dimensional vectors.

interior of the first-best region of high effort. Therefore, unlike pure adverse selection models with both one- and multi-dimensional types, the solution may involve *distortion at all points* (including the top).

It is well known that bunching is a robust property in multi-dimensional settings (Rochet and Choné, 1998). In our setting, the informational rents required to prevent an agent from deviating can be so high that the optimal mechanism offers the agent a very limited number of contracts. 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 output 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.*

Many real-world contracts are tremendously simple. Differently from the predictions of standard adverse selection models, contracting parties offer a limited number of contracts. Moreover, unlike the predictions of standard moral hazard models, similar contracts are offered in fundamentally different environments. As Hart and Holmstrom and Chiappori and Salanie argue in their surveys of the literature:

> The extreme sensitivity to informational variables that comes across from this type of modeling is at odds with reality. Real world schemes are simpler than the theory would dictate and surprisingly uniform across a wide range of circumstances. *(Hart and Holmstrom, 1987, pp. 105)*

> The recent literature ... provides very strong evidence that contractual forms have large effects on behavior. As the notion that "incentives matter" is one of the central tenets of economists of every persuasion, this should be comforting to the community. On the other hand, it raises an old puzzle: if contractual form matters so much, why do we observe such a prevalence of fairly simple contracts? *(Chiappori and Salanie, 2003 , pp. 34)*

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 opportunities for gaming. 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 outputs given efforts is observable (pure moral hazard), the principal is able to perfectly design a 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 outputs are unobservable, large menus of contracts give the agents too many possible deviations, which requires the principal to leave large informational rents. Offering fewer contracts can be an efficient way to prevent gaming by the agents. In fact, in some cases, these informational rents are so large that the optimal mechanism offers the same contract to all agents.

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.[3] Moreover, as in their work, we also contribute to the applied literature by identifying assumptions under which researchers can focus on a simpler set of contracts when solving their models.

Our framework builds on the principal-agent model of Grossman and Hart (1983), which has a natural interpretation in terms of employment relationships. However, we illustrate its applicability beyond this canonical model by considering models of insurance and procurement/regulation featuring both adverse selection and moral hazard.[4] Our model generates new features relative to the one-dimensional pure-adverse-selection models that are the benchmarks in these literatures.

Empirical work in insurance has shown that simultaneous moral hazard and adverse selection is a key feature of many markets.[5] We show that the joint presence of adverse selection and moral hazard substantially changes the conclusions of standard insurance models. For example, the existence of a substantial uninsured population is a major policy issue and lied at the heart of the recent health care reform. We show that *exclusion is always optimal* in our model of insurance. This exclusion result differs from the first-best exclusion condition in the canonical principal-agent model because the reservation utility in insurance is type-dependent. The optimality of exclusion in our model is a consequence of the multidimensionality of types; it contrasts with one-dimensional models where exclusion is not optimal if there are "enough low types" in the population (Stiglitz, 1977; Chade and Schlee, 2012). Thus, our model suggests that the existence of a mass of uninsured consumers is a general property of insurance markets when both adverse selection and moral hazard are present.

In standard moral hazard models, insurance companies offer partial insurance in order to induce consumers to engage in preventive effort. Therefore, contracts in which a partially-insured consumer shirks are (constrained) Pareto inefficient. When adverse selection is also present, it is optimal to offer partial insurance to a mass of types who shirk. Thus, shirking by partially-insured consumers does not necessarily imply that contracts are sub-optimal. We also show that, because of moral hazard, policyholders *under-provide effort* in the sense that the second-best high-effort region is strictly contained in the region of high effort in the absence of insurance.

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, procurement/regulation, and optimal taxation.

[3] 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 (Forthcoming) 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.

[4] In Appendix B, we also present an application of our model to optimal taxation and discuss its relationship with that literature.

[5] See, for example, Karlan and Zinman (2009), Bajari et al. (2012), and Einav et al. (2013).

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. Although no general method for this class of problems exists, we 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) establishes that it is generically optimal to exclude a positive mass of buyers with low valuations. Rochet and Choné (1998) show 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 establish that bunching is 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 do not hold 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 when agents are risk neutral. 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").[6]

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) argues 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) introduce preventive effort in the seminal model of Rothschild and Stiglitz (1976) and characterize the set of separating equilibria. De Meza and Webb (2001) and Jullien et al. (2007) consider models where consumers have private information about their risk aversion and may engage in preventive effort and show that the correlation between risk and coverage may be negative.[7] Similarly, Chiu and Karni

[6]Laffont et al. (1987) consider a natural departure from the nonlinear pricing models of Mussa and Rosen (1978) or Maskin and Riley (1984), by assuming that agents have quadratic utility functions (linear demands) and types are two-dimensional. Rochet and Stole (2002) introduce 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).

[7]In De Meza and Webb (2001), there is a risk-neutral and a risk-averse type of consumer, and insurance firms have positive administrative costs. Jullien et al. (2007) study consumers with CARA utilities and show that the power of

(1998) present 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) study a model where policyholders consume products that affect their loss probabilities and insurers do not observe their tastes for such products. While these papers study 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 the relevant binding constraints and provides a clearer representation of the richness of the incentive problem.[8]

Our paper also contributes to the literature on procurement and regulation. The classic model of Laffont and Tirole (1986, 1993) has 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] We extend their canonical model by allowing effort to affect the regulated firm's costs stochastically, so the regulator's incentive problem cannot be reduced to a pure adverse selection model.[10] The optimal mechanism is then remarkably different.

The structure of the paper is as follows. Section 2 presents the basic framework and Section 3.1 derives some general properties of the solution. Section 3.2 then characterizes the solution and establishes several additional properties under the assumption of risk neutrality, and Section 3.3 obtains conditions under which the mechanism can be implemented with finitely many contracts. Section 4 applies our framework to multidimensional models of insurance (4.1). Then, Section 5 concludes.

Several generalizations and extensions are presented in appendices. Appendix A generalizes the characterization from Section 3.2 to settings where agents may be risk averse. Appendix B applies our framework to an optimal taxation model. For expositional simplicity, the main text focuses on the setup in which the agent's private information concerns his conditional distributions of outputs only. In the Online Appendix I, we generalize the model to allow the agent to have private information about his cost of effort as well.[11]

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 the principal does not observe. The principal does, however, observe the output from the partnership $x \in \{x_L, x_H\}$, which is stochastically affected by the agent's effort. Let

incentives decreases with risk aversion.

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

[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] In the Online Appendix I, we also allow the manager's cost of effort to be private information and show that our results persist.

[11] The benchmark cases of pure moral hazard and pure adverse selection are presented in the Online Appendix II. We present a method for calculating the optimal mechanisms numerically in the Online Appendix III.

p_e denote the probability of output x_H given effort e. We refer to x_H and x_L as high and low outputs, $e = 1$ and $e = 0$ as high and low efforts, and we refer to $\Delta x := x_H - x_L > 0$ as the incremental output.

The agent has private information about the conditional distribution of outputs. Therefore, the agent's type $\mathbf{p} := (p_0, p_1)$ is a vector of conditional probabilities of a high output given each effort. 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 output: $p_1 \geq p_0$. Under MLRP, the type space is contained in the area above the 45-degree line in Figure 1. Let $\boldsymbol{P} := \left\{(p_0, p_1) \in \mathbb{R}^2 : 1 \geq p_1 \geq p_0 \geq 0\right\}$ denote the space of types satisfying MLRP. We assume that the distribution of types f has full support on $\boldsymbol{P}$. Types on the 45-degree line will play a key role in our analysis. Since they have the same output distributions conditional on both efforts, they are not subject to moral hazard. We will refer to them as *diagonal types*.[12]

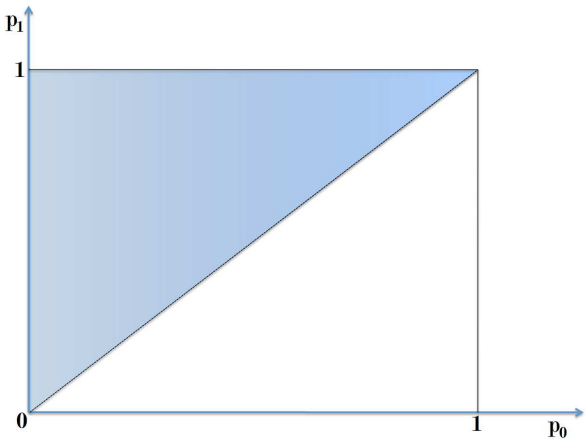

Figure 1: *Type Space (shaded area).*

The agent's utility function is additively separable in money and effort: $u(M) - c(e)$, where the utility from money u is continuously differentiable, increasing, and weakly concave, and the marginal utility function $\dot{u}$ is bounded. The low effort costs zero and the high effort costs $C > 0$:

$$c(e) = \begin{cases} 0 & \text{if } e = 0 \\ C & \text{if } e = 1 \end{cases}.$$

Throughout the main text, we assume that the cost of effort is commonly known. The Online Appendix I generalizes our results to environments in which the agent also has private information about the cost of effort.

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 $W : \boldsymbol{P} \to \mathbb{R}$, a bonus function $B : \boldsymbol{P} \to \mathbb{R}$, and an effort recommendation function $e : \boldsymbol{P} \to \{0, 1\}$. 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.

[12] It is immediate to generalize our results for distributions that do not satisfy MLRP as long as their support contains $\boldsymbol{P}$, by projecting types outside $\boldsymbol{P}$ onto the 45-degree line.

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 $(w, b, e) : \boldsymbol{P} \to \mathbb{R}^2 \times \{0, 1\}$, 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 gets 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 agent's *informational rent*. The agent follows honest and obedient strategies if the following *incentive-compatibility* constraint holds:

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

A mechanism satisfies *individual rationality* if the following participation constraint is satisfied:[13]

$$U(\mathbf{p}) \geq u(0), \quad \forall \mathbf{p} \in \boldsymbol{P}. \tag{IR}$$

A mechanism satisfies *free disposal* if the following monotonicity constraint holds:

$$B(\mathbf{p}) \geq 0, \quad \forall \mathbf{p} \in \boldsymbol{P}. \tag{FD}$$

Free disposal arises if the agent can can costlessly reduce output, or if the principal can secretly borrow from an outside lender in order to inflate output.[14] A mechanism is *feasible* if it satisfies incentive compatibility, individual rationality, and free disposal.

Given a mechanism (w, b, e), the principal gets expected utility

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

Two mechanisms are *equivalent* if they give the same expected utility to the principal and all agent types. A mechanism is *trivial* if it recommends low effort to almost all types.

2.2 Feasibility

In this subsection, we obtain necessary and sufficient conditions for a mechanism to be feasible. First, we establish 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:[15]

[13] This participation constraint assumes that reservation utilities are type independent. In Section 4, we allow for type-dependent reservation utilities in order to study optimal insurance contracts.

[14] Many principal-agent models assume free disposal, including Innes (1990), Acemoglu (1998), Matthews (2001), Dewatripont et al. (2003), Poblete and Spulber (2012), and Chaigneau et al. (2014).

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

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* $p_1 > \mathcal{E}(p_0)$ *for a continuous and non-decreasing function* $\mathcal{E} : [0,1] \to [0,1]$.

Lemma 1 follows from the monotonicity and the continuity of the agent's informational rent. For a given feasible mechanism (w, b, e), we refer to the function $\mathcal{E}$ as the *effort frontier* associated with it.[16] The effort frontier partitions the type space into types who exert low and high efforts:

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

The next lemma establishes necessary conditions for incentive compatibility:

Lemma 2. *Let* (w, b, e) *be a feasible mechanism and let* $\mathcal{E}$ *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*

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

b. $b(p_0, p_1)$ *is constant in* p_1 *for* $p_1 < \mathcal{E}(p_0)$ *and constant in* p_0 *for* $p_1 > \mathcal{E}(p_0)$*;*

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

d. $U(p_1, p_1) = U(p_0, p_1) + C$ *for* $p_1 > \mathcal{E}(p_0)$.

The incentive-compatibility constraints from adverse selection state that reporting one's 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 from this maximization program. Property (c) follows from the participation and free disposal constraints.

While conditions (a)-(c) are implied by adverse selection alone, moral hazard introduces additional incentive-compatibility constraints. In particular, under moral hazard, satisfying the local incentive 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. Property (d) is a necessary condition to prevent global deviations. Because effort is costly and diagonal types have the same conditional distribution over outputs under both high and low efforts, they always pick low 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., $p_1 > \mathcal{E}(p_0)$). Then, as Property (d) states, they get the same utility net of the cost of effort. Properties (a) and (d) imply that, for almost all types in the high-effort region, the contract power is the same as the diagonal type with the same probability of success: $b(p_0, p_1) = b(p_1, p_1)$ for almost all (p_0, p_1) such that $p_1 > \mathcal{E}(p_0)$.

In models of pure adverse selection, (a)-(c) are also sufficient conditions for feasibility. Moral hazard introduces a new necessary condition: Property (d). We now establish that these necessary conditions are also sufficient (given the conventions from footnotes 15 and 16).

[16] Due to the equivalence result of Lemma 1, we focus on mechanisms for which an effort frontier function $\mathcal{E}$ 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).

Lemma 3. *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)-(d) for a continuous and non-decreasing effort frontier function $\mathcal{E}$ satisfying condition (3).*

In the next subsection, we will use these conditions to rewrite feasible mechanisms as one-dimensional objects, which will allow us to characterize optimal mechanisms.

2.3 One-Dimensional Conditions

Fix a mechanism with informational rent U and let $\mathcal{U}(t) := U(t, t)$ denote its *rent projection*. The rent projection associated with the mechanism is a one-dimensional function that specifies the informational rents for all diagonal types. The following lemma establishes that any feasible mechanism is characterized by its rent projection:[17]

Lemma 4. *Let (w, b, e) be a feasible mechanism and let $\mathcal{E}$ and $\mathcal{U}$ denote the effort frontier and rent projection functions associated with it. Then:*

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

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

$$\mathcal{U}(\mathcal{E}(p_0)) = \min\{\mathcal{U}(p_0) + C;\ \mathcal{U}(1)\}. \tag{6}$$

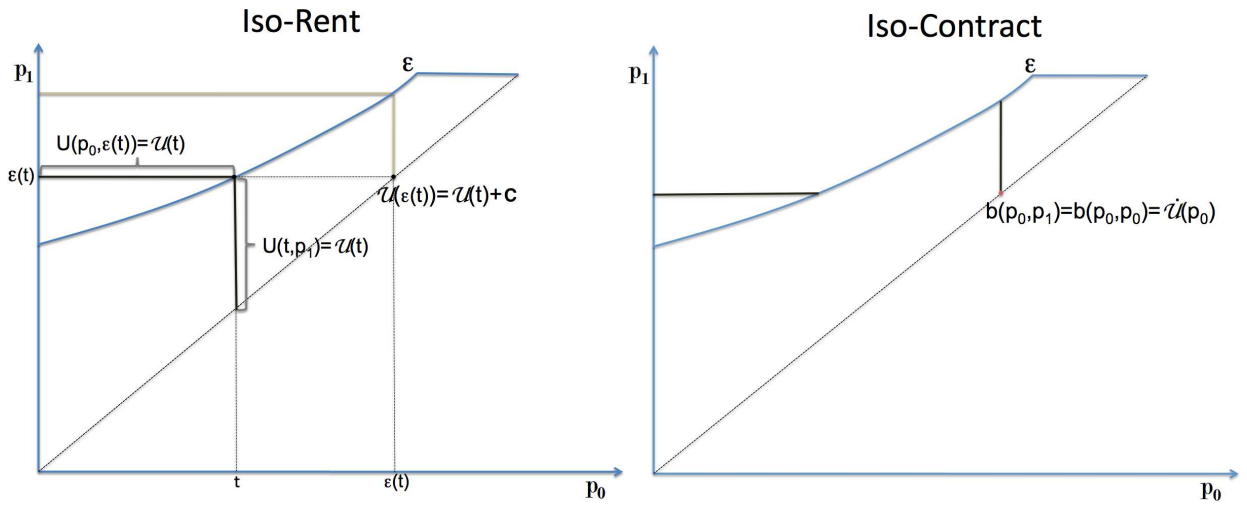

Figure 2: *Types with the same informational rent ('iso-rent') and with the same contract ('iso-contract').*

Using Lemma 4, we can recover the entire two-dimensional mechanism from its one-dimensional rent projection. Equation (6) shows how to obtain the effort frontier from the rent projection. Along the effort frontier, types are indifferent between high and low efforts. By Property (a), rents are constant

[17] 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).

along vertical segments in the low-effort region and along horizontal segments in the high-effort region. Moreover, by Property (d), the rents of types in the high effort region equal those of diagonal types net of the effort cost C. Thus, as Figure 2 illustrates, the effort frontier is obtained by finding the diagonal types whose rents differ by C.[18]

Equation (4) allows us to determine the contract powers from the rent projection. By incentive compatibility, two types with the same contract power b must also have the same fixed payment w. By Property (a), the derivative of the rent projection $\dot{\mathcal{U}}(p_0)$ equals the power of the contracts of diagonal types $b(p_0, p_0)$. Moreover, in the low-effort region, types in the same vertical line get the same contract (Property (b)) and the contract of a diagonal type equals the contract of types in the high effort region with the same probability of success given high effort (Properties (a) and (d)). Thus, the iso-contract curve is a horizontal line segment in the high-effort region and a vertical line segment in the low-effort region. That is, all types with the same probability of success given the (endogenous) recommended effort get the same contract. By Property (a), iso-rent curves have an inverted-L shape with the kink at the effort frontier. Then, using the definition of the informational rent (1), we can recover the fixed component of the mechanism w.

It is more convenient to work with the one-dimensional function $\mathcal{U}$ rather than the original two-dimensional mechanism (w, b, e). We will establish that a mechanism is feasible if and only if its associated rent projection is non-decreasing and convex. Let $\bar{u} := \sup_{x \in \mathbb{R}} u(x)$ denote the highest possible utility attainable to the agent (possibly $+\infty$). It is convenient to introduce the following definition:

Definition 1. *A function $\mathcal{U} : [0,1] \to [0, \bar{u}]$ is called a feasible rent projection if it is non-decreasing and convex.*

The following lemma establishes the equivalence between the feasibility of a mechanism and the feasibility of its rent projection:

Lemma 5. *Let (w, b, e) be a feasible mechanism, and let $\mathcal{U}$ and $\mathcal{E}$ be the rent projection and effort frontier functions associated with it. Then, $\mathcal{U}$ is a feasible rent projection and $(\mathcal{U}, \mathcal{E})$ solves equation (6). Conversely, let $\mathcal{U}$ be a feasible rent projection, suppose that $(\mathcal{U}, \mathcal{E})$ solves equation (6). Let (w, b, e) be given by equations (3), (4) and (5). Then, (w, b, e) is a feasible mechanism.*

Lemma 5 allows us to substitute the feasibility conditions (a)-(d) by conditions on the one-dimensional objects $\mathcal{U}$ and $\mathcal{E}$.[19] In order to characterize optimal mechanisms, we need to rewrite the principal's

[18] When no such type exists (i.e., all diagonal types to the right of p_0 obtain utility lower than $\mathcal{U}(p_0) + C$), all types in the vertical line segment above (p_0, p_0) exert low effort: $\mathcal{E}(p_0) = 1$. This projection method resembles the technique that Laffont et al. (1987) use to determine the boundary condition of the partial differential equation that characterizes incentive-compatible mechanisms in their model.

[19] 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, instruments have the same dimensionality as the type space – namely, there are two instruments (bonus and effort) and types are two dimensional. However, the global moral hazard constraint reduces the dimensionality of the instrument to one through the one-dimensional projection method (i.e., the bonus offered to agents with the same probability of success has to be the same regardless of the effort being made). Laffont et al. (1987) consider a model with two-dimensional types and one-dimensional instruments in which only local incentive constraints are binding.

expected utility (1) in terms of these objects. Let G denote the cost of providing expected utility $\mathcal{U}$ and power $\dot{\mathcal{U}}$ to an agent with probability of success t:

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

Substituting $\mathcal{U}$ and $\mathcal{E}$ in the principal's expected utility (1), yields

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

Applying Fubini's theorem, this expression becomes

$$\begin{aligned}&x_L+\int_0^1\int_t^{\mathcal{E}}\left(t\Delta x-G(\mathcal{U},\dot{\mathcal{U}},t)\right)f\left(t,s\right)dsdt+\int_{\mathcal{E}(0)}^1\int_0^{\mathcal{E}^{-1}}\left(t\Delta x-G(\mathcal{U},\dot{\mathcal{U}},t)\right)f\left(s,t\right)dsdt\\&=x_L+\int_0^1\left(t\Delta x-G(\mathcal{U},\dot{\mathcal{U}},t)\right)F_0\left(t,\mathcal{E}\right)dt+\int_{\mathcal{E}(0)}^1\left(t\Delta x-G(\mathcal{U},\dot{\mathcal{U}},t)\right)F_1\left(\mathcal{E}^{-1},t\right)dt,\end{aligned} \tag{8}$$

where $F_0(t,s) := \int_t^s f(t,z)dz$ and $F_1(s,t) := \int_0^s f\left(z,t\right)dz$, and we are omitting the dependence of the functions $\mathcal{U}$, $\mathcal{E}$ and $\mathcal{E}^{-1}$ on t for notational simplicity.

Although one-dimensional, these programs differ from those from standard 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 diagonal type t to its projection along the effort frontier $\mathcal{E}\left(t\right)$. 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. Since each agent type can pretend to be a 'distant' type by choosing a different effort, these global constraints capture the moral hazard dimension of the problem.[20]

The following proposition establishes the existence of optimal mechanisms.

Proposition 1 (**Existence**). *There exists an optimal mechanism.*

3 Optimal Mechanisms

3.1 General Properties

This subsection presents general properties of optimal mechanisms. Our first proposition establishes that a positive mass of agents do not receive any informational rents:

Proposition 2 (**Zero Rents at the Bottom**). *No mechanism that gives strictly positive informational rents for almost all types is optimal.*

[20]Formally, although the utility function satisfies the single crossing, moral hazard introduces binding global constraints because effort is discrete. In principle, in a framework with continuous effort, it is possible that only local constraints matter. However, even in the pure moral hazard case, the conditions for global incentive constraints not to bind are excessively strong and are not satisfied by any standard output distribution (Rogerson, 1985). Therefore, we conjecture that, even in models with continuous efforts, global incentive constraints will still bind.

Because the rent projection function is nondecreasing, Proposition 2 implies that there exists $\underline{t} > 0$ such that $\mathcal{U}(t) = 0$ for $t \leq \underline{t}$ and $\mathcal{U}(t) > 0$ for $t > \underline{t}$. 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: $w = b = 0$. Then, equation (6) implies that the effort frontier is flat for $t < \underline{t}$ – i.e., $\mathcal{E}(t) = \underline{\mathcal{E}}$ for all $t \in [0, \underline{t}]$ for some $\underline{\mathcal{E}} > \underline{t}$. Figure 3 depicts these results graphically.

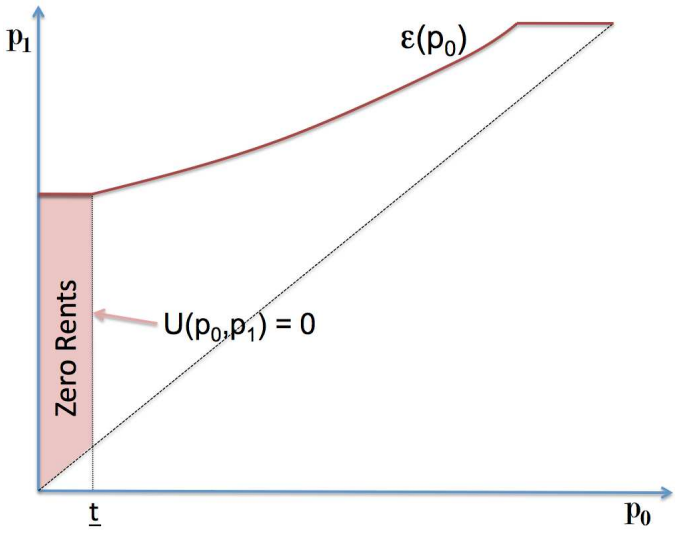

Figure 3: *Zero Rents at the Bottom: Types with $p_0 \leq \underline{t}$ and $p_1 \leq \underline{\mathcal{E}}$ are offered the zero-power contract and get zero rents.*

Our next result concerns the slope of the effort frontier $\mathcal{E}$. The first best effort frontier has a unit slope. By equation (6), the effort frontier in any feasible mechanism satisfies

$$\mathcal{U}(\mathcal{E}(t)) - \mathcal{U}(t) = C$$

for all diagonal types t in which there is effort, i.e., $\mathcal{E}(t) < 1$. The convexity of the rent projection $\mathcal{U}$ then implies that the slope of $\mathcal{E}(t)$ is less than one. That is, the effort frontier function in any feasible mechanism is flatter than the first-best effort frontier. Moreover, by Proposition 2, the effort frontier in any optimal mechanism is flat for t low enough. We formally state this result in the following lemma:

Lemma 6. *Let (w, b, e) be a optimal mechanism and let $\mathcal{E}$ be the effort frontier function associated with it. Then, $\mathcal{E}$ is Lipschitz with constant 1. Moreover, there exists $\underline{t} > 0$ such that $\mathcal{E}(t) = \underline{\mathcal{E}}$ for all $t \leq \underline{t}$.*

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 contract that dominates their reservation utility. We now consider the desirability of exclusion.

Let $\pi(\mathbf{p}) \in \{0, 1\}$ denote the agent's participation decision: type $\mathbf{p}$ does not participate in the mechanism and gets zero utility if $\pi(\mathbf{p}) = 0$, and he participates and gets the utility specified in equation (1) if $\pi(\mathbf{p}) = 1$. 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 gets expected utility

$$U(\mathbf{p}) \equiv \pi(\mathbf{p}) \left[w(\mathbf{p}) + p_{e(\mathbf{p})} b(\mathbf{p}) - c(e(\mathbf{p})) \right], \tag{9}$$

and the principal gets expected utility

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

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 (9). All previous results can be adjusted to the 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, consider first the exclusion rule under perfect information. From the first-best effort region, the principal's expected utility when contracting with type (p_0, p_1) is

$$\max \left\{ x_L + p_0 \Delta x - u^{-1}(0) \, ; \; x_L + p_1 \Delta x - u^{-1}(C) \right\}.$$

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 the lowest type: $(0, 0)$.

When types and effort are not observable, informational rents are non-decreasing in the agent's type. Thus, the principal can only exclude an agent type if all types below him (ordered by their projections on the 45-degree line) also get zero rents. Because the lowest types in the optimal mechanism get zero rents (Proposition 2), the principal can recommend that they do not participate at zero costs. As a result, 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 (1996) for multidimensional screening in the context of a multiproduct monopolist. It strongly relies on the assumption of type-independent reservation utility. We return to this issue when we consider an insurance application (Subsection 4.1), where the reservation utility is type-dependent and exclusion is optimal.[21]

3.2 Risk Neutrality

This section characterizes optimal non-trivial mechanisms when the agent is risk neutral: $u(X) = X$. The optimal mechanism balances effort distortions against informational rents left to the agent. In Appendix A, we generalize the characterization from this section to weakly concave utility functions.

[21] 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.

Let $\mathcal{U}$ be a feasible rent projection and let $\mathcal{E}$ be the effort frontier associated with it. As before, let $\underline{t} := \sup\{t : \mathcal{U}(t) = 0\}$ denote the lowest diagonal type to get positive rents. Let $\underline{\mathcal{E}} := \mathcal{U}^{-1}(C)$ denote the lowest probability of success in the high-effort region, and let $\overline{t} := \inf\{t : \mathcal{E}(t) = 1\}$ denote the point at which the effort frontier hits $p_1 = 1$ (see Figure 4). Let $[\mathcal{E}(t) - t]\Delta x - C$ denote the 'effort distortion at point t.' This term is zero if the mechanism implements the first-best effort frontier at t. It is positive if there is less effort than in the first best and negative if there is more effort than in the first best.

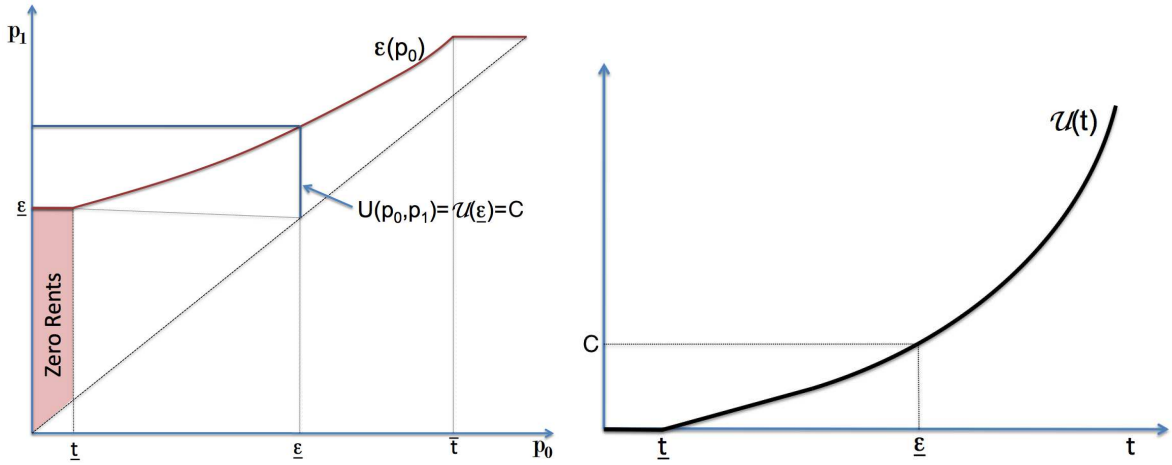

Figure 4: *Effort Frontier Function $\mathcal{E}$ (left) and Rent Projection Function $\mathcal{U}$ (right).*

We will first present a heuristic derivation of the optimality conditions and then state them formally. Suppose we increase the rent projection $\mathcal{U}$ by a "small" amount in a neighborhood of $t > \underline{t}$. Recall that iso-rents have an inverted-L shape with the kink at the effort frontier (see Figure 2). It is instructive to consider the effect on types in the low- and high-effort regions separately. In each case, there is an effect on the effort frontier ('marginal effect') and an effect on types who do not change their effort choices but obtain higher rents ('inframarginal effect').

Consider first the effect on the low-effort region (see graph on the left in Figure 5). Type $(t, \mathcal{E}(t))$ is indifferent between exerting high and low efforts (we will omit t from $\mathcal{E}(t)$ for notational simplicity). Exerting high effort yields expected payoff $\mathcal{U}(\mathcal{E}) - C$, whereas exerting low effort yields $\mathcal{U}(t)$. If we increase $\mathcal{U}(t)$ while leaving $\mathcal{U}(\mathcal{E})$ constant, type $(t, \mathcal{E})$ will strictly prefer to exert low effort. The type who will now be indifferent between high and low efforts, $\left(t, \hat{\mathcal{E}}\right)$, will be above the original one: $\hat{\mathcal{E}} > \mathcal{E}(t)$. Therefore, an increase in the rent projection at t shifts the effort frontier up, reducing the effort region. Recall that, for $t < \overline{t}$, the effort distortion is $(\mathcal{E} - t)\Delta x - C$. The cost of increasing the effort frontier – the 'marginal effect' – is then captured by the distortion per unit of bonus paid to the marginal type $(t, \mathcal{E})$:

$$\frac{(\mathcal{E} - t)\Delta x - C}{\dot{\mathcal{U}}(\mathcal{E})}, \text{ for } t < \overline{t}.$$

Increasing the rent projection at t also involves leaving higher rents to all types in the vertical line segment between (t, t) and $(t, \mathcal{E})$, who still exert low effort but are paid more ('inframarginal effect'). The total mass of those types is $F_0(t, \mathcal{E})$. Since the marginal type $(t, \mathcal{E})$ has mass $f(t, \mathcal{E})$, the cost of leaving higher rents relative to the marginal type is captured by the hazard rate: $\frac{F_0(t,\mathcal{E})}{f(t,\mathcal{E})}$. The total effect on the

low-effort region is then

$$S_0(t,\mathcal{U}) := \begin{cases} -\frac{(\mathcal{E}-t)\Delta x - C}{\dot{\mathcal{U}}(\mathcal{E})} - \frac{F_0(t,\mathcal{E})}{f(t,\mathcal{E})} & \text{if } t < \overline{t} \\ -\frac{F_0(t,1)}{f(t,1)} & \text{if } t \geq \overline{t} \end{cases} \tag{10}$$

(with negative signs because both effects are costs).

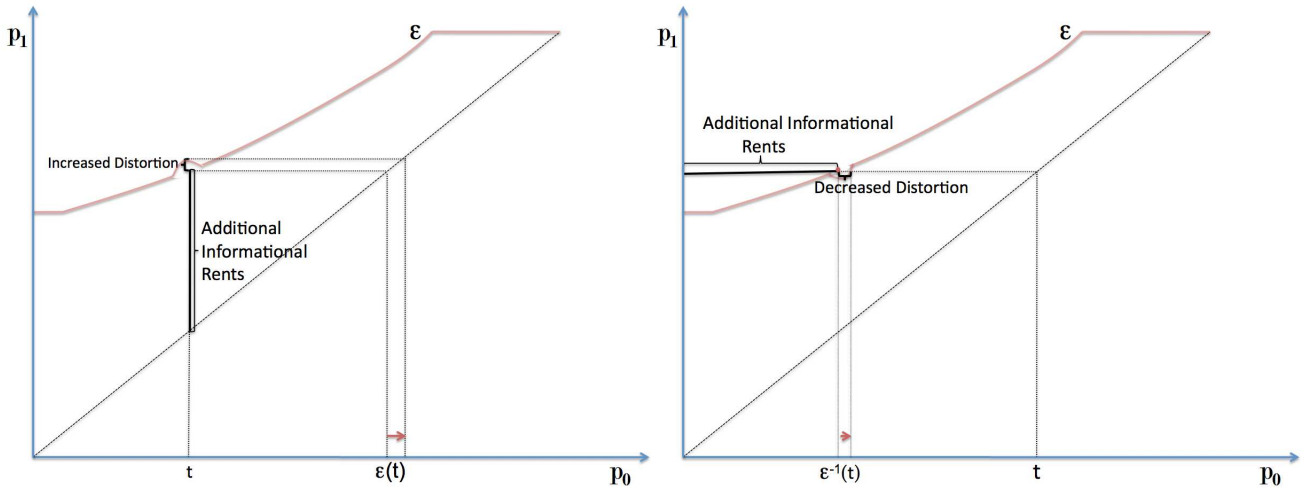

Figure 5: *Effect on the Low-Effort Region (left) and on the High-Effort Region (right).*

Now, consider the effect on the high-effort region (see graph on the right in Figure 5). Recall that, whenever $t > \underline{\mathcal{E}}$, type $\left(\mathcal{E}^{-1}, t\right)$ is indifferent between high and low efforts (when $t \leq \underline{\mathcal{E}}$, no type exerts high effort and this region is empty). His expected payoff from high effort is $\mathcal{U}(t) - C$, whereas his expected payoff from exerting low effort is $\mathcal{U}\left(\mathcal{E}^{-1}\right)$.

Raising $\mathcal{U}(t)$ while keeping $\mathcal{U}\left(\mathcal{E}^{-1}\right)$ unchanged makes type $\left(\mathcal{E}^{-1}, t\right)$ strictly prefers to exert high effort. Thus, the effort frontier shifts to the right (the type who will now be indifferent between both effort levels is $\left(\hat{\mathcal{E}}^{-1}, t\right)$ with $\hat{\mathcal{E}}^{-1} > \mathcal{E}^{-1}$), increasing the region of high effort. The benefit from shifting the effort frontier – i.e., the marginal effect – is the effort distortion per unit of bonus at the marginal type $\left(\mathcal{E}^{-1}, t\right)$:

$$\frac{(t - \mathcal{E}^{-1})\Delta x - C}{\dot{\mathcal{U}}(\mathcal{E}^{-1})}, \quad \text{for } t > \underline{\mathcal{E}}.$$

Increasing the rent projection at t, however, requires leaving rents to all types to the left of $\left(\mathcal{E}^{-1}, t\right)$, who still exert high effort but now obtain higher informational rents (inframarginal effect). The cost of leaving these rents is given by the mass of such inframarginal types relative to the marginal type:

$$\frac{F_1(\mathcal{E}^{-1}, t)}{f(\mathcal{E}^{-1}, t)}, \quad \text{for } t > \underline{\mathcal{E}}.$$

The total effect on the high-effort region is then:

$$S_1(t,\mathcal{U}) := \begin{cases} 0 & \text{if } t \leq \underline{\mathcal{E}} \\ \frac{(t-\mathcal{E}^{-1})\Delta x - C}{\dot{\mathcal{U}}(\mathcal{E}^{-1})} - \frac{F_1(\mathcal{E}^{-1},t)}{f(\mathcal{E}^{-1},t)} & \text{if } t > \underline{\mathcal{E}} \end{cases}. \tag{11}$$

Let $\mathcal{S}(t,\mathcal{U}) := S_0(t,\mathcal{U})\, f(t,\mathcal{E}) + S_1(t,\mathcal{U})\, f\left(\mathcal{E}^{-1},t\right)$ denote the sum of the effects on low- and high-effort regions weighted by their probability densities. $\mathcal{S}(t,\mathcal{U})$ captures the marginal payoff to the principal of increasing the rent projection $\mathcal{U}$ at point t.

Suppose, instead, that we increase the rent projection $\mathcal{U}$ by a "small" amount in a neighborhood of $\underline{t}$ (see Figure 6). Because all such types get zero rents, only the marginal effect remains (i.e., there is no inframarginal effect since there are no informational rents).

Notice that all types $(t,\underline{\mathcal{E}})$ with $t \leq \underline{t}$ get the same contract as $(\underline{\mathcal{E}},\underline{\mathcal{E}})$ and are indifferent between exerting high and low efforts. Thus, their expected payoff from high effort is

$$w(\underline{\mathcal{E}},\underline{\mathcal{E}}) + \underline{\mathcal{E}} b(\underline{\mathcal{E}},\underline{\mathcal{E}}) - C = \mathcal{U}(\underline{\mathcal{E}}) - C.$$

The payoff from low effort is zero – since, by Proposition 2, types (t,t) with $t \leq \underline{t}$ get zero rents. Therefore, an increase in $\mathcal{U}(\underline{\mathcal{E}})$ makes all those types strictly prefer to exert high effort, shifting down the effort frontier. As before, the gain from inducing type $(t,\underline{\mathcal{E}})$ to exert high effort is the ratio between the distortion at t, $(\underline{\mathcal{E}} - t)\,\Delta x - C$, and the power of that type's contract, $\dot{\mathcal{U}}(\underline{\mathcal{E}})$. Integrating the effect over all affected types, gives the marginal effect at $\underline{t}$:

$$\underline{S}(\mathcal{U}) := \frac{\{\underline{\mathcal{E}} - E[t|t \leq \underline{t},\underline{\mathcal{E}}]\}\,\Delta x - C}{\dot{\mathcal{U}}(\underline{\mathcal{E}})} \times F_1(\underline{t},\underline{\mathcal{E}}),$$

where $E[t|t \leq \underline{t},\underline{\mathcal{E}}] := \frac{\int_0^{\underline{t}} t f(t,\underline{\mathcal{E}})dt}{F_1(\underline{t},\underline{\mathcal{E}})}$. Notice that the hazard rate that appears in the expressions of S_0 and S_1 vanishes from $\underline{S}$ since these types do not get informational rents.

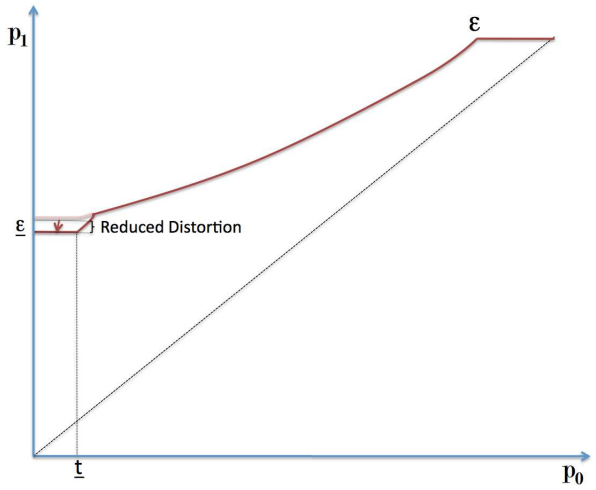

Figure 6: *Effect of a Perturbation at* $\underline{t}$.

Combining all the effects above, we can, in the spirit of Myerson (1981), define the *expected virtual surplus* as

$$\int_0^1 \mathcal{S}(t,\mathcal{U})\,\mathcal{U}(t)\,dt + \underline{S}(\mathcal{U})\,\mathcal{U}(\underline{\mathcal{E}}). \tag{12}$$

Our expected virtual surplus (12) 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 non-adjacent types with binding incentive constraints. The following lemma establishes that any optimal mechanism must maximize the expected virtual surplus among the class of feasible mechanisms.

Lemma 7. *Let $\mathcal{U}$ be an optimal rent projection. Then, for any feasible rent projection $\mathcal{V}$,*

$$\int_0^1 \mathcal{U}(t)\,\mathcal{S}(t,\mathcal{U})\,dt + \mathcal{U}(\underline{\mathcal{E}})\underline{S}(\mathcal{U}) \geq \int_0^1 \mathcal{V}(t)\,\mathcal{S}(t,\mathcal{U})\,dt + \mathcal{V}(\underline{\mathcal{E}})\underline{S}(\mathcal{U}).$$

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

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

- g is *strongly convex* in an interval $[t_1,t_2]\subset[0,1]$ if there exists $m>0$ such that $\dot{g}(y)-\dot{g}(x)\geq m(y-x)$ for all $x,y\in[t_1,t_2]$;
- g has a *kink* at $x_0\in(0,1]$ if $\lim\limits_{x\uparrow x_0}\dot{g}(x)\neq\dot{g}(x_0)$; and
- $[t_1,t_2]\subset[0,1]$ is called a *maximal interval where g is affine* if: *(i)* there exists $m\in\mathbb{R}$ such that $\dot{g}(x)=m$, for all $x\in[t_1,t_2]$, and *(ii)* there is no open interval containing $[t_1,t_2]$ such that $\dot{g}(x)=m$ 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 $[t_1,t_2]\subset[0,1]$, then $\mathcal{S}(t,\mathcal{U})=0$ for almost all $t\in[t_1,t_2]$.*
2. **(bunching conditions)** *Let $[t_1,t_2]\subset[0,1]$ be a maximal interval where $\mathcal{U}$ is affine.*

- *If $\underline{\mathcal{E}}\notin[t_1,t_2]$, then*

$$0\geq t_1\int_{t_1}^{t_2}\mathcal{S}(t,\mathcal{U})\,dt\geq\int_{t_1}^{t_2}t\mathcal{S}(t,\mathcal{U})\,dt\geq t_2\int_{t_1}^{t_2}\mathcal{S}(t,\mathcal{U})\,dt.$$

Moreover, if $\mathcal{U}$ has kink at t_1 (at t_2) and $t_2<1$, then $\int_{t_1}^{t_2}(t-t_1)\mathcal{S}(t,\mathcal{U})dt=0$ ($\int_{t_1}^{t_2}(t-t_2)\mathcal{S}(t,\mathcal{U})dt=0$).[22]

- *If $t_1=\underline{t}$ and $t_2\geq\underline{\mathcal{E}}$, then*

$$\int_{\underline{t}}^{t_2}\mathcal{S}(t,\mathcal{U})dt+\underline{S}(\mathcal{U})\leq 0 \text{ and } \int_{\underline{t}}^{t_2}(t-\underline{\mathcal{E}})\mathcal{S}(t,\mathcal{U})dt\leq 0.$$

Moreover, if $\mathcal{U}$ has kink at $t_2<1$, then

$$\int_{\underline{t}}^{t_2}\mathcal{S}(t,\mathcal{U})dt+\underline{S}(\mathcal{U})=0 \text{ and } \int_{\underline{t}}^{t_2}(t-\underline{\mathcal{E}})\mathcal{S}(t,\mathcal{U})dt=0.$$

[22] *If $t_2=1$ and $\dot{\mathcal{U}}(1)=\Delta x$, then the equalities become inequalities lower or equal.*

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, there exists 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, bunching is determined by the ironing principle, which can be obtained by considering perturbations to the interval of pooled types. Because our model has two-dimensional types, there are two 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 2, 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 $\mathcal{U}$ be an optimal rent projection of a nontrivial mechanism. There exist $\hat{\mathcal{E}} \geq \underline{\mathcal{E}}$ and constant $b > C$* such that *$\underline{t} \in (0, \underline{\mathcal{E}})$ and*

$$\dot{\mathcal{U}}(t) = \begin{cases} 0 \text{ if } t \in [0, \underline{t}) \\ b \text{ if } t \in [\underline{t}, \hat{\mathcal{E}}) \end{cases}.$$

Figure 7 illustrates the result from Proposition 4. Types with sufficiently low probability of success conditional on both low and high efforts ($p_0 \leq \underline{t}$ and $p_1 \leq \underline{\mathcal{E}}$) receive a constant zero payment and exert low effort (Region A). Region B comprises types with intermediate probabilities of success given low efforts. All types in this region are offered the same contract, which involves a payment with a lower fixed component $w < 0$ and power i greater than the cost of effort C.

Recall that, in general, an increase in the rent projection at t raises the effort frontier at point $\mathcal{E}^{-1}(t)$ (through the effect on the high-effort region S_1), reduces the effort frontier at point t, and increases informational rents left to all inframarginal types. Since, no types mapped into diagonal point $\underline{\mathcal{E}}$ exert high effort, the effect on the high-effort region (S_1) vanishes. Thus, the only remaining effects are the reduction of the effort frontier at t and the increase in informational rents left to inframarginal types who exert low effort: S_0. Since both effects are negative, the principal would like to reduce the rent projection as much as possible subject to convexity and the initial effort point $\underline{\mathcal{E}}$. This is achieved by a piecewise linear curve.

The intuition for this result is the following. All types projected into points to the left of $\underline{\mathcal{E}}$ on the 45-degree line exert low effort. Therefore, if we increase their informational rents, they will keep choosing a low effort, keeping the effort region at these points unchanged. However, increasing their informational rents incentivizes types above them to reduce their effort, thereby reducing the effort region at points above $\underline{\mathcal{E}}$. Since both the increase in informational rents and the increased distortion hurt the principal, she will want to leave as little informational rents as possible while preserving the condition that the effort frontier starts at $\underline{\mathcal{E}}$. 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 $\underline{t}$, the principal needs to pay a bonus greater than the incremental cost of effort in order to incentivize types $\left\{\left(t, \hat{\mathcal{E}}\right) : t \leq \underline{t}\right\}$ to exert high effort. The principal then reduces the informational rents left in this region by paying the same

bonus to all those types.

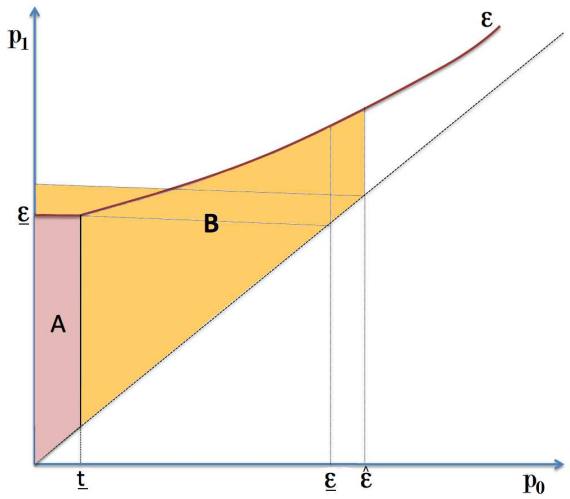

Figure 7: *Two Contracts at the Bottom: Types in Region A receive the same constant payment ($w = b = 0$); types in Region B receive the same contract ($w < 0$, $b > C$).*

We now assume the *bilateral free disposal*:

$$b(\mathbf{p}) \leq \Delta x, \quad \forall \mathbf{p} \in \boldsymbol{P}. \tag{BFD}$$

As Innes (1990) argues, condition (BFD) arises if the principal can reduce output at no cost, or if the agent can secretly borrow from an outside lender to inflate output. Bilateral free disposal (BFD) is equivalent to

$$\dot{\mathcal{U}}(t) \leq \Delta x \quad \forall t \in [0,1]. \tag{13}$$

Thus, a mechanism is BFD-optimal if and only if its associated rent projection and effort frontier functions maximize (8) subject to (6), $\mathcal{U}$ nondecreasing and convex, $\mathcal{U}(0) \geq 0$, and (13).

We now examine the effort distortion relative to the first best. Recall that the first-best effort region under risk neutrality is determined by $(p_1 - p_0)\Delta x \geq C$. 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 C. The first-best effort is implemented by making the agent a residual claimant: $b = \Delta x$. If bonuses are bounded above by the incremental output, $b \leq \Delta x$, the effort region in any mechanism that satisfies bilateral free disposal is contained in the first-best effort region.

We say that the a mechanism *partially sells the firm* if all types pick one of the following two contracts: $(0,0)$ and $(w, \Delta x)$, for some $w \leq 0$. Under a mechanism that partially sells the firm, agents self-select into two categories: "employees" who work for a fixed wage, exert low effort, and are indifferent between participating or not, and "entrepreneurs" who buy the firm for the price $-w$ and become residual claimants. Entrepreneurs choose effort efficiently. Unlike in pure moral hazard models, those with a high enough probability of success given low effort choose to exert low effort despite being offered a variable payment.

Recall that a mechanism has *insufficient effort* if its high-effort region is contained in the interior of the first-best effort region. The next lemma establishes that any optimal mechanism either partially sells the firm or features insufficient effort:

Lemma 8. *Let (w, b, e) be a BFD-optimal mechanism. Then, either there is insufficient effort, or the principal partially sells the firm.*

The intuition behind Lemma 8 is the following. Because distortions close to the optimum have second-order costs, it can only be desirable not to distort at one point if there is no other point with distortions and positive rents (otherwise, the principal can improve by rebalancing the distortions at these two points). Lemma 8 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 ($t \leq \underline{t}$) obtain distorted allocations.

The distortion of all projected types is a consequence of the global incentive constraint, which induces the principal to distort even the allocation of the highest types. Because only local incentive constraints bind in standard one- and multi-dimensional screening models, there is "no distortion at the boundary." In this model, because all types in the high-effort region have binding global incentive-compatibility constraints, the optimal mechanism "distorts the effort frontier at all points" whenever the bilateral free disposal constraint is non-binding, causing the effort region to be in the interior of the first-best effort region.[23]

3.3 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, because of moral hazard, the principal has a one-dimensional instrument, there has to be some bunching in our model. The interesting issue here is whether a *positive mass* of types get the same contract. For example, under 'pure moral hazard' (i.e., when types are observable but effort is not), if two types (p_0, p_1) and $(\hat{p}_0, \hat{p}_1)$ with $\hat{p}_1 \neq p_1$ both choose high effort, then they must pick different contracts. Thus, in the high-effort region, the set of types who get each contract has measure zero. If a strictly convex rent projection $\mathcal{U}$ solved the principal's program, each contract would be taken by the vertical and horizontal projections from Figure 2, which also have zero measure. However, Proposition 2 showed that the convexity constraint binds. As a result, regions of types with positive mass are offered the same contract (both in the regions of high and low effort). 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. We now show that, under some conditions, the optimal mechanism can be implemented with a reduced number of contracts.

[23] As we show in the Online Appendix, our distortion result can be strengthened. In that case, optimal mechanisms generically have a "distortion at all points," in the sense that, for generic distributions of types, the boundary of the effort region coincides with the boundary of the first-best effort region in at most one point. However, because bonuses can exceed the incremental output, it is possible that the optimal mechanism induces excessive effort from some types.

High Cost of Effort and Non-Decreasing Hazard Rate

Let $H(p_0, p_1) := \frac{F_0(p_1,1)+F_1(p_0,p_1)}{f(p_0,p_1)}$ denote the *generalized hazard rate*. 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

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

Because increasing rents allows H to decrease in p_1 as long as it is sufficiently increasing in p_0, it is weaker than strict monotonicity. The uniform distribution, for example, satisfies increasing rents. 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 $\mathcal{E}(p_0) = 1$:

Lemma 9. *Suppose that the distribution of types satisfies increasing rents. The optimal rent projection is a piecewise linear function with at most two pieces on* $\left[\overline{t}, 1\right]$.

The intuition behind Lemma 9 is the following. Recall that the marginal virtual surplus $\mathcal{S}$ consists of a distortion effect and an informational rent effect. By Lemma 6, the slope of the effort frontier is less than one, while the first-best frontier has a unit slope. Thus, the effort distortion is decreasing in t. Under increasing rents, the informational rents are strictly decreasing in t. Consequently, the marginal virtual surplus is strictly decreasing, implying that the principal's benefit from leaving rents decreases in t.

Consider a feasible rent projection that is strictly convex in an interval. Since the marginal virtual surplus is strictly decreasing, there are three possible cases: it may be always positive, always negative, or initially positive and then negative. In all of these cases, it is possible to increase the expected virtual surplus by replacing the original strictly increasing bonus by a piecewise linear one that preserves incentive compatibility. For example, suppose the marginal virtual surplus is negative in the entire interval $[\overline{t}, 1]$. Replacing the rent projection by the piecewise linear function consisting of the maximum of the tangents of the original rent projection at $\overline{t}$ and 1 preserves feasibility. Since this function lies strictly below the original rent projection and the marginal virtual surplus is negative, it attains a higher expected virtual surplus.

In sum, the increasing rents assumption ensures that the principal's benefit from distorting allocations is decreasing in t, implying that the optimal rent projection consists of a bang-bang solution in the interval $[\overline{t}, 1]$. Since the bonus is the slope of the rent projection, there are at most two contracts offered in this interval. Recall that, by Proposition 4, the principal offers two contracts in the interval $[0, \underline{\mathcal{E}}]$ (see Figure 7). The next proposition establishes that $\overline{t} \leq \underline{\mathcal{E}}$ when the incremental output Δx is "not too large" relative to the cost of effort C. Then, these regions overlap and the optimal mechanism features at most three contracts:

Proposition 5 (**Three Contracts**). *Suppose that the distribution of types satisfies increasing rents and let* $\Delta x \leq 2C$. *Then, the optimal mechanism can be implemented with at most three contracts.*

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

Corollary 1 (**Uniform Distribution**). *Suppose that types are uniformly distributed on* $\mathbf{P}$ *and let* $\Delta x \leq 3C$. *Then, the optimal mechanism can be implemented with a finite number of contracts.*

In the Online Appendix III , we present a numerical method for computing the solution of our model. Applying our method to the uniform distribution, we find that, under the conditions of Corollary 1, the optimal mechanism has at most *two contracts*. There is always the fixed-wage contract ($w = b = 0$). Moreover, when $\frac{\Delta x}{C}$ is sufficiently large – i.e., effort is valuable enough –, there is also a contract with a positive bonus ($w < 0,\ b > 0$). In fact, our numerical results from the Online Appendix III show that, for the uniform distribution, offering a small number of contracts is optimal even when $\Delta x > 3C$ (so the condition from Corollary 1 fails to hold). For example, when $\Delta x = 100C$, the optimal mechanism offers four contracts.

Probability of Success Bounded Away from Zero

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

$$\boldsymbol{P}\left(\underline{p}\right) = \left\{(p_0, p_1) \in \boldsymbol{P} :\ \underline{p} \leq p_0 \leq p_1\right\},$$

where $\underline{p} \in [0, 1)$, and we assume that the distribution of types $f(p_0, p_1)$ has full support on $\boldsymbol{P}\left(\underline{p}\right)$. It is straightforward to adapt our previous characterization for this modified type space.

Proposition 6 (**Two Contracts**). *Suppose* $f(p_0, p_1)$ *is non-increasing in* p_0, *and let* $\underline{p} \geq \frac{\Delta x - C}{\Delta x + C}$. *Then, the optimal mechanism can be implemented with at most two contracts.*

Propositions 5 and 6 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 5) or is "sufficiently bounded away from zero" (Proposition 6), the principal prefers to offer a small number of contracts, reducing the informational rents that have to be left to the agent.

4 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, and procurement and regulation.

4.1 Insurance

Unlike the framework considered previously, insurance models typically have type-dependent participation constraints since riskier types have a lower opportunity cost of remaining uninsured. In this

subsection, we drop the type-independence assumption to study the provision of insurance by a monopolist.[24]

Consider a monopolistic insurance firm (principal) that offers insurance to consumers (agents) who have 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. 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). The insurance firm has a continuous prior distribution f over types with full support on the set of distributions satisfying MLRP: $\boldsymbol{P}$. A type-(p_0, p_1) consumer who does not purchase insurance gets expected utility

$$V(p_0, p_1) := \max_{e \in \{0,1\}} p_e u(I) + (1 - p_e) u(I - L) - c(e).$$

We assume that policies satisfy bilateral free disposal, so that indemnities are non-negative and do not exceed the value of the loss:

$$0 \leq B(p_0, p_1) \leq L, \quad \text{for all } (p_0, p_1) \in \boldsymbol{P}. \tag{14}$$

The first inequality must be satisfied if consumers can hide a loss from the insurance company, in which case indemnity payments cannot be negative. The second inequality must hold if consumers can costlessly generate a loss, so that 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 9), we obtain the following participation constraint for the insurance model:

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

Thus, an insurance mechanism is *feasible* if it satisfies incentive compatibility (IC), participation (IR INS), and bilateral free disposal (14). The insurer's problem is to pick a feasible insurance mechanism that maximizes its expected profits (2). It is straightforward to adapt Proposition 1 to establish existence of an optimal insurance mechanism.

Any mechanism in which some types are excluded is equivalent to a mechanism in which the principal offers the *zero-coverage contract* to all excluded types: $W = I - L$, $B = L$. 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 7 (**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 $p_1 \geq \bar{p}_0 + \frac{C}{u(I) - u(I-L)}$.*

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,

[24] The pure adverse selection model of insurance provision by a monopolist was studied by Stiglitz, 1977 for two types and Chade and Schlee, 2012 for a continuum of types.

Chade and Schlee (2012, Proposition 2) show 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 3.1 that when reservation utilities are not type-dependent, exclusion is not optimal (as long as there is no exclusion in the first best). Proposition 7 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, when consumers can hide a loss from the insurer, 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

$$p_1 \geq p_0 + \frac{C}{u(I) - u(I-L)}. \tag{15}$$

Since excluded types are uninsured, the effort frontier for them coincides with the uninsured effort frontier (15). 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 8 (**Strict Distortion Relative to No Insurance**). *Let $\mathcal{E}$ 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 7. Then, $\mathcal{E}(p_0) > p_0 + \frac{C}{u(I)-u(I-L)}$ 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 (15) holds. On the other hand, high effort is efficient *from the principal's perspective* if $p_1 \geq p_0 + \frac{C}{L}$. The later corresponds to the first-best frontier in our model, since we are assuming that the principal has all the bargaining power.

When the agent has a lower incremental utility from the loss than the principal – i.e., $u(I) - u(I-L) \leq L$ – he picks a lower effort than the principal would demand if effort were observable. Combining with Proposition 8, 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 (15), which lies below the first-best frontier.

Remark 2. Our model can potentially contribute to the current policy debate on insurance reform. In particular, one of the main rationales of the recent Affordable Care Act was the need to reduce the large uninsured population. Proposition 7 shows that exclusion may be an unavoidable property of markets with both moral hazard and averse selection. Our model also shows that shirking is not necessarily a sign of poorly designed incentives. When the support of the conditional distributions is rich enough (such as

in our model), the principal can only incentivize some types to exert effort if she allows other types to pick the same high-powered incentives and shirk.

Because the participation constraint in insurance binds at the top rather than at the bottom, we cannot apply the argument from Proposition 4 and the optimal mechanism may have separation at the bottom. In the Online Appendix IV, we show that, when the first-best effort region is empty, the firm offers a single contract with full insurance to an interval containing the riskiest types ('the bottom').

4.2 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 is not observed by the regulator and can be either high ($e = 1$) or low ($e = 0$). 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: $(p_0, p_1) \in \boldsymbol{P}$. The firm's manager has cost C from exerting high effort and 0 from exerting low effort.

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

$$U = w + p_e b - Ce. \tag{16}$$

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 $\lambda > 0$. Thus, the net surplus of consumers/taxpayers is

$$S - (1 + \lambda) [w + p_e b + c_H - p_e (c_H - c_L)] .$$

A utilitarian regulator maximizes the sum of the consumers' net surplus and the expected utility of the firm's manager (16):

$$S - (1 + \lambda) [w + p_e b + c_H - p_e (c_H - c_L)] + U. \tag{17}$$

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:

$$x_H := S - (1+\lambda)c_L \text{ and } x_L := S - (1+\lambda)c_H.$$

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 (17), we can rewrite the regulator's objective function as

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

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

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

subject to $U \geq 0$. The first-best mechanism leaves zero rents to the firm's manager and requires a high effort whenever $p_1 \geq p_0 + (1+\lambda)\frac{C}{\Delta x}$.[25]

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, $\boldsymbol{P}$, described by the continuous density f.

The results from Subsection 3.2 can then be adapted to this framework. For example, in any optimal mechanism, only two contracts are offered to all types with low enough probability of success (see Proposition 4 and Figure 7). Those with low probability of success (Region A) get a cost-plus contract *($w = b = 0$)*, exert low effort, and obtain zero rents. Thus, any optimal mechanism must contain a cost-plus contract, which is accepted by firms with low enough probabilities of cutting costs. Types with intermediate probabilities of success (Region B) get a uniform contract with positive power and obtain positive rents. The following proposition states the other main results for the regulation model:

Proposition 9 (**Optimal Regulation**). *There exists optimal mechanism, which has the following properties:*

1. *Exclusion is optimal if and only if exclusion is first-best optimal;*
2. *The mechanism either offers only a fixed-price ($b = \Delta x$) and a cost-plus ($w = b = 0$) contract, or it has insufficient effort; and*
3. *If the distribution of types satisfies increasing rents and $\Delta x \leq 2C$, the optimal mechanism can be implemented with at most three contracts.*

[25] There are two differences between this model and the framework from Section 3.2. First, each dollar left to the agent costs $1+\lambda$ 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 $c(e)$, the principal subtracts it by $(1+\lambda)c(e)$.

In generic optimal mechanisms, price caps are suboptimal. The characterization of the optimal mechanism (Theorem 1) and the result on finite mechanisms when probabilities are bounded away from zero (Proposition 6) can also be easily adapted for the regulation model.

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

In our model, the principal 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.

Proceeding as in our analysis of unobservable effort costs, our approach can be used to study models with more than two effort levels. As with unobservable costs, the effort frontier becomes a multidimensional object when there are more than two efforts. Nevertheless, the diagonal – i.e., the set of types with the same probability of success conditional on all efforts – is still a one-dimensional object. Since the informational rents and efforts of all types are still determined by the (one-dimensional) rent projection along the diagonal, we can apply the same calculus of variations approach to obtain necessary conditions for an optimal mechanism.

Our approach cannot, however, easily accomodate models with $N > 2$ outputs. Since the diagonal corresponds to the set of types with the same probability of each output conditional on high and low efforts, the rent projection along the diagonal is an $(N-1)$-dimensional object. Thus, with more than two outputs, the projection along the diagonal does not lead to a one-dimensional program.

In addition, our analysis can be extended in two ways. First, the dual approach used on the optimal taxation model naturally leads to a Rawlsian planner (see Appendix B). 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 numerical methods. While we develop such method for our model in the Online Appendix III, we believe that developing such methods for more general models

could provide additional insights into the properties of optimal mechanisms.

Appendix

A Risk Aversion

This appendix generalizes the characterization of optimal mechanisms obtained in the risk-neutral case (Theorem 1) for weakly concave utility functions. The generalizations of the marginal virtual surpluses at the low-effort region, high-effort region, and in the region of types who get zero rent when the utility function is weakly concave are:[26]

$$\begin{aligned} S_0(t,\mathcal{U}) &:= \begin{cases} -\frac{(\mathcal{E}-t)\Delta x-(G(\mathcal{E})-G)}{\dot{\mathcal{U}}(\mathcal{E})} - \frac{\partial G}{\partial \mathcal{U}}\frac{F_0(t,\mathcal{E})}{f(t,\mathcal{E})} & \text{if } t < \overline{t} \\ -\frac{\partial G}{\partial \mathcal{U}}\frac{F_0(t,1)}{f(t,1)} & \text{if } t \geq \overline{t} \end{cases}, \\ S_1(t,\mathcal{U}) &:= \begin{cases} 0 & \text{if } t \leq \underline{\mathcal{E}} \\ \frac{(t-\mathcal{E}^{-1})\Delta x-(G-G(\mathcal{E}^{-1}))}{\dot{\mathcal{U}}(\mathcal{E}^{-1})} - \frac{\partial G}{\partial \mathcal{U}}\frac{F_1(\mathcal{E}^{-1},t)}{f(\mathcal{E}^{-1},t)} & \text{if } t > \underline{\mathcal{E}} \end{cases}, \text{ and} \\ \underline{S}(\mathcal{U}) &:= \frac{(\underline{\mathcal{E}}-E[t|t\leq \underline{t},\underline{\mathcal{E}}])\Delta x-G(\underline{\mathcal{E}})}{\dot{\mathcal{U}}(\underline{\mathcal{E}})}F_1(\underline{t},\underline{\mathcal{E}}), \end{aligned}$$

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

S_0 and S_1 differ from their risk-neutral counterparts (10) and (11) in that now the hazard rates are multiplied by the partial derivative $\partial G/\partial \mathcal{U}$. In the risk neutral case, each util left to the agent costs one dollar to the principal. Therefore, the informational rent is determined solely by the mass of types who receives these rents relative to the type on the effort frontier (i.e., the hazard rate). Under risk aversion, each util 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 $\partial G/\partial \mathcal{U}$. The expression for $\underline{S}$, however, remains unchanged relative to the risk neutral case since these types do not obtain any informational rents. As in the risk-neutral case, let $\mathcal{S}(t,\mathcal{U}) \equiv S_0(t,\mathcal{U})\, f(t,\mathcal{E}) + S_1(t,\mathcal{U})\, f\left(\mathcal{E}^{-1},t\right)$ denote the marginal virtual surplus weighted by its probability density.

When the agent is risk averse, the cost of providing utility $\mathcal{U}$ also depends on the power of the contract $\dot{\mathcal{U}}$. Thus, the relative cost of increasing the power at 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 low-effort region and the hazard rate of types who get the contract on the high-effort region. It is, therefore, useful to define each of these marginal costs as

$$C_0(t,\mathcal{U}) := \begin{cases} \frac{\partial G}{\partial \dot{\mathcal{U}}}\frac{F_0(t,\mathcal{E})}{f(t,\mathcal{E})} & \text{if } t < \overline{t} \\ \frac{\partial G}{\partial \dot{\mathcal{U}}}\frac{F_0(t,1)}{f(t,1)} & \text{if } t \geq \overline{t} \end{cases},$$

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

[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.

and to define the marginal cost of providing power weighted by its probability density as

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

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 $[t_1,t_2] \subset [0,1]$ such that $\underline{\mathcal{E}} \notin [t_1,t_2]$, then*
$$\mathcal{S}(t,\mathcal{U}) + \frac{d}{dt}\left\{\mathcal{C}(t,\mathcal{U})\right\} = 0,$$
for almost all $t \in [t_1,t_2]$.

2. **(bunching conditions)** *Let $[t_1,t_2] \subset [0,1]$ be a maximal interval where $\mathcal{U}$ is affine.*

- *If $\underline{\mathcal{E}} \notin [t_1,t_2]$, then*
$$0 \geq t_1 \int_{t_1}^{t_2} \mathcal{S}\left(t,\mathcal{U}\right) dt \geq \int_{t_1}^{t_2} t\mathcal{S}\left(t,\mathcal{U}\right) dt \geq t_2 \int_{t_1}^{t_2} \mathcal{S}\left(t,\mathcal{U}\right) dt.$$
Moreover, if $\mathcal{U}$ has kink at t_1 (at t_2) and $t_2 < 1$, then $\int_{t_1}^{t_2}(t-t_1)\mathcal{S}(t,\mathcal{U})dt = 0$ ($\int_{t_1}^{t_2}(t-t_2)\mathcal{S}(t,\mathcal{U})dt = 0$).[27]

- *If $t_1 = \underline{t}$ and $t_2 \geq \underline{\mathcal{E}}$, then*
$$\int_{\underline{t}}^{t_2} \mathcal{S}(t,\mathcal{U})dt + \underline{S}\left(\mathcal{U}\right) \leq 0, \text{ and } \int_{\underline{t}}^{t_2}(t-\underline{\mathcal{E}})\mathcal{S}(t,\mathcal{U})dt \leq 0.$$
Moreover, if $\mathcal{U}$ has kink at $t_2 < 1$, then
$$\int_{\underline{t}}^{t_2} \mathcal{S}(t,\mathcal{U})dt + \underline{S}\left(\mathcal{U}\right) = 0, \text{ and } \int_{\underline{t}}^{t_2}(t-\underline{\mathcal{E}})\mathcal{S}(t,\mathcal{U})dt = 0.$$

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.

B Optimal Taxation

We now show how our model can be applied in an optimal taxation context. This brings our paper closer to the literature of optimal taxation models with multidimensional taxpayer types.

[27] *If $t_2 = 1$ and $\dot{\mathcal{U}}(1) = \Delta x$, then the equalities become inequalities lower or equal.*

The seminal model of Mirrlees (1971) and most of the literature that followed, assumes that taxpayers differ only through a one-dimensional productivity parameter. Although, in reality, taxpayer heterogeneity is multidimensional, the difficulty in characterizing the solution of such screening programs has been a substantial barrier in the analysis of optimal taxes with multidimensional taxpayer types. Accordingly, most of the literature either assumes a discrete number of types, or uses numerical simulations.[28] A few recent notable exceptions are Kleven et al. (2009), Choné and Laroque (2010), Rothschild and Scheuer (2014), and Rothschild and Scheuer (2013), who study 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.

Consider a Rawlsian tax agency (principal) that 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]

This model can be interpreted as studying 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). This model can also be interpreted as a model of optimal income taxes in the spirit of Mirrlees (1971), although, in this case, the assumption of two outputs may be harder to justify. In the Mirrleesian framework, taxpayers 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 cannot be reduced to a pure adverse selection problem. 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 in Section 2, a mechanism $(w, b, e) : \boldsymbol{P} \to \mathbb{R}^2 \times \{0, 1\}$ specifies

[28] 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 consider models with two types in each of two dimensions, which can be suitably mapped into one-dimensional models with four types. For example, Boadway et al. (2002) study optimal income taxes and Cremer et al. (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.

[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) studies optimal taxation in a model where incomes are uncertain, although he restricts 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.

the agent's utility in case of low output w, the power of the contract b, and the effort recommendation e. The tax agency designs a mechanism that maximizes the utility of the lowest type, $w(0,0)$, among mechanisms that satisfy incentive compatibility (ICIC), free disposal (FD), and the resource constraint

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

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 main 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 from Appendix A derives the optimality conditions.

Adapting Proposition 2, it follows that types in a non-degenerate region at the bottom of the distribution $\mathbf{p} \in [0, \underline{t}) \times [0, \underline{\mathcal{E}}) \cap \boldsymbol{P}$ 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 3.2. Proposition 4 establishes that types in the intermediate region, $\mathbf{p} \in [\underline{t}, \underline{\mathcal{E}}) \times [0,1] \cap \boldsymbol{P}$, also face a uniform tax rate (although their tax rate is lower than 100%).

Proposition 10 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 5, 6 and Corollary 1 determine conditions under which optimal tax system can be implemented using a finite number of tax brackets.

[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.

C Relaxed BFD and Partially Selling the Firm

In Section 3 we assumed *bilateral free disposal (BFD)*. We now generalize condition (BFD) by allowing the bonus upper bound to be any fixed positive number:

$$B(\mathbf{p}) \leq K, \quad \forall \mathbf{p} \in \boldsymbol{P}, \tag{BB}$$

where $K > 0$. We can show that several results valid under (BFD) still hold under the more general condition (BB). An easy inspection of the proofs[34] of Propositions 1, 2, 3, 4 and 6; Lemmata 6 and 7; and Theorems 1 and 2 shows that they are easily extended to this more general case. In the Online Appendix I) we allow the agent to have private information about his cost of effort. Again, we can also extend the corresponding results of that appendix to this more general.

Under uniform distribution, $C = 1$ and condition (BB) for $K = 5 > \Delta x = 3$, our numerical method (described in the Online Appendix III) gives that the optimal mechanism can be implemented by only two contracts: zero bonus and positive bonus contracts. Figure 8 depicts the optimal mechanism and shows that the positive bonus is greater than Δx. The intuition is that paying bonus greater than the incremental output leads to three effects: higher rents to all agent's types who choose the positive bonus contract; low-effort effect (distortion) increase; and high-effort effect (distortion) reduction. Hence, paying high bonus for the top types allows the principal to induce higher level of effort of bottom types, which improves the principal's profit.

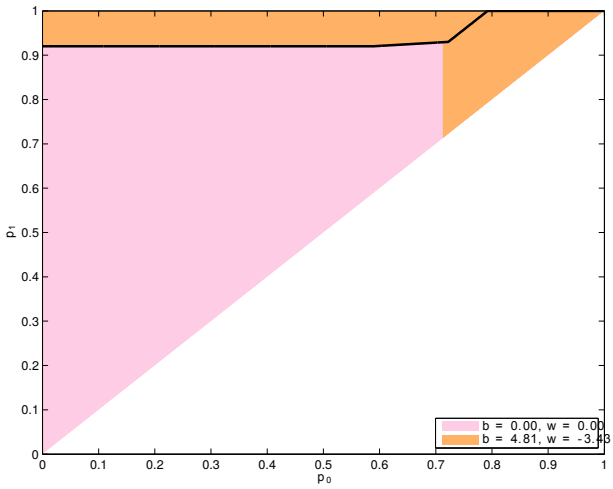

Figure 8: *Optimal mechanism for uniform distribution and* $\Delta x = 3c$.

Lemma 10. *Then, either there is insufficient effort, or the principal partially sells the firm.*

Lemma 8 shows that the BDF-optimal mechanism involves insufficient effort or selling partially the firm. We now examine the effort distortion relative to the first best when $K \neq \Delta x$.

Definition 3. Let (w, b, e) be a feasible mechanism and let $\mathcal{E}$ be the associated effort frontier. We say that there is *strict distortion* if $\mathcal{E}(t) \neq t + \frac{C}{\Delta x}$ whenever $\mathcal{E}(t) < 1$ except for at most one t.

[34]Propositions 5 and Corollary 1 can be also extended if we substitute Δx for K in their statement.

Proposition 10 below shows that there exists strict distortion when $K \neq \Delta x$. Let $(\mathcal{D}, ||\cdot||_\infty)$ be the space of continuous density functions $f : \boldsymbol{P} \to \mathbb{R}_+$ 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 10 (**Strict Distortion**). *Suppose that the agent is risk neutral and $K \neq \Delta x$. Generically, there exists strict distortion at the optimal mechanism.*

Proposition 10 raises the question whether strict distortion is generically true for $K = \Delta x$. We now present a sufficient condition for which partially selling the firm is optimal.[35] Let us assume that there is *no-rent at the top*, i.e., suppose that the density of types satisfies:

$$\frac{F_1(p_0, 1)}{f(p_0, 1)} = 0,$$

for all $p_0 \in [0, 1]$, where $F_1(p_0, p_1) = \int_0^{p_0} f(z, p_1)dz$ was defined in the text.

Proposition 11 (**Partially Selling the Firm**). *Suppose that the agent is risk neutral and the no-rent at top holds. Then, the BDF-optimal mechanism is implemented by partially selling the firm.*

Let us give two examples where partially selling the firm is optimal. The first one explores Proposition 11 and the second one explores Corollary 1 and our numerical method.

Example 1. Consider the density of types given by

$$f(p_0, p_1) \cong (1 - p_1)^{A - p_0},$$

where $A > 1$ is a constant. Note that

$$\frac{F_1(p_0, p_1)}{f(p_0, p_1)} = \int_0^{p_0} \frac{f(z, p_1)}{f(p_0, p_1)} dz = \int_0^{p_0} (1 - p_1)^{(p_0 - z)} dz$$

and, for each $p_0 \in [0, 1]$, the integrand converge to zero when $p_1 \to 1$. By the dominated convergence theorem, we have that no-rent at top condition holds for this distribution. Therefore, partially selling the firm implements the BDF-optimal mechanism.

For the uniform distribution, $C = 1$ and $\Delta x \in [1, 3]$, applying Corollary 1 and our numerical method, we can show that there exists a cutoff $\bar{\Delta x} \in (1, 3)$ such that the BDF-optimal mechanism is implemented by partially selling the firm if and only if $\Delta x \in (\bar{\Delta x}, 3]$. For $\Delta x \in [1, \bar{\Delta x}]$ the optimal mechanism is given by the trivial contract.

D Proofs

The long but straightforward proofs of Lemmata 1 and 3 can be found in the Online Appendix V.

[35] This condition in particular implies that the density must be zero at types with $p_1 = 1$. In this case we are assuming that the full support assumption almost everywhere with respect to the Lesbegue measure. This condition does not define an open set with respect to the uniform convergence metric. However, for every density that satisfies it and neighborhood of this density, we can find a large set of densities in the neighborhood that satisfies the condition.

Proof of Lemma 2

(a) The informational rent function can be written as

$$U(p_0, p_1) = \max_{\hat{\boldsymbol{p}} \in \boldsymbol{P}} \max_{\hat{e} \in \{0,1\}} \{w(\hat{\boldsymbol{p}}) + p_{\hat{e}} b(\hat{\boldsymbol{p}}) - c_{\hat{e}}\},$$

which is convex since it is the upper envelope of linear functionals. Convexity implies in differentiability almost everywhere and, from the envelope theorem,

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

at all points of differentiability.

(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 $b(\mathbf{p}) \geq 0$ for all $\mathbf{p}$ (including $\mathbf{p} = (0,0)$). Analogously, the participation constraint implies $U(0,0) \geq 0$.

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

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

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

Combine these two conditions to obtain

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

Therefore,

$$\begin{aligned} U(p_1, p_1) &= w(p_1, p_1) + p_1 b(p_1, p_1) \\ &= w(p_0, p_1) + p_1 b(p_0, p_1) + C \\ &= U(p_0, p_1) + C. \end{aligned}$$

Proof of Lemma 4

By property (a), $\mathcal{U}$ is differentiable a.e. and $\dot{\mathcal{U}}(p_0) = b(p_0, p_0)$ at all points of differentiability. By property (b), $b(p_0, p_1) = b(p_0, p_0) = \dot{\mathcal{U}}(p_0)$ for almost all (p_0, p_1) with $p_1 \leq \mathcal{E}(p_0)$, while, by (a) and (d), $b(p_0, p_1) = b(p_1, p_1) = \dot{\mathcal{U}}(p_1)$ for almost all (p_0, p_1) with $p_1 > \mathcal{E}(p_0)$. Thus,

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

for almost all $(p_0, p_1) \in \boldsymbol{P}$.

Properties (a) and (d) imply that $U(p_0,p_1) = U(p_0,p_0) = \mathcal{U}(p_0)$ if $p_1 \leq \mathcal{E}(p_0)$ and $U(p_0,p_1) = U(p_1,p_1) - C = \mathcal{U}(p_1) - C$ if $p_1 > \mathcal{E}(p_0)$. Therefore,

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

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

$$w(p_0,p_1) = \begin{cases} \mathcal{U}(p_0) - p_0\dot{\mathcal{U}}(p_0) & \text{if } p_1 \leq \mathcal{E}(p_0) \\ \mathcal{U}(p_1) - p_1\dot{\mathcal{U}}(p_1) & \text{if } p_1 > \mathcal{E}(p_0) \end{cases}.$$

Property (d) and the continuity of U yield

$$\mathcal{U}(\mathcal{E}(p_0)) = \mathcal{U}(p_0) + C \tag{18}$$

for almost all p_0 with $\mathcal{E}(p_0) < 1$. Since the high-effort region is non-empty (the mechanism is nontrivial), $\mathcal{E}(0) < 1$. Then, by continuity of U, we must have

$$\begin{aligned} U\left(\overline{t}, \mathcal{E}\left(\overline{t}\right)\right) &= U\left(\mathcal{E}\left(\overline{t}\right), \mathcal{E}\left(\overline{t}\right)\right) - C \\ &= U(1,1) - C \\ &= \mathcal{U}(1) - C. \end{aligned}$$

Moreover, because $U\left(\overline{t}, \mathcal{E}\left(\overline{t}\right)\right) = U\left(\overline{t}, \overline{t}\right) = \mathcal{U}\left(\overline{t}\right)$ (properties (a) and (d)) and $\mathcal{U}$ is increasing (property (a)), it follows that $\mathcal{U}(p_0) \geq \mathcal{U}(1) - C$ for all $p_0 \geq \overline{t}$. Combining this last inequality with (18), we obtain $\mathcal{U}(\mathcal{E}(p_0)) = \min\{\mathcal{U}(p_0) + C; \mathcal{U}(1)\}$.

Proof of Lemma 5

Lemma 4 establishes the mapping between $(\mathcal{E},\mathcal{U})$ and (w,b,e). From Lemma 3, any incentive-compatible mechanism (w,b,e) induces an effort frontier $\mathcal{E}$. Moreover, using equation (1) and $\mathcal{U}(t) := U(t,t)$, we can calculate the rent projection associated with it. Conversely, given an effort frontier and a rent projection $(\mathcal{E},\mathcal{U})$, we can recover the nontrivial mechanism (w,b,e) (at almost all points) using Lemma 4. Using the expressions from Lemma 4, it is straightforward to check that properties (a)-(d) from Lemma 2 are satisfied if and only if $\mathcal{U}$ is nondecreasing and convex, $\mathcal{U}(0) \geq 0$, and equation (6) is satisfied.

Proof of Proposition 1

For each feasible mechanism in this space, let $\mathcal{U}$ and $\mathcal{E}$ denote the rent projection and effort frontier functions associated with it. We need the following auxiliary lemma:

Lemma 11. *Let (f_n) be a sequence of real convex functions defined on $[0,1]$ converging pointwise to f. Then, f_n' converges almost everywhere to f'.*

Proof. Let I be the intersection of all $x \in [0,1]$ such that f_n and f are differentiable, for all n. Since f_n

and f are convex functions, $[0,1]\backslash I$ has zero Lesbegue measure. Then, for each $x \in I$ we have that

$$f_n(x) - f_n(x-h) \leq f_n'(x).h \leq f_n(x+h) - f_n(x),$$

for all $h > 0$ sufficiently small. Taking the limit in n we also have that

$$f(x) - f(x-h) \leq l.h \leq f(x+h) - f(x),$$

for each limit point l of the sequence $(f_n'(x))$ and each $h > 0$ sufficiently small. Since f is differentiable at x and h is arbitrary, then l must be $f'(x)$, i.e., the sequence $(f_n'(x))$ converges to $f'(x)$. □

Define the space of admissible rent functions

$$\boldsymbol{\mathcal{U}} = \{\mathcal{U} : [0,1] \to \mathbb{R};\ \mathcal{U}(0) = 0,\ \text{non-decreasing and convex function}\},$$

which is non-empty and compact with respect to the weak topology (i.e., this is the weakest topology such that a sequence $(\mathcal{U}_n)$ converges to $\mathcal{U}$ if and only if $(\mathcal{U}_n(t))$ converges to $\mathcal{U}(t)$ in all points in which $\mathcal{U}$ is continuous). Let $(\mathcal{U}_n)$ be a sequence in $\boldsymbol{\mathcal{U}}$ weakly converging to $\mathcal{U} \in \boldsymbol{\mathcal{U}}$. By the previous lemma and the definition of $\mathcal{E}_N$, the sequence $(\mathcal{U}_n,\dot{\mathcal{U}_n},\mathcal{E}_n)$ converges pointwise to $(\mathcal{U},\dot{\mathcal{U}},\mathcal{E})$. By the Lebesgue Dominated Convergence Theorem (see Rudin, 1986, pp. 26), the limit of principal's objective function (8) evaluated at $(\mathcal{U}_n)$ converges to its value at $\mathcal{U}$.

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 $(\mathcal{U}_n)$ be a sequence in $\boldsymbol{\mathcal{U}}$ such that the sequence of the principal's payoff evaluated at each $\mathcal{U}_n$ converges to its supremum.

Notice that, from (7) and the convexity of u^{-1},

$$G(\mathcal{U},\dot{\mathcal{U}},t) \leq u^{-1}(\mathcal{U}),$$

for every $\mathcal{U} \in \boldsymbol{\mathcal{U}}$. There are two cases to consider:

(i) $\mathcal{U}_n(t) \leq u^{-1}(t\Delta x)$, for all $t \in [0,1]$. Then, there exists a subsequence $(\mathcal{U}_{n_k}(t))$ converging to $\mathcal{U}$ that attains the supreme.

(ii) Without loss of generality, we can assume that there exists $\bar{t}_n \in [0,1)$ such that $\mathcal{U}_n(\bar{t}_n) = u^{-1}(\bar{t}_n\Delta x)$ and $\dot{\mathcal{U}}_n(t)$ is constant for $t \geq \bar{t}_n$. Suppose that $\mathcal{U}_n(1) \to \infty$ and $\bar{t}_n \to \bar{t} < 1$. Then, $\mathcal{U}_n(t) > u^{-1}(t\Delta x)$, for all $t \in [\bar{t}_n, 1]$ the integrand of (8) would diverge to $-\infty$ on the interval $[\bar{t}, 1]$, which is a contradiction. Therefore, $(\mathcal{U}_n(1))$ would converge to finite value or $\bar{t}_n \to 1$.

In both cases, we can define a limit point of $(\mathcal{U}_n)$.

Proof of Proposition 2

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

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

The mechanism induced by $\mathcal{U}_\alpha$ uniformly reduces the rent of all types by α and types in $[0,\underline{t}_\alpha]\times[0,\underline{\mathcal{E}}_\alpha]\cap \boldsymbol{P}$ have zero rent, where $\underline{t}_\alpha$ and $\underline{\mathcal{E}}_\alpha$ are defined as

$$\mathcal{U}(\underline{t}_\alpha)=\alpha \text{ and } \mathcal{U}(\underline{\mathcal{E}}_\alpha)-\alpha=C.$$

Notice that $\mathcal{U}_\alpha$ satisfies the constraints of Program (P') and, therefore, the mechanism associated with it is feasible.

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

$$\frac{d\underline{\mathcal{E}}_\alpha}{d\alpha}=\frac{1}{\dot{\mathcal{U}}(\underline{\mathcal{E}}_\alpha)}\geq 0.$$

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

$$G_\alpha(t)=\begin{cases} G(\mathcal{U}(t)-\alpha,\dot{\mathcal{U}}(t),t) & \text{if } t>\underline{t}_\alpha \\ u^{-1}(0) & \text{if } t\leq \underline{t}_\alpha \end{cases}.$$

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

$$\Pi_\alpha:=\int_0^1 (t\Delta x-G_\alpha(t))\,F_0(t,\mathcal{E}_\alpha)dt+\int_{\underline{\mathcal{E}}_\alpha}^1 (t\Delta x-G_\alpha(t))\,F_1(\mathcal{E}^{-1},t)dt,$$

where we are using the fact that neither the effort frontier changes for all $t\geq\underline{t}_\alpha$ nor its inverse $\mathcal{E}^{-1}$ for all $t\geq\underline{\mathcal{E}}_\alpha$.

Take the derivative of Π_α with respect to α and evaluate at 0:

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

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

where we omit the arguments of G and its derivative. Notice that the first and third terms are strictly positive, the second term is zero because $\underline{t}_0=0$, and the fourth term is zero because $F_1(0,\underline{\mathcal{E}}_0)=0$. Therefore, the derivative of Π_α is positive at 0, which implies that, for sufficiently small $\alpha>0$, principal strictly prefers the mechanism induced by $\mathcal{U}_\alpha$ to the one induced by $\mathcal{U}$.Proof of Proposition 3

Suppose $x_L\geq u^{-1}(0)$ 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: $w=u^{-1}(0)$, $b=0$. For any other type, the payoff from this contract is 0 under low effort and $-C$ under high effort. Thus, no type can benefit by deviating to this contract. For each of these types, the principal gets $x_L+p_0\Delta x-u^{-1}(0)$ (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 $x_L\geq u^{-1}(0)$. 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 $x_L < u^{-1}(0)$ and suppose there exists an optimal mechanism with no exclusion a.e.. By Proposition 2, there exist $\underline{t} > 0$ and $\underline{\mathcal{E}} > \underline{t}$ such that all types $(p_0, p_1) \leq (\underline{t}, \underline{\mathcal{E}})$ are offered the trivial contract: $w = u^{-1}(0)$, $b = 0$. Consider the alternative mechanism that recommends non-participation to all types a set $(p_0, p_1) \leq (\alpha, \alpha)$ for

$$\alpha \equiv \min\left\{\underline{t};\ \frac{u^{-1}(0) - x_L}{\Delta x}\right\} > 0. \tag{19}$$

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

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

where the last inequality follows from (19). Thus, the principal obtains a strictly higher payoff under this new mechanism, which contradicts the optimality of the original one.

Proofs of Lemma 7 and Theorem 1

The lemma is an immediate consequence of Lemma 12 (presented in the proof of Theorem 2), whereas the theorem follows from Theorem 2 for the risk-neutral case.

Proof of Proposition 4

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

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

Note that $\mathcal{U}(t) = \mathcal{V}(t)$ for all $t \geq \underline{\mathcal{E}}$ and $\mathcal{U}(\underline{\mathcal{E}}) = C$. Since the rent projection function $\mathcal{V}$ is also feasible, Lemma 7 gives

$$\int_0^{\underline{\mathcal{E}}} \left[\frac{(\mathcal{E}(t) - t)\Delta x - C}{\dot{\mathcal{U}}(\mathcal{E}(t))} f(t, \mathcal{E}(t)) + F_0(t, \mathcal{E}(t))\right] [\mathcal{U}(t) - \mathcal{V}(t)]\, dt \leq 0. \tag{20}$$

Since $\frac{(\mathcal{E}(t)-t)\Delta x - C}{\dot{\mathcal{U}}(\mathcal{E}(t))} f(t, \mathcal{E}(t)) \geq 0$, the term inside the first brackets is positive. Moreover, because $\mathcal{U}$ is convex, $\mathcal{U}(t) \geq \mathcal{V}(t)$ for all $t \in [0, \underline{\mathcal{E}}]$. Hence, the continuity of $\mathcal{U}$ and $\mathcal{V}$ and condition (20) yield $\mathcal{U}(t) = \mathcal{V}(t)$ for all $t \in [0, \underline{\mathcal{E}}]$.

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

Moreover,

$$\mathcal{U}(\underline{\mathcal{E}}) = \int_{\underline{t}}^{\underline{\mathcal{E}}} \dot{\mathcal{U}}(\underline{\mathcal{E}})\, dt = (\underline{\mathcal{E}} - \underline{t})\, \dot{\mathcal{U}}(\underline{\mathcal{E}}).$$

Combining these two conditions yields

$$\dot{\mathcal{U}}(\underline{\mathcal{E}}) = \frac{C}{\underline{\mathcal{E}} - \underline{t}} > C,$$

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

Proof of Lemma 8

We have argued in the text that the optimal effort region is contained in the first-best effort region:

$$\mathcal{E}(t) \geq t + \frac{C}{\Delta x} \tag{21}$$

for all $t \leq 1 - \frac{C}{\Delta x}$. We will show that it is contained in the *interior* of the first-best effort region.

Since $\mathcal{E}(t)$ is strictly increasing in the region where $\mathcal{E}(t) < 1$ and constant when $\mathcal{E}(t) = 1$, its inverse is always well defined for $t < \overline{t}$. We adopt the following convention: $\mathcal{E}^{-1}(t) \equiv \inf\left\{\hat{t} : \mathcal{E}\left(\hat{t}\right) \geq t\right\}$. Thus, $\mathcal{E}^{-1} : \left[\underline{\mathcal{E}}, \overline{t}\right] \to [0,1]$ is a strictly increasing function. The following claims will be useful in the proof:

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

Proof. Applying equation (21) to $\mathcal{E}^{-1}(t)$, yields

$$\mathcal{E}^{-1}(t) \leq t - \frac{C}{\Delta x}. \tag{22}$$

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

Since, by Lemma 6,

$$\dot{\mathcal{E}^{-1}}(t) = \frac{\dot{\mathcal{U}}(t)}{\dot{\mathcal{U}}\left(\mathcal{E}^{-1}(t)\right)} \text{ a.e.}, \tag{23}$$

the convexity of $\mathcal{U}$ implies that $\dot{\mathcal{E}^{-1}}(t) \geq 1$ a.e. Therefore, $\mathcal{E}^{-1}\left(\tilde{t}\right) = \mathcal{E}_f^{-1}\left(\tilde{t}\right)$ and $\dot{\mathcal{E}^{-1}}(t) \geq \dot{\mathcal{E}_f^{-1}}(t)$ (a.e.). It then follows that

$$\mathcal{E}^{-1}(t) \geq \mathcal{E}_f^{-1}(t) = t - \frac{C}{\Delta x}, \text{ for all } t \geq \tilde{t}.$$

Combining with inequality (22), yields $\mathcal{E}^{-1}(t) = t - \frac{C}{\Delta x}$ for all $t \geq \tilde{t}$.

From equation (6), $\mathcal{U}\left(t - \frac{C}{\Delta x}\right) = \mathcal{U}(t) - C$ for all $t \geq \tilde{t}$. Moreover, from equation (23), we must have $\dot{\mathcal{U}}(t) = \dot{\mathcal{U}}(\mathcal{E}^{-1}(t))$ a.e., which implies that there exist constants $\beta > 0$ and $\alpha \in \mathbb{R}$ such that $\mathcal{U}(t) = \beta t + \alpha$

for almost all $t \geq \tilde{t}$. Combining these two statements, yields

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

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

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

Proof. The result is immediate if $\mathcal{E}(\tilde{t}) = 1$. Let $\mathcal{E}\left(\tilde{t}\right) < 1$. By Lemma 6, $\dot{\mathcal{E}}(t) = \frac{\dot{\mathcal{U}}(t)}{\dot{\mathcal{U}}(\mathcal{E}(t))}$ for almost all t such that $\mathcal{E}(t) < 1$. Because $\dot{\mathcal{U}}(t) = \dot{\mathcal{U}}(\mathcal{E}(t))$ for $t \geq \tilde{t}$, it follows that $\dot{\mathcal{E}}(t) = 1$ for almost all $t \geq \tilde{t}$ such that $\mathcal{E}(t) < 1$. By continuity of $\mathcal{E}$ (Lemma 1), $\mathcal{E}(t) = \mathcal{E}(\tilde{t}) - \tilde{t} + t$ whenever $\mathcal{E}(t) < 1$. For $\mathcal{E}(t) = 1$, the result is immediate. □

Suppose, in order to obtain a contradiction, that the statement in the lemma is false. Recall that the domain of $\mathcal{E}^{-1}$ is $[\underline{\mathcal{E}}, 1]$. Then, by condition (21), there must exist a type $t \in [\underline{\mathcal{E}}, 1]$ for which $\mathcal{E}^{-1}(t) = t - \frac{C}{\Delta x}$. Denote the infimum of such types by

$$\tilde{t} \equiv \inf\left\{t \in [0,1] :\ \mathcal{E}^{-1}(t) = t - \frac{C}{\Delta x}\right\} \in [\underline{\mathcal{E}}, 1].$$

By Claim 1, $\mathcal{E}^{-1}(t) = t - \frac{C}{\Delta x}$ and $\dot{\mathcal{U}}(t) = \Delta x$ for all $t \geq \tilde{t}$. There are two cases: $\tilde{t} = \underline{\mathcal{E}}$ and $\tilde{t} > \underline{\mathcal{E}}$.

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

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

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

which is the substitution of $\mathcal{U}$ by the envelope of tangent lines at points $\tilde{t} - \delta$ and $\tilde{t}$ of the function $\mathcal{U}$ on the interval $[\tilde{t} - \delta,\ \tilde{t}]$. By the definition of $\tilde{t}$, $\dot{\mathcal{U}}(t) < \Delta x$,[36] convexity of $\mathcal{U}$ and the hypothesis that $\mathcal{U}$ does not have kink at $\tilde{t}$, $\mathcal{V}(t) < \mathcal{U}(t)$ for all $t \in (\tilde{t} - \delta, \tilde{t})$. Hence,

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

which contradicts the optimality condition of Lemma 7. Hence, there is kink at $\tilde{t}$. Then, by Theorem 1, $\int_{\tilde{t}}^{1} \mathcal{S}(t,\mathcal{U})dt = 0$, which contradicts $\mathcal{S}(t,\mathcal{U}) < 0$ on $[\tilde{t}, 1]$.

[36] Notice that if $\tilde{t} = 1$, then $\dot{\mathcal{U}}(t) < \Delta x = \dot{\mathcal{U}}(1)$ for all $t < 1$ and, because $\dot{\mathcal{U}}$ is a càdlàg function, $\lim_{t \to 1} \dot{\mathcal{U}}(t) = \dot{\mathcal{U}}(1)$.

Proof of Proposition 10

Let us assume that the optimal mechanism is not trivial (otherwise, the result would be straightforward). First if $K < \Delta x$, the optimal mechanism features necessarily insufficient effort provision (partially selling the firm can not be achieved).

Suppose that $K > \Delta x$. If the optimal bonus at the top is strictly below the incremental output, then again necessarily the optimal mechanism features insufficient effort provision. On the other hand, if the optimal bonus at the top is strictly above the incremental output, then the optimal effort frontier is strictly below the first-best effort frontier at the top. Since the slope of the first-best frontier is one and the slope of the optimal effort frontier is lower or equal to one (see Lemma 6), then these frontiers must cross at most once. Hence, this last case then features over effort at the top and eventually insufficient effort at the bottom. To complete the proof we will show that partially selling the firm (i.e., setting the positive bonus equals to the incremental output) is generically not optimal.

Let $r = \frac{C}{\Delta x}$. 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 8, partially selling the firm must be optimal. This optimal mechanism is then characterized by the following rent projection function $\mathcal{U}(t) = \max\{\Delta x(t - \underline{t}), 0\}$, for some $\underline{t} \in (0, 1)$. We also have that $\underline{\mathcal{E}} = \underline{t} + r < 1$. From Theorem 1, the necessary optimality bunching conditions are then given by:

$$\begin{array}{l} \int_{\underline{t}}^{1} F_0(t, t+r)dt + \int_{\underline{t}+r}^{1} F_1(t-r, t)dt = \underline{t}F_1(\underline{t}, \underline{t}+r) - \int_0^{\underline{t}} tf(t, \underline{t}+r)dt, \text{ and} \\ \int_{\underline{t}}^{1} (t-\underline{t})F_0(t, t+r)dt + \int_{\underline{t}+r}^{1} (t-\underline{t})F_1(t-r, t)dt = r\left(\underline{t}F_1(\underline{t}, \underline{t}+r) - \int_0^{\underline{t}} tf(t, \underline{t}+r)dt\right). \end{array}$$

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

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

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 $\underline{t} \in (0, 1)$ such that $H(\underline{t}, f) = 0$ (i.e., $\underline{t}$ 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 $H(\underline{t}, \cdot) : \mathcal{D} \to \mathbb{R}^2$ exists and is onto.*

Notice that $H(\underline{t}, \cdot)$ is a linear mapping from $L_\infty(\boldsymbol{P})$ 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 $L_\infty(\boldsymbol{P})$ such that the vectors $\{H(\underline{t}, f_1), H(\underline{t}, f_2)\} \subset \mathbb{R}^2$ are linearly independent. Consider $\alpha > 0$ sufficiently small and define $h_\alpha(t, s) = \mathbf{1}_{[t \leq \underline{t} - \alpha]}(t, s)$, where $\mathbf{1}_A$ is the indicator of the set A. Then,

$$F_0^\alpha(t, s) = \int_t^s h_\alpha(t, x)dx = \begin{cases} s - t & \text{if } t \leq \underline{t} - \alpha \\ 0 & \text{otherwise} \end{cases}, \text{ and}$$

$$F_1^\alpha(t, s) = \int_0^t h_\alpha(x, s)dx = \begin{cases} t & \text{if } t \leq \underline{t} - \alpha \\ \underline{t} - \alpha & \text{otherwise} \end{cases}.$$

Now we can compute:

$$\begin{array}{l} H_1(\underline{t}, h_\alpha) = \int_{\underline{t}+r}^{1} (\underline{t} - \alpha) dt - \int_0^{\underline{t}-\alpha} t dt - \int_{\underline{t}-\alpha}^{\underline{t}} (\underline{t} - \alpha) dt \\ H_2(\underline{t}, h_\alpha) = \int_{\underline{t}+r}^{1} t(\underline{t} - \alpha) dt - (\underline{t} + r) \left(\int_0^{\underline{t}-\alpha} t dt + \int_{\underline{t}-\alpha}^{\underline{t}} (\underline{t} - \alpha) dt \right) \end{array}.$$

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

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

and the second derivative gives its curvature:

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

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

$$h_\alpha(t, s) = \begin{cases} 1 \text{ if } t \leq \underline{t} - \alpha \\ 0 \text{ if } t \geq \underline{t}, \end{cases}$$

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 $\alpha > 0$ there exists $\tilde{f} \in \mathcal{D}$ such that $||f - \tilde{f}||_\infty < \alpha$ 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 $H(\underline{t}, f) = (0, 0)$. 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 $A(t, x, y) = H(t, f + x_1 h_1 + x_2 h_2)$ has Jacobean with respect to variables (x_1, x_2) at the point $(\underline{t}, 0, 0)$ given by

$$\begin{bmatrix} H(\underline{t}, h_1) & H(\underline{t}, h_2) \end{bmatrix} = \begin{bmatrix} e_1' & e_2' \end{bmatrix},$$

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 $\alpha > 0$ such that $A(t, f + x_1 h_1 + x_2 h_2) = (a_1, a_2)$ if and only if $x_i = \epsilon_i(t, a_1, a_2)$ where $\epsilon_i : [\underline{t} - \delta, \underline{t} + \delta] \times [-\alpha, \alpha]^2 \to \mathbb{R}^2$ are smooth functions. Notice that $H(t, f) \neq (0, 0)$, for all $t \in K := [0, 1] - (\underline{t} - \delta, \underline{t} + \delta)$. By continuity of H and the compactness of K, we can find $(x_1, x_2) \notin \{(\epsilon_1(t, 0, 0), \epsilon_2(t, 0, 0)); t \in [\underline{t} - \delta, \underline{t} + \delta]\}$ with a sufficiently small norm such that $H(t, f + x_1 h_1 + x_2 h_2) \neq (0, 0)$, for all $t \in [0, 1]$.

Define $\tilde{f} = f + h$, where $h = x_1 h_1 + x_2 h_2$. Notice that, since h is a bounded function we can choose $|\alpha| > 0$ sufficiently small such that $f + \alpha h$ is strictly positive function. Finally, normalizing $\tilde{f}$ we have a density and get the result.

If the number of solutions of the equation $H(\underline{t}, f) = (0, 0)$ 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 $\underline{t}_n \in (0, 1)$. Take a subsequence such that $(\underline{t}_{n_k})$ converges to $\underline{t} \in [0, 1]$. It is easy to see that $\Pi(\mathcal{U}_{n_k}, f_{n_k})$ converges to $\Pi(\mathcal{U}, f)$, where $\mathcal{U}_n(t) = \max\{\Delta x(t - \underline{t}_n), 0\}$ and $\mathcal{U}(t) = \max\{\Delta x(t - \underline{t}), 0\}$, 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 or Proposition 11

Let $\mathcal{U}$ and $\mathcal{E}$ be the rent projection and the effort frontier of a BDF-optimal mechanism. Suppose, by absurd, that the optimal mechanism is such that $\lim_{t\uparrow 1} \dot{\mathcal{U}}(t) < \Delta x$. Consider the following perturbation of the optimal mechanism. Take any $\delta > 0$ sufficiently small and $\bar{b} \in \left(\lim_{t\uparrow 1} \dot{\mathcal{U}}(t), \Delta x\right)$. Define the following feasible rent projection:

$$\mathcal{V}(t) = \max\left\{\mathcal{U}(t), \bar{b}(t - 1 + \delta) + \mathcal{U}(1 - \delta)\right\}.$$

By (10) and (11)

$$S_0(t,\mathcal{U}) = -\frac{F_0(t, \mathcal{E})}{f(t, \mathcal{E})} \text{ and } S_1(t,\mathcal{U}) = \frac{(t - \mathcal{E}^{-1})\Delta x - C}{\dot{\mathcal{U}}(\mathcal{E}^{-1})} - \frac{F_1(\mathcal{E}^{-1}, t)}{f(\mathcal{E}^{-1}, t)},$$

for all $t \in [1 - \delta, 1]$. Under the no-rent at top condition we have that

$$\lim_{t\to 1} S_0(t,\mathcal{U}) = 0 \text{ and } \lim_{t\to 1} S_1(t,\mathcal{U}) = \frac{(1 - \mathcal{E}^{-1}(1))\Delta x - C}{\dot{\mathcal{U}}(\mathcal{E}^{-1}(1))} > 0,$$

because $\dot{\mathcal{U}}(\mathcal{E}^{-1}(1)) < \bar{b} < \Delta x$. Applying Lemma 7 to the perturbation $\mathcal{V}$ we must have that

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

However, taking $\delta > 0$ sufficiently small we have that $\mathcal{S}(t,\mathcal{U}) > 0$ and $\mathcal{U}(t) - \mathcal{V}(t) < 0$, for all $t \in [1-\delta, 1]$, which contradicts the previous inequality. Therefore, $\lim_{t\uparrow 1} \dot{\mathcal{U}}(t) = \Delta x$ and $\mathcal{E}^{-1}(1) = 1 - \frac{C}{\Delta x}$. Applying the same arguing made in the proof of Proposition 8 we conclude that partially selling the firm is optimal.

Proof of Lemma 9

For $t \geq \overline{t}$,

$$\mathcal{S}(t,\mathcal{U}) = \frac{(t - \mathcal{E}^{-1})\Delta x - C}{\dot{\mathcal{U}}(\mathcal{E}^{-1})} - H(\mathcal{E}^{-1}, t).$$

Recall that the distortion is always non-negative, $(t-\mathcal{E}^{-1})\Delta x - C \geq 0$; the slope of the inverse effort frontier satisfies $\dot{\mathcal{E}^{-1}} \geq 1$ at all points of differentiability (Lemma 6); and the rent projection $\mathcal{U}$ is convex (Lemma 5). Using the signs of the partial derivatives of H implied by increasing rents, it follows that

$$\frac{d}{dt}\left(\mathcal{S}(t,\mathcal{U})\right) = \frac{d}{dt}\left(\frac{(t-\mathcal{E}^{-1})\Delta x - C}{\dot{\mathcal{U}}(\mathcal{E}^{-1})} - H(\mathcal{E}^{-1},t)\right)$$

$$= -\frac{\left(\dot{\mathcal{E}^{-1}}-1\right)\Delta x}{\dot{\mathcal{U}}(\mathcal{E}^{-1})} - \left[\frac{\left(t-\mathcal{E}^{-1}\right)\Delta x - C}{\dot{\mathcal{U}}(\mathcal{E}^{-1})}\right]\frac{\ddot{\mathcal{U}}(\mathcal{E}^{-1})}{\dot{\mathcal{U}}(\mathcal{E}^{-1})}\dot{\mathcal{E}^{-1}} - H_1(\mathcal{E}^{-1},t)\dot{\mathcal{E}^{-1}} - H_2(\mathcal{E}^{-1},t) < 0$$

for all $t \geq \overline{t}$ in which $\frac{\mathcal{S}(t,\mathcal{U})}{f(\mathcal{E}^{-1},t)}$ is differentiable (where we write $H_1(t,s) \equiv \frac{\partial H}{\partial t}(t,s)$ and $H_2(t,s) \equiv \frac{\partial H}{\partial s}(t,s)$), showing that $\frac{\mathcal{S}(t,\mathcal{U})}{f(\mathcal{E}^{-1},t)}$ is a strictly decreasing function of t. Because $\frac{\mathcal{S}(t,\mathcal{U})}{f(\mathcal{E}^{-1},t)}$ is strictly decreasing in t and $f\left(\mathcal{E}^{-1},t\right) > 0$, there are three possible cases:

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

Consider the convex and piecewise linear function

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

Notice that, because $\mathcal{U}(1) = \mathcal{V}(1)$, and $\bar{t}$ is determined by $\mathcal{U}(\bar{t}) = \mathcal{U}(1) - \Delta c$, it follows that $\bar{t}$ is the same under both $\mathcal{U}$ and $\mathcal{V}$. Notice that $\mathcal{V}$ is also feasible. Since $\mathcal{U}$ is optimal, by Lemma 7,

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

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

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

Consider the convex and piecewise linear function

$$\mathcal{V}(t) = \begin{cases} \mathcal{U}(t) & \text{if } t \leq \overline{t} \\ \mathcal{U}(1) + \frac{\mathcal{U}(1)-\mathcal{U}(\overline{t})}{1-\overline{t}}(t-1) & \text{if } t > \overline{t} \end{cases},$$

which, as in case (i), coincides with $\mathcal{U}$ for $t \leq \overline{t}$ and is a feasible rent projection. Proceeding exactly as in case (i) establishes that $\mathcal{U}(t) = \mathcal{V}(t)$, for all $t \in [\overline{t},1]$.

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

Consider the feasible rent projection

$$\mathcal{V}(t) = \begin{cases} \mathcal{U}(t) & \text{if } t \leq \overline{t} \\ \max\left\{\mathcal{U}(\overline{t}) + \frac{\mathcal{U}(\tilde{t})-\mathcal{U}(\overline{t})}{\tilde{t}-\overline{t}}(t-\overline{t});\ \mathcal{U}(1) + \dot{\mathcal{U}}(1)(t-1)\right\} & \text{if } t > \overline{t} \end{cases}.$$

Since $\mathcal{U}(t) = \mathcal{V}(t)$ on $t \leq \overline{t}$, Lemma 7 yields

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

Because $\mathcal{U}(t) \leq \mathcal{V}(t)$ on $[\overline{t}, \tilde{t}]$ and $\mathcal{U}(t) \geq \mathcal{V}(t)$ on $[\tilde{t}, 1]$, 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 $t \in [\overline{t}, 1]$. We conclude that $\mathcal{U}$ must have at most two pieces on the interval $[\overline{t}, 1]$.

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

Proof of Proposition 5

(i) We claim that $\frac{\Delta x}{C} \leq 2$ implies that $\underline{\mathcal{E}} \geq \overline{t}$. Because $\mathcal{U}$ is increasing, it is enough to show that $\mathcal{U}(\underline{\mathcal{E}}) \geq \mathcal{U}(\overline{t})$. By condition (6), $\mathcal{U}(\underline{\mathcal{E}}) = C$ and $\mathcal{U}(\overline{t}) = \mathcal{U}(1) - C$, so that

$$\mathcal{U}(\underline{\mathcal{E}}) \geq \mathcal{U}(\overline{t}) \iff \mathcal{U}(1) \leq 2C.$$

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

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

where the last inequality follows from the assumption that $\Delta x \leq 2C$.

(ii) Follows from (i) and equation (4).

Proof of Corollary 1

See Online Appendix V.

Proof of Proposition 6

We have that

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

since, by hypothesis, $f(x,s) \geq f(t,s)$, for all $x \in [0,t]$. Recall that the effect on the low-effort region is always non-positive: $S_0(t,\mathcal{U}) \leq 0$. Let us investigate the effect on the high-effort region. For any $t > \underline{\mathcal{E}}$, we have

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

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

$$\frac{\Delta x}{C} t - 1 \leq \left(1 + \frac{\Delta x}{C}\right) \mathcal{E}^{-1}.$$

This condition is implied by the following inequality

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

which is equivalent to the condition in the statement of the proposition. Given the optimal rent projection $\mathcal{U}$, let $\mathcal{V}(t) = \max\{0, \dot{\mathcal{U}}(\underline{\mathcal{E}})(t - \underline{\mathcal{E}}) + C\}$, where $\mathcal{U}(\underline{\mathcal{E}}) = C$. By Lemma 7, we must have

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

Since $\mathcal{V}(t) \leq \mathcal{U}(t)$, it follows that $\mathcal{U}(t) = \mathcal{V}(t)$ for all $t \in [0, 1]$, establishing the result.

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

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

Proof of the lemma.

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

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

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

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

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

$$\Pi(\mathcal{U} + \alpha h) \leq \Pi(\mathcal{U})$$

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

$$\lim_{\alpha \downarrow 0} \frac{\Pi(\mathcal{U} + \alpha h) - \Pi(\mathcal{U})}{\alpha} \leq 0.$$

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

$$\mathcal{U}(\mathcal{E}_\alpha(t)) + \alpha h(\mathcal{E}_\alpha(t)) = \mathcal{U}(t) + \alpha h(t) + C$$

for all $t \in [0, \overline{t}_\alpha]$, where $\overline{t}_\alpha$ solves $\mathcal{U}(\overline{t}_\alpha) + \alpha h(\overline{t}_\alpha) = \mathcal{U}(1) + \alpha h(1) - C$. Taking the total derivative of this expression with respect to α and evaluating at 0, we obtain

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

for all $t \leq \overline{t}$.

Analogously, its inverse, $\mathcal{E}_\alpha^{-1}$, satisfies an analogous functional equation:

$$\mathcal{U}\left(\mathcal{E}_\alpha^{-1}(t)\right) + \alpha h(\mathcal{E}_\alpha^{-1}(t)) = \mathcal{U}(t) + \alpha h(t) - C,$$

for all $t \in [\underline{\mathcal{E}}_\alpha, 1]$, where $\mathcal{U}(\underline{\mathcal{E}}_\alpha) + \alpha h(\underline{\mathcal{E}}_\alpha) = C$. Again, taking the total derivative of this expression with respect to α and evaluating at 0, we get:

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

for all $t \geq \underline{\mathcal{E}}$. Applying the same procedure with respect to $\underline{\mathcal{E}}_\alpha$ yields

$$\left.\frac{\partial \underline{\mathcal{E}}_\alpha}{\partial \alpha}\right|_{\alpha=0} = -\frac{h(\underline{\mathcal{E}})}{\dot{\mathcal{U}}(\underline{\mathcal{E}})}.$$

Then,

$$\left.\frac{\partial \mathcal{E}_\alpha}{\partial \alpha}\right|_{\alpha=0} = \frac{h(t) - h(\mathcal{E})}{\dot{\mathcal{U}}(\mathcal{E})}, \quad \left.\frac{\partial \mathcal{E}_\alpha^{-1}}{\partial \alpha}\right|_{\alpha=0} = \frac{h(t) - h(\mathcal{E}^{-1})}{\dot{\mathcal{U}}(\mathcal{E}^{-1})}, \quad \left.\frac{\partial \underline{\mathcal{E}}_\alpha}{\partial \alpha}\right|_{\alpha=0} = -\frac{h(\underline{\mathcal{E}})}{\dot{\mathcal{U}}(\underline{\mathcal{E}})}, \text{ and} \tag{24}$$

$$\dot{\mathcal{U}}_\alpha = \dot{\mathcal{U}}(t) + \alpha \dot{h}(t). \tag{25}$$

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

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

Using conditions (24), we obtain

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

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

$$\begin{gathered}
\int_0^{\overline{t}} (t\Delta x - G) \tfrac{h(\mathcal{E})}{\dot{\mathcal{U}}(\mathcal{E})} f(t, \mathcal{E}) dt = \int_0^{\underline{t}} t\Delta x \tfrac{h(\mathcal{E})}{\dot{\mathcal{U}}(\underline{\mathcal{E}})} f(t, \underline{\mathcal{E}}) dt + \int_{\underline{\mathcal{E}}}^1 \left(\mathcal{E}^{-1}\Delta x - G\left(\mathcal{E}^{-1}\right) \right) \tfrac{h(t)}{\dot{\mathcal{U}}(t)} f(\mathcal{E}^{-1}, t) \dot{\mathcal{E}^{-1}}(t) dt \\
\int_{\underline{\mathcal{E}}}^1 (t\Delta x - G) \tfrac{h(\mathcal{E}^{-1})}{\dot{\mathcal{U}}(\mathcal{E}^{-1})} f(\mathcal{E}^{-1}, t) dt = \int_0^{\overline{t}} \left(\mathcal{E}\Delta x - G(\mathcal{E}) \right) \tfrac{h(t)}{\dot{\mathcal{U}}(t)} f(t, \mathcal{E}) \dot{\mathcal{E}}(t) dt.
\end{gathered}$$

Using condition (25) yields:

$$\begin{aligned}\tfrac{d}{d\alpha}\Pi_\alpha\big|_{\alpha=0} = \; & -\int_0^1 \left(\tfrac{\partial G}{\partial \mathcal{U}}F_0(t,\mathcal{E})h(t) + \tfrac{\partial G}{\partial \dot{\mathcal{U}}}F_0(t,\mathcal{E})\dot{h}(t)\right)dt \\ & -\int_{\underline{\mathcal{E}}}^1 \left(\tfrac{\partial G}{\partial \mathcal{U}}F_1(\mathcal{E}^{-1},t)h(t) + \tfrac{\partial G}{\partial \dot{\mathcal{U}}}F_1(\mathcal{E}^{-1},t)\dot{h}(t)\right)dt \\ & -\int_0^{\bar{t}} \tfrac{(\mathcal{E}-t)\Delta x-(G(\mathcal{E})-G)}{\dot{\mathcal{U}}(\mathcal{E})}f(t,\mathcal{E})h(t)dt \\ & +\int_{\underline{\mathcal{E}}}^1 \tfrac{(t-\mathcal{E}^{-1})\Delta x-(G-G(\mathcal{E}^{-1}))}{\dot{\mathcal{U}}(\mathcal{E}^{-1})}f(\mathcal{E}^{-1},t)h(t)dt \\ & +\left(-\int_0^t t\Delta x f(t,\underline{\mathcal{E}})dt + (\underline{\mathcal{E}}\Delta x - G(\underline{\mathcal{E}}))F_1(\underline{t},\underline{\mathcal{E}})\right)\tfrac{h(\underline{\mathcal{E}})}{\dot{\mathcal{U}}(\underline{\mathcal{E}})}.\end{aligned}$$

This establishes the result. Notice that, in the case of Lemma 7, substituting $\mathcal{U}-\mathcal{U}(\mathcal{E}^{-1}) = C$ and $\mathcal{U}(\mathcal{E})-\mathcal{U} = C$ into the equation above, gives the result claimed in the statement of the lemma. ■

The proof of the theorem will use the following lemma, whose proof is presented in the Online Appendix VI.

Lemma 13. *Let $f \in L_\infty[t_1,t_2]$ satisfying $\int_{t_1}^{t_2} f(t)g(t)dt = 0$, for all $g \in C([t_1,t_2])$ such that $\int_{t_1}^{t_2} g(t)dt = 0$. Then, f is a constant function a.e.*

Proof of the theorem.

(1) Notice that $\mathcal{S}(t,\mathcal{U})$ is an integrable function on $[t_1,t_2]$ (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 $t \notin (t_1,t_2)$. Since $\mathcal{U}$ is strongly convex on $[t_1,t_2]$, $\mathcal{U}+\alpha h$ is a strongly convex function if $|\alpha|$ is sufficiently small. Performing the variational calculus (given by the previous theorem) for such feasible direction, we get

$$\int_{t_1}^{t_2}\mathcal{S}(t,\mathcal{U})h(t)dt - \int_{t_1}^{t_2}\mathcal{C}(t,\mathcal{U})\dot{h}(t)dt = 0.$$

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

$$\int_{t_1}^{t_2}\left[\int_0^t \mathcal{S}(x,\mathcal{U})dx + \mathcal{C}(t,\mathcal{U})\right]\dot{h}(t)dt = 0.$$

Since the function inside the brackets of the above integral is càdlàg, $\dot{h}$ is a generic continuous function. By Lemma 13,

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

is constant on $[t_1,t_2]$. Since this function is a.e. differentiable (since $\dot{\mathcal{U}}$ is a.e. differentiable), we have that

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

a.e. on $[t_1,t_2]$.

(2) We have two possible feasible perturbations that we can do with the rent projection function on the interval $[t_1,t_2]$: translations and rotations. Let us start with the translations and consider the case

$\underline{\mathcal{E}} \notin [t_1, t_2]$ and $t_2 < 1$. We have that there exist $\beta > 0$ and $\alpha \in \mathbb{R}$ such that $\mathcal{U}(t) = \beta t + \alpha$, for all $t \in [t_1, t_2]$. 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 12, we get

$$\int_{t_{1\delta}}^{t_{2\delta}} \mathcal{S}(t,\mathcal{U}) h_\delta(t) dt - \int_{t_{1\delta}}^{t_{2\delta}} \mathcal{C}(t,\mathcal{U}) \dot{h}_\delta(t) dt \geq 0,$$

where $h_\delta = \mathcal{U} - \mathcal{V}_\delta$, $t_{1\delta}$ and $t_{2\delta}$ 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 $[t_1, t_2]$ for sufficiently small $\delta > 0$). Let $t'_{1\delta} \geq t_{1\delta}$ and $t'_{2\delta} \leq t_{2\delta}$ 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 $[t_1, t_2]$ for sufficiently small $\delta > 0$). It is easy to check that $\lim_{\delta \to 0} t_{1\delta} = \lim_{\delta \to 0} t'_{1\delta} = t_1$ and $\lim_{\delta \to 0} t_{2\delta} = \lim_{\delta \to 0} t'_{2\delta} = t_2$. Therefore, since $h_\delta(t) = -\delta$, for all $t \in [t_{1\delta}, t_{2\delta}]$,

$$\begin{array}{l} \frac{1}{\delta} \int_{t_{1\delta}}^{t'_{1\delta}} \mathcal{S}(t,\mathcal{U}) h_\delta(t) dt - \frac{1}{\delta} \int_{t_{1\delta}}^{t'_{1\delta}} \mathcal{C}(t,\mathcal{U}) \dot{\mathcal{U}}(t) dt + \\ \frac{1}{\delta} \int_{t'_{2\delta}}^{t_{2\delta}} \mathcal{S}(t,\mathcal{U}) h_\delta(t) dt - \frac{1}{\delta} \int_{t'_{2\delta}}^{t_{2\delta}} \mathcal{C}(t,\mathcal{U}) \dot{\mathcal{U}}(t) dt + \int_{t'_{1\delta}}^{t'_{2\delta}} \mathcal{S}(t,\mathcal{U}) dt \geq 0. \end{array}$$

Notice that

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

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

$$\int_{t_1}^{t_2} \mathcal{S}(t,\mathcal{U}) dt = \lim_{\delta \to 0} \int_{t'_{1\delta}}^{t'_{2\delta}} \mathcal{S}(t,\mathcal{U}) dt \geq \lim_{\delta \to 0} \inf \frac{1}{\delta} \left(\int_{t_{1\delta}}^{t'_{1\delta}} \mathcal{C}(t,\mathcal{U}) \dot{\mathcal{U}}(t) dt + \int_{t'_{2\delta}}^{t_{2\delta}} \mathcal{C}(t,\mathcal{U}) \dot{\mathcal{U}}(t) dt \right) \geq 0.$$

Therefore, the first result holds.

Suppose that $\mathcal{U}$ has kink at t_1 and at t_2. Given $\delta > 0$ sufficiently small, define the following rent projection function:

$$\mathcal{V}_\delta(t) = \begin{cases} \max\{(\beta - \delta)(t - \mathcal{U}(t_1)) + \mathcal{U}(t_1), \beta t + \alpha - \delta, (\beta + \delta)(t - \mathcal{U}(t_2)) + \mathcal{U}(t_2)\} & \text{if } t \in [t_1, t_2] \\ \mathcal{U}(t) & \text{if otherwise} \end{cases}$$

which is obviously feasible for δ sufficiently small. Define $t_{1\delta}$ and $t_{2\delta}$ the solutions of $(\beta - \delta)(t - \mathcal{U}(t_1)) + \mathcal{U}(t_1) = \beta t + \alpha - \delta$ and $\beta t + \alpha - \delta = (\beta + \delta)(t - \mathcal{U}(t_2)) + \mathcal{U}(t_2)$, respectively. It is easy to see that $\lim_{\delta \to 0} t_{1\delta} = t_1$ and $\lim_{\delta \to 0} t_{2\delta} = t_2$. Therefore, since $h_\delta(t) = \delta$ for all $t \in [t_{1\delta}, t_{2\delta}]$,

$$\begin{array}{l} \frac{1}{\delta} \int_{t_1}^{t_{1\delta}} \mathcal{S}(t,\mathcal{U}) h_\delta(t) dt - \int_{t_1}^{t_{1\delta}} \mathcal{C}(t,\mathcal{U}) dt + \\ \frac{1}{\delta} \int_{t_{2\delta}}^{t_2} \mathcal{S}(t,\mathcal{U}) h_\delta(t) dt + \int_{t_{2\delta}}^{t_2} \mathcal{C}(t,\mathcal{U}) dt - \int_{t_{1\delta}}^{t_{2\delta}} \mathcal{S}(t,\mathcal{U}) dt \geq 0. \end{array}$$

As 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, $\int_{t_1}^{t_2} \mathcal{S}(t,\mathcal{U})dt \leq 0$. Combining these two inequalities gives the desired result.

Next, consider rotations and $\underline{\mathcal{E}} \notin [t_1, t_2]$ and $t_2 < 1$. Given $\delta > 0$ sufficiently small, define the following rent projection function:

$$\mathcal{V}_\delta(t) = \max\left\{\mathcal{U}(t),\ (\beta+\delta)(t-t_1) + \mathcal{U}(t_1)\right\},$$

which represents a small anti-clockwise rotation of the affine function $\mathcal{U}$ on $[t_1, t_2]$ at point $(t_1, \mathcal{U}(t_1))$ in the plane type versus informational rent. This perturbation is feasible. Applying Lemma 12, we obtain

$$\int_{t_1}^{t_{2\delta}} \mathcal{S}(t,\mathcal{U})h(t)dt \geq 0,$$

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

$$\int_{t_1}^{t_2} \mathcal{S}(t,\mathcal{U})(t-t_1)dt \geq 0.$$

Analogously, we can make a small clockwise rotation of $\mathcal{U}$ on $[t_1, t_2]$ at point $(t_2, \mathcal{U}(t_2))$ and conclude that

$$\int_{t_1}^{t_2} \mathcal{S}(t,\mathcal{U})(t-t_2)dt \leq 0.$$

If $\mathcal{U}$ has kink at t_1 (at t_2), then we can do also a small anti-clockwise (clockwise) rotation at t_2 (at t_1) and get the equality. If $\mathcal{U}$ has kink at both at t_1 and t_2, using that $\int_{t_1}^{t_2} \mathcal{S}(t,\mathcal{U})dt = 0$, we conclude the last equality for this case.

The cases where $t_1 = \underline{t}$ and $t_2 \geq \underline{\mathcal{E}}$ or $t_2 = 1$ are analogous. The only difference in the first case is that we have to consider the rotation at the point $(\underline{\mathcal{E}}, C)$ to eliminate the point effect from $\underline{\mathcal{E}}$ in the condition of Lemma 12. And in the second case, only clockwise rotation at t_1 and at t_1 are allowed if $\beta = \Delta x$.

Proof of Proposition 7

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

$$\mathcal{V}(t) := tu(I) + (1-t)\,u(I-L). \tag{26}$$

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:

$$\mathcal{U}(t) \geq \mathcal{V}(t). \tag{27}$$

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

Lemma 14. *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 $\mathcal{V}$ for $t = 1$ at condition (27) yields

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

Since $K = L$, $B \leq L$. 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 15. *In any optimal mechanism, the set of diagonal types that do not participate is an interval of the form* $(\tilde{t}, 1]$ *for some* $\tilde{t} \in [0, 1)$.

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 $\tilde{t} \in [0, 1)$ such that $\mathcal{U}(t) \geq \mathcal{V}(t)$ if and only if $t \leq \tilde{t}$. □

Expressing the utility of off-the-diagonal types using the projection into the diagonal, Lemma 15 implies that types will prefer not to participate if $p_0 \geq \tilde{t}$, or $p_1 \geq \mathcal{E}_U(\tilde{t})$.

Lemma 16. *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 15, $\dot{\mathcal{V}}(t) = u(I) - u(I - L)$, and $\dot{\mathcal{U}}(t) = b(t, t)$ is a non-decreasing function of t. Moreover, by Lemma 14, $\dot{\mathcal{U}}(t) \leq \dot{\mathcal{V}}(t)$. 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 14, $b(t, t) \leq u(I) - u(I - L)$, for all t. Because b is non-decreasing, there are two possible cases:

- there exists $\alpha > 0$ such that $b(t,t) = u(I) - u(I-L)$ for all $t > 1-\alpha$, 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 $t > 1-\alpha$, where $\alpha > 0$. By Lemma 16, we must have

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

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

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

where $\tilde{t}$ is the last type who participates.

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

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

Note that the perturbation preserves $\dot{\mathcal{U}}$ and $\mathcal{E}$. Let $\tilde{t}_\alpha$ denote the highest diagonal type who participates:

$$\mathcal{U}\left(\tilde{t}_\alpha\right) - \alpha = \mathcal{V}\left(\tilde{t}_\alpha\right).$$

(Note that, by Lemma 16, $\tilde{t}_0 = 1$). Substituting the expression for $\mathcal{V}$, yields

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

Total differentiation gives:

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

Therefore, this perturbation excludes a positive mass of types. We will show that, for small α, 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

$$\Pi_\alpha = \int_0^{\tilde{t}_\alpha} \left(t\Delta x - G(\mathcal{U}-\alpha, \dot{\mathcal{U}}, t)\right) F_0(t, \mathcal{E}) dt + \int_{\underline{\mathcal{E}}}^{\min\left\{1; \tilde{t}_\alpha + \frac{C}{u(x_H)-u(x_L)}\right\}} (t\Delta x - G(\mathcal{U}-\alpha, \dot{\mathcal{U}}, t)) F_1(\mathcal{E}^{-1}, t) dt.$$

Since $\tilde{t}_0 = 1$, it follows that $1 < \tilde{t}_\alpha + \frac{C}{u(x_H)-u(x_L)}$ for α small enough. Differentiating with respect to α, yields

$$\frac{\partial \Pi_\alpha}{\partial \alpha} = \left(t\Delta x - G(\mathcal{U}-\alpha, \dot{\mathcal{U}}, t)\right) F_0(t, \mathcal{E})\Big|_{t=\tilde{t}_\alpha} \frac{\partial \tilde{t}_\alpha}{\partial \alpha}$$

$$+ \int_0^{\tilde{t}_\alpha} \frac{\partial G}{\partial \mathcal{U}}(\mathcal{U}-\alpha, \dot{\mathcal{U}}, t) F_0(t, \mathcal{E}) dt + \int_{\underline{\mathcal{E}}}^1 \frac{\partial G}{\partial \mathcal{U}}(\mathcal{U}-\alpha, \dot{\mathcal{U}}, t) F_1(\mathcal{E}^{-1}, t) dt.$$

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

Moreover, $\lim_{\alpha \downarrow 0} \tilde{t}_\alpha = 1$ and

$$t\Delta x - G(\mathcal{U}-\alpha, \dot{\mathcal{U}}, t)\Big|_{t=1} = \Delta x - u^{-1}(\mathcal{U}(1)).$$

By Lemma 16, $\mathcal{U}(1) = u(x_H)$. Therefore,

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

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

Proof of Proposition 8

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

Lemma 17. *Let $\tilde{t}$ be the first diagonal type to be excluded: $\mathcal{U}(t) > \mathcal{V}(t)$ for $t < \tilde{t}$ and $\mathcal{U}(\tilde{t}) = \mathcal{V}(\tilde{t})$. Then, $b(t,t) < u(I) - u(I-L)$ for all $t < \tilde{t}$.*

Proof. The proof follows from the fact that $\mathcal{U}$ is convex with slope $\dot{\mathcal{U}}(t) = b(t,t)$, whereas $\mathcal{V}$ is affine with slope $\dot{\mathcal{V}}(t) = u(I) - u(I-L)$ (see the proof of Lemma 15). □

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

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

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

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

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

$$p_1 \geq \mathcal{E}^{-1}(p_1) + \frac{C}{b(p_1, p_1)} > \mathcal{E}^{-1}(p_1) + \frac{C}{u(I) - u(I - L)},$$

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

$$\mathcal{E}(\hat{p}_0) > \hat{p}_0 + \frac{C}{u(I) - u(I - L)}.$$

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

Proof of Proposition 9

The proof of existence of an optimal mechanism is analogous to the proof of Proposition 1.

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

(2) Using item (1), the proof is analogous to the proof of Lemma 8 and Proposition 8.

(3) The proof is analogous to the proof of Proposition 5.

References

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

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

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

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, 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, 177–200.

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

Carroll, G. (Forthcoming): "Robustness and Linear Contracts," *American Economic Review*.

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

Chaigneau, P., A. Edmans, and D. Gottlieb (2014): "The Value of Informativeness for Contracting," Tech. rep., HEC Montreal, LBS, and Wharton.

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*, 81, 1935–1971.

Chiappori, P.-A. and B. Salanie (2003): "Testing Contract Theory: A Survey of Some Recent Work," in *Advances in Economics and Econometrics*, ed. by M. Dewatripont, L. P. Hansen, and S. T. Turnovsky, Cambridge: Cambridge University Press, vol. 1.

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

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

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

De Meza, D. and D. C. Webb (2001): "Advantageous Selection in Insurance Markets," *RAND Journal of Economics*, 32, 249–62.

Dewatripont, M., P. Legros, and S. A. Matthews (2003): "Moral Hazard and Capital Structure Dynamics," *Journal of the European Economic Association*, 1, 890–930.

Diamond, P. A. (1998): "Optimal Income Taxation: an Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 88, 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, 8–27.

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

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

Edmans, A. and X. Gabaix (2011): "Tractability in Incentive Contracting," *Review of Financial Studies*, 24, 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, 7–45.

Hart, O. D. and B. Holmstrom (1987): "The Theory of Contracts," in *Advances in Economic Theory, Fifth World Congress*, ed. by T. Bewley, Cambridge: Cambridge University Press.

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

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

Judd, K. and C.-L. Su (2006): "Optimal Income Taxation with Multidimensional Taxpayer Types," Tech. rep.

Jullien, B., B. Salanie, and F. Salanie (2007): "Screening Risk-Averse Agents under Moral Hazard: Single-Crossing and the CARA Case," *Economic Theory*, 30, 151–69.

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

Kleven, H. J., C. T. Kreiner, and E. Saez (2009): "The Optimal Income Taxation of Couples," *Econometrica*, 77, 537–60.

Laffont, J.-J. and D. Martimort (2002): *The Theory of Incentives: The Principal-Agent Model*, Princeton University Press.

Laffont, J.-J., E. Maskin, and J.-C. Rochet (1987): *Optimal Nonlinear Pricing with Two-Dimensional Characteristics* -, Minneapolis: University of Minnesota Press, 256–66.

Laffont, J.-J. and J. Tirole (1986): "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, 94, 614–641.

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

Maskin, E. and J. Riley (1984): "Monopoly with Incomplete Information," *RAND Journal of Economics*, 15, 171–96.

Matthews, S. A. (2001): "Renegotiating Moral Hazard Contracts under Limited Liability and Monotonicity," *Journal of Economic Theory*, 97, 1–29.

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

——— (1972): "On Producer Taxation," *Review of Economic Studies*, 39, 105–11.

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

Mussa, M. and S. Rosen (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18, 301–17.

Myerson, R. B. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73.

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

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

Piketty, T. (1997): "La Redistribution Fiscale Face au Chômage," *Revue Française d'Économie*, 12, 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, Amsterdam: Elsevier-North Holland, vol. 5.

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, 215–34.

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

Rochet, J.-C. and P. Choné (1998): "Ironing, Sweeping, and Multidimensional Screening," *Econometrica*, 66, 783–826.

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

——— (2003): *The Economics of Multidimensional Screening* -, Econometric Society Monographs, advances in economics and econometrics: theory and applications - ed.

Rothschild, C. and F. Scheuer (2013): "Redistributive Taxation in the Roy Model," *Quarterly Journal of Economics*, 128, 623–668.

——— (2014): "Optimal Taxation with Rent-Seeking," Tech. rep., Middlebury College and Stanford University.

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

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

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

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

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

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

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

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

Online Appendix I: Private Information on Costs

Statement of the Problem

In this appendix, we assume that the agent also has private information about his cost of effort. Thus, we assume that the cost C is privately known by the agent. Therefore, the agent's type is now (p_0, p_1, C).

The principal's beliefs about the agent's type is represented by a continuous density h over types on the set $\boldsymbol{P} \times [\underline{C}, \overline{C}]$, where $\underline{C} > 0$. We assume that, for all $C \in [\underline{C}, \overline{C}]$, the conditional distribution

$$f(\boldsymbol{p}|C) = \frac{h(\boldsymbol{p}, C)}{h_C(C)}$$

has full support on $\boldsymbol{P}$, where $h_C(C) = \int_{\boldsymbol{P}} h(\boldsymbol{p}, C) d\boldsymbol{p}$ is the marginal distribution of C.

A mechanism in utility terms is a function $(w, b, e) : \boldsymbol{P} \times [\underline{C}, \overline{C}] \to \mathbb{R}^2 \times \{0, 1\}$. Given a mechanism (w, b, e) , a type-$(\mathbf{p}, C)$ agent obtains expected utility

$$U(\mathbf{p}, C) \equiv w(\mathbf{p}, C) + p_{e(\mathbf{p},C)} b(\mathbf{p}, C) - c_{e(\mathbf{p},C)}. \tag{28}$$

We can easily define the incentive compatibility constraint (IC), individual rationality (IR), free disposal (FD) and feasible mechanisms for this extended model. Using the iterated expected law, the principal's expected utility is

$$\begin{aligned} &E_C \Big[\int_{\boldsymbol{P}} \ \big\{ p_{e(\mathbf{p},C)} \left[x_H - u^{-1}\left(w(\mathbf{p}, C) + b(\mathbf{p}, C) \right) \right] \\ &+ \left(1 - p_{e(\mathbf{p},C)} \right) \left[x_L - u^{-1}\left(w(\mathbf{p}, C) \right) \right] \big\} f(\mathbf{p}|C) d\mathbf{p} \Big], \end{aligned}$$

where $E_C[\cdot]$ represents the expectation operator with respect to the marginal distribution h_C. Notice that, conditional on C, the inner integral corresponds exactly to the principal's expected utility (1) in the text. We can also define equivalent and optimal mechanisms in the same fashion.

Feasible Mechanisms

We now show how the characterization results we derived for the model with known costs extend to this more general framework. The first set of results are the necessary and sufficient conditions for a mechanism to be feasible.

Lemma 18. *For any feasible mechanism, there exists an equivalent mechanism (w, b, e) such that $e(p_0, p_1, C) = 1$ if and only if $p_1 > \mathcal{E}(p_0, C)$ for a continuous and non-decreasing function $\mathcal{E} : [0, 1] \times [\underline{C}, \overline{C}] \to [0, 1]$.*

For a given mechanism, Lemma 18 defines the effort frontier as associated with it. For a given feasible mechanism (w, b, e), we refer to the function $\mathcal{E}$ as the *effort frontier* associated with it. Conditional on C, the effort frontier partitions the type space into types who exert low and high efforts:

$$e(p_0, p_1, C) = 1 \iff p_1 > \mathcal{E}(p_0, C). \tag{29}$$

The following lemma establishes necessary conditions for incentive compatibility:

Lemma 19. *Let (w, b, e) be a feasible mechanism and let $\mathcal{E}$ and U be the effort frontier and informational rent functions associated with it. Then:*

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

$$\nabla U(p_0, p_1, C) = \begin{cases} (b(p_0, p_1, C), 0, 0) & \text{if } p_1 < \mathcal{E}(p_0, C) \\ (0, b(p_0, p_1, C), -1) & \text{if } p_1 > \mathcal{E}(p_0, C) \end{cases};$$

b. $b(p_0, p_1, C)$ is constant in C, constant in p_1 for $p_1 < \mathcal{E}(p_0, C)$ and constant in p_0 for $p_1 > \mathcal{E}(p_0, C)$;

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

d. $U(p_1, p_1, C) = U(p_0, p_1, C) + C$ for $p_1 > \mathcal{E}(p_0, C)$.[37]

Lemma 19 extends Lemma 2. Properties (a) and (b) are the local first- and second-order conditions of the agent's maximization program. Notice that the rent function does not depend on C on the low effort region, which will allows us to extend the one-dimensional projection method that follows. Property (c) gives the participation and free disposal constraints for the lowest type.

We also have an analogous version of Lemma 3 which says that conditions (a)-(d) are also sufficient for feasibility.

Lemma 20. *Fix a mechanism (w, b, e), and let U denote the associated informational rent function defined according to equation (28). The mechanism is feasible if and only if it satisfies conditions (a)-(d) for an effort frontier function $\mathcal{E}$ satisfying condition (29).*

One-Dimensional Conditions

The next step is to define the one-dimensional conditions. The key observation is that for any given feasible mechanism, we can define the rent projection associated with this mechanism in the same way we did for the for the model with known costs, i.e., the function $\mathcal{U} : [0, 1] \to \mathbb{R}$ defined as $\mathcal{U}(t) := U(t, t, C)$. The reason is that as we remarked just after Lemma 19, for types in the low effort region (e.g., (t, t, C)) the rent function does not depend on C. That said, we can easily replicate all the results of Subsection 2.3 for the extended model with the convenient adaptions. The following lemma extends Lemma 4 and establishes that any non-trivial mechanism is characterized by the one-dimensional functions $\mathcal{U}$ and $\mathcal{E}$:

Lemma 21. *Let (w, b, e) be a nontrivial feasible mechanism and let $\mathcal{E}$ and $\mathcal{U}$ denote the effort frontier and rent projection functions associated with it. Then:*

$$b(p_0, p_1, C) = \begin{cases} \dot{\mathcal{U}}(p_0) & \text{if } p_1 < \mathcal{E}(p_0, C) \\ \dot{\mathcal{U}}(p_1) & \text{if } p_1 > \mathcal{E}(p_0, C) \end{cases} \quad (a.e.), \tag{30}$$

$$w(p_0, p_1, C) = \begin{cases} \mathcal{U}(p_0) - p_0\dot{\mathcal{U}}(p_0) & \text{if } p_1 < \mathcal{E}(p_0, C) \\ \mathcal{U}(p_1) - p_1\dot{\mathcal{U}}(p_1) & \text{if } p_1 > \mathcal{E}(p_0, C) \end{cases} \quad (a.e.), \quad \text{and} \tag{31}$$

$$\mathcal{U}(\mathcal{E}(p_0, C)) = \min\{\mathcal{U}(p_0) + C; \mathcal{U}(1)\}. \tag{32}$$

[37] Note that by the observation above, $U(p_1, p_1, C)$ does not depend on C since type (p_1, p_1, C) belongs to the low effort region.

The following lemma establishes the equivalence between the feasibility of a mechanism and the feasibility of its rent projection:

Lemma 22 (**One-Dimensional Characterization of Feasibility**). *Let (w, b, e) be a feasible mechanism, and let $\mathcal{U}$ and $\mathcal{E}$ be the rent projection and effort frontier functions associated with it. Then, $\mathcal{U}$ is a feasible rent projection and $(\mathcal{U}, \mathcal{E})$ solves equation (32). Conversely, let $\mathcal{U}$ be a feasible rent projection, let $\mathcal{E}$ be defined by the solution of equation (32), and let (w, b, e) be given by equations (29), (30) and (31). Then, (w, b, e) is a feasible mechanism.*

Given the cost G of providing projected rent $\mathcal{U}$ and power $\dot{\mathcal{U}}$ to the agent with type (t, t, C) previously defined, the principal's payoff becomes

$$\begin{array}{l} x_L + E_C \left[\int_0^1 \left(t\Delta x - G(\mathcal{U}, \dot{\mathcal{U}}, t) \right) F_0 \left(t, \mathcal{E} | C \right) dt \right. \\ \left. + \int_{\mathcal{E}(0,C)}^1 \left(t\Delta x - G(\mathcal{U}, \dot{\mathcal{U}}, t) \right) F_1 \left(\mathcal{E}^{-1}, t | C \right) dt \right], \end{array} \tag{33}$$

where $F_0(t, s|C) \equiv \int_t^s f(t, z|C) dz$ and $F_1(s, t|C) \equiv \int_0^s f\left(z, t|C\right) dz$ and $\mathcal{E}^{-1}\left(t, C\right) = \sup \left\{ p_0 : \mathcal{E}\left(p_0, C\right) \leq t \right\}$. We are omitting the dependence of the functions $\mathcal{U}$, $\mathcal{E}$ and $\mathcal{E}^{-1}$ on t and C for notational simplicity. The program (P) for the extended model can be rewritten as the maximization of the objective function (33) subject to (32), $\mathcal{U}$ nondecreasing and convex, and $\mathcal{U}\left(0\right) \geq 0$.

General Properties

Using the representation of the principal's payoff as an iterated expectation (33), we will establish general properties of the optimal mechanism conditionally on C.

Proposition 12 (**Zero Rents at the Bottom**). *No mechanism that gives strictly positive informational rents for almost all types is BDF-optimal.*

The next proposition extends Proposition 1.

Proposition 13 (**Existence**). *There exists an optimal mechanism.*

Lemma 23. *Let (w, b, e) be an optimal mechanism and let $\mathcal{E}$ be the effort frontier function associated with it. Then, $\mathcal{E}$ is continuous, differentiable (a.e.), and $\dot{\mathcal{E}} \leq 1$ at all points of differentiability (where the dot represents the derivative with respect to t). Moreover, there exists $\underline{t} > 0$ such that $\mathcal{E}\left(t, C\right) = \mathcal{E}(0, C)$ for all $t < \underline{t}$ and C.*

Proposition 14 (**Exclusion**). *It is optimal to exclude a strictly positive mass of types if and only if exclusion of types is first-best optimal.*

Risk Aversion

As in the main text, let

$$\underline{\mathcal{E}}(C) := \mathcal{U}^{-1}\left(C\right), \quad \underline{t} := \sup \ \left\{ t : \mathcal{U}\left(t\right) = 0 \right\}, \text{ and } \overline{t}(C) := \inf \left\{ t : \mathcal{E}(t, C) = 1 \right\}$$

denote the lowest projected type for which there is high effort, the lowest projected type with positive rents, and the projected type for which the effort frontier hits $p_1 = 1$. Let $\mathcal{S}(t,\mathcal{U}) := S_0(t,\mathcal{U}) f(t,\mathcal{E}) + S_1(t,\mathcal{U}) f\left(\mathcal{E}^{-1},t\right)$ denote the sum of the effects on low-effort region S_0 and on the high-effort region S_1 weighted by their probability densities, where:

$$S_0(t,\mathcal{U}|C) := \begin{cases} -\frac{(\mathcal{E}-t)\Delta x-(G(\mathcal{E})-G)}{\dot{\mathcal{U}}(\mathcal{E})} - \frac{\partial G}{\partial \mathcal{U}} \frac{F_0(t,\mathcal{E}|C)}{f(t,\mathcal{E}|C)} & \text{if } t < \overline{t}(C) \\ -\frac{\partial G}{\partial \mathcal{U}} \frac{F_0(t,1|C)}{f(t,1|C)} & \text{if } t \geq \overline{t}(C) \end{cases},$$

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

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

$$\underline{S}(\mathcal{U}|C) := \frac{\left(\underline{\mathcal{E}}(C) - E\left[t|t \leq \underline{t},C\right]\right)\Delta x - G(\underline{\mathcal{E}}(C))}{\dot{\mathcal{U}}(\underline{\mathcal{E}}(C))} F_1(\underline{t},\underline{\mathcal{E}}(C)|C)$$

denote the marginal effect at $\underline{t}$, where $E\left[t|t \leq \underline{t},C\right] := \frac{\int_0^{\underline{t}} t f(t,\underline{\mathcal{E}}(C)|C)dt}{F_1(\underline{t},\underline{\mathcal{E}}(C)|C)}$. Let

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

denote the weighted marginal cost of providing power, where

$$C_0(t,\mathcal{U}|C) := \begin{cases} \frac{\partial G}{\partial \dot{\mathcal{U}}} \frac{F_0(t,\mathcal{E}|C)}{f(t,\mathcal{E}|C)} \text{ if } t < \overline{t}(C) \\ \frac{\partial G}{\partial \dot{\mathcal{U}}} \frac{F_0(t,1|C)}{f(t,1|C)} \text{ if } t \geq \overline{t}(C) \end{cases}, \text{ and}$$

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

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

Lemma 24. *Let $\mathcal{U}$ be the rent projection associated with an optimal mechanism. Then, for any feasible $\mathcal{V} : [0,1] \to \mathbb{R}$,*

$$\begin{aligned} &\int_0^1 \left[\mathcal{U}(t) - \mathcal{V}(t)\right] E_C\left[\mathcal{S}(t,\mathcal{U}|C)\right] dt - \int_0^1 \left[\dot{\mathcal{U}}(t) - \dot{\mathcal{V}}(t)\right] E_C\left[\mathcal{C}(t,\mathcal{U}|C)\right] dt \\ &+ E_C\left[\left[\mathcal{U}(\underline{\mathcal{E}}(C)) - \mathcal{V}(\underline{\mathcal{E}}(C))\right] \underline{S}(\mathcal{U}|C)\right] \geq 0. \end{aligned}$$

The following theorem determines the necessary optimality conditions:

Theorem 3 (Optimal Mechanisms under Risk Aversion). *Let $\mathcal{U}$ be an optimal rent projection. Suppose that $[t_1,t_2] \subset [0,1]$ is a non-degenerate interval such that $\underline{\mathcal{E}}(C) \notin [t_1,t_2]$, for all C.*

1. **(pointwise condition)** *If $\mathcal{U}$ is strongly convex in $[t_1,t_2]$, then*

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

for almost all $t \in [t_1, t_2]$.

2. **(bunching conditions)** *Let* $[t_1, t_2]$ *be a maximal interval where* $\mathcal{U}$ *is affine. Then*

$$0 \geq t_1 \int_{t_1}^{t_2} E_C \left[\mathcal{S}\left(t, \mathcal{U}|C\right)\right] dt \geq \int_{t_1}^{t_2} t E_C \left[\mathcal{S}\left(t, \mathcal{U}|C\right)\right] dt \geq t_2 \int_{t_1}^{t_2} E_C \left[\mathcal{S}\left(t, \mathcal{U}|C\right)\right] dt.$$

If $\mathcal{U}$ *has kink at* t_1 *(at* t_2*), then* $\int_{t_1}^{t_2} (t - t_1) E_C \left[\mathcal{S}(t, \mathcal{U}|C)\right] dt = 0$ *(*$\int_{t_1}^{t_2} (t - t_2) E_C \left[\mathcal{S}(t, \mathcal{U}|C)\right] dt = 0$*).*

Risk Neutrality

Next, we generalize the results from Section 3.2:

Proposition 15 (**Two Contracts at the Bottom**). *Let* $\mathcal{U}$ *be an optimal rent projection of a nontrivial mechanism. There exist* $\hat{\mathcal{E}} \geq \underline{\mathcal{E}}(\underline{C})$ *and constant* $b \in (\underline{C}, \Delta x]$ such that $\underline{t} \in (0, \underline{\mathcal{E}}(\underline{C}))$ *and*

$$\dot{\mathcal{U}}(t) = \begin{cases} 0 \text{ if } t \in [0, \underline{t}) \\ b \text{ if } t \in [\underline{t}, \hat{\mathcal{E}}) \end{cases}.$$

Lemma 25. *Let* $\mathcal{E}$ *be the effort frontier function associated with it an optimal mechanism. Then,* $\mathcal{E}\left(t, C\right) \geq t + \frac{C}{\Delta x}$ *whenever* $\mathcal{E}\left(t, C\right) < 1$.

Finite Mechanisms

For this subsection, suppose that the conditional output probabilities $\boldsymbol{p}$ are independent from the incremental costs of effort C and let $f\left(\mathbf{p}\right)$ denote the marginal distribution of output probabilities $\boldsymbol{p}$. Let $\mathcal{E}$ be the rent projection of an optimal mechanism, and let $\underline{\mathcal{E}} = \mathcal{E}(0, \underline{C})$ and $\overline{\mathcal{E}} = \mathcal{E}(0, \overline{C})$.

Lemma 26. *Suppose that* $f(\mathbf{p})$ *satisfies increasing rents. There exists* $\hat{t} \in [\overline{\mathcal{E}}, 1]$ *such that the optimal rent projection is a piecewise linear function with at most two pieces on* $\left[\hat{t}, 1\right]$.

Proposition 16 (**Two Contracts at Top**). *Suppose that* $f(\mathbf{p})$ *satisfies increasing rents and let* $\Delta x \leq 2\overline{C}$. *Then* $\mathcal{E}$ *is piecewise linear with two pieces on* $[0, \underline{\mathcal{E}}]$ *and at most two pieces on* $[\overline{\mathcal{E}}, 1]$.

As in Section 3.3, let $\boldsymbol{P}(\underline{p})$ denote the modified type space in which the probability of success is bounded below by $\underline{p}$.

Proposition 17 (**One Contract at Top**). *Suppose that* $f(p_0, p_1)$ *is non-increasing in* p_0 *and has full support on* $\boldsymbol{P}(\underline{p})$, *and let* $\underline{p} \geq \frac{\overline{C}}{\underline{C}} \frac{\Delta x - \underline{C}}{\Delta x + \overline{C}}$. *Then, for the BDF-optimal mechanism,* $\mathcal{E}$ *is piecewise linear with two pieces on* $[0, \underline{\mathcal{E}}]$ *and with at most one piece on* $[\overline{\mathcal{E}}, 1]$.

Applications

We now extend some properties of the applications of the model to insurance and procurement/regulation. First, we establish that it is optimal to exclude a positive mass of types:

Proposition 18 (**Exclusion in Insurance**). *There exists* $\bar{p}_0 < 1$ *such that it is optimal to exclude type* (p_0, p_1, C) *if and only if* $p_0 \geq \bar{p}_0$ *or* $p_1 \geq \bar{p}_0 + \frac{C}{u(I) - u(I - L)}$.

Next, we establish that insured agents exert "less effort" than if they were uninsured::

Proposition 19 (**Strict Distortion Relative to No Insurance**). *Let $\mathcal{E}$ 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 18. Then, $\mathcal{E}(p_0, C) > p_0 + \frac{C}{u(I)-u(I-L)}$ for all $p_0 < \bar{p}_0$.*

The following proposition summarizes the results of the procurement/regulation model when the regulated firm also has private information about the incremental cost of effort:

Proposition 20 (**Optimal Regulation**). *There exists an optimal mechanism, which has the following properties:*

1. *There exists $\hat{\mathcal{E}} > \underline{\mathcal{E}}$ and $\underline{t} \in (0, \underline{\mathcal{E}}(\underline{C}))$ such that*

   - *All types $\mathbf{p} \in [0, \underline{t}) \times [0, \underline{\mathcal{E}}(C)) \cap \boldsymbol{P}$ get a cost-plus contract ($w = 0$, $b = 0$), exert zero effort, and get zero rents;*
   - *All types $\mathbf{p} \in [\underline{t}, \hat{\mathcal{E}}] \times [0, 1] \cap \boldsymbol{P}$ exerting low effort and $\mathbf{p} \in [0, 1] \times [\underline{\mathcal{E}}(\underline{C}), \hat{\mathcal{E}}] \cap \boldsymbol{P}$ exerting high effort get a uniform contract with positive power ($w < 0$, $b \in (\underline{C}, \Delta x]$) and get positive rents.*

2. *Exclusion is optimal if and only if exclusion is first-best optimal; and*

3. There is weak insufficient effort.

Proofs

The proofs of Lemmas 18, 19, 20, 21, 22 and 25, and of Propositions 13, 14, **??**, 18 and 19 are analogous to the corresponding ones in the case of known costs.

Proof of Proposition 12

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

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

The mechanism induced by the rent function $\mathcal{U}_\alpha$ uniformly reduces the rent by α of all types $(\boldsymbol{p}, C)$ such that $\boldsymbol{p}$ in $([0, \underline{t}_\alpha] \times [0, \underline{\mathcal{E}}_\alpha(C)]) \cap \boldsymbol{P}$ which have zero rent, where $\underline{t}_\alpha$ and $\underline{\mathcal{E}}_\alpha(C)$ are defined as

$$\mathcal{U}(\underline{t}_\alpha) = \alpha \text{ and } \mathcal{U}(\underline{\mathcal{E}}_\alpha(C)) - \alpha = C.$$

It is immediate that $\mathcal{U}_\alpha$ satisfies the constraints of the principal's program and, therefore, the mechanism it induces is feasible.

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

$$\frac{d\underline{\mathcal{E}}_\alpha}{d\alpha} = \frac{1}{\dot{\mathcal{U}}(\underline{\mathcal{E}}_\alpha)} \geq 0.$$

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

$$G_\epsilon(t) = \begin{cases} G(\mathcal{U}(t) - \alpha, \dot{\mathcal{U}}(t), t) & \text{if } t > \underline{t}_\alpha \\ u^{-1}(0) & \text{if } t \leq \underline{t}_\alpha \end{cases}.$$

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

$$\Pi_\alpha := E_C \left[\int_0^1 (t\Delta x - G_\alpha(t))\, F_0(t, \mathcal{E}_\alpha | C) dt + \int_{\underline{\mathcal{E}}_\alpha}^1 (t\Delta x - G_\alpha(t))\, F_1(\mathcal{E}^{-1}, t | C) dt \right],$$

where we are using the fact that neither the effort frontier changes for all $t \geq \underline{t}_\alpha$ nor its inverse $\mathcal{E}^{-1}$ for all $t \geq \underline{\mathcal{E}}_\alpha$.

Take the derivative of Π_α with respect to α and evaluate at 0:

$$\left.\frac{d\Pi_\alpha}{d\alpha}\right|_{\alpha=0} = E_C \left[\int_0^1 \frac{\partial G}{\partial \mathcal{U}} F_0(t, \mathcal{E} | C) dt + \int_0^{\underline{t}_0} (t\Delta x - G_0)\, f(t, \mathcal{E} | C)\, \left.\frac{d\underline{\mathcal{E}}_\alpha}{d\alpha}\right|_{\alpha=0} dt \right]$$

$$+ E_C \left[\int_{\underline{\mathcal{E}}_0}^1 \frac{\partial G}{\partial \mathcal{U}} F_1(\mathcal{E}^{-1}, t | C) dt - (\underline{\mathcal{E}}_0 \Delta x - G_0(\underline{\mathcal{E}}_0)) F_1(0, \underline{\mathcal{E}}_0 | C)\, \left.\frac{d\underline{\mathcal{E}}_\alpha}{d\alpha}\right|_{\alpha=0} \right],$$

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 $\underline{t}_0 = 0$ and the fourth is zero since $F_1(0, \underline{\mathcal{E}}_0 | C) = 0$. Therefore, the derivative of Π_α is positive at 0 which implies that principal strictly prefers the mechanism induced by $\mathcal{U}_\alpha$ than the one induced by $\mathcal{U}$ for sufficiently small $\alpha > 0$.

Proof of Lemma 24

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

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

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

$$\Pi(\mathcal{U}) = E_C \left[\int_0^1 (t\Delta x - G(\mathcal{U}, \dot{\mathcal{U}}, t)) F_0(t, \mathcal{E} | C) dt + \int_{\underline{\mathcal{E}}(C)}^1 (t\Delta x - G(\mathcal{U}, \dot{\mathcal{U}}, t)) F_1(\mathcal{E}^{-1}, t | C) dt \right],$$

where $\mathcal{E}$ is obtained from equation (32). Because $\mathcal{U}$ is optimal and $\mathcal{U} + \alpha h$ is feasible, we must have

$$\Pi(\mathcal{U} + \alpha h) \leq \Pi(\mathcal{U}),$$

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

$$\lim_{\alpha \downarrow 0} \frac{\Pi(\mathcal{U} + \alpha h) - \Pi(\mathcal{U})}{\alpha} \leq 0.$$

By equation (32), the effort frontier associated with $\mathcal{U} + \alpha h$, $\mathcal{E}_\alpha$, is defined as the solution to the

following functional equation:

$$\mathcal{U}\left(\mathcal{E}_\alpha(t,C)\right)+\alpha h(\mathcal{E}_\alpha(t,C))=\mathcal{U}\left(t\right)+\alpha h\left(t\right)+C$$

for all $t\in[0,\overline{t}_\alpha]$, where $\overline{t}_\alpha(C)$ solves $\mathcal{U}(\overline{t}_\alpha)+\alpha h(\overline{t}_\alpha)=\mathcal{U}(1)+\alpha h(1)-C$. Taking the total derivative of this expression with respect to α and evaluating at 0, we obtain

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

for all $t\leq\overline{t}$.

Analogously, its inverse, $\mathcal{E}_\alpha^{-1}$, satisfies an analogous functional equation:

$$\mathcal{U}\left(\mathcal{E}_\alpha^{-1}(t,C)\right)+\alpha h(\mathcal{E}_\alpha^{-1}(t,C))=\mathcal{U}\left(t\right)+\alpha h\left(t\right)-C,$$

for all $t\in[\underline{\mathcal{E}}_\alpha,1]$, where $\mathcal{U}(\underline{\mathcal{E}}_\alpha)+\epsilon h(\underline{\mathcal{E}}_\alpha)=C$. Again, taking the total derivative of this expression with respect to α and evaluating at 0, we get:

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

for all $t\geq\underline{\mathcal{E}}$. Applying the same procedure with respect to $\underline{\mathcal{E}}_\alpha$ yields

$$\left.\frac{\partial\underline{\mathcal{E}}_\alpha}{\partial\alpha}\right|_{\alpha=0}=-\frac{h(\underline{\mathcal{E}})}{\dot{\mathcal{U}}(\underline{\mathcal{E}})}.$$

Then,

$$\left.\frac{\partial\mathcal{E}_\alpha}{\partial\alpha}\right|_{\alpha=0}=\frac{h\left(t\right)-h(\mathcal{E})}{\dot{\mathcal{U}}\left(\mathcal{E}\right)},\ \left.\frac{\partial\mathcal{E}_\alpha^{-1}}{\partial\alpha}\right|_{\alpha=0}=\frac{h\left(t\right)-h(\mathcal{E}^{-1})}{\dot{\mathcal{U}}\left(\mathcal{E}^{-1}\right)},\ \left.\frac{\partial\underline{\mathcal{E}}_\alpha}{\partial\alpha}\right|_{\alpha=0}=-\frac{h(\underline{\mathcal{E}})}{\dot{\mathcal{U}}(\underline{\mathcal{E}})},\tag{34}$$

$$\text{and }\dot{\mathcal{U}}_\alpha(t)=\dot{\mathcal{U}}\left(t\right)+\alpha\dot{h}\left(t\right).\tag{35}$$

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

$$\left.\frac{d\Pi_\alpha}{d\alpha}\right|_{\alpha=0}=\lim_{\alpha\downarrow0}\frac{\Pi(\mathcal{U}+\alpha h)-\Pi(\mathcal{U})}{\alpha}.$$

Using conditions (34), we obtain

$$\begin{aligned}\left.\tfrac{d\Pi_\alpha}{d\alpha}\right|_{\alpha=0}=E_C\ \ &\Big[-\textstyle\int_0^1\left\{\frac{\partial G}{\partial\mathcal{U}}h(t)+\frac{\partial G}{\partial\dot{\mathcal{U}}}\dot{h}(t)\right\}F_0(t,\mathcal{E}|C)dt\\&-\textstyle\int_{\underline{\mathcal{E}}}^1\left\{\frac{\partial G}{\partial\mathcal{U}}h(t)+\frac{\partial G}{\partial\dot{\mathcal{U}}}\dot{h}(t)\right\}F_1(\mathcal{E}^{-1},t|C)dt\\&+\textstyle\int_0^{\overline{t}}(t\Delta x-G)\frac{h(t)-h(\mathcal{E})}{\dot{\mathcal{U}}(\mathcal{E})}f(t,\mathcal{E}|C)dt\\&+\textstyle\int_{\underline{\mathcal{E}}}^1(t\Delta x-G)\frac{h(t)-h(\mathcal{E}^{-1})}{\dot{\mathcal{U}}(\mathcal{E}^{-1})}f(\mathcal{E}^{-1},t|C)dt\\&+(\underline{\mathcal{E}}\Delta x-G(\underline{\mathcal{E}})\tfrac{h(\underline{\mathcal{E}})}{\dot{\mathcal{U}}(\underline{\mathcal{E}})}F_1(\underline{t},\underline{\mathcal{E}}|C)\Big].\end{aligned}$$

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

$$\begin{aligned}\int_0^{\bar{t}}(t\Delta x - G)\tfrac{h(\mathcal{E})}{\dot{\mathcal{U}}(\mathcal{E})}f(t,\mathcal{E}|C)dt &= \int_0^{\underline{t}} t\Delta x \tfrac{h(\underline{\mathcal{E}})}{\dot{\mathcal{U}}(\underline{\mathcal{E}})}f(t,\underline{\mathcal{E}}|C)dt \\ &\quad + \int_{\underline{\mathcal{E}}}^1 \left(\mathcal{E}^{-1}\Delta x - G\left(\mathcal{E}^{-1}\right)\right)\tfrac{h(t)}{\dot{\mathcal{U}}(t)}f(\mathcal{E}^{-1},t|C)\dot{\mathcal{E}^{-1}}(t,C)dt \\ \int_{\underline{\mathcal{E}}}^1 (t\Delta x - G)\tfrac{h(\mathcal{E}^{-1})}{\dot{\mathcal{U}}(\mathcal{E}^{-1})}f(\mathcal{E}^{-1},t|C)dt &= \int_0^{\bar{t}} \left(\mathcal{E}\Delta x - G\left(\mathcal{E}\right)\right)\tfrac{h(t)}{\dot{\mathcal{U}}(t)}f(t,\mathcal{E}|C)\dot{\mathcal{E}}(t,C)dt.\end{aligned}$$

Using condition (35) yields:

$$\begin{aligned}\tfrac{d}{d\alpha}\Pi_\alpha\big|_{\alpha=0} = \; & E_C\Big[-\int_0^1\left(\tfrac{\partial G}{\partial \mathcal{U}}F_0(t,\mathcal{E}|C)h(t) + \tfrac{\partial G}{\partial \mathcal{U}}F_0(t,\mathcal{E}|C)\dot{h}(t)\right)dt \\ & - \int_{\underline{\mathcal{E}}}^1\left(\tfrac{\partial G}{\partial \mathcal{U}}F_1(\mathcal{E}^{-1},t|C)h(t) + \tfrac{\partial G}{\partial \mathcal{U}}F_1(\mathcal{E}^{-1},t|C)\dot{h}(t)\right)dt \\ & - \int_0^{\bar{t}}\tfrac{(\mathcal{E}-t)\Delta x-(G(\mathcal{E})-G)}{\dot{\mathcal{U}}(\mathcal{E})}f(t,\mathcal{E}|C)h(t)dt \\ & + \int_{\underline{\mathcal{E}}}^1\tfrac{(t-\mathcal{E}^{-1})\Delta x-(G-G(\mathcal{E}^{-1}))}{\dot{\mathcal{U}}(\mathcal{E}^{-1})}f(\mathcal{E}^{-1},t|C)h(t)dt \\ & + \Big(-\int_0^{\underline{t}}t\Delta x f(t,\underline{\mathcal{E}}|C)dt + (\underline{\mathcal{E}}\Delta x - G(\underline{\mathcal{E}}))F_1(\underline{t},\underline{\mathcal{E}}|C)\Big)\tfrac{h(\underline{\mathcal{E}})}{\dot{\mathcal{U}}(\underline{\mathcal{E}})}\Big].\end{aligned}$$

This establishes the result. Notice that, in the case of Lemma 24, substituting $\mathcal{U} - \mathcal{U}(\mathcal{E}^{-1}) = C$ and $\mathcal{U}(\mathcal{E}) - \mathcal{U} = C$ into the equation above, gives the result claimed in the statement of the lemma.

Proof of Theorem 3

(1) Notice that $E_C\left[\mathcal{S}(t,\mathcal{U}|C)\right]$ is an integrable function on $[t_1,t_2]$ (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 $t\notin(t_1,t_2)$. Since $\mathcal{U}$ is strongly convex on $[t_1,t_2]$, $\mathcal{U}+\alpha h$ is a strongly convex function if $|\alpha|$ is sufficiently small. Since $\underline{\mathcal{E}}(C)\notin[t_1,t_2]$, for all C, performing the variational calculus (given by the previous theorem) for such feasible direction, we get

$$\int_{t_1}^{t_2} E_C\left[\mathcal{S}(t,\mathcal{U}|C)\right]h(t)dt - \int_{t_1}^{t_2} E_C\left[\mathcal{C}(t,\mathcal{U}|C)\right]\dot{h}(t)dt = 0.$$

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

$$\int_{t_1}^{t_2} E_C\left[\int_0^t \mathcal{S}(x,\mathcal{U}|C)dx + \mathcal{C}(t,\mathcal{U}|C)\right]\dot{h}(t)dt = 0.$$

Since the function inside the brackets of the above integral is càdlàg, $\dot{h}$ is a generic continuous function. Lemma 13 implies that

$$E_C\left[\int_0^t \mathcal{S}(x,\mathcal{U}|C)dx + \mathcal{C}(t,\mathcal{U}|C)\right]$$

is constant on $[t_1,t_2]$. Since this function is a.e. differentiable (since $\dot{\mathcal{U}}$ is a.e. differentiable), we have that

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

a.e. on $[t_1, t_2]$.

(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 $\underline{\mathcal{E}}(C) \notin [t_1, t_2]$, for all C. We have that there exist $\beta > 0$ and $\alpha \in \mathbb{R}$ such that $\mathcal{U}(t) = \beta t + \alpha$, for all $t \in [t_1, t_2]$. 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 24, we get

$$E_C\left[\int_{t_{1\delta}}^{t_{2\delta}} \mathcal{S}(t,\mathcal{U}|C)h_\delta(t)dt - \int_{t_{1\delta}}^{t_{2\delta}} \mathcal{C}(t,\mathcal{U}|C)\dot{h}_\delta(t)dt\right] \geq 0,$$

where $h_\delta = \mathcal{U} - \mathcal{V}_\delta$, $t_{1\delta}$ and $t_{2\delta}$ 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 $[t_1, t_2]$ for sufficiently small $\delta > 0$). Let $t'_{1\delta} \geq t_{1\delta}$ and $t'_{2\delta} \leq t_{2\delta}$ 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 $[t_1, t_2]$ for sufficiently small $\delta > 0$). It is easy to check that $\lim_{\delta\to 0} t_{1\delta} = \lim_{\delta\to 0} t'_{1\delta} = t_1$ and $\lim_{\delta\to 0} t_{2\delta} = \lim_{\delta\to 0} t'_\delta = t_2$. Therefore, since $h_\delta(t) = -\delta$ for all $t \in [t_{1\delta}, t_{2\delta}]$,

$$\begin{array}{l} E_C\left[\frac{1}{\delta}\int_{t_{1\delta}}^{t'_{1\delta}} \mathcal{S}(t,\mathcal{U}|C)h_\delta(t)dt - \frac{1}{\delta}\int_{t_{1\delta}}^{t'_{1\delta}} \mathcal{C}(t,\mathcal{U}|C)\dot{\mathcal{U}}(t)dt + \right. \\ \left. \frac{1}{\delta}\int_{t'_{2\delta}}^{t_{2\delta}} \mathcal{S}(t,\mathcal{U}|C)h_\delta(t)dt - \frac{1}{\delta}\int_{t'_{2\delta}}^{t_{2\delta}} \mathcal{C}(t,\mathcal{U}|C)\dot{\mathcal{U}}(t)dt + \int_{t'_{1\delta}}^{t'_{2\delta}} \mathcal{S}(t,\mathcal{U}|C)dt\right] \geq 0. \end{array}$$

Notice that

$$\begin{array}{l} \left|\frac{1}{\delta}\int_{t_{1\delta}}^{t'_{1\delta}} \mathcal{S}(t,\mathcal{U}|C)h_\delta(t)dt\right| \leq \frac{t'_{1\delta}-t_{1\delta}}{\delta}\sup\ \{|\mathcal{S}(t,\mathcal{U}|C)h_\delta(t)|\,; t \in [t_{1\delta}, t'_{1\delta}]\} \\ \leq (t'_{1\delta} - t_{1\delta})\sup\ \{|\mathcal{S}(t,\mathcal{U}|C)|\,; t \in [t_{1\delta}, t'_{1\delta}]\} \end{array}$$

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

$$\begin{array}{l} \int_{t_1}^{t_2} E_C\left[\mathcal{S}(t,\mathcal{U}|C)\right]dt = \lim_{\delta\to 0}\int_{t'_{1\delta}}^{t'_{2\delta}} E_C\left[\mathcal{S}(t,\mathcal{U}|C)\right]dt \\ \geq \lim\inf_{\delta\to 0}\ \frac{1}{\delta}\left(\int_{t_{1\delta}}^{t'_{1\delta}} E_C\left[\mathcal{C}(t,\mathcal{U}|C)\right]\dot{\mathcal{U}}(t)dt + \int_{t'_{2\delta}}^{t_{2\delta}} E_C\left[\mathcal{C}(t,\mathcal{U}|C)\right]\dot{\mathcal{U}}(t)dt\right) \geq 0. \end{array}$$

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:

$$\mathcal{V}_\delta(t) = \begin{cases} \max\{(\beta-\delta)(t-\mathcal{U}(t_1)) + \mathcal{U}(t_1), \beta t + \alpha - \delta, (\beta+\delta)(t-\mathcal{U}(t_2)) + \mathcal{U}(t_2) & \text{if } t \in [t_1, t_2] \\ \mathcal{U}(t) & \text{if otherwise} \end{cases}$$

which is obviously feasible for δ sufficiently small. Define a_δ and b_δ the solutions of $(\beta-\delta)(t-\mathcal{U}(t_1)) + \mathcal{U}(t_1) = \beta t + \alpha - \delta$ and $\beta t + \alpha - \delta = (\beta+\delta)(t-\mathcal{U}(t_2)) + \mathcal{U}(t_2)$, respectively. It is easy to see that

$\lim_{\delta\to 0} t_{1\delta} = t_1$ and $\lim_{\delta\to 0} t_{2\delta} = t_2$. Therefore, since $h_\delta(t) = \delta$ for all $t \in [t_{1\delta}, t_{2\delta}]$,

$$\begin{array}{l} \frac{1}{\delta}\int_{t_1}^{t_{1\delta}} E_C\left[\mathcal{S}(t,\mathcal{U}|C)\right] h_\delta(t)dt - \int_{t_1}^{t_{1\delta}} E_C\left[\mathcal{C}(t,\mathcal{U}|C)\right] dt + \\ \frac{1}{\delta}\int_{t_{2\delta}}^{b} E_C\left[\mathcal{S}(t,\mathcal{U}|C)\right] h_\delta(t)dt + \int_{t_{2\delta}}^{t_2} E_C\left[\mathcal{C}(t,\mathcal{U}|C)\right] dt - \int_{t_{1\delta}}^{t_{2\delta}} E_C\left[\mathcal{S}(t,\mathcal{U}|C)\right] dt \geq 0. \end{array}$$

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 $\int_{t_1}^{t_2} E_C\left[\mathcal{S}(t,\mathcal{U}|C)\right] dt \leq 0$. Putting the two inequalities together we get our result.

Next, consider rotations and $\underline{\mathcal{E}}(C) \notin [t_1, t_2]$, for all C. Given $\delta > 0$ sufficiently small, define the following rent projection function:

$$\mathcal{V}_\delta(t) = \max\left\{\mathcal{U}(t),\ (\beta+\delta)(t-t_1) + \mathcal{U}(t_1)\right\},$$

which represents a small anti-clockwise rotation of the affine function $\mathcal{U}$ on $[t_1, t_2]$ at point $(t_1, \mathcal{U}(t_1))$ in the plane type versus informational rent. This perturbation is feasible. Applying Lemma 24, we obtain

$$\int_{t_1}^{t_{2\delta}} E_C\left[\mathcal{S}(t,\mathcal{U}|C)\right] h(t)dt \geq 0,$$

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

$$\int_a^b E_C\left[\mathcal{S}(t,\mathcal{U}|C)\right](t-t_1)dt \geq 0.$$

Analogously, we can make a small clockwise rotation of $\mathcal{U}$ on $[t_1, t_2]$ at point $(t_2, \mathcal{U}(t_2))$ and conclude that

$$\int_{t_1}^{t_2} E_C\left[\mathcal{S}(t,\mathcal{U}|C)\right](t-t_2)dt \leq 0.$$

If $\mathcal{U}$ has kink at t_1 (at t_2), then we can do also a small anti-clockwise (clockwise) rotation at t_2 (at t_1) and get the equality. If $\mathcal{U}$ has kink at both at t_1 and t_2, using that $\int_{t_1}^{t_2} E_C\left[\mathcal{S}(t,\mathcal{U}|C)\right] dt = 0$, we conclude the last equality for this case.

The case where $t_1 = \underline{t}$ and $t_2 \geq \underline{\mathcal{E}}$ is analogous. The only difference is that we have to consider the rotation at the point $(\underline{\mathcal{E}}, C)$ to eliminate the point effect from $\underline{\mathcal{E}}$ in the condition of Lemma 24.

Proof of Proposition 15

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

$$\mathcal{V}(t) = \begin{cases} \max\left\{\mathcal{U}(\underline{\mathcal{E}}) + \dot{\mathcal{U}}(\underline{\mathcal{E}})(t-\underline{\mathcal{E}}),\ 0\right\} & \text{if } t < \underline{\mathcal{E}} \\ \mathcal{U}(t) & \text{if } t \geq \underline{\mathcal{E}} \end{cases},$$

where $\mathcal{U}(\underline{\mathcal{E}}) = \underline{C}$.

Note that $\mathcal{U}(t) = \mathcal{V}(t)$ for all $t \geq \underline{\mathcal{E}}$. Since the rent projection function $\mathcal{V}$ is also feasible, Lemma 24 gives

$$E_C\left[\int_0^{\underline{\mathcal{E}}}\left[\frac{(\mathcal{E}(t,\Delta c)-t)\Delta x - C}{\dot{\mathcal{U}}(\mathcal{E}(t,C))} f(t,\mathcal{E}(t,C)|\Delta c) + F_0(t,\mathcal{E}(t,C)|C)\right](\mathcal{U}(t)-\mathcal{V}(t))dt\right] \leq 0. \tag{36}$$

By Lemma 25, $\frac{(\mathcal{E}(t,\Delta c)-t)\Delta x - C}{\dot{\mathcal{U}}(\mathcal{E}(t,C))} f(t,\mathcal{E}(t,C)|C) \geq 0$, 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 $t \in [0,\underline{\mathcal{E}}]$. Hence, the continuity of $\mathcal{U}$ and $\mathcal{V}$ and condition (36) yield that $\mathcal{U}(t) = \mathcal{V}(t)$, for all $t \in [0,\underline{\mathcal{E}}]$.

Recall that $\mathcal{U}(t) = 0$ for all $t \leq \underline{t}$. Therefore, the power of the contract for all types who get projected to a diagonal type $t < \underline{t}$ is $b(t,t,C) = \dot{\mathcal{U}}(t) = 0$, and, by (IR), they get $w = 0$. Types who get projected to a diagonal type $t \in (\underline{t},\underline{\mathcal{E}})$ get the constant power $b(\underline{\mathcal{E}},\underline{\mathcal{E}},C) = \dot{\mathcal{U}}(\underline{\mathcal{E}})$. From equation (32), we have $\mathcal{U}(\underline{\mathcal{E}}) = \underline{C}$. Moreover,

$$\mathcal{U}(\underline{\mathcal{E}}) = \int_{\underline{t}}^{\underline{\mathcal{E}}} \dot{\mathcal{U}}(\underline{\mathcal{E}})\, dt = (\underline{\mathcal{E}} - \underline{t})\dot{\mathcal{U}}(\underline{\mathcal{E}}).$$

Combining these two conditions yields

$$\dot{\mathcal{U}}(\underline{\mathcal{E}}) = \frac{\underline{C}}{\underline{\mathcal{E}} - \underline{t}} \leq \underline{C},$$

where the inequality uses the fact that $\underline{\mathcal{E}} - \underline{t} \leq 1$ (since $\underline{t}$ and $\underline{\mathcal{E}}$ 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 $t < \underline{t}$ would prefer to deviate to this contract).

Proof of Lemma 26

Let $\overline{t} = \inf\left\{t : \mathcal{E}\left(t,\overline{C}\right) \geq 1\right\}$, $\overline{\mathcal{E}} = \mathcal{E}(0,\overline{C})$ and $\hat{t} = \max\left\{\overline{t},\overline{\mathcal{E}}\right\}$. Since $\boldsymbol{p}$ and C are independent, note that for $t \geq \hat{t}$,

$$\frac{\mathcal{S}(t,\mathcal{U}|C)}{f(\mathcal{E}^{-1},t)} = \frac{(t-\mathcal{E}^{-1})\Delta x - C}{\dot{\mathcal{U}}(\mathcal{E}^{-1})} - H(\mathcal{E}^{-1},t).$$

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

$$\frac{d}{dt}\left(\frac{\mathcal{S}(t,\mathcal{U}|C)}{f(\mathcal{E}^{-1},t)}\right) = \frac{d}{dt}\left(\frac{(t-\mathcal{E}^{-1})\Delta x - C}{\dot{\mathcal{U}}(\mathcal{E}^{-1})} - H(\mathcal{E}^{-1},t)\right)$$

$$= -\frac{\left(\dot{\mathcal{E}^{-1}} - 1\right)\Delta x}{\dot{\mathcal{U}}(\mathcal{E}^{-1})} - \left[\frac{\left(t-\mathcal{E}^{-1}\right)\Delta x - C}{\dot{\mathcal{U}}(\mathcal{E}^{-1})}\right]\frac{\ddot{\mathcal{U}}(\mathcal{E}^{-1})}{\dot{\mathcal{U}}(\mathcal{E}^{-1})}\dot{\mathcal{E}^{-1}} - H_1(\mathcal{E}^{-1},t)\dot{\mathcal{E}^{-1}} - H_2(\mathcal{E}^{-1},t) < 0,$$

for almost all $t \geq \hat{t}$ (where $H_1(t,s) \equiv \frac{\partial H}{\partial t}(t,s)$ and $H_2(t,s) \equiv \frac{\partial H}{\partial s}(t,s)$). Therefore, $\frac{\mathcal{S}(t,\mathcal{U}|C)}{f(\mathcal{E}^{-1},t)}$ is a strictly increasing function of t.

Since $\frac{\mathcal{S}(t,\mathcal{U}|C)}{f(\mathcal{E}^{-1},t)}$ is strictly decreasing, there are three possible cases:

(i) $E_C[\mathcal{S}(t,\mathcal{U}|C)] < 0$ for all $t \in [\hat{t},1]$.

Consider the following convex and piecewise linear function:

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

where $\dot{\mathcal{U}}_{-}(\hat{t})=\lim_{t\uparrow\hat{t}}\dot{\mathcal{U}}(t)$, which is feasible. Notice that $\mathcal{U}(t)=\mathcal{V}(t)$ for $t\leq\hat{t}$. Since $\mathcal{U}$ is optimal, by Lemma 24,

$$\int_{\hat{t}}^{1}\left[\mathcal{U}(t)-\mathcal{V}(t)\right]E_C\left[\mathcal{S}(t,\mathcal{U}|C)\right]dt\geq 0.$$

Because $E_C\left[\mathcal{S}(.,\mathcal{U}|C)\right]$, $\mathcal{U}$, and $\mathcal{V}$ are continuous functions and $\mathcal{U}(t)\geq\mathcal{V}(t)$ for all $t\in[\hat{t},1]$, we must have that $\mathcal{U}(t)=\mathcal{V}(t)$, for all $t\in[\hat{t},1]$.

(ii) $E_C\left[\mathcal{S}(t,\mathcal{U}|C)\right]>0$ for all $t\in[\hat{t},1]$.

Consider the following convex and piecewise linear function:

$$\mathcal{V}(t)=\begin{cases}\mathcal{U}(t), & \text{if } t\leq\hat{t}\\ \mathcal{U}(1)+\frac{\mathcal{U}(1)-\mathcal{U}(\hat{t})}{1-\hat{t}}(t-1) & \text{if } t>\hat{t}\end{cases},$$

which is feasible. As in case (i), $\mathcal{V}$ coincides with $\mathcal{U}$ for $t\leq\hat{t}$. Using Lemma 24, we obtain

$$\int_{\hat{t}}^{1}\left[\mathcal{U}(t)-\mathcal{V}(t)\right]E_C\left[\mathcal{S}(t,\mathcal{U}|C)\right]dt\geq 0.$$

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

(iii) There exists $\tilde{t}\in[\hat{t},1]$ such that $E_C\left[\mathcal{S}(t,\mathcal{U}|C)\right]\lesseqgtr 0$ if and only if $t\gtreqless\tilde{t}$.

Consider the following convex and piecewise linear function:

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

which is feasible. Since $\mathcal{U}(t)=\mathcal{V}(t)$ on $t\leq\hat{t}$, Lemma 24 implies

$$\int_{\hat{t}}^{1}\left[\mathcal{U}(t)-\mathcal{V}(t)\right]E_C\left[\mathcal{S}(t,\mathcal{U}|C)\right]dt\geq 0.$$

Because $\mathcal{U}(t)\leq\mathcal{V}(t)$ on $[\hat{t},\tilde{t}]$ and $\mathcal{U}(t)\geq\mathcal{V}(t)$ on $[\hat{t},1]$, and $E_C\left[\mathcal{S}(t,\mathcal{U}|C)\right]$, $\mathcal{U}$ and $\mathcal{V}$ are continuous functions, it follows that $\mathcal{U}(t)=\mathcal{V}(t)$, for all $t\in[\hat{t},1]$. We conclude that $\mathcal{U}$ must have at most two pieces on the interval $[\hat{t},1]$.

Proof of Proposition 16

Let $\overline{t}=\inf\left\{t:\mathcal{E}\left(t,\overline{C}\right)\geq 1\right\}$ and $\overline{\mathcal{E}}=\mathcal{E}(0,\overline{C})$. We claim that $\frac{\Delta x}{\overline{C}}\leq 2$ implies that $\overline{\mathcal{E}}\geq\overline{t}$. Because $\mathcal{U}$ is increasing, it is enough to show that $\mathcal{U}(\overline{\mathcal{E}})\geq\mathcal{U}(\overline{t})$. By condition (32), $\mathcal{U}(\overline{\mathcal{E}})=\overline{C}$ and $\mathcal{U}(\overline{t})=\mathcal{U}(1)-\overline{C}$,

so that

$$\mathcal{U}(\overline{\mathcal{E}}) \geq \mathcal{U}(\overline{t}) \iff \mathcal{U}(1) \leq 2\overline{C}.$$

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

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

where the last inequality follows from the assumption. From Proposition 15 we have that $\mathcal{U}$ has two pieces on the interval $[0, \underline{\mathcal{E}}]$ and, from Lemma 26, $\mathcal{U}$ has at most two pieces on the interval $[\hat{t}, 1]$, where $\hat{t} = \max\{\overline{t}, \overline{\mathcal{E}}\} = \overline{\mathcal{E}}$. The result then follows.

Proof of Proposition 17

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., $E_C[S_0(t,\mathcal{U}|C)] \leq 0$. Let us investigate the effect on the high effort region. For any $t > \overline{\mathcal{E}} = \mathcal{E}(0,\overline{C})$, we have

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

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

$$\frac{\Delta x}{C}t - 1 \leq \left(1 + \frac{\Delta x}{C}\right)\mathcal{E}^{-1}.$$

This condition is implied by the following inequality

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

which is equivalent to the condition in the statement of the proposition. Therefore, given the optimal rent projection function $\mathcal{U}$, let $\mathcal{V}(t) = \min\{\mathcal{U}(t), \dot{\mathcal{U}}(\overline{\mathcal{E}})(t-\overline{\mathcal{E}}) + \overline{C}\}$, where $\mathcal{U}(\overline{\mathcal{E}}) = \overline{C}$. By Lemma 24, we must have that

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

Since $\mathcal{V}(t) \leq \mathcal{U}(t)$, we must have that $\mathcal{U}(t) = \mathcal{V}(t)$, for all $t \in [0,1]$. From Proposition 15 we have that $\mathcal{U}$ has two pieces on the interval $[0, \underline{\mathcal{E}}]$. Then, the result immediately follows.

Proof of Proposition 20

The proof of existence of optimal mechanisms is analogous to the proof of Proposition 14.

(1) The proof that there are two contracts at bottom is analogous to Proposition 15.

(2) The proof is analogous to the proof of Proposition 14.

(3) The proof is analogous to the proof of Lemma 25.

Online Appendix II: 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.[38]

Pure Moral Hazard

There is a continuum of agents in the population with different productivities: $\mathbf{p} \in \boldsymbol{P}$ is distributed according to the probability distribution function f with full support. Unlike the model from 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:[39]

$$\Delta x > u^{-1}(C) - u^{-1}(0). \tag{37}$$

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 $u^{-1}(0)$. 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 (37).

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 (37) 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.

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 output given effort. We,

[38] Our results contrast with the ones from Caillaud et al. (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.

[39] 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.

therefore, follow the model from the main text in assuming that the cost of effort is commonly known. 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 outputs, 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 the 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 outputs given efforts $\{\boldsymbol{p}(\cdot|e) : \mathbf{X} \to \mathbb{R} | e \in \mathbf{E}\}$. This formulation allows for infinite-dimensional types. However, when there are two outputs 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 outputs 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 $\mathbf{P}$.[40]

A direct mechanism $\{(w_{\mathbf{p}}(x), e(\mathbf{p})) : \mathbf{p} \in \mathbf{P}, x \in \mathbf{X}\}$ specifies a payment function $w_{\mathbf{p}}(\cdot) : \mathbf{X} \to \mathbb{R}$ 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:

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

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

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:

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

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

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

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

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

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

observable:

Proposition 21. *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 $\tilde{w}_{\boldsymbol{p}}(x) = c(e(\mathbf{p}))$ for all $\mathbf{p}$. This mechanism satisfies (IC AS) and satisfies (IR) with equality. Moreover, since the payments are constant in outputs, 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 37) 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 $\underline{t}$ obtain payoffs strictly above their reservation utilities (see Figure 3).

Online Appendix III: Numerical Method

For the numerical simulations, we work with a semi-discrete approach, in which the type space consists of n horizontal lines in $\boldsymbol{P}$. Formally, fix a finite set P_1 with n elements lying between 0 and 1. The type space is

$$\{(p_0, p_1) \in \mathbf{P} : \ p_1 \in P_1\}.$$

Because diagonal types are still present for all p_1, most results from the model with type space $\mathbf{P}$ can be easily adapted to this framework. For notational simplicity, let $x_L = 0$. The principal's problem is to find a rent projection $\mathcal{U}$ and an inverse effort frontier ξ to maximize:

$$\begin{aligned} W = &\sum_{s^i \in P_1} \int_0^{\xi(s^i)} [s^i \Delta x - \mathcal{U}(s^i)] f(t, s^i) dt \\ &+ \sum_{s^i \in P_1} \int_{\xi(s^i)}^{s^i} [t \Delta x - \mathcal{U}(t)] f(t, s^i) dt \end{aligned} \tag{P1}$$

subject to $\mathcal{U}$ non-negative, continuous, increasing, and convex, together with the effort condition

$$\xi(s) = \begin{cases} \mathcal{U}^{-1}(\mathcal{U}(s) - C) & \text{if } \mathcal{U}(s) > C \\ \min\left\{\max\left\{\mathcal{U}^{-1}(0)\right\}, \xi^{FI}(s)\right\} & \text{if } \mathcal{U}(s) = C \\ 0 & \text{otherwise,} \end{cases}$$

where ξ^{FI} is the first-best inverse effort recommendation.

It is straightforward to prove that, since there is only a finite number of constraints on $\mathcal{U}$ given an effort frontier, any feasible $\mathcal{U}$ is dominated by a piecewise linear function, which can be represented by a finite number of parameters.

Our numerical approach to solve this problem is as follows. For a given number j of contracts (the number of pieces in $\mathcal{U}$), we solve for its j breakpoints and slopes. Since internalizing the effort condition entails a discontinuity in the optimization problem, we solve for j breakpoints ($\tau^1 \leq \tau^2 \leq ... \leq \tau^j$), $j-1$ slope increments ($0 \leq z^i$, $i = 1, ..., j-1$), and p_1^* which is the smaller $p_1 \in P_1$ for which there is some $\mathbf{p} = (p_0, p_1)$ for which effort is recommended. Given n and p_1^*, we solve the sub-problem

$$\begin{aligned} \max_{\substack{0 \leq \tau^1 \leq p_1^* \\ \tau^1 \leq ... \leq \tau^j \\ 0 \leq z_i}} W = & \sum_{i=1}^{n} \int_0^{\xi(s^i)} [s^i \Delta x - \mathcal{U}\left(s^i\right)] f(t, s^i) dt \\ & + \sum_{i=1}^{n} \int_{\xi(s^i)}^{s^i} [t\Delta x - \mathcal{U}(t)] f(t, s^i) dt \end{aligned} \tag{P2}$$

subject to

$$\begin{aligned} \mathcal{U}(x) &= \textstyle\sum_{i=1}^{j} b^i (x - \tau^i)^+ \\ b^i &= \begin{cases} \frac{C}{p_1^* - \tau^1} & \text{if } i = 1 \\ b^{i-1} + z^i & \text{otherwise} \end{cases} \\ \xi(s) &= \begin{cases} \mathcal{U}(s) - C & \text{if } \mathcal{U}(s) > C \\ \min\left\{\tau^1, \xi^{FI}(p_1^*)\right\} & \text{if } \mathcal{U}(s) = C \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

This program is readily solved by standard numerical optimization packages (as KNITRO) when the problem dimensionality is low, as we have found in our examples. The strategy for solving the original problem is to start with $j = 1$, solving Program (P2), for all $p_1^* \in P_1$, and increasing j in case any improvement was found in relation to the previous best solution (in the case $j = 1$, the solution is trivial). Figure 9 depicts the optimal contracts when $\Delta x = 100$ and $C = 1$ for the uniform distribution. As we can see the optimal mechanism offers four contracts.

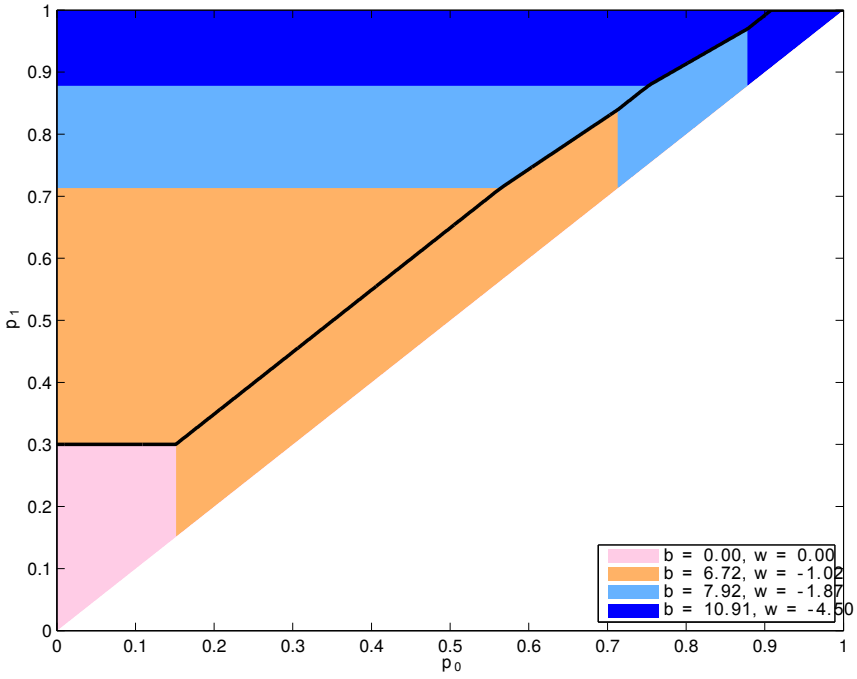

Figure 9: *BFD-optimal mechanism for uniform distribution and* $\Delta x = 100$.

Online Appendix IV: Full Insurance at the Bottom

We now show that, when the first-best effort region is empty, the firm offers a single contract with full insurance to an interval containing the riskiest types. Because in insurance the participation constraint binds at the top rather than at the bottom, we cannot apply the argument from Proposition 4.

Starting from a feasible rent projection $\mathcal{U}$, suppose the insurance firm decides to fully insure all types in an initial interval. That is, suppose the firm replaces $\mathcal{U}$ by $\max\{\mathcal{U}(t), \mathcal{U}(\alpha)\}$ for some $\alpha > 0$. There are three effects: (1) a lower power reduces the region of effort; (2) it increases the informational rents of all types in this interval; and (3) the lower power allows the firm to charge a higher risk premium since consumers are risk averse. When the first-best effort region is empty, the first effect is positive. Moreover, for small α, the first effect has a higher order of magnitude than the other two. Thus, in the optimal mechanism, there is an initial interval of types that get the same full insurance contract.

Proposition 22 (**Full Insurance at the Bottom**). *Let $u^{-1}(C) - u^{-1}(0) \geq L$ (i.e., the first-best high-effort region is empty) and let $\mathcal{U}$ be an optimal rent projection. Then, $\dot{\mathcal{U}}(t) = 0$ for all $t \in [0, \underline{t}]$ for some $\underline{t} > 0$.*

Proof. The result is trivially true if $\mathcal{U}(t) = 0$ for all t. Suppose $\mathcal{U}(t) \neq 0$ for some t and, for each $\alpha > 0$, let

$$\mathcal{V}_\alpha(t) := \max\{\mathcal{U}(\alpha),\ \mathcal{U}(t)\}.$$

Note that $\mathcal{V}_\alpha$ is a feasible rent projection since it is obtained by perturbing $\mathcal{U}$ in a way that preserves convexity and does not violate the participation constraint.

Apply Theorem 12 for $\mathcal{V}_\alpha$ to obtain

$$a(\alpha) := \int_0^\alpha [\mathcal{U}(t) - \mathcal{V}_\alpha(t)]\, \mathcal{S}(t,\mathcal{U})\, dt - \int_0^\alpha \dot{\mathcal{U}}(t)\mathcal{C}(t,\mathcal{U})dt \geq 0.$$

The function $a(\cdot)$ is differentiable at almost all α. Its derivative, where it exists, equals

$$a'(\alpha) = \dot{\mathcal{U}}(\alpha)\left(-\int_0^\alpha S(t,\mathcal{U})dt + C(\alpha,\mathcal{U})\right).$$

At almost all t, the derivative of $\mathcal{C}(t,\mathcal{U})$ with respect to t equals

$$\frac{d}{dt}\mathcal{C}(t,\mathcal{U}) = \frac{d}{dt}\left(\frac{\partial G}{\partial \dot{\mathcal{U}}}\right)\frac{F_0(t,\mathcal{E})}{f(t,\mathcal{E})} + \frac{\partial G}{\partial \dot{\mathcal{U}}}\frac{d}{dt}\left(\frac{F_0(t,\mathcal{E})}{f(t,\mathcal{E})}\right).$$

Since $\lim_{t\downarrow 0} \frac{\partial G}{\partial \dot{\mathcal{U}}} = \lim_{t\downarrow 0} \frac{F_0(t,\mathcal{E})}{f(t,\mathcal{E})} = 0$, it follows that $\lim_{t\to 0} \frac{d}{dt}\mathcal{C}(t,\mathcal{U}) = 0$.

Divide the term

$$-\int_0^\alpha S(t,\mathcal{U})dt + C(\alpha,\mathcal{U})$$

by $\alpha > 0$ and consider its limit as $\alpha \to 0$. Since $\lim_{\alpha\to 0} \frac{C(\alpha,\mathcal{U})}{\alpha} = 0$, this limit is $-\mathcal{S}(0,\mathcal{U})$. Note also that

$$-\mathcal{S}(0,\mathcal{U}) = \frac{\underline{\mathcal{E}}L - (G(\underline{\mathcal{E}}) - G(0))}{\dot{\mathcal{U}}(\underline{\mathcal{E}})} < 0$$

because

$$G(\underline{\mathcal{E}}) - G(0) \geq u^{-1}(C) - u^{-1}(0) > \underline{\mathcal{E}}L$$

by the convexity of u^{-1} and the assumption that $u^{-1}(C) - u^{-1}(0) \geq L$. Thus, if $\dot{\mathcal{U}}$ is strictly positive in an interval around $t = 0$, $a'(\cdot) < 0$ a.e. in this interval, contradicting $a(0) = 0$ and $a(\alpha) \geq 0$ for all α. □

Online Appendix V: Omitted Proofs

Before presenting the formal proof, we discuss the intuition behind Lemma 1. Suppose a feasible mechanism recommends that type $\mathbf{p} = (p_0, p_1)$ exerts high effort, and consider a type $\hat{\mathbf{p}} = (p_0, \hat{p}_1)$ with $\hat{p}_1 > p_1$. Type $\hat{\mathbf{p}}$ has the same distribution of outputs conditional on low effort as $\mathbf{p}$, but has a higher probability of high output conditional on high effort. Therefore, $\hat{\mathbf{p}}$ has an even higher incentive to exert high effort. Similarly, suppose that the mechanism recommends that type $\mathbf{p} = (p_0, p_1)$ exerts low effort, and consider some type $\hat{\mathbf{p}} = (\hat{p}_0, p_1)$ for some $\hat{p}_0 > p_0$. Incentive compatibility implies that $\hat{\mathbf{p}}$ will have a higher incentive to exert low effort than type $\mathbf{p}$ has.

The continuity of $\mathcal{E}$ 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 10 illustrates the argument. The arrows indicate the direction of growth of the informational rent function U. Since U is continuous, if the distances between points $\boldsymbol{a}$ 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 $\mathcal{E}$, we must have $U(\mathbf{c}) > U(\mathbf{a})$, and because the informational rent is constant in p_1 in the region below $\mathcal{E}$, 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.

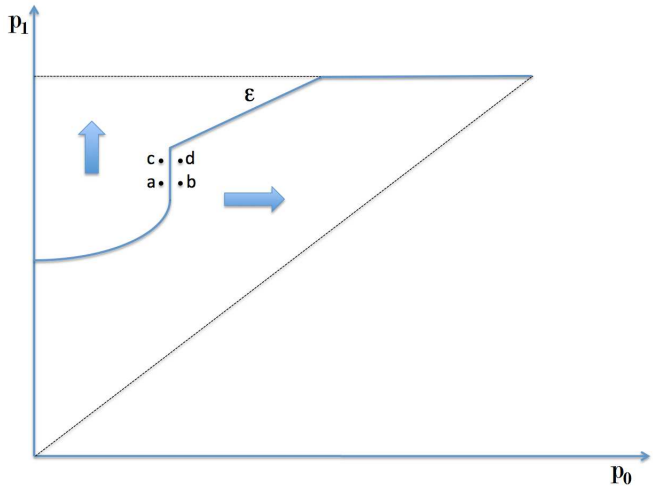

Figure 10: *Intuition behind Lemma 1 (continuity of $\mathcal{E}$).*

For simplicity, we will use the following notation throughout the proofs below. Given a mechanism, let $\boldsymbol{P}_0$ and $\boldsymbol{P}_1$ denote the set of types for which the low and high efforts are recommended.

Proof of Lemma 1

The proof proceeds by a series of claims. The two first claims establish that the vertical and horizontal sections of the sets $\boldsymbol{P}_0$ and $\boldsymbol{P}_1$ are intervals.

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

Proof. Let $\mathbf{p} = (p_0, p_1) \in \boldsymbol{P}_1$ and suppose that $\hat{\mathbf{p}} = (p_0, \hat{p}_1) \in \boldsymbol{P}_0$. Incentive compatibility implies that

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

Combining these inequalities, we obtain $(p_1 - \hat{p}_1)\, b(\mathbf{p}) \geq 0$. Since $\mathbf{p} \in \boldsymbol{P}_1$, 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 $(p_0, p_1) \in \boldsymbol{P}_0$ and $\hat{p}_0 > p_0$, then $(\hat{p}_0, p_1) \in \boldsymbol{P}_0$.

Proof. Let $\mathbf{p} = (p_0, p_1) \in \boldsymbol{P}_0$ and suppose that $\hat{\mathbf{p}} = (\hat{p}_0, p_1) \in \boldsymbol{P}_1$. Incentive compatibility implies that

$$\begin{aligned} w(\mathbf{p}) + p_0 b(\mathbf{p}) &\geq w(\hat{\mathbf{p}}) + p_1 b(\hat{\mathbf{p}}) - C, \text{ and} \\ w(\hat{\mathbf{p}}) + p_1 b(\hat{\mathbf{p}}) - C &\geq w(\mathbf{p}) + \hat{p}_0 b(\mathbf{p}). \end{aligned} \tag{38}$$

Combining these inequalities, we obtain $(\hat{p}_0 - p_0)\, b(\mathbf{p}) = 0$, which, because $\hat{p}_0 > p_0$, implies that $b(\mathbf{p}) = 0$. Substituting back, yields $w(\mathbf{p}) = w(\hat{\mathbf{p}}) + p_1 b(\hat{\mathbf{p}}) - C$. 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 type $\mathbf{p}$'s contract to type $\hat{\mathbf{p}}$ as well. First, we verify that new mechanism is also feasible. 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 $\mathbf{p} = (p_0, p_1) \in \boldsymbol{P}_0$ and $\hat{\mathbf{p}} = (\hat{p}_0, p_1) \in \boldsymbol{P}_1$ 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

$$w(p_0, p_1) = w(\hat{p}_0, p_1) + p_1 b(\hat{p}_0, p_1) - C$$

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

It follows directly from Claims 3 and 4 that there exists a non-decreasing function $\mathcal{E}:[0,1]\to\mathbb{R}_+$ such that $(p_0,p_1)\in\boldsymbol{P}_1$ if and only if $p_1\geq\mathcal{E}(p_0)$. The next claim establishes that this function is continuous.

Claim 5. For every feasible mechanism, there exists an equivalent mechanism with the following property: $(p_0,p_1)\in\boldsymbol{P}_1$ if and only if $p_1\geq\mathcal{E}(p_0)$ for a non-decreasing and continuous function $\mathcal{E}:[0,1]\to[0,1]$.

Proof. The existence of such a non-decreasing function $\mathcal{E}$ follows straight from Claims 3 and 4. It remains to be shown that $\mathcal{E}$ is continuous. Suppose, in order to obtain a contradiction, that $\mathcal{E}$ is discontinuous at a point p_0. Since $\mathcal{E}$ is bounded and non-decreasing, there exist $\mathcal{E}_+>\mathcal{E}_-$ such that

$$\mathcal{E}_+=\lim_{p\downarrow p_0}\mathcal{E}(p)\ \text{ and }\ \mathcal{E}_-=\lim_{p\uparrow p_0}\mathcal{E}(p).$$

From the definition of $\mathcal{E}$, $(p_0,p_1)\in\boldsymbol{P}_1$ for all $p_1\in[\mathcal{E}_-,\mathcal{E}_+]$. Moreover, for any $\delta>0$ and $p_1\in[\mathcal{E}_-,\mathcal{E}_+]$, it follows that $(p_0+\delta,p_1)\in\boldsymbol{P}_0$.

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

$$U(p_0,p_1)=\lim_{\delta\to0}U(p_0,p_1+\delta),\quad\forall p_1\in[\mathcal{E}_-,\mathcal{E}_+].$$

Let $\bar{\mathcal{E}}=\frac{\mathcal{E}_++\mathcal{E}_-}{2}$. Note that types $\left(p_0+\delta,\bar{\mathcal{E}}\right)$ and $\left(p_0+\delta,\bar{\mathcal{E}}+\alpha\right)$ both belong to $\boldsymbol{P}_0$ for any $\delta>0$ and $\alpha\in\left[0,\frac{\mathcal{E}_++\mathcal{E}_-}{2}\right]$. Then, using the incentive-compatibility constraint of type $\left(p_0+\delta,\bar{\mathcal{E}}\right)$, we obtain

$$\begin{aligned}U\left(p_0+\delta,\bar{\mathcal{E}}\right)&\geq w\left(p_0+\delta,\bar{\mathcal{E}}+\alpha\right)+(p_0+\delta)\,b\left(p_0+\delta,\bar{\mathcal{E}}+\alpha\right)\\&=U\left(p_0+\delta,\bar{\mathcal{E}}+\alpha\right).\end{aligned}$$

Similarly, the incentive-compatibility constraint of type $\left(p_0+\delta,\bar{\mathcal{E}}+\alpha\right)$ yields

$$\begin{aligned}U\left(p_0+\delta,\bar{\mathcal{E}}+\alpha\right)&\geq w\left(p_0+\delta,\bar{\mathcal{E}}\right)+(p_0+\delta)\,b\left(p_0+\delta,\bar{\mathcal{E}}\right)\\&=U\left(p_0+\delta,\bar{\mathcal{E}}\right).\end{aligned}$$

Combining both inequalities, we obtain

$$U\left(p_0+\delta,\bar{\mathcal{E}}\right)=U\left(p_0+\delta,\bar{\mathcal{E}}+\alpha\right),\tag{39}$$

for any $\delta>0$ and $\alpha\in\left[0,\bar{\mathcal{E}}\right]$.

Moreover, from the incentive-compatibility constraint of type $\left(p_0,\bar{\mathcal{E}}+\alpha\right)\in\boldsymbol{P}_1$, we have

$$\begin{aligned}U\left(p_0,\bar{\mathcal{E}}+\alpha\right)&\geq w\left(p_0,\bar{\mathcal{E}}\right)+\left(\bar{\mathcal{E}}+\alpha\right)b\left(p_0,\bar{\mathcal{E}}\right)-C\\&=U\left(p_0,\bar{\mathcal{E}}\right)+\alpha b\left(p_0,\bar{\mathcal{E}}\right),\end{aligned}$$

and because $b(\mathbf{p})>0$ for any $\boldsymbol{p}\in\boldsymbol{P}_1$

$$U\left(p_0,\bar{\mathcal{E}}+\alpha\right)>U\left(p_0,\bar{\mathcal{E}}\right).\tag{40}$$

Equation (39) implies that

$$\lim_{\delta \downarrow 0} U\left(p_0+\delta, \bar{\mathcal{E}}\right)=\lim_{\delta \downarrow 0} U\left(p_0+\delta, \bar{\mathcal{E}}+\alpha\right), \tag{41}$$

and, by the continuity of U,

$$\lim_{\delta \downarrow 0} U\left(p_0+\delta, \bar{\mathcal{E}}\right)=U\left(p_0, \bar{\mathcal{E}}\right), \text{ and} \tag{42}$$

$$\lim_{\delta \downarrow 0} U\left(p_0+\delta, \bar{\mathcal{E}}+\alpha\right)=U\left(p_0, \bar{\mathcal{E}}+\alpha\right). \tag{43}$$

Combining equations (41)-(43), we obtain $U\left(p_0, \bar{\mathcal{E}}+\alpha\right)=U\left(p_0, \bar{\mathcal{E}}\right)$, which contradicts inequality (40). □

Proof of Lemma 3

Let (w, b, e) be a mechanism for which there exists a continuous and non-decreasing function $\mathcal{E}$ satisfying condition (3). For such a mechanism, let $U: \boldsymbol{P} \rightarrow \mathbb{R}_{+}$ denote the informational rent function as defined in equation (2). Lemma 3 is a direct consequence of the following result, which establishes that conditions (a)-(d) from Lemma 2 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 $\mathcal{E}:[0,1] \rightarrow[0,1]$. Let U be as defined in equation (1). Suppose that conditions (a)-(d) are satisfied. Then, (w, b, e) is a feasible mechanism.

Proof of the Claim. We need to establish that a mechanism satisfying conditions (a)-(d) for a continuous and nondecreasing $\mathcal{E}$ satisfies incentive-compatibility (IC), individual-rationality (IR), and free disposal (FD). Condition (b) implies that $b(\mathbf{p}) \geq b(0,0)$ for all $\mathbf{p}$. Then, by condition (c), (FD) holds. Moreover, conditions (a) and (c) imply that $U(\mathbf{p}) \geq 0$ 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 $\boldsymbol{P}_0$ and $\boldsymbol{P}_1$ separately. There are 4 possible deviations in each region: taking a contract designed to types in regions $\boldsymbol{P}_0$ or $\boldsymbol{P}_1$ and exerting efforts 0 or 1. First, let $\boldsymbol{p}=(p_0, p_1) \in \boldsymbol{P}_0$ (i.e. $p_1 \leq \mathcal{E}(p_0)$).

Case 1: Reporting type $\mathbf{q} \in \boldsymbol{P}_0$ 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 $\mathbf{q} \in \boldsymbol{P}_0$ and choosing $e=1$.

We have to verify that the following inequality is satisfied:

$$U(\boldsymbol{p})=w(\mathbf{p})+p_0 b(\mathbf{p}) \geq w(\mathbf{q})+p_1 b(\mathbf{q})-C.$$

Since type $(0, \mathcal{E}(p_0)) \in \boldsymbol{P}_1$ and, from condition (a), $U(\boldsymbol{p})=U(0, \mathcal{E}(p_0))$, the previous inequality is equivalent to

$$U(0, \mathcal{E}(p_0))=w(0, \mathcal{E}(p_0))+\mathcal{E}(p_0) b(0, \mathcal{E}(p_0))-C \geq w(\mathbf{q})+p_1 b(\mathbf{q})-C \tag{44}$$

for all $\boldsymbol{q} \in \boldsymbol{P}_0$. Note that this is the incentive-compatibility constraint preventing type $(0, \mathcal{E}(p_0)) \in \boldsymbol{P}_1$ from getting the contract designed for $\mathbf{q} \in \boldsymbol{P}_0$ 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 $\mathbf{q} \in \boldsymbol{P}_1$ and choosing $e = 0$.

We have to show that

$$w(\mathbf{p}) + p_0 b(\mathbf{p}) \geq w(\mathbf{q}) + p_0 b(\mathbf{q}). \tag{45}$$

Conditions (a) and (d) imply that, for almost all $\mathbf{q} \in \boldsymbol{P}_1$, $b(\mathbf{q}) = b(q_1, q_1)$ and $w(\mathbf{q}) = w(q_1, q_1)$. Then, for all such $\mathbf{q}$, we have

$$w(\boldsymbol{q}) + p_0 b(\boldsymbol{q}) = w(q_1, q_1) + p_0 b(q_1, q_1).$$

Because $(q_1, q_1) \in \boldsymbol{P}_0$, the result from Case 1 implies that inequality (45) holds for all such $\mathbf{q}$ (which holds a.e.).

It remains to be shown that (45) 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 $\mathbf{q} \in \boldsymbol{P}_1$, 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 (45) implies that

$$\begin{aligned} w(\mathbf{p}) + p_0 b(\mathbf{p}) &\geq w(q_0, q_1^n) + p_0 b(q_0, q_1^n) \\ &= U(q_0, q_1^n) + (p_0 - q_1^n) b(q_0, q_1^n). \end{aligned}$$

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,

$$w(\mathbf{p}) + p_0 b(\mathbf{p}) \geq U(q_0, q_1^n) + (p_0 - q_1^n) b(q_0, \hat{q}_1).$$

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

$$\begin{aligned} w(\mathbf{p}) + p_0 b(\mathbf{p}) &\geq w(q_0, \hat{q}_1) + \hat{q}_1 b(q_0, \hat{q}_1) + (p_0 - \hat{q}_1) b(q_0, \hat{q}_1) \\ &= w(q_0, \hat{q}_1) + p_0 b(q_0, \hat{q}_1), \end{aligned}$$

which concludes the proof.

Case 4: Reporting type $\mathbf{q} \in \boldsymbol{P}_1$ and choosing $e = 1$.

From standard single-crossing arguments, we have:

$$w(0, \mathcal{E}(p_0)) + \mathcal{E}(p_0) b(0, \mathcal{E}(p_0)) - C \geq w(\mathbf{q}) + \mathcal{E}(p_0) b(\mathbf{q}) - C. \tag{46}$$

From condition (a), it follows that

$$w(0, \mathcal{E}(p_0)) + p_0 b(0, \mathcal{E}(p_0)) = w(p_0, p_1) + p_0 b(p_0, p_1)$$

for all $(p_0, p_1) \in \boldsymbol{P}_0$. Moreover, since U is continuous, we have

$$\begin{aligned} w\left(0, \mathcal{E}\left(p_0\right)\right) + \mathcal{E}\left(p_0\right) b\left(0, \mathcal{E}\left(p_0\right)\right) - C &= w\left(0, \mathcal{E}\left(p_0\right)\right) + p_0 b\left(0, \mathcal{E}\left(p_0\right)\right) \\ &= w\left(p_0, p_1\right) + p_0 b\left(p_0, p_1\right). \end{aligned}$$

Substituting in (46), we obtain

$$\begin{aligned} w\left(p_0, p_1\right) + p_0 b\left(p_0, p_1\right) &\geq w\left(\mathbf{q}\right) + \mathcal{E}\left(p_0\right) b\left(\mathbf{q}\right) - C \\ &\geq w\left(\mathbf{q}\right) + p_1 b\left(\mathbf{q}\right) - C, \end{aligned}$$

where the last inequality uses the fact that $p_1 \leq \mathcal{E}\left(p_0\right)$ (since $(p_0, p_1) \in \boldsymbol{P}_0$).

This concludes the possible deviations for types in $\boldsymbol{P}_0$. Now, let $\boldsymbol{p} = (p_0, p_1) \in \boldsymbol{P}_1$ (i.e., $p_1 > \mathcal{E}\left(p_0\right)$). Again, the possible deviations can be grouped into 4 possible cases.

Case 5: Reporting type $\mathbf{q} \in \boldsymbol{P}_1$ and choosing $e = 1$.

This result follows from standard single-crossing arguments taking effort as fixed.

Case 6: Reporting type $\mathbf{q} \in \boldsymbol{P}_1$ and choosing $e = 0$.

From Case 3, the following condition holds:

$$w\left(p_0, \mathcal{E}\left(p_0\right)\right) + p_0 b\left(p_0, \mathcal{E}\left(p_0\right)\right) \geq w\left(\mathbf{q}\right) + p_0 b\left(\mathbf{q}\right). \tag{47}$$

Case 5 and condition (a) implies that

$$\begin{aligned} w\left(\boldsymbol{p}\right) + p_1 b\left(\boldsymbol{p}\right) - C &\geq w\left(0, \mathcal{E}\left(p_0\right)\right) + \mathcal{E}\left(p_0\right) b\left(0, \mathcal{E}\left(p_0\right)\right) - C \\ &= w\left(p_0, \mathcal{E}\left(p_0\right)\right) + p_0 b\left(p_0, \mathcal{E}\left(p_0\right)\right). \end{aligned}$$

Then, inequality (47) yields

$$w\left(\boldsymbol{p}\right) + p_1 b\left(\boldsymbol{p}\right) - C \geq w\left(\mathbf{q}\right) + p_0 b\left(\mathbf{q}\right),$$

which concludes the proof of this case.

Case 7: Reporting type $\mathbf{q} \in \boldsymbol{P}_0$ and choosing $e = 0$.

Let $\mathcal{E}^{-1}\left(p_1\right) = \sup\left\{p_0 : \mathcal{E}\left(p_0\right) \leq p_1\right\}$. From Case 1, we have

$$w\left(\mathcal{E}^{-1}\left(p_1\right), p_1\right) + \mathcal{E}^{-1}\left(p_1\right) b\left(\mathcal{E}^{-1}\left(p_1\right), p_1\right) \geq w\left(\mathbf{q}\right) + \mathcal{E}^{-1}\left(p_1\right) b\left(\mathbf{q}\right). \tag{48}$$

From the continuity of U, we have

$$w\left(\mathcal{E}^{-1}\left(p_1\right), p_1\right) + \mathcal{E}^{-1}\left(p_1\right) b\left(\mathcal{E}^{-1}\left(p_1\right), p_1\right) = w\left(\mathcal{E}^{-1}\left(p_1\right), p_1\right) + p_1 b\left(\mathcal{E}^{-1}\left(p_1\right), p_1\right) - C.$$

Substituting in inequality (48), yields

$$w\left(\mathcal{E}^{-1}\left(p_1\right), p_1\right) + p_1 b\left(\mathcal{E}^{-1}\left(p_1\right), p_1\right) - C \geq w\left(\mathbf{q}\right) + \mathcal{E}^{-1}\left(p_1\right) b\left(\mathbf{q}\right). \tag{49}$$

However, condition (a) implies that, for all $p_0 < \mathcal{E}^{-1}(p_1)$,

$$w\left(\mathcal{E}^{-1}(p_1), p_1\right) + p_1 b\left(\mathcal{E}^{-1}(p_1), p_1\right) - C = w(\boldsymbol{p}) + p_1 b(\boldsymbol{p}) - C, \text{ and}$$

$$w(\mathbf{q}) + \mathcal{E}^{-1}(p_1)\, b(\mathbf{q}) \geq w(\mathbf{q}) + p_0 b(\mathbf{q}).$$

Substituting in (49), we obtain:

$$w(\boldsymbol{p}) + p_1 b(\boldsymbol{p}) - C \geq w(\mathbf{q}) + p_0 b(\mathbf{q}),$$

which concludes the proof of this case.

Case 8: Reporting type $\mathbf{q} = (q_0, q_1) \in \boldsymbol{P}_0$ and choosing $e = 1$.

Since $(p_1, p_1) \in \boldsymbol{P}_0$, standard single-crossing arguments establish that

$$w(p_1, p_1) + p_1 b(p_1, p_1) \geq w(\mathbf{q}) + p_1 b(\mathbf{q}).$$

Conditions (a) and (d) yield:

$$w(0, p_1) + p_1 b(0, p_1) = w(p_1, p_1) + p_1 b(p_1, p_1).$$

Substituting in the previous inequality and subtracting C, we obtain:

$$w(0, p_1) + p_1 b(0, p_1) - C \geq w(\boldsymbol{q}) + p_1 b(\boldsymbol{q}) - C.$$

However, from condition (d), we have

$$w(\boldsymbol{p}) + p_1 b(\boldsymbol{p}) - C = w(0, p_1) + p_1 b(0, p_1) - C$$

for all $p_0 < \mathcal{E}^{-1}(p_1)$. Thus,

$$w(\boldsymbol{p}) + p_1 b(\boldsymbol{p}) - C \geq w(\boldsymbol{q}) + p_1 b(\boldsymbol{q}) - C,$$

which concludes the proof.

Proof of Lemma 6

We claim that $\mathcal{E}(t + \Delta t) - \mathcal{E}(t) \leq \Delta t$, for all $t, \Delta t \geq 0$ such that $\mathcal{E}(t + \Delta t) < 1$. Indeed,

$$\begin{aligned}
\mathcal{U}^{-1}(\mathcal{U}(t + \Delta t) + C) - \mathcal{U}^{-1}(\mathcal{U}(t) + C) &\leq \dot{\mathcal{U}}^{-1}(\mathcal{U}(t) + C)\left[\mathcal{U}(t + \Delta t) - \mathcal{U}(t)\right] \\
&\leq \dot{\mathcal{U}}^{-1}(\mathcal{U}(t) + C)\dot{\mathcal{U}}(t)\Delta t \\
&\leq \dot{\mathcal{U}}^{-1}(\mathcal{U}(t))\dot{\mathcal{U}}(t)\Delta t = \Delta t
\end{aligned}$$

where the first inequality is a consequence of the subgradient inequality of $\mathcal{U}^{-1}$ at $\mathcal{U}(t) + C$, the second is the supergradient inequality of $\mathcal{U}$ at t, and the third is a consequence of concavity of $\mathcal{U}^{-1}$. By the definition of $\mathcal{E}$, we get the result. It thus follows that $\mathcal{E}$ is Lipschitz and, in particular, differentiable

almost everywhere with $\dot{\mathcal{E}} \leq 1$ at all points of differentiability.

Proof of Corollary 1

By Proposition 4, $\mathcal{U}$ is piecewise linear in $[0, \underline{\mathcal{E}}]$. Since the uniform distribution satisfies increasing rents, it is also piecewise linear in $[\overline{t}, 1]$ (Lemma 9). It remains to be shown that $\mathcal{U}$ is piecewise linear on $(\underline{\mathcal{E}}, \overline{t})$.

We claim that $\mathcal{E}(\underline{\mathcal{E}}) \geq \overline{t}$. Because $\mathcal{U}$ is increasing, it suffices to show that $\mathcal{U}(\overline{t}) \leq \mathcal{U}(\mathcal{E}(\underline{\mathcal{E}}))$. By equation (6),

$$\mathcal{U}(\mathcal{E}(\underline{\mathcal{E}})) = \mathcal{U}(\underline{\mathcal{E}}) + \Delta c = 2C.$$

Since $\mathcal{U}(\overline{t}) = \mathcal{U}(1) - C$, we need to show that $\mathcal{U}(1) \leq 3C$. 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 $\Delta x \leq 3C$.

Since $\mathcal{U}$ is piecewise linear on $[0, \underline{\mathcal{E}}] \cup [\overline{t}, 1]$ and the image of $[\underline{\mathcal{E}}, \overline{t}]$ by $\mathcal{E}^{-1}$ and by $\mathcal{E}$ are contained in $[0, \underline{\mathcal{E}}]$ and $[\overline{t}, 1]$, respectively, we can define a partition of the interval $[\underline{\mathcal{E}}, \overline{t}]$ such that the functions $\dot{\mathcal{U}}(\mathcal{E}^{-1})$ and $\dot{\mathcal{U}}(\mathcal{E})$ are constant in each interval of the partition. Let $[t_1, t_2] \subset [\underline{\mathcal{E}}, \overline{t}]$ be an element of the partition and let $\dot{\mathcal{U}}(\mathcal{E}^{-1}(t)) = \beta_0$ and $\dot{\mathcal{U}}(\mathcal{E}(t)) = \beta_1$ for all $t \in [t_1, t_2]$. Then,

$$\mathcal{S}(t,\mathcal{U}) = \frac{1}{2}\left[-\frac{(\mathcal{E}-t)\Delta x - C}{\beta_1} + \frac{(t-\mathcal{E}^{-1})\Delta x - C}{\beta_0} - \mathcal{E} + t - \mathcal{E}^{-1}\right],$$

where we have substituted the expressions for F_0 and F_1 under the uniform distribution. Differentiating with respect to t (and ignoring the $\frac{1}{2}$ term), yields:

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

Substituting $\dot{\mathcal{E}^{-1}}(t) = \frac{\dot{\mathcal{U}}(t)}{\beta_0}$ and $\dot{\mathcal{E}}(t) = \frac{\dot{\mathcal{U}}(t)}{\beta_1}$, 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 $[t_1, t_2]$. Thus, $\mathcal{S}(t,\mathcal{U})$ is an increasing function of t on $[t_1, t_2]$. Then, by the same procedure as in the proof of Lemma 9, it follows that $\mathcal{U}$ is piecewise linear on $[t_1, t_2]$. Since the partition is finite, we have that $\mathcal{U}$ is piecewise linear on $[0, 1]$.

Proof of Lemma 13

Proof. By Theorem 3.14 of (1986, pp. 69), we know that the space of real continuous functions $C([t_1, t_2])$ is dense in the space of integral functions $L_1([t_1, t_2])$ and, by the Stone-Weierstrass Theorem, every function in $C([t_1, t_2])$ is the uniform limit of a sequence of polynomial functions. Therefore, the hypothesis of the lemma implies that $\int_{t_1}^{t_2} f(t)g(t)dt = 0$, for all $g \in L_1[t_1, t_2]$ such that $\int_{t_1}^{t_2} g(t)dt = 0$.

Notice that $L_2[t_1, t_2] \subset L_1[t_1, t_2]$. Consider the closed subspace $H = \left\{g \in L_2[t_1, t_2]; \int_{t_1}^{t_2} g(t)dt = 0\right\}$ of

$L_2[t_1, t_2]$. Notice that the orthogonal subspace of H in $L_2[t_1, t_2]$, $H^\perp$, is the space of constant functions.[41] Indeed, the constant functions are obviously contained in $H^\perp$ and, for each $g \in L_2[t_1, t_2]$, we have that

$$g = \left(g - \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} g(t)dt \right) + \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} g(t)dt,$$

where $g - \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} g(t)dt \in H$, which implies that $H^\perp$ is generated by the constant functions. Therefore, $f \in H^\perp$. □

[41] As usual for $L_p[t_1, t_2]$ space, a function g is constant when $g(t) = k$, a.e., for some $k \in \mathbb{R}$.