Simple Contracts with Adverse Selection and Moral Hazard*

Daniel Gottlieb and Humberto Moreira†

May, 2017

Abstract

We study a principal-agent model with moral hazard and adverse selection. Risk-neutral agents with limited liability have arbitrary private information about the distribution of outputs and the cost of effort. We show that, under a multiplicative separability condition, the optimal mechanism offers a single contract. This condition is always satisfied when output is binary. If, in addition, the principal’s payoff must satisfy free disposal and the distribution of outputs has the monotone likelihood ratio property, the mechanism offers a single debt contract. Our results persist if the output distribution is “approximately multiplicatively separable” or if the agent is “approximately risk neutral.” Our model suggests that offering a single contract may be optimal in environments with adverse selection and moral hazard when agents have limited liability.

Keywords: principal-agent problem; contract theory; mechanism design.

*We thank Eduardo Azevedo, Dirk Bergemann, Vinicius Carrasco, Sylvain Chassang, Gonzalo Cisternas, Alex Edmans, Mehmet Ekmekci, Eduardo Fuingold, Leandro Gorno, Faruk Gul, Jason Hartline, Tibor Heumann, Bengt Holmström, Johannes Hörner, Ohad Kadan, Lucas Maestri, George Mailath, David Martimort, Stephen Morris, Roger Myerson, Larry Samuelson, Yuliy Sannikov, Jean Tirole, Rakesh Vohra, John Zhu, Alessandro Pavan (editor), and three anonymous referees for comments and suggestions. We also thank seminar audiences at Arizona State University, Boston College, FGV, HEC Montréal, Johns Hopkins University, Princeton University, PUC-Rio, UBC, Universidad de Chile, University of Pennsylvania, University of Pittsburgh/Carnegie Mellon University, the Wharton School, Yale University, and the BYU Computational Public Economics, 2013 LAMES, 2013 SBE, 2014 IWGTS, 2014 ESEM, 2015 AEA meetings, and 2015 ESWC. Rafael Mourão provided outstanding research assistance. Gottlieb gratefully acknowledges financial support from the Dorinda and Mark Winkelman Distinguished Scholar Award. Moreira acknowledges FAPERJ and CNPq for financial support.

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

1 Introduction 1

2 Two Outputs 5
   2.1 Statement of the Problem 5
   2.2 Benchmarks 6
   2.3 Contract Simplicity 8

3 Multiple Outputs 10

4 Free Disposal and Debt Contracts 15

5 Procurement and Regulation 16

6 Robustness 19

7 Conclusion 21

Appendix 23
   Proofs 23
      Theorem 1 23
      Theorem 2 27
      Proposition 1 36
      Proposition 2 39
   Examples 54
   Multiplicative Separability 55

References 64

Online Appendix 67
1 Introduction

Most real-world contracts are much simpler than theory predicts. Differently from standard adverse selection models, contracting parties offer a limited number of contracts, often a single one. Unlike in standard moral hazard models, similar contracts are offered in fundamentally different environments. As Hart and Holmstrom (1987) and Chiappori and Salanie (2003) 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)

In this paper, we propose an answer to this puzzle based on the interaction between adverse selection, moral hazard, and limited liability. Most contracting situations have 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. Moreover, virtually all contracting parties have limited liability. Entrepreneurs raising capital from investors, for example, enjoy limited liability as the value of their equity cannot fall below zero. Minimum wage laws often enforce limited liability in employment contracts.

We consider a principal-agent relationship with bilateral risk neutrality and limited liability. The agent selects an unobserved “effort,” which may consist of a single or multiple tasks. The agent also has private information, in an arbitrary way, about the distribution of outputs and about effort costs, resulting in a model where types and efforts are multidimensional (possibly infinite dimensional) and unordered. We show that the interaction between adverse selection, moral hazard, and limited liability imposes severe screening costs.

With binary outcomes, the optimal mechanism offers a single contract to all agents regardless of the type space or the distribution of types. The intuition for this result is as follows. Limited liability ensures that the agent’s participation constraint does not bind. Therefore, no agent type would decides not to participate when the principal drops a contract from a menu with multiple contracts. We show that there are two effects from removing all contracts except for the one with the highest bonus. First, efficiency increases since agents are paid a higher bonus.
Second, because agents are risk neutral, they always pick the contract that maximizes their expected payment conditional on their effort. Thus, holding effort fixed, limiting the choice of contracts offered to the agent reduces their expected payments. Since both effects are positive, the principal always gains with this substitution.

We generalize this result to settings with multiple outputs under a multiplicative separability condition. This condition – satisfied, for example, under the spanning condition of Grossman and Hart (1983) – is equivalent to assuming that agents rank the “power” of all contracts equally. Additionally, if the output distribution satisfies the monotone likelihood ratio property and the principal’s payoff must be monotone (“free disposal”), the optimal mechanism consists of the principal taking a single debt contract or, equivalently, giving all agents the same call option. We also show that these results are not knife edge, in the sense that it is still generically optimal to offer a single contract if the distribution is close to multiplicatively separable or if the agent’s utility function is close to linear and the distributions of types overlap.

More broadly, our paper shows that offering flexibility to agents through menus of contracts can hurt the principal. Because the agent is risk neutral, he always chooses the contract with the highest expected payment conditional on his effort, which is precisely the contract with the highest cost to the principal. That is, holding the agent’s effort fixed, reducing the number of contracts always increases the principal’s profits. In particular, when the principal can identify the contract with the highest power (i.e. when multiplicative separability holds), she can simultaneously reduce informational rents and increase efficiency by removing all other contracts.

Although the framework we study has been widely applied to financial contracting, it has many other applications. One such application is procurement and regulation. Despite the central role that menus of contracts play in the theory of procurement and regulation, they are rarely observed in practice.1 Accordingly, many papers try to identify conditions for simple procurement contracts to be close to optimal.2 We generalize the classic model of Laffont and Tirole (1986, 1993) by allowing effort to affect the regulated firm’s costs stochastically and assuming that the firm has limited liability. We then obtain conditions for the optimal mechanism to offer a single contract and for the optimal contract to be a price cap. Since limited liability constraints are a key aspect of most procurement contracts (see, e.g., Burguet et al., 2012), our model provides an explanation for the absence of menus of contracts in procurement.

Our results also contribute to an applied literature by identifying assumptions under which

---

1For example, Bajari and Tadelis (2001) argue that “the descriptive engineering and construction management literature (...) suggests that menus of contracts are not used. Instead, the vast majority of contracts are variants of simple fixed-price (FP) and cost-plus (C+) contracts.” In her survey of the literature, Netz (2000) argues that “price-cap regulation is the most commonly discussed and used form of incentive regulation.”

2Using the Laffont-Tirole framework, Rogerson (2003) and Chu and Sappington (2007) show that a pair of simple contracts can achieve a large fraction of the surplus under a certain range of parametric settings – 75 or 73 percent when costs follow either uniform or power distributions, respectively – for quadratic costs. Bajari and Tadelis (2001) assume that there is a fixed cost of specifying each state of nature in the contract to rationalize the simplicity of observed contracts.
researchers can focus on a simpler set of contracts when solving their models, which simplifies the task of obtaining comparative statics results. For example, we show that, under multiplicative separability, there is no loss of generality in restricting attention to a single debt contract when one introduces adverse selection in the pure moral hazard model of Innes (1990).

**Related Literature**

We consider a principal-agent relationship with bilateral risk neutrality and limited liability, as is commonly studied in corporate finance (c.f. Tirole, 2005). We build on this standard environment by adding adverse selection in an arbitrary way and allowing effort to be multidimensional. Our work is related to a literature that identifies conditions for contracts to take the form of debt and for equilibria to have complete pooling.

In a single-task setting with pure moral hazard and free disposal, Innes (1990) shows that contracts take the form of debt if the distribution of output satisfies the monotonicity of the likelihood ratio property. Our main focus is on the lack of menus of contracts, which, of course, can only be addressed by introducing adverse selection. Nevertheless, our Theorem 3, which obtains conditions for a single debt contract to be optimal, extends their result to settings that also have adverse selection under multiplicative separability.

In a signaling model of financial contracting (pure adverse selection), Nachman and Noe (1994) show that there is complete pooling if and only if firm types are strictly ordered by conditional stochastic dominance. When firms are ordered by conditional stochastic dominance, investors face a lemons problem: while they would like to offer better terms to healthier firms, less healthy firms are more likely to accept each contract. Our papers emphasize different forces that may lead to pooling. In their model, pooling occurs when the distribution of types induces a market breakdown for all but the worse contract. In our model, pooling happens because of moral hazard and limited liability: giving flexibility to agents requires the principal to leave excessive rents. For example, when output is binary, complete pooling occurs in our model for any parameters of the model (i.e. regardless of whether types are ordered). Similarly, Demarzo and Duffie (1999) consider a signaling model of security design and show that, under a uniform-worst case condition and a free disposal constraint, equilibrium contracts take the form of debt. Using this model, several authors studied whether intermediaries pool different assets in equilibrium. The conclusion depends on whether the security is designed before or after

---

3Poblete and Spulber (2012) generalize the analysis of Innes (1990) by introducing a critical ratio notion that captures the returns to providing incentives for effort and showing that it is optimal to offer a debt contract if this ratio is increasing.

4See Jewitt et al. (2008) for a general analysis of moral hazard models with limited liability. Moral hazard models with bilateral risk neutrality, limited liability, and free disposal include, for example, Matthews (2001), Dewatripont et al. (2003), Poblete and Spulber (2012), and Chaigneau et al. (2014b). Adverse selection models in this setting include Nachman and Noe (1994), Demarzo and Duffie (1999), DeMarzo (2005), and DeMarzo et al. (2005).
firms learn about the asset’s profitability (c.f., DeMarzo (2005), Biais and Mariotti (2005), and Farhi and Tirole (2015)).

In a one-dimensional screening setting (pure adverse selection), Guesnerie and Laffont (1984) showed that optimal mechanisms are “non-responsive” when the first-best allocation is decreasing. This occurs because optimality clashes with incentive compatibility, which requires allocations to be non-decreasing. The reason for pooling in our model is different from non-responsiveness. For example, if the agent only has private information about the distribution of output, the first best is increasing and therefore implementable in a pure adverse selection environment. Nevertheless, with multiplicative separability, the principal offers at most one contract (see footnote 12). More related to our work, Ollier and Thomas (2013) substitute the traditional (interim) participation constraint by an ex-post constraint in a one-dimensional model with binary outcomes. They show that, under conditions that ensure that the first-order approach holds, there is no benefit from screening.

As argued previously, our application to procurement and regulation builds on Laffont and Tirole (1986, 1993). In their model, there is both adverse selection and moral hazard. However, because the link between effort, types, and output is deterministic, the model can be reduced to a pure adverse selection model. For this reason, they are often referred to as models with ‘false moral hazard’ (c.f. Laffont and Martimort, 2002). We allow effort to affect the regulated firm’s costs stochastically so the problem cannot be reduced to a pure adverse selection model. Picard (1987), Melumad and Reichelstein (1989), and Caillaud et al. (1992) also introduce noise in the relationship between output and effort and show that, under certain conditions, the principal can achieve the same utility as in the absence of noise. Therefore, unlike in our model, they find that there is no cost from moral hazard. Our model differs from theirs in two ways. First, we also allow the agent to have private information about the distribution of output, while they assume that all private information concerns the cost of effort. Second, they do not assume that agents have limited liability. Limited liability also prevents the principal from eliminating moral hazard at no cost.

In Section 2, we present the model with two outputs and discuss the benchmark cases of pure adverse selection and pure model hazard. In Section 3, we generalize the results for multiple outputs. In Section 4, we introduce the free disposal constraint and obtain conditions for the optimality of debt. In Section 5, we present the application of our model to procurement and regulation. Section 6 discusses the robustness of the results to MS and risk neutrality. Then, Section 7 concludes. All proofs are in the appendix.
2 Two Outputs

2.1 Statement of the Problem

We start with a two-output model. There is a risk-neutral principal and a risk-neutral agent with limited liability. The agent has private information about the environment, captured by a type $\theta \in \Theta$. From the principal’s perspective, types are distributed according to a distribution $\mu$. We will discuss the assumptions on $\Theta$ and $\mu$ below.

The agent exerts an effort $e \in E$, which costs $c^\theta_e$. The space of possible efforts $E$ is a compact metric space. Effort can consist of a single task ($E \subset \mathbb{R}$) or multiple tasks ($E \subset \mathbb{R}^N$). The least-costly effort has a non-positive cost: $\min_{e \in E} c^\theta_e \leq 0$. This condition is satisfied in standard frameworks where the lowest effort costs zero, as well as in more general multi-task frameworks that allow the agent to derive private benefits from certain actions.\(^5\)

The principal does not observe the effort chosen by the agent. She does, however, observe the output from the partnership $x \in \{x_L, x_H\}$, which is stochastically affected by the agent’s effort. We refer to $x_H$ as a high output or as success, to $x_L$ as a low output or failure, and to $\Delta x := x_H - x_L > 0$ as the incremental output. Given effort $e$, a high output happens with probability $p^\theta_e$.

The type space $\Theta$ may be discrete or continuous, and types may be finite- or infinite-dimensional. Each type is fully characterized by the pair of functions $(p^\theta_e, c^\theta_e)$ specifying the probability of success and the cost of each effort.\(^6\) For example, if effort is binary, each type can be described by the four-dimensional vector $(p^\theta_0, p^\theta_1, c^\theta_0, c^\theta_1)$. If effort is continuous, each type is described by the infinite-dimensional function $(p^\theta_e, c^\theta_e): E \rightarrow \mathbb{R}^2$. Our model does not require the agent to have private information about all of these dimensions, of course. The case where the cost of effort $e$ is common knowledge, for example, is accommodated by letting $c^\theta_e$ be constant in $\theta$. Note that we do not impose any order on the space of types and efforts. Success probabilities and costs may be non-monotone functions and, moreover, types and effort may be complements, substitutes, or neither in terms of probabilities and costs.

By the revelation principle, we can focus on direct mechanisms. Let $\mathcal{B}(\Theta)$ denote the Borel $\sigma$-field of $\Theta$. A direct mechanism is a triple of $\mathcal{B}(\Theta)$-measurable functions $(s, b, e): \Theta \rightarrow \mathbb{R}^2 \times E$, consisting of fixed payments $s$ (or salaries), bonuses $b$, and effort recommendations $e$. An agent who reports type $\theta$ agrees to exert effort $e(\theta)$ and receives $s(\theta)$ in case of failure and $s(\theta) + b(\theta)$ in case of success. A pair of payments $s(\theta)$ and $b(\theta)$ is called a contract.

---

\(^5\)Allowing the cost of the lowest effort to be positive makes participation random as in Rochet and Stole (2002) and is beyond the scope of this paper.

\(^6\)We write $p^\theta_e$ to denote the function $e \mapsto p^\theta_e$ that keeps $\theta$ constant and varies $e$. Similarly, $p^\theta_\cdot_e$ refers to $\theta \mapsto p^\theta_\cdot_e$. The same notation is used for other functions.
Given a mechanism \((s, b, e)\), a type-\(\theta\) agent gets payoff

\[
U(\theta) := s(\theta) + p_{e(\theta)}^\theta b(\theta) - c_{e(\theta)}^\theta.
\]  
(1)

The mechanism must satisfy the following incentive compatibility (IC) and participation (IR) constraints:

\[
U(\theta) \geq s(\hat{\theta}) + p_{e(\hat{\theta})}^\theta b(\hat{\theta}) - c_{e(\hat{\theta})}^\theta, \quad \forall \theta, \hat{\theta}, \hat{e}.
\]  
(IC)

\[
U(\theta) \geq 0, \quad \forall \theta.
\]  
(IR)

The agent is protected by limited liability (LL), which prevents payments from being negative:

\[
s(\theta) \geq 0 \text{ and } s(\theta) + b(\theta) \geq 0, \quad \forall \theta.
\]  
(LL)

An optimal mechanism maximizes the principal’s expected profit

\[
\int_\Theta \left\{ p_{e(\theta)}^\theta [x_H - (s(\theta) + b(\theta))] + \left(1 - p_{e(\theta)}^\theta\right) [x_L - s(\theta)] \right\} d\mu(\theta)
\]  
(2)

among mechanisms that satisfy IC, IR, and LL. Two mechanisms are equivalent if, for almost all types, they give the same payoffs to both the principal and the agent.

To ensure the existence of an optimal mechanism, we make the following technical assumptions, which are satisfied by all standard agency models:

**Assumption 1.** \(\Theta\) is a measurable space. \(\mu\) is a probability measure on \(\mathcal{B}(\Theta)\). For each \(\theta \in \Theta\), \(p_{e(\theta)}^\theta\) and \(c_{e(\theta)}^\theta\) are continuous functions and, for each \(e \in E\), \(p_e^\theta\) and \(c_e^\theta\) are \(\mathcal{B}(\Theta)\)-measurable functions.

### 2.2 Benchmarks

We first consider the benchmark cases of pure moral hazard and pure adverse selection. We show that, in both cases, the principal typically offers a different contract to each type. Therefore, the uniform contract result that we will obtain in Section 2.3 requires both moral hazard and adverse selection. For simplicity, we focus on the single-task case \((E \subset \mathbb{R})\) and assume that the probability of high output \(p_e^\theta\) and the cost of effort are both non-decreasing in \(e\) (i.e. effort increases the probability of success at a cost). We normalize the lowest effort to zero.

---

7Setting the reservation utility equal to zero is a normalization. The important assumption is that there exists an effort with low enough cost, so that any mechanism satisfying IC and LL also satisfies the IR (see footnote 5 and Lemma 1).
Pure Moral Hazard

Suppose the principal observes the agent’s type but does not observe effort. Without limited liability, the principal can implement the first best by “selling the firm” to each agent – i.e., paying a bonus equal to the incremental output \( b(\theta) = \Delta x \) and offering a fixed payment that extracts the entire surplus \( s(\theta) = c^0_{e(\theta)} - p^0_{e(\theta)} \Delta x \), where \( e(\theta) \) is the first-best effort. With limited liability, the principal needs to leave rents to the agent if she wants to sell the firm. Then, it is profitable to distort the bonus downward, reducing the agent’s effort. Moreover, limited liability binds, so the agent gets a zero fixed payment.

Optimal contracts with and without limited liability vary in opposite dimensions: while, without limited liability, they have the same bonus \( (b = \Delta x) \) and different salaries, optimal contracts with limited liability have the same salary \( (s = 0) \) and different bonuses. In both cases, however, the principal offers different contracts to different types. Moreover, these mechanisms are no longer feasible if types are unobservable. If offered contracts with the same bonus and different salaries, all types would select the one with the highest salary. Similarly, if offered the same salary and different bonuses, they would all pick the contract with the highest bonus. The principal can still screen unobservable types by varying both the salary and the bonus. In fact, without limited liability, this is typically optimal. Our main result shows that, with limited liability, the principal prefers not to offer a menu of contracts. Instead, the optimal mechanism offers a single contract despite the presence of many different types.

Pure Adverse Selection

Now suppose the principal observes the agent’s effort but not his type. If effort costs are common knowledge (that is, \( \theta \) only affects the conditional probability of success), the principal can implement the first best by fully reimbursing the cost of each effort. The agent would be

\[ s(\theta) = c^0_{e(\theta)} - p^0_{e(\theta)} \Delta x, \]

\[ e(\theta) \]

\[ b(\theta) = \Delta x \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]

\[ x_L \]

\[ x_H \]

\[ p^0_H \]

\[ p^0_L \]

\[ c^0_{e(\theta)} \]

\[ p^0_{e(\theta)} \]

\[ \Delta x \]
indifferent between all efforts and would therefore accept to choose the principal’s preferred one.\textsuperscript{12} The first best can no longer be implemented when the agent has private information about effort costs: If the principal offered to fully reimburse the effort costs of all types, they would all pretend to be the types with the highest costs. Conversely, if the probability of success is common knowledge (\(\theta\) only affects the cost of effort), the optimal mechanism posts a payment for each (observed) effort and agents choose their efforts based on their privately-known costs.

As the next example illustrates, offering a menu of contracts can be optimal if the agent has private information about both probabilities and costs (but effort is still observable):

**Example 1.** There are two efforts (0 and 1, or “low” and “high”) and two types (A and B). The effort costs are \(c^A_1 = 1\), \(c^B_1 = \frac{2}{3}\), and \(c^A_0 = c^B_0 = 0\). Given a high effort, the probability of success for type A is \(p^A_1 = \frac{2}{3}\) and for type B is \(p^B_1 = \frac{1}{3}\). We assume that the project fails with a high enough probability if they exert low effort and take \(x_H - x_L\) to be large enough for the principal to want to implement high effort from both types.

In the appendix, we show that the optimal mechanism offers the following payments: \(s(A) = 0\), \(b(A) = \frac{3}{2}\), \(s(B) = \frac{2}{3}\), and \(b(B) = 0\). Notice that the principal uses salaries and bonuses to screen types: type A, which has a higher probability of success, gets zero salary and a positive bonus, whereas type B, which has a lower probability of success, gets a positive salary but no bonus. Moreover, this mechanism is no longer feasible if effort is not observable, since both types would choose \(e = 0\).

### 2.3 Contract Simplicity

We can now state the simplicity result with binary outcomes, which establishes that the principal offers a single contract:

**Theorem 1.** There exists an optimal mechanism that offers a single contract \((s, b)\) to all types, with \(s = 0\) and \(b < \Delta x\). Moreover, any optimal mechanism is equivalent to a mechanism that offers the same contract to all types.

The proof is based on three lemmas. The first one shows that IC and LL imply that IR never binds. This follows from the fact that the agent can always guarantee himself a non-negative payoff by picking the lowest effort and collecting the non-negative payments. The second lemma shows that any mechanism that pays a bonus greater than the incremental output to some type cannot be optimal. Any such mechanism gives the principal a payoff that is lower than if she offered all types a constant payment of zero.

\textsuperscript{12}Note that, unlike in Guesnerie and Laffont (1984), the first-best allocation in this case is non-decreasing and is therefore feasible when the agent is not subject to moral hazard. Complete pooling in our model is due to the interaction between adverse selection, moral hazard, and limited liability, not because of non-responsiveness.
The last lemma establishes that for any mechanism with bonuses lower than the incremental output, if multiple contracts are being offered, the principal can improve by offering all types the contract with the highest bonus. Since IR never binds, all agents pick this single contract once other contracts are removed. There are two effects from this migration to the highest-powered contract: an increase in efficiency and a reduction of rents. The higher bonus induces agents to pick efforts with a higher probability of success, which raises the total surplus, although at a higher cost to the agent (efficiency effect). On the other hand, because agents are risk neutral, they pick the contract that maximizes expected payments conditional on their effort choice. Thus, holding effort fixed, reducing the set of contracts being offered decreases expected payments to the agents (rent extraction effect). Combining the two effects, the principal’s payoff gain is the increased probability of success times the incremental output net of the bonus paid to the agent:

$$(p_{\tilde{e}} - p_e)(\Delta x - b),$$

where $e$ is the agent’s old effort, $\tilde{e}$ is the agent’s new effort, and $p_{\tilde{e}} \geq p_e$. Since the incremental output exceeds the bonus, both terms are positive. The proof then concludes by showing that any optimal mechanism is equivalent to a mechanism that offers the same contract to all types and that an optimal mechanism exists.

Limited liability and risk neutrality play an important role in Theorem 1. Limited liability ensures that agents do not leave the mechanism if their contract is removed. Without it, the participation constraint would bind for some type. Then, removing low-powered contracts would induce some types to prefer not to participate. Risk neutrality implies that, holding effort fixed, the principal and the agent split a pie of a fixed size. Since the agent always picks the contract with the highest expected payment, providing more freedom of choice to the agent can only hurt the principal (holding effort fixed). With risk aversion, different bonuses also affect the size of the pie since lower bonuses insure the agent better. Then, removing all but the highest-powered contract improves efficiency but worsens risk sharing.

Theorem 1 greatly simplifies the analysis of the optimal mechanism by allowing us to rewrite the principal’s program as a standard optimization problem with a single instrument $b \in [0, \Delta x]$. It is then straightforward to obtain comparative statics results. For example, using a supermodularity argument, we can show that the optimal bonus and the success probabilities of all types are increasing in the incremental output $\Delta x$. Moreover, the probability of success is distorted downwards relative to the first best. Formally, letting $e^{FB}$ denote a first-best effort, we have:

$$e^{FB}(\theta) \in \arg\max_{e} x_L + p_e^\theta \Delta x - c_e^\theta \quad \text{and} \quad e(\theta) \in \arg\max_{e} p_e^\theta b - c_e^\theta.$$ 

Since $b < \Delta x$, it follows by a revealed-preferences argument that $p_e^{\theta(\theta)} \leq p_e^{\theta e^{FB}(\theta)}$, with strict
inequality for some type if the type space is sufficiently rich. Finally, notice that Theorem 1 does not depend on the distribution of types or other parameters of the model.

It is straightforward to generalize Theorem 1 to the case of multiple outputs when contracts are restricted to two-part tariffs, where the fixed part corresponds to a wage that is paid independent of the output and the variable part corresponds to equity payments that are linear in the firm’s output. The restriction to two-part tariffs can be motivated by arbitrage opportunities when the principal deals with multiple agents who can costlessly redistribute outputs between themselves. In the next section, we study optimal contracts without this restriction.

3 Multiple Outputs

We now generalize the model to allow for multiple outputs. Let \( X := \{x_1, ..., x_N\} \) be the set of possible (real-valued) outputs with \( x_1 < ... < x_N \). As before, the agent’s private information is described by a type \( \theta \in \Theta \). Types are distributed according to a probability measure \( \mu \) on \( B(\Theta) \). Notice that our formulation allows types to be finite- or infinite-dimensional, and their distribution may be discrete or continuous.

The agent chooses an unobservable effort \( e \) from the compact metric space \( E \). A type-\( \theta \) agent who exerts effort \( e \) produces output \( x_i \) with probability \( p_e^\theta (x_i) := \Pr(x = x_i|\theta, e) \). Let \( c_e^\theta \) denote type \( \theta \)'s cost of effort \( e \). As before, we assume that the least-costly effort has a non-positive cost: \( \min_e c_e^\theta \leq 0 \) for all \( \theta \).

A contract is a function that specifies a transfer to the agent conditional on each possible output. A mechanism specifies a contract and an effort recommendation for each type. That is, a mechanism is a pair of measurable functions \( w : \Theta \times X \to \mathbb{R} \) and \( e : \Theta \to E \), so that a type-\( \theta \) agent is recommended effort \( e(\theta) \) and gets paid \( w^\theta (x) \) in case of output \( x \).

Given a mechanism \( (w, e) \), a type-\( \theta \) agent gets expected payoff

\[
U(\theta) := \sum_{i=1}^{N} w^\theta (x_i) p_e^\theta (x_i) - c_e^\theta.
\]

As in the two-output case, the mechanism has to satisfy the following IC, IR, and LL constraints:

\[
U(\theta) \geq \sum_{i=1}^{N} w^\theta (x_i) p_e^\theta (x_i) - c_e^\theta, \quad \forall \theta, \hat{\theta}, \hat{e}, \tag{IC}
\]

\[
U(\theta) \geq 0, \quad \forall \theta, \tag{IR}
\]

\[
w^\theta (x_i) \geq 0, \quad \forall \theta, i. \tag{LL}
\]

An optimal mechanism maximizes the principal’s expected profit
\[
\int_{\Theta} \sum_{i=1}^{N} (x_i - w^\theta(x_i)) P_{\theta}(x_i) d\mu(\theta)
\]  

among mechanisms that satisfy IC, IR, and LL.

The main question we address in this section is whether there exists an optimal mechanism that offers the same contract to all types. The example below shows that, without additional restrictions, the answer is no.

**Example 2.** There are \(N - 1 \geq 3\) types and \(N\) states: \(\Theta = \{1, \ldots, N-1\}\) and \(X = \{x_1, \ldots, x_N\}\). There are two efforts, \(E = \{0, 1\}\), and all types have the same effort costs: \(c^\theta_0 = 0\) and \(c^\theta_1 = (N - 2)/(N - 1)\) for all \(\theta\). The conditional probability of each output is

\[
P^\theta_1(x) = \begin{cases} 
0.5 & \text{if } x = x_{\theta+1} \\
\frac{1}{2(N-1)} & \text{if } x \neq x_{\theta+1}
\end{cases}
\text{ and } 
P^\theta_0(x) = \begin{cases} 
0.5 & \text{if } x = x_1 \\
\frac{1}{2(N-1)} & \text{if } x \neq x_1
\end{cases}
\]

For each type, the lowest output \((x_1)\) is the most likely outcome with low effort, whereas \(x_{\theta+1}\) is the most likely outcome with high effort. Suppose that \(x_1\) is low enough, so that it is optimal for the principal to implement high effort from all types.

In the appendix, we show that the optimal mechanism offers the following contracts:

\[
w^\theta(x) = \begin{cases} 
2 & \text{if } x = x_{\theta+1} \\
0 & \text{if } x \neq x_{\theta+1}
\end{cases}
\]

Each type’s contract pays 2 if output matches that agent’s type and zero otherwise. Therefore, the optimal mechanism consists of \(N - 1\) different contracts, one for each type.

The example illustrates the main problem in generalizing Theorem 1 to multiple outputs. With two outputs, the only way to incentivize effort is to pay a higher bonus. The power of any contract is determined by its bonus.\(^\text{13}\) With multiple outputs, there is one bonus associated with each incremental output, so contract power has, in general, only a partial order. A contract that convinces one type to exert effort may be ineffective in incentivizing another type. In the example, the cheapest way to incentivize type \(\theta\) to exert effort was to pay a bonus in case of output \(x_{\theta}\).

To rule out cases such as the one in Example 2, where types disagree over the effectiveness of incentives, we need to ensure that types order the power of different contracts in the same way. Formally, we need the following property to hold. For any \(w\) and \(\tilde{w}\) be two contracts satisfying

\(^{13}\)Similarly, when the contract space is restricted to two-part tariffs, we can unequivocally rank the power of two contracts by their slopes.
LL, if there exists $\theta_0 \in \Theta$ for which

$$\sum_{i=1}^{N} w(x_i) \left[ p^\theta(x_i) - p^\theta_0(x_i) \right] = \sum_{i=1}^{N} \tilde{w}(x_i) \left[ p^\theta(x_i) - p^\theta_0(x_i) \right]$$

for all $e, \tilde{e} \in E$, then

$$\sum_{i=1}^{N} w(x_i) \left[ p^\theta(x_i) - p^\theta_{\tilde{e}}(x_i) \right] = \sum_{i=1}^{N} \tilde{w}(x_i) \left[ p^\theta(x_i) - p^\theta_{\tilde{e}}(x_i) \right]$$

for all $\theta \in \Theta$ and all $e, \tilde{e} \in E$. This condition states that if one type has the same incentive to exert all efforts under contracts $w$ and $\tilde{w}$, so do all other types. In other words, types agree on the incentives provided by each contract. Of course, they may still pick different efforts depending on their distributions and effort costs.

Although intuitive, this is a somewhat convoluted assumption. In Appendix C, we show that this assumption is equivalent to the following *multiplicative separability* (MS) condition:

**Definition 1.** A cumulative distribution $F^\theta_e$ satisfies multiplicative separability (MS) if there exist functions $H : X \to \mathbb{R}$ and $I : E \times \Theta \to \mathbb{R}$ such that

$$F^\theta_e(x) + I(e, \theta)H(x) = F^\theta_{\tilde{e}}(x) + I(\tilde{e}, \theta)H(x) \quad \forall e, \tilde{e}, \theta, x.$$  \hspace{1cm} (4)

Multiplicative separability always holds when there are only two outputs. It also holds under the Linearity of the Distribution Function Condition (Grossman and Hart (1983) and Hart and Holmstrom (1987)), which is obtained by taking $E = [0, 1]$, $I(e, \theta) \in [0, 1]$, and $H(x) = F_0(x) - F_1(x)$ for some distributions $F_0$ and $F_1$:

$$F^\theta_e(x) = I(e, \theta)F_1(x) + [1 - I(e, \theta)]F_0(x).$$  \hspace{1cm} (5)

This condition is commonly used in pure moral hazard models, along with the convexity of costs, to justify the first-order approach. Importantly, however, none of our results assume the validity of the first-order approach.\textsuperscript{14}

As we show in Appendix C, MS has a simple geometric interpretation. Suppose there are three outputs, so the probabilities conditional on each type and effort lie on the two-dimensional space. 

\textsuperscript{14}There is no relationship between MS and Holmstrom’s (1979) sufficient statistic result. For example, MS always holds with binary outputs. However, as long as the likelihood ratio $p^\theta_e(x)$ is not constant in $\theta$ for some efforts $e_L$ and $e_H$, $x$ will not be a sufficient statistic for $e$ given $(x, \theta)$. Moreover, because types also affect the cost of effort, the optimal pure-moral-hazard contract is also a function of $\theta$ even if types do not affect the likelihood ratio (see footnote 9).
Figure 1: MS with three outputs.

simplex. MS holds if and only if

\[ p_\theta^e(x_2) - p_\theta^\tilde{e}(x_2) = \phi \left[ p_\theta^e(x_3) - p_\theta^\tilde{e}(x_3) \right] \quad \forall \theta, e, \tilde{e} \]

for some constant $\phi$. Figure 1 illustrates this condition graphically. Each point in the graph corresponds to the probability distribution conditional on a type and an effort. We draw two points with the same color if they correspond to the same type but different efforts. MS requires the lines connecting these points to have the same slope ($\phi$). More generally, MS requires all efforts to have the same proportional effect on the probability of each outcome for all types.

The following technical conditions, which generalize Assumption 1, guarantee the existence of an optimal mechanism:

**Assumption 2.** $\Theta$ is a measurable space, $c^\theta_e$ and $p^\theta_e(x_i)$ are continuous functions of $(e, \theta)$, and $p^\theta_e(x_i) \geq p$ for some $p > 0$.

The next theorem shows that, whenever MS holds, the optimal mechanism offers the same contract to all types.\(^{15}\) Different types may still choose different efforts depending on their output distributions and effort costs.

\(^{15}\)MS can be slightly weakened since it does not need to hold for all efforts, only the ones that principal may want to implement. For example, Theorem 2 remains unchanged if (4) fails at points where $F^\theta(x)$ and $c^\theta$ are locally concave (since such efforts are not implementable).
**Theorem 2.** Suppose MS holds. There exists an optimal mechanism that offers a single contract to all types. Moreover, any optimal mechanism is equivalent to a mechanism that offers the same contract to all types.

With two outputs, substituting all contracts by the one with the highest bonus had two effects. First, holding effort fixed, it reduced the expected payment to all types. Second, it increased the probability of success, which raised the principal’s profit because the bonus was lower than the incremental output. The first of these effects remains unchanged with multiple outputs: holding effort fixed, the principal and the agent split a pie of fixed size (since they are both risk neutral). Therefore, for any fixed effort, removing a contract from the mechanism cannot hurt the principal.

The main difficulty with multiple outputs is with the second effect. High-powered contracts may be unprofitable for two reasons. First, a high-powered contract may incentivize an inefficient effort. Second, because some bonuses may exceed the incremental output (i.e., \(x_{i+1} - x_i < w(x_{i+1}) - w(x_i)\) for some \(i\)), improving the distribution of outputs may decrease the principal’s profits.\(^\text{16}\) There are three possible cases. If we start with a mechanism in which the principal would like to encourage effort for all types, replacing contracts by the one with the highest power (which exists by MS) increases profits. If we start with a mechanism in which the principle would like to discourage effort for all types, it is profitable to replace all contracts by the one with the lowest power (which also exists by MS). Lastly, if the principal would like to encourage effort by some types and discourage effort by some other types, it is possible to identify a single contract that leaves all effort incentives constant. Therefore, the second effect is also non-negative.

Recall that, with two outputs, the agent gets zero in the low state and a positive payment in the high state. The proposition below generalizes this result for arbitrary outputs when MS holds:

**Proposition 1.** Suppose MS holds. Let \(w^*\) be an optimal contract and let \(e(\theta)\) be an optimal effort for type \(\theta\) when offered contract \(w^*\). Then, either (i) \(w^*(x_i) = 0\) for all \(i \notin \arg \max \left\{ \int_{\theta_1} p_{\theta_1}^\theta \frac{d\mu(\theta)}{h_i} \right\} \), or (ii) \(w^*(x_i) = 0\) for all \(i \notin \arg \min \left\{ \int_{\theta_1} p_{\theta_1}^\theta \frac{d\mu(\theta)}{h_i} \right\} \).

To understand Proposition 1, recall that in models of pure moral hazard, payments strongly depend on the likelihood ratio between the effort being implemented and the effort to which the agent is most tempted to deviate (see Holmstrom (1979); Chaigneau et al. (2014a)). Since agents are risk neutral in our model, if there were no adverse selection, each agent’s optimal

\(^\text{16}\)If the principal has access to a free disposal technology, as we discuss in Section 4, bonuses cannot exceed the incremental output so the second concern discussed above is not an issue. Then, the MS condition can be generalized. In Appendix D, we show that a finite number of contracts implements the optimal mechanism under a Generalized Multiplicative Separability (GMS) condition, which assumes that distributions are convex combinations of a finite family of distributions that is ordered by first-order stochastic dominance. The distribution in Example 2 satisfies GMS.
contract would pay zero in all states but the one with the highest likelihood ratio. Under MS, the distribution of outputs satisfies
\[
\frac{p^\theta_{\bar{e}, i}}{p^\theta_{e, i}} - 1 = - \left[ I(\bar{e}, \theta) - I(e, \theta) \right] \frac{h_i}{p^\theta_{e, i}},
\]
so the likelihood ratio between efforts \(e\) and \(\bar{e}\) for type \(\theta\) is maximized either at the state with the highest or with the lowest ratio \(\frac{p^\theta_{\bar{e}, i}}{h_i}\) (depending on whether \(I(\bar{e}, \theta) < I(e, \theta)\) or \(I(\bar{e}, \theta) > I(e, \theta)\), respectively). With adverse selection and MS, Theorem 2 shows that the principal offers the same contract to all types. Then, instead of offering each type a contract that pays in the state that either maximizes or minimizes the ratio \(\frac{p^\theta_{\bar{e}, i}}{h_i}\), the optimal single contract pays in the state that maximizes or minimizes the *average* ratio of among all types \(\int_\Theta p^\theta_{e, i}(\theta) d\mu(\theta)\). In particular, an optimal contract can only pay a positive amount in multiple states if they have the same average ratios. Therefore, for generic distributions, optimal contracts pay zero in all but one state.

### 4 Free Disposal and Debt Contracts

We now introduce the following *free disposal* (FD) constraint in the multiple-output environment from Section 3:

\[
y - w^\theta(y) \geq x - w^\theta(x)
\]

for all \(\theta \in \Theta\) and all \(x, y \in X\) with \(y \geq x\). FD requires the principal’s profit to be non-decreasing.\(^{17}\) As Innes (1990) argues, FD can be seen as an additional incentive constraint if the principal can costlessly reduce output or if the agent can borrow from outside lenders to inflate output. An optimal FD mechanism maximizes the principal’s expected profit (3) among mechanisms that satisfy IC, IR, LL, and FD.

In this section, we consider the optimality of debt contracts. The principal gets a *debt contract* if his payments \(x - w(x)\) equal min \(\{x, \bar{x}\}\) for some face value \(\bar{x}\), or, equivalently, if the agent is given a call option \(w(x) = \max\{x - \bar{x}, 0\}\). Incentive-compatible mechanisms cannot offer more than one debt contract, since agents would always pick the one with the lowest face value. Therefore, for a mechanism to offer debt contracts only, it needs to offer the same contract to all types. Accordingly, we assume that MS holds, so the principal offers a single contract.

The existing literature established that, in the single-task version of the model \((E \subset \mathbb{R})\) with pure moral hazard, the monotone likelihood ratio property (MLRP) is sufficient for the optimality of debt. In this one-dimensional effort model, a probability mass function \(p^\theta_e\) satisfies MLRP if, for any \(e_L, e_H\) with \(e_L < e_H\), the ratio \(\frac{p^\theta_{e_H}(x)}{p^\theta_{e_L}(x)}\) is increasing in \(x\). Intuitively, MLRP means that

\(^{17}\)FD still allows the agent’s payment \(w^\theta(x)\) to be non-monotonic. The result from this section continues to hold if we also impose monotonicity on the agent’s side or limited liability on the principal, since debt contracts, which will be shown to be optimal, automatically satisfy these constraints.
the “evidence” in favor of higher efforts increases with output. MLRP plays an important role
on the monotonicity of contracts (Holmstrom (1979); Grossman and Hart (1983)). Accordingly,
we work with the following notion of MLRP:

**Definition 2.** A distribution satisfying MS has the monotone likelihood ratio property if it
can be written as in equation (4) and \( \frac{p_{e_{H}(x)}}{p_{e_{L}(x)}} \) is increasing in \( x \) for any \( e_{L}, e_{H}, \) and \( \theta \) with
\[ I(e_{H}, \theta) < I(e_{L}, \theta). \]

Our next result establishes that when the distributions satisfy MLRP, not only is it optimal
to offer only one contract, but this contract takes the form of debt:

**Theorem 3.** Suppose MS holds and the distributions satisfy MLRP. There exists an optimal FD
mechanism that gives the principal a single debt contract. Moreover, any optimal FD mechanism
is equivalent to a mechanism that gives the principal a single debt contract.

The intuition for the optimality of debt is reminiscent of Innes (1990). With MLRP, higher
outputs are more “indicative” of higher effort. Therefore, transferring payments from lower to
higher outputs relaxes the IC constraints.\(^{18}\)

In the canonical pure moral hazard model, there is a single IC, preventing the agent from
choosing a lower effort. Here, because of adverse selection, there is one IC for each type, which
prevents each type from picking a different effort. However, because of multiplicative separability,
the ICs of all types are aligned, in the sense that if a perturbation raises one type’s incentives to
exert effort, it must also raise the incentives of all types. This observation allows us to summarize
all the constraints of the program into a single one, as in the pure moral hazard model.

5 **Procurement and Regulation**

In this section, we adapt our main framework to a setting of procurement and regulation that
builds on the classic framework of Laffont and Tirole (1986, 1993). Their model can be reduced
to a pure adverse selection model because effort affects the regulated firm’s cost deterministically.
Our model incorporates moral hazard in their model by allowing effort to affect the firm’s cost
stochastically.

A regulated firm produces an indivisible good that generates a consumer surplus of \( S > 0 \) at
a random monetary cost \( C \in \{ C_{1}, ..., C_{N} \} \) with \( C_{1} < ... < C_{N} \). The firm’s manager chooses a
cost-reducing effort \( e \in E \), which is not observed by the regulator. Let \( p_{e}^{\theta} \) denote the probability
that the firm’s cost equals \( C \) conditional on type \( \theta \) and effort \( e \).

\(^{18}\)In Appendix D we provide a generalization of the MS condition (GMS) which allows for a larger class of
distribution. We show that if a strong version of MLRP holds and contracts satisfy FD, the optimal mechanism
is again implemented by a single debt contract.
The firm’s manager has cost of effort \( e_\theta \), with \( \min_\theta e_\theta \leq 0 \). As argued previously, this is satisfied if the lowest effort costs zero or if the manager gets private benefits out of some activities. The firm’s manager has private information about both his ability to cut costs (i.e., the conditional distribution of costs \( p_\theta \)) and the cost of effort \( c_\theta \). The regulator observes the monetary cost \( C \) incurred by the firm but not the manager’s effort \( e \). As an accounting convention, assume that the regulator reimburses the firm’s monetary costs in addition to paying the firm an amount conditional on the observed cost \( C \).

A contract is a function that specifies a transfer to the firm conditional on each possible cost \( C \). A mechanism is a pair of measurable functions \( w : \{C_1, ..., C_N\} \times \Theta \rightarrow \mathbb{R} \) and \( e : \Theta \rightarrow \mathbb{R} \) specifying, for each reported type, a recommended effort and a transfer for each cost realization. Given a mechanism \((w, e)\), a type-\( \theta \) manager gets payoff

\[
U(\theta) := \sum_{i=1}^{N} w^\theta(C_i) p^\theta_{e(\theta)}(C_i) - c^\theta_{e(\theta)}.
\tag{6}
\]

As usual, the mechanism must satisfy the IC and IR constraints

\[
U(\theta) \geq \sum_{i=1}^{N} w^\theta(C_i) p^\theta_{e(\hat{\theta})}(C_i) - c^\theta_{e(\hat{\theta})}, \quad \forall \theta, \hat{\theta}, \hat{e},
\tag{IC}
\]

\[
U(\theta) \geq 0, \quad \forall \theta.
\tag{IR}
\]

The manager is protected by limited liability, so that payments are non-negative:

\[
w^\theta(C) \geq 0, \quad \forall C.
\tag{LL}
\]

We impose the technical conditions from Assumption 2 (with \( C \) instead of \( x \)).

Since, by the accounting convention described above, the regulator fully reimburses the firm’s cost realization, the regulator’s expected payment to type \( \theta \) equals

\[
\sum_{i=1}^{N} \left[ C_i + w^\theta(C_i) \right] p^\theta_{e(\theta)}(C_i).
\]

Because the government uses distortionary taxation to raise public funds, the regulator faces a shadow cost of public funds \( \lambda > 0 \). The net surplus of consumers/taxpayers is

\[
S - (1 + \lambda) \sum_{i=1}^{N} \left[ C_i + w^\theta(C_i) \right] p^\theta_{e(\theta)}(C_i).
\tag{7}
\]

A utilitarian regulator maximizes the sum of the expected utility of the firm’s manager (6) and
the consumers’ net surplus \( S \):

\[
S - (1 + \lambda) \sum_{i=1}^{N} C_i p_e^\theta (C_i) - c_e^\theta - \lambda \sum_{i=1}^{N} w^\theta (C_i) p_e^\theta (C_i).
\]

As the last term of this expression shows, because taxation is distortionary \((\lambda > 0)\), leaving rents the the regulated firm is costly. Moreover, because each dollar reimbursed to the firm has an additional cost of \( \lambda \) due to distortionary taxation, cutting the firm’s cost increases social surplus by \( 1 + \lambda \). The first-best effort minimizes \((1 + \lambda) \sum_{i=1}^{N} C_i p_e^\theta (C_i) + c_e^\theta \).

The main difference between this model and the principal-agent model considered previously is that, while the principal only cares about her own payoff, a utilitarian regulator also cares about the manager’s payoffs. Therefore, the regulator also internalizes the manager’s effort cost so their preferences are not perfectly misaligned. However, to avoid distortionary taxation, the regulator would still like to leave as little rents as possible to the firm’s manager.

We can now state the simplicity result for the case of binary outcomes:

**Proposition 2.** Suppose there are two possible monetary costs \((N = 2)\). There exists an optimal mechanism that offers the same contract to all types. Moreover, for any optimal mechanism, there exists an equivalent mechanism that offers the same contract to all types.

The proof of Proposition 2 is similar to the one from Theorem 1, with one key difference. Because the regulator internalizes the cost of effort, to show that the bonus must be lower than the one that implements the first best, we substitute contracts by the one that “sells the firm” to the manager (rather than the contract that always pays zero, as in Theorem 1).

Next, we consider the case of multiple outputs \( N > 2 \). As in Section 4, contracts must satisfy the following free disposal constraint:

\[
w^\theta (C_j) - w^\theta (C_i) \leq C_i - C_j
\]

for all \( j \) and \( i \) with \( i > j \). FD requires the firm’s compensation for cutting costs not to exceed the amount cut. It must be satisfied, for example, if the firm’s manager can secretly inflate firm costs. An optimal FD mechanism maximizes the principal’s expected profit (3) among mechanisms that satisfy IC, IR, LL, and FD.

\(^{19}\)If types were observable and the firm did not have limited liability, the regulator would implement the first best by making the firm the residual claimant of the social gain from cutting costs and extracting the entire surplus:

\[
w^\theta (C_j) = (1 + \lambda) \left[ \sum_{i=1}^{N} C_i p_e^{FB(\theta)} (C_i) - C_j \right],
\]

where \( e^{FB}(\theta) \) is the first-best effort. This violates LL because \( w \) is non-degenerate and has mean zero.
As in Section 4, we assume that the cost distributions satisfy MS. Notice that, differently from output in the principal-agent model, a higher cost decreases the principal’s payoff. We therefore say that the cost distribution satisfies MLRP if, for any \( e_L, e_H, \) and \( \theta \) with \( I(e_H, \theta) > I(e_L, \theta) \), \( \frac{p^{\theta}_{e_H}(C)}{p^{\theta}_{e_L}(C)} \) is decreasing in \( C \). That is, higher costs are indicative of “less effort” (more precisely, they are indicative of a lower \( I(e, \theta) \)).

Adapting Theorems 2 and 3, we obtain the following results:

**Proposition 3.** Suppose MS holds.

a) There exists an optimal mechanism that offers a single contract to all types. Moreover, any optimal mechanism is equivalent to a mechanism that offers the same contract to all types.

b) Suppose, in addition, that the distributions satisfy MLRP. Then, there exists an optimal FD mechanism that offers to all types the contract \( w(C) = \max\{\bar{C} - C; 0\} \), for some \( \bar{C} \). Moreover, for any optimal FD mechanism, there exists an equivalent mechanism that offers to all types the contract \( w(C) = \max\{\bar{C} - C; 0\} \), for some \( \bar{C} \).

The optimal contract in part (b) specifies a “reasonable cost” \( \bar{C} \) and reimburses the regulated firm for any cost cuts beyond \( \bar{C} \). Since, by our accounting convention, the regulator pays the firm’s cost directly, the firm’s revenues under this contract equal:

\[
w(C) + C = \max\{\bar{C}, C\}.
\]

This reimbursement rule consists of a standard price cap except that, because of limited liability, the regulator must bailout firms with cost realizations above the cap.\(^{20}\) Price caps are the most common form of incentive regulation. They are used, for example, by the U.S. Federal Communications Commission (FCC) to regulate the telephone industry. Price caps are often used in procurement as well. For example, prospective reimbursement systems commonly used in health care specify an amount \( \bar{C} \) based on what a service should cost, and let providers keep cost savings \( \bar{C} - C \) to themselves. They are used, for example, by Medicare.

## 6 Robustness

In this section we study the robustness of the result on the optimality of a single contract to the assumptions of multiplicative separability and risk neutrality. Generalizing the setup from Section 3, we show that it is generically optimal to offer a single contract when the distributions of types is close to multiplicatively separable. Then, we consider an extension of the setup from

\(^{20}\)More specifically, if the realized cost is below the price cap \( \bar{C} \), the firm is profitable and the reimbursement rule is the same as with a standard price cap. However, because the firm is protected by limited liability, the regulator cannot force it to remain active if its profits are low enough (normalized to zero here). Therefore, the reimbursement rule above is a price cap with the additional feature that the regulator must bailout the firm if the losses from the price cap are substantial.
Section 2 and show that it is optimal to offer a single contract when the agent’s utility function is close to linear and the distributions of types overlap.

Robustness with respect to multiplicative separability

Theorem 2 relies on the multiplicative separability of the output distribution, which is a very strong assumption. One may question whether its conclusion still holds when the distribution is only approximately multiplicatively separable. We now show that the single-contract result is not knife-edge, in the sense that it is robust to perturbations to the distribution.

To simplify the analysis, we assume that the effort space and the type space are both finite. Consider a set of economies indexed by a vector of probabilities and costs:

\[ \{p_\theta^{e,i}, c_\theta^e : e \in E, \theta \in \Theta, x_i \in X\}. \]

Since these are finite-dimensional vectors, we take any of the equivalent norms of Euclidean space. Proposition 4 shows that, generically, if the distribution is “close” to multiplicatively separable, it is optimal to offer a single contract that pays a bonus in a single state:

**Proposition 4.** Let $E$ and $\Theta$ be finite. For almost all economies satisfying MS, there is a neighborhood around it for which the optimal mechanism is unique and offers a single contract to all types. Moreover, this contract pays zero in all but one state.

Recall that when the distribution satisfies MS and there are no ties in the likelihood ratio, the optimal contract pays zero in all but one state. The proof of proposition 4 establishes that if we apply a small perturbation to the output distribution or the agent’s cost, the set of binding constraints in the principal’s problem remains unchanged. Then, it is optimal to pay zero in all but the same state as before the perturbation, although the amount paid in that state must be adjusted to incorporate the change in the incentive constraints.

Robustness with respect to risk neutrality

Since the proof of Theorem 1 relies on the risk neutrality assumption, one may question whether adding a small amount of risk aversion would undermine its conclusion. While a general analysis of optimal mechanisms under risk aversion is beyond the scope of this paper, we will study whether the conclusion from Theorem 1 remains valid when the agent is “sufficiently risk neutral.”

For tractability, suppose that efforts and types are both binary: $E = \{H, L\}$ and $\Theta = \{A, B\}$. As usual, we assume that high effort ($H$) is more likely to succeed but is more costly: $p_H^\theta \geq p_L^\theta$ and $c_H^\theta \geq c_L^\theta = 0$. Without loss of generality, we will identify by $B$ the type with the highest benefit-cost of effort ratio:

\[ \frac{p_H^B - p_L^B}{c_H^B} > \frac{p_H^A - p_L^A}{c_H^A}. \]  

(8)
Unlike in the rest of the paper, we allow the agent to be risk averse. Formally, a type-$\theta$ agent has Bernoulli utility over money $x$ and effort $e$ given by $u(x) - c_\theta e$, where $u$ is twice differentiable, strictly increasing, weakly concave, has a uniformly bounded derivative, and satisfies $u(0) \geq \min_{e \in E} c_\theta e$. Let $v(\bar{u}) := u^{-1}(\bar{u})$ denote the cost of providing utility $\bar{u}$ to the agent.

**Proposition 5.** Consider the model with a risk averse agent and binary outputs, efforts, and types. Suppose that $p_{B_H} > p_{A_L} > p_{B_L}$. There exists $\epsilon > 0$ such that, whenever $\sup \|v'(x) - v'(0)\| < \epsilon$, the unique optimal mechanism offers a single contract $(s_v, b_v)$ to all types, with $s_v = 0$ and $b_v < \Delta x$.

Recall that with risk neutrality, removing all but the contract with the highest power allows the principal to profit from all types. Risk aversion introduces a cost of forcing some types to get contracts with higher power by worsening risk sharing. Proposition 5 provides conditions under which the loss from exposing agents to higher risk is lower than the gain from increasing efficiency and reducing informational rents, so it is still optimal to offer a single contract to all types. The conditions are that as agents are sufficiently risk neutral, and the distributions of both types overlap. When the distributions of both types do not overlap and the optimal mechanism recommends high effort from some type, it is generically optimal to offer multiple contracts even if the utility function is arbitrarily close to linear. We provide a formal proof of this result in the online appendix.

7 Conclusion

The observation that many contracts are simple and relatively uniform across different sectors is an old puzzle in contract theory. While standard adverse selection models predict that agents will be offered large menus of contracts, contracting parties typically offer a limited number of contracts, often a single one. While standard moral hazard models predict that contracts should be fine tuned to the likelihood ratio of output, similar contracts are offered in different environments.

We argue that these two features endogenously emerge in a general model of moral hazard and adverse selection if contracts must satisfy limited liability. With binary outcomes, the principal always offers a single contract regardless of any parameters of the model. The joint presence of moral hazard and adverse selection is key for this result. When either types or effort are observed, the principal typically prefers to offer different contracts to different types. With multiple outputs, it is optimal to offer two contracts if the distribution of output satisfies a separability condition. Moreover, if the marginal distribution satisfies MLRP, this optimal FD contract is a debt contract for the principal.

---

21This last assumption ensures that, as in the risk neutral case, the participation constraint does not bind.
Our paper shows an important downside from giving flexibility to agents by offering menus of contracts. This is particularly stark with bilateral risk neutrality where, holding effort fixed, the agent always selects the most expensive contract to the principal. Then, reducing the number of contracts offered to the agent always increases the principal’s profits (for a fixed effort). If, in addition, we can identify a most efficient contract from the menu (such as when output is binary or when the distribution is multiplicatively separable and ordered), the principal can always improve by eliminating other contracts.

The simplicity results rely on the presence of limited liability and bilateral risk neutrality. Limited liability ensures that increasing the power of a contract will not force the agent to abandon the mechanism. Bilateral risk neutrality means that, holding effort fixed, principal and agent are perfectly misaligned so that the principal always benefits by reducing the agent’s flexibility. With risk aversion, their misalignment is no longer perfect because there are potential gains from risk-sharing. While risk neutrality is a reasonable assumption in many settings (such as the optimal compensation of wealthy managers, procurement contracts, or the regulation of large companies), there are many other settings where they are not (for example, insurance contracts or sharecropping). When the agent is sufficiently risk averse, our results no longer hold. In our companion paper (Gottlieb and Moreira 2014), we study optimal mechanisms in the binary-outcome model when agents are risk averse and when there are no limited liability constraints. Although we obtain some simplicity results, optimal mechanisms are considerably more complex than they are here.
Appendix

A. Proofs

Proof of Theorem 1

We first verify that participation is implied by incentive compatibility and limited liability:

Lemma 1. Let \((s, b, e)\) be a mechanism that satisfies IC and LL. Then, it satisfies IR.

Proof. The IC preventing \(\theta\) from deviating to \(e\) states that \(U(\theta) \geq s(\theta) + p_\theta^b(\theta) - c_\theta^e\). Then, because \(s(\theta)\) and \(b(\theta) + s(\theta)\) are non-negative (LL) and \(\min_{e \in E} c_\theta^e \leq 0\), it follows that \(U(\theta) \geq 0\).

The proof of the theorem will use the following result, which states that the bonus does not exceed the incremental output:

Lemma 2. Consider a mechanism \((s, b, e)\) in which \(b(\theta) > \Delta x\) for some \(\theta\). Then, there exists another mechanism that gives the principal a greater payoff than \((s, b, e)\).

Proof. This is a special case of Lemma 8.

The next lemma, which is the main step for proving Theorem 1, shows that any feasible mechanism is weakly dominated by a mechanism that offers a single contract to all types:

Lemma 3. Let \((s, b, e)\) be a mechanism satisfying IC, IR, and LL. There exists a mechanism that offers a single contract \((0, b^\ast)\) to all types and gives the principal a (weakly) greater payoff than \((s, b, e)\).

Proof. Let \((s, b, e)\) be a mechanism that satisfies IC, IR, and LL. By Lemma 2, for any mechanism that offers a bonus greater than \(\Delta x\) for some type, there exists another mechanism offering bonuses lower than \(\Delta x\) to all types that gives the principal a higher payoff. Thus, there is no loss of generality in assuming that \(b(\hat{\theta}) \leq \Delta x\) for all \(\hat{\theta} \in \Theta\). Fix a type \(\theta \in \Theta\).

Let \(b^\ast \equiv \sup \left\{ b(\hat{\theta}) : \hat{\theta} \in \Theta \right\}\) and \(s^\ast \equiv \inf \left\{ s(\hat{\theta}) : \hat{\theta} \in \Theta \right\}\) denote the “highest” bonus and the “lowest” fixed payment in the mechanism. The result is immediate if \(b^\ast = 0\). Suppose that \(b^\ast > 0\). If \(s^\ast > 0\), reducing all fixed payments uniformly by \(s^\ast\) would keep all the constraints satisfied and increase the principal’s payoff. Therefore, we can assume that \(s^\ast = 0\). Moreover, either there exists \(\hat{\theta}\) such that \(s(\hat{\theta}) = 0, b(\hat{\theta}) = b^\ast\) (i.e., \((0, b^\ast)\) is offered in the mechanism), or \((0, b^\ast)\) is a limit point of \(\left\{ (s(\hat{\theta}), b(\hat{\theta})) : \hat{\theta} \in \Theta \right\}\).

Consider the alternative mechanism that offers the contract \((0, b^\ast)\) to all types and let

\[ e^\ast(\theta) \in \arg \max_{e \in E} p_\theta^e b^\ast - c_\theta^e, \]
which exists because the objective function is continuous and $E$ is compact. The principal’s payoff from type $\theta$ in the original mechanism is

$$x_L - s(\theta) + p_{e(\theta)}^\theta (\Delta x - b(\theta)).$$

(9)

Her payoff in the alternative mechanism is

$$x_L + p_{e^*(\theta)}^\theta (\Delta x - b^*).$$

(10)

Since the contract $(0, b^*)$ either belongs to, or is a limit point of the original mechanism, no agent can be better off by switching to $(0, b^*)$ while holding the recommended effort fixed:

$$s(\theta) + p_{e(\theta)}^\theta b(\theta) - c_{e(\theta)}^\theta \geq p_{e(\theta)}^\theta b^* - c_{e(\theta)}^\theta.$$

(This inequality follows from type $\theta$’s incentive constraint while holding $e(\theta)$ fixed). Adding $c_{e(\theta)}^\theta$ to both sides, it follows that the expected payment in the alternative mechanism cannot exceed the one from the original mechanism if effort is held constant:

$$s(\theta) + p_{e(\theta)}^\theta b(\theta) \geq p_{e(\theta)}^\theta b^*.$$

(11)

That is, if the agent chooses not to change effort ($e^*(\theta) = e(\theta)$), the principal obtains a higher payoff in the alternative mechanism (10) than in the original one (9). Allowing effort to change, the principal’s payoff in the alternative mechanism minus the payoff in the original mechanism becomes

$$p_{e^*(\theta)}^\theta (\Delta x - b^*) - p_{e(\theta)}^\theta (\Delta x - b(\theta)) + s(\theta) = (p_{e^*(\theta)}^\theta - p_{e(\theta)}^\theta)(\Delta x - b^*) + s(\theta) + p_{e(\theta)}^\theta b(\theta) - p_{e(\theta)}^\theta b^*$$

$$\geq (p_{e^*(\theta)}^\theta - p_{e(\theta)}^\theta)(\Delta x - b^*),$$

(12)

where the inequality follows from (11).

By assumption, $\Delta x \geq b^*$. We claim that $p_{e^*(\theta)}^\theta \geq p_{e(\theta)}^\theta$. To wit, by the incentive constraint of type $\theta$ in the original mechanism,

$$s(\theta) + p_{e(\theta)}^\theta b(\theta) - c_{e(\theta)}^\theta \geq s(\theta) + p_{e^*(\theta)}^\theta b(\theta) - c_{e^*(\theta)}^\theta,$$

and, by the definition of $e^*(\theta)$,

$$p_{e^*(\theta)}^\theta b^* - c_{e^*(\theta)}^\theta \geq p_{e(\theta)}^\theta b^* - c_{e(\theta)}^\theta.$$
Rearranging both inequalities, we can write them as:

\[(p_{e^*(\theta)}^\theta - p_{e(\theta)}^\theta)b^* \geq c_{e^*(\theta)}^\theta - c_{e(\theta)}^\theta \geq (p_{e^*(\theta)}^\theta - p_{e(\theta)}^\theta)b(\theta).\]

Since \(b^*\) is the supremum of bonuses, \(b^* \geq b(\theta)\). Therefore, since \(b^* > 0\), \(p_{e^*(\theta)}^\theta \geq p_{e(\theta)}^\theta\). Hence, (12) implies that the principal’s payoff from each type \(\theta\) is higher in the alternative mechanism even when the agent chooses a different effort.

The proof of existence follows arguments similar to Page (1992) and is given in the online appendix. To conclude the proof, we now establish that any optimal mechanism is equivalent to a mechanism that offers a single contract.

**Lemma 4.** Any optimal mechanism is equivalent to a mechanism that offers the same contract to all types.

**Proof.** Let \((s, b, e)\) be an optimal mechanism. Using the notation of the proof of Lemma 3, consider a mechanism \((0, b^*, e^*)\) that gives the principal a (weakly) greater payoff than \((s, b, e)\) and \(b^*\) is a constant function. Using the same argument as in the proof of Lemma 3, we can show that

\[s(\theta) + p_{e(\theta)}^\theta b(\theta) - c_{e(\theta)}^\theta \geq p_{e^*(\theta)}^\theta b^* - c_{e^*(\theta)}^\theta \geq p_{e^*(\theta)}^\theta b^* - c_{e^*(\theta)}^\theta\]  

(13)

and

\[p_{e^*(\theta)}^\theta(\Delta x - b^*) - p_{e(\theta)}^\theta(\Delta x - b(\theta)) + s(\theta) = (p_{e^*(\theta)}^\theta - p_{e(\theta)}^\theta)(\Delta x - b^*) + s(\theta) + p_{e(\theta)}^\theta b(\theta) - p_{e(\theta)}^\theta b^* \geq (p_{e^*(\theta)}^\theta - p_{e(\theta)}^\theta)(\Delta x - b^*) \geq 0.\]  

(14)

If the first inequality of (13) is strict, then the first inequality of (14) is also strict. Since \((s, b, e)\) is an optimal mechanism, we must have

\[s(\theta) + p_{e(\theta)}^\theta b(\theta) - c_{e(\theta)}^\theta = p_{e^*(\theta)}^\theta b^* - c_{e^*(\theta)}^\theta\]

and

\[p_{e(\theta)}^\theta(\Delta x - b(\theta)) - s(\theta) = p_{e^*(\theta)}^\theta(\Delta x - b^*)\]

for all almost all \(\theta \in \Theta\), i.e., the agent’s and principal’s payoffs in mechanisms \((s, b, e)\) and \((0, b^*, e^*)\) are the same almost surely.

Before presenting the proof of Theorems 2 and 3 and Proposition 1, it is helpful to introduce some notation. We will write \(p_{e,i}^\theta\) for \(p_e^\theta(x_i)\) and \(w_i\) for \(w(x_i)\). It will also be helpful to write variables in terms of increments: \(h_i \equiv H(x_i) - H(x_{i-1})\), \(\Delta x_i \equiv x_i - x_{i-1}\) and \(\Delta w_i \equiv w_i - w_{i-1}\) for \(i \geq 1\); \(H(x_0) = 0\), \(h_1 := H(x_1)\), \(\Delta x_0 \equiv x_0\) and \(\Delta w_0 \equiv w_0\).
Proof of Theorem 2

Let $\Delta x \equiv x_N - x_1$. The following result will be important in order to establish existence of an optimal mechanism, by allowing us to restrict the set of possible contracts to a compact set.

**Lemma 5.** Let $(w, e)$ be a mechanism satisfying IC and LL. Suppose that $w_i^\theta \geq \frac{\Delta x}{p}$ for some $\theta$ and $i$. Then $(w, e)$ is not optimal.

**Proof.** The proof has two steps. First, it shows that if an incentive-compatible mechanism offers a high enough payment in one state, then every other contract must also have a high enough payment in some state (otherwise, everyone would prefer the former contract). Second, it shows that any contract that makes a high enough payment in some state is dominated by than the null contract.

**Step 1.** Suppose the mechanism offers a contract $\bar{w} = (\bar{w}_1, ..., \bar{w}_N)$ with $\bar{w}_i \geq \frac{\Delta x}{p}$ for some output $i$. Let $\theta$ be a type that picks $w$ and exerts effort $e$. Then, this type's incentive-compatibility constraint gives:

$$
\sum_{i=1}^{N} w_i p_{e,i}^\theta - c_e^\theta \geq \sum_{i=1}^{N} \bar{w}_i p_{e,i}^\theta - c_e^\theta.
$$

Since $\max\{w_1, ..., w_N\} \geq E[w|\theta, e]$ and $\bar{w}_i \geq 0$ for all $i$, this inequality implies the following:

$$
\max_{i\in\{1, ..., N\}} \{w_i\} \geq p\bar{w}_i \geq \frac{\Delta x}{p},
$$

where the last inequality uses $\bar{w}_i \geq \frac{\Delta x}{p}$.

**Step 2.** We now show that this mechanism gives the principal a lower payoff than offering the contract that always pays zero to all types. Since the probability is bounded below by $p$ and payments are non-negative, the principal's payoff from offering $w = (w_1, ..., w_N)$ to type $\theta$ is

$$
\sum_{i=1}^{N} (x_i - w_i) p_{e(\theta),i} \leq x_N - pw_i, \ \forall i.
$$

Let $e_0 \in \arg\max c_e^\theta$. The principal's payoff from offering type $\theta$ a zero payment in all states is $\sum_{i=1}^{N} x_i p_{e,\theta,i} > x_0$. Combining these two inequalities, we obtain the following necessary condition for $w$ to give a higher payoff to the principal than the null contract:

$$
w_i < \frac{\Delta x}{p}, \ \forall i.
$$

Thus, if

$$
w_i \geq \frac{\Delta x}{p}
$$

for some $i$, then the principal is strictly better offering the null contract. \hfill \square
The proof will also use the fact that any contract satisfies IC and LL also satisfies IR because \( \min_e c_e^\theta \leq 0 \). Before presenting the proof, it is useful to introduce some terminology. For simplicity, we abuse notation and write \( p_{e,i}^\theta \) for \( p_e^\theta(x_i) \) and \( w_i \) for \( w(x_i) \). The payoff of a type-\( \theta \) agent who exerts effort \( e \) and gets contract \( w \) equals

\[
 v_e^\theta(w) := \sum_{i=1}^N w_i p_{e,i}^\theta - c_e^\theta. 
\]  

(15)

Analogously, the principal’s payoff from such type is

\[
 u_e^\theta(w) := \sum_{i=1}^N (x_i - w_i) p_{e,i}^\theta.
\]

(16)

By MS, a type-\( \theta \) agent who switches from effort \( \tilde{e} \) to \( e \) while keeping the same contract \( w \) gains

\[
 v_e^\theta(w) - v_{\tilde{e}}^\theta(w) = -[I(e, \theta) - I(\tilde{e}, \theta)] \sum_{i=1}^N w_i h_i + c_{\tilde{e}}^\theta - c_e^\theta.
\]

(17)

In turn, this switch changes the principal’s payoff by

\[
 u_e^\theta(w) - u_{\tilde{e}}^\theta(w) = -[I(e, \theta) - I(\tilde{e}, \theta)] \sum_{i=1}^N (x_i - w_i) h_i.
\]

(18)

If \( I(e, \theta) \geq I(\tilde{e}, \theta) \) the principal (weakly) gains from shifting effort from \( \tilde{e} \) to \( e \) if and only if

\[
 \sum_{i=1}^N (x_i - w_i) h_i \leq 0.
\]

(19)

Similarly, the agent’s expected payment (weakly) increases from this shift in efforts if and only if \( \sum_{i=1}^N w_i h_i \leq 0 \).

The proof shows that the principal can (weakly) profit by removing all but two contracts from any feasible menu of contracts. Contracts for which the principal would like to “encourage effort” – i.e., equation (19) holds – are substituted by the contract with the highest incentives. Contracts for which the principal wants to “discourage effort” are substituted by the one with the lowest incentives. Establishing this result requires two steps. First, we verify that this substitution increases the principal’s profits if types pick the contract intended for them. Second, we verify that each type prefers to pick the contract intended for him.

**Proof of the theorem.** Let \((w,e)\) be a feasible mechanism. Let \( \mathcal{M} := \{w^\theta : \theta \in \Theta\} \) denote the set of all contracts in this mechanism. By Lemma 5, there is no loss of generality in assuming that \( \mathcal{M} \) is bounded. Its closure, \( \bar{\mathcal{M}} \), is compact. There are three cases to consider:
**Case 1** $\sum_{i=1}^{N} (x_i - w_i^0)h_i \geq 0$, for all $\theta \in \Theta$. Let $w^+ \in \arg \max_{w \in \mathcal{M}} \sum_{i=1}^{N} w_i h_i$, which exists because $\mathcal{M}$ is compact and the objective function is a continuous linear functional. Let $e^+(\theta)$ be an effort that maximizes the agent’s payoff under contract $w^+$ (see the online appendix for existence). Then, the agent’s payoff with the effort chosen in the original mechanism, $e(\theta)$, cannot exceed the agent’s payoff with effort $e^+(\theta)$:

$$v_{e^+(\theta)}(w^+) \geq v_{e(\theta)}(w^+),$$

which, by MS, can be written as

$$[I(e(\theta), \theta) - I(e^+(\theta), \theta)] \sum_{i=1}^{N} w_i^0 h_i \geq c_{e^+(\theta)}^0 - c_{e(\theta)}^0. \quad (20)$$

Similarly, because $e(\theta)$ is the agent’s effort choice with contract $w^0$,

$$[I(e(\theta), \theta) - I(e^+(\theta), \theta)] \sum_{i=1}^{N} w_i^0 h_i \leq c_{e^+(\theta)}^0 - c_{e(\theta)}^0. \quad (21)$$

Combining (20) and (21), we obtain

$$[I(e(\theta), \theta) - I(e^+(\theta), \theta)] \sum_{i=1}^{N} w_i^0 h_i \leq c_{e^+(\theta)}^0 - c_{e(\theta)}^0 \leq [I(e(\theta), \theta) - I(e^+(\theta), \theta)] \sum_{i=1}^{N} w_i^0 h_i. \quad (22)$$

By the definition of $w^+$,

$$\sum_{i=1}^{N} w_i^+ h_i \geq \sum_{i=1}^{N} w_i^0 h_i.$$

Therefore, if $\sum_{i=1}^{N} w_i^+ h_i \neq 0$, it follows from (22) that $I(e(\theta), \theta) \geq I(e^+(\theta), \theta)$. If $\sum_{i=1}^{N} w_i^+ h_i = 0$, we can take $e^+(\theta) = e(\theta)$, for all $\theta$, and again $I(e(\theta), \theta) \geq I(e^+(\theta), \theta)$.

We now establish that replacing contract $w^0$ by $w^+$ increases the principal’s payoff from type $\theta$. As in the proof of Theorem 1, we first show that, holding effort fixed, the principal is better off with the substitution of contracts. Since $w^+$ is the limit of sequence in $\mathcal{M}$, the agent’s utility is continuous, and the original mechanism is incentive compatible, it follows that

$$v_{e(\theta)}^0(w^0) \geq v_{e(\theta)}(w^+). \quad (23)$$

Substitute the expression for the agent’s payoff, multiply both sides by $-1$, and add $\sum_{i=1}^{N} x_i p_{e(\theta),i}^0$ to both sides to write:

$$\sum_{i=1}^{N} (x_i - w_i^+) p_{e(\theta),i}^0 \geq \sum_{i=1}^{N} (x_i - w_i^0) p_{e(\theta),i}^0, \quad (24)$$

28
which states that, holding effort $e(\theta)$ fixed, the principal gets a higher profit with contract $w^+$ than with $w^\theta$.

To show that the change in effort also benefits the principal, notice that since $I(e(\theta), \theta) \geq I(e^+(\theta), \theta)$ and $w^+ \in \mathcal{M}$ (so that $\sum_{i=1}^N (x_i - w_i^+) h_i \geq 0$), the following inequality holds:

$$\left[I(e(\theta), \theta) - I(e^+(\theta), \theta)\right] \sum_{i=1}^N (x_i - w_i^+) h_i \geq 0.$$ 

Use MS to rewrite this inequality as

$$\sum_{i=1}^N (x_i - w_i^+) p_{e^+(\theta),i} \geq \sum_{i=1}^N (x_i - w_i^+) p_{e(\theta),i},$$

which states that the principal gains from the change in effort.

Combining (24) and (25) establishes that the principal’s profit from $\theta$ with the new contract exceeds her profit with the original contract:

$$u_{e^+(\theta)}(w^+) \geq u_{e(\theta)}(w^\theta).$$

By construction, the mechanism $(w^+, e^+)$ is incentive compatible and satisfies LL. Moreover, from the previous argument, it raises the principal’s payoff point-wise (i.e. it raises the principal’s payoff conditional on each type).

**Case 2** $\sum_{i=1}^N (x_i - w_i^\theta)h_i \leq 0$, for all $\theta \in \Theta$. The proof of case 2 is similar to the one of case 1, except that, instead of substituting all contracts by the one that maximizes $\sum_{i=1}^N w_i h_i$, we substitute them by the one that minimizes this expression. Formally, let $w^- \in \arg\min_{w \in \mathcal{M}} \sum_{i=1}^N w_i h_i$ and let $e^-(\theta)$ be an effort that maximizes the agent’s payoff under contract $w^-$. As in case 1, incentive compatibility gives

$$\left[I(e(\theta), \theta) - I(e^-(\theta), \theta)\right] \sum_{i=1}^N w_i^\theta h_i \leq c_{e^-(\theta)}^\theta - c_{e(\theta)}^\theta \leq \left[I(e(\theta), \theta) - I(e^-(\theta), \theta)\right] \sum_{i=1}^N w_i^- h_i.$$ 

Since $w^- \in \mathcal{M}$, it satisfies

$$\sum_{i=1}^N w_i^- h_i \leq \sum_{i=1}^N w_i^\theta h_i,$$

so that, by inequality (26), it follows from same argument as in case 1 that $I(e^-(\theta), \theta) \geq I(e(\theta), \theta)$. As in case 1, incentive compatibility implies that, holding effort $e(\theta)$ fixed, the
principal’s profit is higher with contract $w^{-}$ than with $w_{\theta}$:

$$\sum_{i=1}^{N} (x_{i} - w_{i}^{-})p_{e(\theta),i}^{\theta} \geq \sum_{i=1}^{N} (x_{i} - w_{i}^{\theta})p_{e(\theta),i}^{\theta}. \quad (27)$$

Next, we show that the change in effort also benefits the principal. Because $I(e^{-}(\theta), \theta) \geq I(e(\theta), \theta)$, and because $w^{-} \in \bar{M}$, the following inequality holds:

$$\left[ I(e(\theta), \theta) - I(e^{-}(\theta), \theta) \right] \sum_{i=1}^{N} (x_{i} - w_{i}^{-})h_{i} \geq 0.$$  

Use MS to rewrite this inequality as

$$\sum_{i=1}^{N} (x_{i} - w_{i}^{-})p_{e^{-}(\theta),i}^{\theta} \geq \sum_{i=1}^{N} (x_{i} - w_{i}^{-})p_{e(\theta),i}^{\theta}, \quad (28)$$

which shows that the principal gains from the change in effort. Combining (24) and (28) establishes that the principal’s profit from $\theta$ with the new contract exceeds her profit with the original contract: $u_{e^{-}(\theta)}(w^{-}) \geq u_{e(\theta)}(w^{\theta})$. Therefore, the mechanism $(w^{-}, e^{-})$ is incentive compatible, satisfies LL, and increases the principal’s payoff point-wise relative to the original mechanism.

**Case 3)** There exist $\theta_{+}, \theta_{-} \in \Theta$ for which $\sum_{i=1}^{N} (x_{i} - w_{i}^{\theta_{+}})h_{i} \geq 0 \geq \sum_{i=1}^{N} (x_{i} - w_{i}^{\theta_{-}})h_{i}$. First, we establish that, because of the risk neutrality and limited liability, introducing “scaled down versions” of the contracts from the original mechanism preserves incentive compatibility, meaning that no type would benefit from deviating to such a contract. More precisely, let

$$\mathcal{N} = \{\alpha w^{\theta}; \theta \in \Theta \text{ and } \alpha \in [0, 1]\}$$

denote the menu of contracts obtained by introducing scaled down versions of all contracts in $\mathcal{M}$. Then, for all $\theta, \hat{\theta} \in \Theta, e \in E$ and $\alpha \in [0, 1]$,

$$v_{e(\theta)}^{\theta}(w^{\theta}) \geq \sum w_{i}^{\hat{\theta}}p_{e,i}^{\hat{\theta}} - c_{e}^{\hat{\theta}} \geq \sum (\alpha w_{i}^{\hat{\theta}}) p_{e,i}^{\theta} - c_{e}^{\theta},$$

where the first inequality follows from incentive compatibility of the original mechanism and the second inequality follows from $w_{i}^{\hat{\theta}} \geq \alpha w_{i}^{\theta} \geq 0$ (by LL and the fact that $\alpha \leq 1$). Therefore, there is no loss of generality in assuming that the principal offers the menu of contracts $\mathcal{N}$ rather than $\mathcal{M}$.

30
Let $w^0 \in \mathcal{N}$ be a contract that satisfies
\[
\sum_{i=1}^{N} (x_i - w^0_i) h_i = 0. \tag{29}
\]

We claim that $w^0$ exists. Indeed, suppose first that $\sum_{i=1}^{N} x_i h_i \geq 0$. Then, because $\sum_{i=1}^{N} x_i h_i \leq \sum_{i=1}^{N} w^\theta_i h_i$, there exists $\alpha^0 \in [0, 1]$ such that
\[
\sum_{i=1}^{N} x_i h_i = \alpha^0 \sum_{i=1}^{N} w^\theta_i h_i.
\]

Similarly, suppose that $\sum_{i=1}^{N} x_i h_i \leq 0$. Then, because $\sum_{i=1}^{N} x_i h_i \geq \sum_{i=1}^{N} w^\theta_i h_i$, there exists $\alpha^0 \in [0, 1]$ such that
\[
\sum_{i=1}^{N} x_i h_i = \alpha^0 \sum_{i=1}^{N} w^\theta_i h_i.
\]

As in cases 1 and 2, incentive compatibility implies that, holding effort fixed, the principal’s profit is higher with contract $w^0$ than with $w^\theta$:
\[
\sum_{i=1}^{N} (x_i - w^0_i) p^\theta_{e(\theta), i} \geq \sum_{i=1}^{N} (x_i - w^\theta_i) p^\theta_{e(\theta), i}. \tag{30}
\]

Let $e^0(\theta)$ be an effort that maximizes the type $\theta$’s payoff under contract $w^0$. We claim that changing efforts from $e(\theta)$ to $e^0(\theta)$ does not affect the principal’s profit. To see this, multiply both sides of equation (29) by $I(e(\theta), \theta) - I(e^0(\theta), \theta)$ to write:
\[
[I(e(\theta), \theta) - I(e^0(\theta), \theta)] \sum_{i=1}^{N} (x_i - w^0_i) h_i = 0.
\]

Using MS, we can rewrite this equality as
\[
\sum_{i=1}^{N} (x_i - w^0_i) p^\theta_{e(\theta), i} = \sum_{i=1}^{N} (x_i - w^0_i) p^\theta_{e(\theta), i}, \tag{31}
\]

which shows that the principal gets the same payoff with both effort profiles.

Combining (30) and (31) establishes that the mechanism $(w^0, e^0)$ is incentive compatible, satisfies LL, and raises the principal’s payoff point-wise relative to the original mechanism.

Next, we establish the equivalence claim. Let $(w, e)$ be any optimal mechanism and construct $(\bar{w}, \bar{e})$ as done previously. We need to show that, for almost all types, the principal and the agent get the same payoffs in both mechanisms. As before, there are three possible cases:
\[ \sum_{i=1}^{N} [x_i - w_i^\theta] h_i \geq 0 \text{ for all } \theta; \sum_{i=1}^{N} [x_i - w_i^{\theta^+}] h_i \leq 0 \text{ for all } \theta \in \Theta; \text{ or } \sum_{i=1}^{N} (x_i - w_i^{\theta^+}) h_i \geq 0 \geq \sum_{i=1}^{N} (x_i - w_i^{\theta^-}) h_i \text{ for some } \theta^+ \text{ and } \theta^- \]. We will consider the first case (the other two are analogous).

By construction, \( u_{e^+(\theta)}(w^+) \geq u_{e(\theta)}(w^\theta) \) for all \( \theta \). By the optimality of \((w,e)\), this inequality cannot hold on a set of types with positive measure. Therefore, the principal must be indifferent between these mechanisms except for a set of types with measure zero. Now consider the agent’s payoff. Since \( w^+ \) is the limit of sequence in \( \mathcal{M} \), we can proceed as in (23) to obtain:

\[ v_{e(\theta)}(w^\theta) \geq v_{e^+(\theta)}(w^+) \]  

for all \( \theta \). Therefore, we must have

\[ v_{e(\theta)}(w^\theta) \geq v_{e^+(\theta)}(w^+) \geq v_{e(\theta)}(w^+) \]

for all \( \theta \) (where the last inequality follows from incentive compatibility of \((\bar{w},\bar{e})\)). Let \( \tilde{\Theta} \) be the set of types for which \( v_{e(\theta)}(w^\theta) > v_{e(\theta)}(w^+) \).

Using the expression for the agent’s utility and rearranging, we obtain

\[ u_{e(\theta)}^\theta(w^+) = \sum_{i=1}^{N} (x_i - w_i^{\theta^+}) p_{e(\theta),i}^\theta \geq \sum_{i=1}^{N} (x_i - w_i^{\theta}) p_{e(\theta),i}^\theta = u_{e(\theta)}(w^\theta), \]

with strict inequality exactly on \( \tilde{\Theta} \). Use MS to write

\[ u_{e^+(\theta)}^\theta(w^+) - u_{e(\theta)}^\theta(w^+) = [I(e(\theta), \theta) - I(e^+(\theta), \theta)] \sum_{i=1}^{N} (x_i - w_i^{\theta^+}) h_i \geq 0, \]

where, as in the previous part of the proof, the inequality follows from \( I(e(\theta), \theta) \geq I(e^+(\theta), \theta) \) and \( \sum_{i=1}^{N} (x_i - w_i^{\theta^+}) h_i \geq 0 \). Combine these two inequalities to obtain \( u_{e^+(\theta)}^\theta(w^+) \geq u_{e(\theta)}^\theta(w^\theta) \) for all \( \theta \), with strict inequality exactly on \( \tilde{\Theta} \). Again, by the optimality of mechanism \((w,e)\), \( \tilde{\Theta} \) must have zero measure, which concludes the proof.

**Proof of Proposition 1**

When the principal offers a single contract, we can use MS to rewrite IC as

\[ [I(e, \theta) - I(e(\theta), \theta)] \sum_{i=1}^{N} w_i h_i \geq c_{e(\theta)} - c_e, \forall e. \]
Notice that, if it is optimal for the agent to pick effort \( e(\theta) \) when he is offered contract \( w \), it is also optimal to do so it when offered any contract \( \tilde{w} \) with \( \sum_{i=1}^{N} w_i h_i = \sum_{i=1}^{N} \tilde{w}_i h_i \). We are now ready to present the proof of the proposition:

**Proof.** Let \( w^* \) be an optimal contract, let \( e(\theta) \) be the effort chosen by type \( \theta \) when offered this contract, and let \( \bar{p}_i \equiv \int_{\Theta} p^\theta_{e(\theta),i} d\mu(\theta) \) denote the (marginal) probability of state \( i \). Then, for \( K := \sum_{i=1}^{N} w^*_i h_i \), \( w^* \) must also solve the following program:

\[
\min_w \sum_{i=1}^{N} w_i \bar{p}_i
\]

subject to

\[
\sum_{i=1}^{N} w_i h_i = K, \quad \text{(33)}
\]

\[
w_i \geq 0, \text{ for all } i = 1, ..., N. \quad \text{(LL)}
\]

As argued above, (33) ensures that effort \( e(\theta) \) is still optimal for the agent. This is a restricted program: any contract that satisfies these constraints is feasible, but not every feasible contract satisfies these constraints. Since \( w^* \) is optimal among all feasible contracts, it must also solve this more restricted program that only includes a subset of feasible contracts.

The first-order conditions of this program are:

\[
-\bar{p}_j + \lambda h_j \leq 0 \quad \text{(34)}
\]

for all \( j \) with \( = \) if \( w^*_j > 0 \). There are three cases: \( \lambda > 0 \), \( \lambda < 0 \), and \( \lambda = 0 \). First, suppose that \( \lambda > 0 \) and let \( w^*_j > 0 \) in some state \( i \). Then, by (34),

\[
\frac{h_j}{\bar{p}_j} \leq \frac{h_i}{\bar{p}_i} = \frac{1}{\lambda}
\]

for all \( j \), so that \( w^*_j = 0 \) whenever \( j \notin \arg \max_j \left\{ \frac{h_j}{\bar{p}_j} \right\} \). Next, suppose that \( \lambda < 0 \) and let \( w^*_i > 0 \) in some state \( i \). Then, by (34), \( w^*_j = 0 \) whenever \( j \notin \arg \min_j \left\{ \frac{h_j}{\bar{p}_j} \right\} \). Lastly, notice that when \( \lambda = 0 \), condition (34) implies that \( w^*_j = 0 \) for all \( j \) (because \( \bar{p}_j > 0 \) for all \( j \)). In all three cases, the claim in the proposition is true. \( \Box \)

**Proof of Theorem 3**

It is straightforward to adapt the proof of Theorem 2 to show that it is still optimal to offer a single contract when free disposal is imposed. Therefore, there is no loss of generality in
assuming that the optimal mechanism offers a single contract.

Notice that whenever MS holds, we can write

\[ p^\theta_{e,i} + I(e, \theta)h_i = p^\theta_{\tilde{e},i} + I(\tilde{e}, \theta)h_i, \quad \forall e, \tilde{e}, \theta, i. \]

Rearranging this expression, gives

\[ \frac{p^\theta_{e,i}}{p^\theta_{\tilde{e},i}} - 1 = - \left[ I(\tilde{e}, \theta) - I(e, \theta) \right] \frac{h_i}{p^\theta_{e,i}}. \]

Thus, MLRP holds if and only if \( \frac{h_i}{p^\theta_{e,i}} \) is increasing in \( i \) for any \( e, \theta \).

Recall that, as shown in the proof of Proposition 1, if it is optimal for the agent to pick effort \( e(\theta) \) when offered contract \( w \), it is also optimal to do so it when offered any contract \( \tilde{w} \) with \( \sum_{i=1}^N w_i h_i = \sum_{i=1}^N \tilde{w}_i h_i \). We are now ready to present the proof:

**Proof of the theorem.** Let \( w^* \) be an optimal contract and let \( e(\theta) \) denote the effort chosen by type \( \theta \) when offered this contract. Then, for \( K = \sum_{i=1}^N w^*_i h_i \), this contract must solve the following program:

\[
\min_w \sum_{i=1}^N w_i \int_\Theta p^\theta_{e(\cdot),i} d\mu(\theta)
\]

subject to

\[
\sum_{i=1}^N w_i h_i = K, \quad (IC')
\]

\[
w_i \geq 0, \quad (LL)
\]

\[
x_i - x_{i-1} \geq w_i - w_{i-1}. \quad (M)
\]

As argued above, the first constraint ensures that effort \( e(\theta) \) is still optimal for the agent. The second and third constraints are LL and FD. This is a restricted program: any contract that satisfies these constraints is feasible but not every feasible contract satisfies these constraints. Therefore, since \( w^* \) is optimal among all feasible contracts, it must also solve this more restricted program that only includes a subset of feasible contracts.

Let \( \bar{p}_i \equiv \int_\Theta p^\theta_{e(\cdot),i} d\mu(\theta) \) denote the marginal distribution of outputs induced by effort \( e(\cdot) \). It is convenient to rewrite the program above in terms of increments:

\[
\min_{\{\Delta w_i\}} \sum_{j=0}^N \bar{p}_j \sum_{i=0}^j \Delta w_i \quad (35)
\]
subject to
\[
\sum_{j=0}^N h_j \sum_{i=0}^j \Delta w_i = K, \quad (\text{IC}')
\]
\[
\sum_{i=0}^j \Delta w_i \geq 0, \quad \forall j \quad (LL_j)
\]
\[
\Delta x_j - \Delta w_j \geq 0, \quad \forall j. \quad (M_j)
\]

We claim that if $LL_j$ holds with equality, then $M_j$ holds with strict inequality and vice-versa. To see this, notice that if $LL_j$ holds with equality, then
\[
\sum_{i=0}^j \Delta w_i = \sum_{i=0}^{j-1} \Delta w_i + \Delta w_j = 0 \implies \Delta w_j \leq 0 < \Delta x_j,
\]
showing that $M_j$ holds with inequality. Conversely, if $M_j$ holds with equality, then
\[
\Delta x_j = \Delta w_j \implies \Delta w_j + \sum_{i=0}^{j-1} \Delta w_i \geq \Delta x_j > 0,
\]
so $LL_j$ holds with inequality.

The necessary first-order conditions associated with program (35) are
\[
-\sum_{j=1}^N \bar{p}_j + \lambda^{IC} \sum_{j=1}^N h_j + \sum_{j=1}^N \mu^{LL}_j - \mu^{M}_i = 0, \quad \forall i \quad (36)
\]
along with the usual complementary slackness conditions.

Let $\xi_j \equiv \bar{p}_j - \lambda^{IC} h_j$ and notice that
\[
\xi_j > 0 \iff \frac{1}{\lambda^{IC}} > \frac{h_j}{\bar{p}_j}.
\]

Notice that MLRP implies that $\frac{h_j}{p_{e(\theta),j}}$ is increasing in $j$ for all $\theta$, so that $\int_{\theta} \frac{h_j}{p_{e(\theta),j} \mu(\theta)} = \frac{h_j}{\bar{p}_j}$ is also increasing in $j$. Hence, there exists a unique $k \in \{1, \ldots, N\}$ such that $\xi_j > (\leq)0$ for all $j < (\geq)k$.

Substituting $\xi_i$ in (36), we obtain
\[
\xi_i = \mu^{LL}_i - \mu^{M}_i + \mu^{M}_{i+1}, \quad i < N
\]
\[
\xi_N = \mu^{LL}_N - \mu^{M}_N.
\]
There are three cases: (1) $\xi_N > 0$, (2) $\xi_N < 0$, and (3) $\xi_N = 0$.

**Case 1) $\xi_N > 0$.** In this case, $LL_N$ binds and $M_N$ does not. Because $\xi_i$ crosses 0 once from above, $\xi_i > 0$ for all $i$ and, by induction, none of the free disposal constraints bind. As a result, the solution is $w_i = 0$ for all $i$.

**Case 2) $\xi_N < 0$.** In this case, $M_N$ binds. For $N-1$, we have

$$\xi_{N-1} = \mu_{N-1}^{LL} - \mu_{N-1}^{M} - \xi_N \geq \mu_{N-1}^{LL} - \mu_{N-1}^{M}.$$ 

If $N - 1 \geq k$ so that $\xi_{N-1} \leq 0$, it then follows that $\mu_{N-1}^{LL} - \mu_{N-1}^{M} \leq 0$ so that $M_{N-1}$ binds (and $LL_{N-1}$ doesn’t). Inductively, it follows that $M_i$ binds for all $i \geq k$.

Let $j$ be such that $LL_j$ binds (and, therefore $\mu_j^{LL} = 0$). By the previous argument, it must be the case that $j < k$ so that $\xi_j > 0$. We have that

$$0 < \xi_{j-1} = \mu_{j-1}^{LL} - \mu_{j-1}^{M} + \mu_j^{M} = \mu_{j-1}^{LL} - \mu_{j-1}^{M}.$$ 

Then, we must have that $LL_{j-1}$ binds and $M_{j-1}$ doesn’t. Hence, there exists $i^* \in \{0, ..., N\}$ such that $LL_i$ binds on all $i \leq i^*$ and $M_i$ binds on all $i > i^*$. The contract must therefore be an option with a price between $x_{i^*}$ and $x_{i^*+1}$.

**Case 3) $\xi_N = 0$.** In this case, $\mu_N^{LL} = \mu_N^{M}$. Since we have previously shown that we cannot have both constraints holding with equality, we must have $\mu_N^{LL} = \mu_N^{M} = 0$. Substituting at the condition for the previous output gives

$$\xi_{N-1} = \mu_{N-1}^{LL} - \mu_{N-1}^{M} > 0,$$

where we used the fact that $\xi_i$ crosses 0 once from above and $\mu_N^{M} = 0$. Thus, $\mu_{N-1}^{LL} > 0 = \mu_{N-1}^{M}$ so that $LL_{N-1}$ binds. Substituting inductively shows that all other LL constraints also bind. Hence, the solution in this case is an option contract with a strike price $x^* \in \{x_{N-1}, x_N\}$.

The last part of the proposition follows similar argument of the proof of Theorem 2.

**Proof of Propositions 2 and 3**

In order to rewrite the procurement model using similar terminology as in Section 3, perform the change of variables:

$$x_i := S - (1 + \lambda)C_i.$$
We will write contracts in terms of the taxpayer’s net surplus $x$, instead of the firm’s production cost $C$ by letting $W^θ(x) := w^θ \left( \frac{S-x}{1+\lambda} \right)$. The distribution of the taxpayer’s net surplus $x$ is determined by $G^θ_{e}(x_i) := F^θ_e \left( \frac{S-x_i}{1+\lambda} \right)$. Notice that if $F^θ_e$ satisfies MS, so does $G^θ_{e}$. With some abuse of notation, we will write $p^θ_{e}$ for the probability distribution function associated with the cumulative distribution function $G^θ_{e}$.

The regulator’s payoff is
\[
\sum_{i=1}^{N} \left( x_i - \lambda W^θ_{i} \right) p^θ_{e(\theta),i} - c^θ_{e(\theta)},
\]
whereas the type-$θ$ manager’s payoff is
\[
\sum_{i=1}^{N} W^θ_{i} p^θ_{e(\theta),i} - c^θ_{e(\theta)},
\]
where we write $W^θ_{i}$ for $W^θ(x_i)$, for all $i$ and $θ$.

**Proof of Proposition 2**

Let $s(θ) \equiv W_1(θ)$ and $b(θ) \equiv W_2(θ) - W_1(θ)$. As in Lemma 1, it can be verified that any mechanism that satisfies IC and LL also satisfies IR. The proof will follow from two lemmas, which are similar to Lemma 2 and 3. The first one shows that any optimal mechanism pays a bonus lower than the one that implements the first-best effort, distorting the probability of success downward:

**Lemma 6.** Consider a mechanism $(s, b, e)$ in which $b(θ) \geq \frac{Δx}{1+\lambda}$ for some $θ$. Then, there exists another mechanism that gives the principal a greater payoff than $(s, b, e)$.

**Proof.** The regulator’s payoff from offering contract $(s, b) = (0, \frac{Δx}{1+\lambda})$ to type $θ$ is
\[
x_1 + p^θ_{e*(θ)} \frac{Δx}{1+\lambda} - c^θ_{e*(θ)},
\]
where $e^*(θ) \in \arg \max_{e \in E} x_1 + p^θ_{e} Δx - (1 + \lambda) c^θ_{e}$ is a first-best effort.

We claim that (38) exceeds (37) when $b(θ) \geq \frac{Δx}{1+\lambda}$. To see this, notice that
\[
b(θ) \geq \frac{Δx}{1+\lambda} \iff p^θ_{e*(θ)} \frac{Δx}{1+\lambda} - c^θ_{e*(θ)} \geq p^θ_{e(θ)} \left( Δx - \lambda b(θ) \right) - c^θ_{e(θ)}
\]
\[
\iff p^θ_{e*(θ)} \frac{Δx}{1+\lambda} - c^θ_{e*(θ)} \geq p^θ_{e(θ)} \left( Δx - \lambda b(θ) \right) - c^θ_{e(θ)}
\]
\[
\iff x_1 + p^θ_{e*(θ)} \frac{Δx}{1+\lambda} - c^θ_{e*(θ)} \geq x_1 - \lambda s(θ) + p^θ_{e(θ)} \left( Δx - \lambda b(θ) \right) - c^θ_{e(θ)};
\]
where the first line follows from algebraic manipulations, the second follows from $e^*(θ)$ being optimal for the agent when offered a bonus of $\frac{Δx}{1+\lambda}$, and the third line adds $x_1$ to both sides and subtracts $\lambda s(θ) > 0$ from the expression on the right hand side.
This would conclude the proof if \( (s, b, e) \) is such that \( b(\theta) \geq \frac{\Delta x}{1+\lambda} \) for all \( \theta \). Therefore, suppose there also exist types with \( b(\theta) < \frac{\Delta x}{1+\lambda} \). From the same argument as above, the principal profits from substituting the contract of any type with \( b(\theta) \geq \frac{\Delta x}{1+\lambda} \) by \((0, \Delta x/(1+\lambda))\). We now verify that this substitution is also profitable for those with with \( b(\theta) < \frac{\Delta x}{1+\lambda} \).

Let \( \theta_L \) and \( \theta_H \) be types with \( b(\theta_L) < \frac{\Delta x}{1+\lambda} \leq b(\theta_H) \) and, without loss of generality, suppose that \( s(\theta_H) = 0 \). Then

\[
\begin{align*}
b(\theta_H) &\geq \frac{\Delta x}{1+\lambda} \iff p^\theta_e(\theta_L) \frac{\Delta x}{1+\lambda} - c^\theta_e(\theta_L) \geq p^\theta_e(\theta_L) (\Delta x - \lambda b(\theta_H)) - c^\theta_e(\theta_L) \\
\iff p^\theta_e^r(\theta_L) \frac{\Delta x}{1+\lambda} - c^\theta_e^r(\theta_L) &\geq p^\theta_e(\theta_L) (\Delta x - \lambda b(\theta_H)) - c^\theta_e(\theta_L) \\
\iff p^\theta_e^r(\theta_L) \frac{\Delta x}{1+\lambda} - c^\theta_e^r(\theta_L) &\geq x_1 + p^\theta_e(\theta_L) \frac{\Delta x}{1+\lambda} - c^\theta_e(\theta_L) \geq x_1 - \lambda s(\theta_L) + p^\theta_e(\theta_L) (\Delta x - \lambda b(\theta_L)) - c^\theta_e(\theta_L); \end{align*}
\]

where the first line follows from algebraic manipulations; the second from the fact that \( e^r(\theta_L) \) is the agent’s optimal effort when given a bonus of \( \frac{\Delta x}{1+\lambda} \); the third follows from IC – holding effort \( e(\theta_L) \) fixed, type \( \theta_L \) must prefer \( (s(\theta_L), b(\theta_L)) \) to \((0, b(\theta_H)) \); and the last line follows from algebraic manipulations. Therefore, the regulator profits from substituting the contracts of all types by \((0, \frac{\Delta x}{1+\lambda})\) if there exists a type with \( b(\theta) \geq \frac{\Delta x}{1+\lambda} \).

\[\square\]

The second lemma shows that removing all but the contract with the highest bonus increases the regulator’s payoff:

**Lemma 7.** Let \((s, b, e)\) be a mechanism satisfying IC and LL. There exists a mechanism that offers a single contract \((0, b^*)\) to all types and gives the principal a weakly greater payoff than \((s, b, e)\).

**Proof.** Let \((s, b, e)\) be a mechanism that satisfies IC and LL. By the previous lemma, there is no loss of generality in assuming that \( b(\theta) < \frac{\Delta x}{1+\lambda} \) for all types.

Fix a type \( \theta \in \Theta \). Let \( b^* \equiv \sup \{b(\hat{\theta}) : \hat{\theta} \in \Theta\} \) and \( s^* \equiv \inf \{s(\hat{\theta}) : \hat{\theta} \in \Theta\} \). Because a uniform reduction in fixed payments increases the regulator’s payoff while retaining IC and LL, we can assume, without loss of generality, that \( s^* = 0 \). Moreover, either there exists \( \hat{\theta} \) such that \( s(\hat{\theta}) = 0 \), \( b(\hat{\theta}) = b^* \), or \((0, b^*)\) is a limit point of \( \{(s(\hat{\theta}), b(\hat{\theta})) : \hat{\theta} \in \Theta\}\).

Consider the alternative mechanism that offers the contract \((0, b^*)\) to all types and let

\[
e^r(\theta) \in \arg \max_{e \in E} p^\theta_e b^* - c^\theta_e. \tag{39}\]

Incentive compatibility establishes that \( b^* \geq b(\theta) \) implies \( p^\theta_{e^r(\theta)} \geq p^\theta_{e(\theta)} \). Since, by the previous lemma, \( \frac{\Delta x}{1+\lambda} \geq b^* \), we have

\[
[p^\theta_{e^r(\theta)} - p^\theta_{e(\theta)}] (\Delta x - (1+\lambda)b^*) \geq 0. \tag{40}\]

38
By (39), it follows that
\[ p_{e^*(\theta)} b^* - c_{e^*(\theta)} \geq p_{e(\theta)} b^* - c_{e(\theta)}. \]
Thus, adding \( p_{e^*(\theta)} b^* - c_{e^*(\theta)} - p_{e(\theta)} b^* + c_{e(\theta)} \geq 0 \) to the term on the LHS at (40) establishes that
\[ [p_{e^*(\theta)} - p_{e(\theta)}] (\Delta x - \lambda b^*) + [p_{e(\theta)} - p_{e^*(\theta)}] b^* + p_{e^*(\theta)} b^* - c_{e^*(\theta)} - p_{e(\theta)} b^* + c_{e(\theta)} \geq 0, \]
which can be simplified to
\[ [p_{e^*(\theta)} - p_{e(\theta)}] (\Delta x - \lambda b^*) + c_{e(\theta)} - c_{e^*(\theta)} \geq 0. \quad (41) \]
As in Lemma 3, use IC to write:
\[ s(\theta) + p_{e(\theta)} b(\theta) \geq p_{e(\theta)} b^*. \quad (42) \]
The regulator’s gain from substituting mechanisms equals
\[
p_{e^*(\theta)} (\Delta x - \lambda b^*) - p_{e(\theta)} (\Delta x - \lambda b(\theta)) + \lambda s(\theta) + c_{e(\theta)} - c_{e^*(\theta)} \\
= [p_{e^*(\theta)} - p_{e(\theta)}] (\Delta x - \lambda b^*) + \lambda \left( s(\theta) + p_{e^*(\theta)} (b(\theta) - b^*) \right) + c_{e(\theta)} - c_{e^*(\theta)} \\
\geq [p_{e^*(\theta)} - p_{e(\theta)}] (\Delta x - \lambda b^*) + c_{e(\theta)} - c_{e^*(\theta)} \geq 0,
\]
where the first inequality follows from (42) and the second follows from (40). Thus, that the principal improves by substituting all contracts by \((0, b^*)\).

The proposition follows from Lemma 7.

**Proof of Proposition 3**

**Part a)** Let \((w, e)\) be a feasible mechanism. It is straightforward to adapt Lemma 5 to show that there is no loss of generality in assuming that contracts are uniformly bounded. As in the proof of Theorem 2, let \(\mathcal{M} := \{W^\theta : \theta \in \Theta\}\) denote the set of all contracts in this mechanism, and let \(\overline{\mathcal{M}}\) denote its closure, which is compact.

Let
\[ W^- \in \arg \min_{W \in \overline{\mathcal{M}}} \sum_{i=1}^{N} W_i h_i, \quad (43) \]
and, for each type, let
\[ e^-(\theta) \in \arg \max_{e} \sum_{i=1}^{N} W^-_i p_{e,i}^\theta - c_e^\theta. \quad (44) \]
Existence of $W^-$ and $e^-(\theta)$ follow from the arguments in Theorem 2.

Use IC to write

\[ [I (e(\theta), \theta) - I (e^-(\theta), \theta)] \sum_{i=1}^{N} W_i^\theta h_i \leq c^\theta_{e^-(\theta)} - c^\theta_{e(\theta)} \leq [I (e(\theta), \theta) - I (e^-(\theta), \theta)] \sum_{i=1}^{N} W_i^- h_i. \]  

(45)

Since $W^-$ solves (43), it satisfies

\[ \sum_{i=1}^{N} W_i^- h_i \leq \sum_{i=1}^{N} W_i^\theta h_i, \]

so that, by (45), $I (e^-(\theta), \theta) \geq I (e(\theta), \theta)$. That is, offering $W^-$ yields a distribution of net surplus $x$ that first-order stochastically dominates any other contract in $\mathcal{M}$.

We first show that, holding effort $e(\theta)$ fixed, the regulator’s payoff is higher with contract $W^-$ than with $W^\theta$. Since $W^-$ is the limit of sequence in $\mathcal{M}$, the agent’s utility is continuous, and the original mechanism is incentive compatible, it follows that

\[ \sum_{i=1}^{N} W_i^\theta p_{e(\theta),i}^\theta - c^\theta_{e(\theta)} \geq \sum_{i=1}^{N} W_i^- p_{e(\theta),i} - c^\theta_{e(\theta)} \]

With some algebraic manipulations, we can rewrite this inequality as

\[ \sum_{i=1}^{N} (x_i - \lambda W_i^-) p_{e(\theta),i} - c^\theta_{e(\theta)} \geq \sum_{i=1}^{N} (x_i - \lambda W_i^\theta) p_{e(\theta),i} - c^\theta_{e(\theta)}, \]  

(46)

which shows that, holding effort constant, the regulator obtains a higher payoff with $W^-$ than with $W^\theta$ for $e(\theta)$ fixed.

Next, we show that changing effort from $e(\theta)$ to $e^-(\theta)$ also increases the regulator’s payoff. Let $\Delta W_i^- = W_i^- - W_{i-1}^-$, so that

\[ \sum_{i=0}^{N} \left( \frac{x_i}{1 + \lambda} - W_i^- \right) h_i = -\sum_{i=1}^{N+1} \left( \frac{\Delta x_i}{1 + \lambda} - \Delta W_i^- \right) H(x_i) \leq 0, \]  

(47)

where the equality uses summation by parts, $x_0 = x_1 = W_0^-$, $x_{N+1} = x_N = W_{N+1}^-$, and the inequality uses $H(x) \geq 0$ for all $x$ and FD. Use (47) to obtain:

\[ [I (e(\theta), \theta) - I (e^-(\theta), \theta)] \sum_{i=1}^{N} x_i h_i \geq (1 + \lambda) [I (e(\theta), \theta) - I (e^-(\theta), \theta)] \sum_{i=1}^{N} W_i^- h_i \geq c^\theta_{e^-(\theta)} - c^\theta_{e(\theta)} + \lambda [I (e(\theta), \theta) - I (e^-(\theta), \theta)] \sum_{i=1}^{N} W_i^- h_i, \]  

(48)

where the first inequality follows from $I (e^-(\theta), \theta) \geq I (e(\theta), \theta)$ and some algebraic manipula-
tions, whereas the second inequality follows from the fact that \( e^{-}(\theta) \) maximizes type \( \theta \)'s effort under contract \( W^{-} \) (program 44). Rearranging (48), we obtain:

\[
- \left[ I \left( e^{-}(\theta), \theta \right) - I \left( e(\theta), \theta \right) \right] \sum_{i=1}^{N} (x_{i} - \lambda W_{i}^{-}) h_{i} \geq c_{e^{-}(\theta)} - c_{e(\theta)}.
\]

Using MS, this inequality can be written as

\[
\sum_{i=1}^{N} \left( x_{i} - \lambda W_{i}^{-} \right) p_{e^{-}(\theta),i} - c_{e^{-}(\theta)} \geq \sum_{i=1}^{N} \left( x_{i} - \lambda W_{i}^{-} \right) p_{e(\theta),i} - c_{e(\theta)},
\]

which establishes that the change in effort from \( e(\theta) \) to \( e^{-}(\theta) \) increases the regulator’s payoff. Combining inequalities (46) and (49) concludes the proof.

**Part b)** If \((W, e)\) is an optimal FD mechanism, it must solve the following program

\[
\min_{W} \sum_{i=1}^{N} W_{i} \int_{\Theta} p_{e(\theta),i} d\mu(\theta)
\]

subject to

\[
\sum_{i=1}^{N} W_{i} h_{i} = K, \quad \text{(IC')} \\
W_{i} \geq 0, \quad \text{(LL)} \\
\frac{x_{i} - x_{i-1}}{1 + \lambda} \geq W_{i} - W_{i-1}. \quad \text{(M)}
\]

Using the same arguments as in proof of Theorem 3, it follows that there exists \( \bar{x} \) such that

\[
w(x) = \begin{cases} 
0 & \text{if } x \leq \bar{x} \\
\frac{x - \bar{x}}{1 + \lambda} & \text{if } x > \bar{x}
\end{cases}
\]

Rewriting in terms of costs, we obtain \( w(C) = \max \{ \bar{C} - C; \ 0 \} \), where \( \bar{C} \equiv \frac{S - \bar{x}}{1 + \lambda} \).

**Proof of Proposition 5**

After a normalization, we can assume without loss of generality that \( u(0) = 0 \) and that there exists \( \delta > 0 \) such that \( \delta \geq v'(x) \geq 1 \), for all \( x \geq 0 \). Since \( v'(\cdot) \) is bounded and incentive compatibility and limited liability constraints remain the same after multiplying by a positive constant, we can assume without loss that \( v'(x) \geq 1 \). Notice that if \( v'(0) = 0 \) (or equivalently, \( u'(0) = \infty \)), then we cannot do such normalization.

With some abuse of notation, \( s \) is the utility of the salary and \( b \) is the bonus in utility terms. The proof is based on several lemmas. The first is a generalization of Lemma 2.
**Lemma 8.** Consider a mechanism \((s, b, e)\) in which \(v(s(\theta) + b(\theta)) - v(s(\theta)) > \delta \Delta x\) for some \(\theta\). Then, there exists another mechanism that gives the principal a greater payoff than \((s, b, e)\).

**Proof.** In any optimal mechanism, the limited liability constraint must bind for some type. Otherwise, reducing the fixed payment to all types by a uniform amount maintains feasibility and increases the principal’s payoff.

Let \((s, b, e)\) be an optimal mechanism with \(v(s(\theta) + b(\theta)) - v(s(\theta)) > \delta \Delta x\) for some type \(\theta\). If all types have bonuses greater than the incremental output \(v(s(\theta) + b(\theta)) - v(s(\theta)) > \delta \Delta x\), the principal’s payoff is strictly lower than the one she obtains by offering \(s = b = 0\) to all types, which contradicts optimality.

Suppose there exists a type \(\tilde{\theta}\) who picks a contract with \(v(s(\tilde{\theta}) + b(\tilde{\theta})) - v(s(\tilde{\theta})) \leq \delta \Delta x\). Because LL must bind for some type \(\theta_H\) and the previous argument, this type must pick a contract \((0, b_H)\) with \(b_H > \Delta x\) (otherwise, the contract \((0, b_H)\) with \(b_H < \Delta x\) would be first-order stochastically dominated by the contract that pays \(s(\theta) \geq 0\) and \(\delta b(\theta) \geq v(s(\theta) + b(\theta)) - v(s(\theta)) > \delta \Delta x > \Delta x\), which is offered by assumption). Therefore, the mechanism includes at least one contract \((s_L, b_L)\) with \(v(s_L + b_L) - v(s_L) \leq \delta \Delta x\) and contract \((0, b_H)\) with \(b_H > \Delta x\). We claim that this mechanism gives the principal a lower profit on all types than offering \(s = b = 0\) to all types.

For each \(\theta\), let \(\bar{e}(\theta) \in \arg \min_{e \in E} e^\theta_e\). The principal’s profit from a type \(\tilde{\theta}\) who chooses contract \((s_H, b_H)\) and exerts effort \(e_H \in E\) is

\[x_L + p_{e_H}^\tilde{\theta} (\Delta x - v(b_H)).\]

Replacing this contract by \(s = b = 0\) changes the principal’s payoff to \(x_L + p_{\bar{e}(\tilde{\theta})}^\tilde{\theta} \Delta x\), increasing profits by

\[p_{\bar{e}(\tilde{\theta})}^\tilde{\theta} \Delta x + p_{e_H}^\tilde{\theta} (v(b_H) - \Delta x) > 0,\]

since \(v(b_H) \geq b_H > \Delta x\).

The principal’s profit from a type \(\theta\) who chooses \((s_L, b_L)\) and exerts effort \(e_L \in E\) is

\[x_L - v(s_L) + p_{e_L}^\theta (\Delta x - (v(s_L + b_L) - v(s_L))).\]

By the incentive compatibility constraint of a type who picks \((s_L, b_L)\) and the fact that \(b_H > \Delta x\),

\[p_{e_L}^\theta \Delta x < p_{e_L}^\theta b_H \leq s_L + p_{e_L}^\theta b_L.\]

Adding \(x_L\) to all terms in this inequality and rearranging, gives

\[x_L + p_{e_L}^\theta \Delta x - (v(s_L) + p_{e_L}^\theta (v(s_L + b_L) - v(s_L))) \leq x_L + p_{e_L}^\theta \Delta x - v(s_L + p_{e_L}^\theta b_L) \leq x_L + p_{e_L}^\theta \Delta x - (s_L + p_{e_L}^\theta b_L) < x_L.\]
Add \( p^\theta_{e(\theta)} \Delta x \geq 0 \) to the expression on the right to obtain:

\[
x_L + p^\theta_{e\theta} \Delta x - (v(s_L) + p^\theta_{e\theta}(v(s_L + b_L) - v(s_L))) < x_L + p^\theta_{e(\theta)} \Delta x.
\]

The term on the right is the principal’s profit from the constant-payment contract \((s = b = 0)\) whereas the term on the left is the profit from the original contract. Hence, this replacement also raises profits from any type who chooses a contract with \( v(s(\theta) + b(\theta)) - v(s(\theta)) \leq \delta \Delta x \). \( \square \)

The next lemma shows that the optimal bonus is lower than \( \Delta x \) if it is optimal to offer a unique contract to all types.

**Lemma 9.** Suppose it is optimal to offer the same contract \((s, b)\) to all types. Then \( v(s + b) - v(s) < \Delta x \).

**Proof.** Let \((s, b, e)\) be a mechanism that offers the same contract \( s \geq 0 \) and \( v(s + b) - v(s) \geq \Delta x \) to all types, and consider an alternative mechanism, in which the principal offers \( s = b = 0 \) to all types. Let \( \hat{e}(\theta) \in \arg \min_c c^\theta_e \) denote an incentive-compatible effort under this alternative mechanism. Notice that

\[
x_L - v(s) + p^\theta_{e(\theta)} (\Delta x - (v(s + b) - v(s))) \leq x_L < x_L + p^\theta_{e(\theta)} \Delta x,
\]

where the first inequality uses \( v(s) \geq 0 \) and \( \Delta x \leq v(s+b) - v(s) \) and the second inequality follows from \( p^\theta_e > 0 \) for all \( \theta, e \). But the expression on the left is the principal’s profit in the original mechanism and the one on the right is her profit in the alternative mechanism. Therefore, the principal profits on each type by replacing the mechanism. \( \square \)

Let

\[
\mathcal{V} = \{v : [0, \infty) \to \mathbb{R} \text{ convex, non-decreasing, } \delta \geq v'(x) \geq 1, \forall x \geq 0, \text{ and } v(0) = 0\}.
\]

For each utility function \( v \in \mathcal{V} \) and effort profile \( e \in E^{\Theta} \), define the feasibility correspondence

\[
\Psi(v, e) = \{ (s, b) \in \mathbb{R}^\#_+ \times \mathbb{R}^\#; \forall \hat{e} \in E, \forall \theta, \hat{\theta} \in \Theta, s(\theta) + b(\theta) \geq 0 \text{ and } s(\theta) + p^\theta_{e(\theta)} b(\theta) - \hat{c}^\theta_e \geq s(\hat{\theta}) + p^\theta_{e(\hat{\theta})} b(\hat{\theta}) - \hat{c}^\theta_e \},
\]

and the policy correspondence \( \Gamma(v, e) \subset \Psi(v, e) \) of the cost minimization problem:

\[
\Gamma(v, e) = \arg \min_{(s, b) \in \Psi(v, e)} \sum_{\theta} \mu^\theta \left[ (1 - p^\theta_{e(\theta)}) v(s(\theta)) + p^\theta_{e(\theta)} v(s(\theta) + b(\theta)) \right]
\]

and \( V(v, e) \) its optimal value. Let us consider the uniform approximation in every compact set of \([0, \infty)\) for the space \( \mathcal{V} \), one of the equivalent norms of the Euclidean space \( \mathbb{R}^\#_+ \times \mathbb{R}^\# \) and the discrete topology for \( E^\# \) (i.e., the topology that contains all subset of \( E^\# \)).
Lemma 10. $V$ is a continuous function and $\Gamma$ is a upper semi-continuous at $(v,e)$ such that the interior of $\Psi(v,e)$ is non-empty.

Proof. We can assume without loss of generality that any feasible contract $(s,b)$ belong to $[0,L]\#\Theta \times [-L,L]\#\Theta$, for some $L > 0$. Indeed, by Lemma 8,

$$|b(\theta)| \leq |v(s(\theta) + b(\theta)) - v(s(\theta))| \leq \delta \Delta x.$$ 

Let $(s_0,b_0) \in \Psi(v,e)$ and

$$V_0 = \sum_{\theta} \mu^\theta \left[ (1 - p^\theta e(\theta)) v(s_0(\theta)) + p^\theta e(\theta) v(s_0(\theta) + b_0(\theta)) \right].$$

Therefore, we can restrict the set of feasible mechanism $(s,b)$ that have value lower than or equal to $V_0$. In particular, for such mechanism we have

$$p v(s(\theta) + b(\theta)) \leq V_0,$$

where $p > 0$ is lower bound of the probability. Hence,

$$s(\theta) \leq u \left( \frac{V_0}{p} \right) - b(\theta) \leq u \left( \frac{V_0}{p} \right),$$

which concludes the proof of our claim.

Notice that $\Psi$ is a non-empty and compacted value correspondence. The proof is a straightforward application of the Maximum Theorem (see Berge, 1963). For this, we only have to notice that $\Psi$ is trivially a continuous correspondence and the objective function of the minimization problem is also continuous.

Now we will characterize the optimal mechanism for each effort recommendation.

Implementing high effort from both types

Suppose that the principal wants to implement high effort from both types, i.e., one wants to implement $e(A) = e(B) = H$ with $s^A = s^B = 0$ and $b^A = b^B = b > 0$. Let us denote

$$b^*_A = \frac{c^A_H}{p^A_H - p^A_L}$$

the uniform bonus for types $A$ and $B$.

Lemma 11. If $p^B_H > p^A_L$, then there exists $\epsilon > 0$ such that, whenever $v'(b^*_A) - v'(0) < \epsilon$, offering the contract $(0,b^*_A)$ to all types is optimal.
Proof. The proof is constructive. We start by writing the principal’s cost minimization program:

\[
\min_{s,b} \mu \left[ (1 - p_H^A) v(s^A) + p_H^A v(s^A + b^A) \right] + (1 - \mu) \left[ (1 - p_H^B) v(s^B) + p_H^B v(s^B + b^B) \right]
\]

subject to

\[
b^A \geq \frac{c_H^A}{p_H^A - p_L^A} \tag{50}
\]

\[
s^A + p_H^A b^A \geq s^B + p_H^A b^B \tag{51}
\]

\[
s^A + p_H^A b^A \geq s^B + p_H^A b^B + c_H^A \tag{52}
\]

\[
b^B \geq \frac{c_H^B}{p_H^B - p_L^B} \tag{53}
\]

\[
s^B + p_H^B b^B \geq s^A + p_H^B b^B \tag{54}
\]

\[
s^B + p_H^B b^B \geq s^A + p_H^B b^A + c_H^B \tag{55}
\]

Since (8) holds, inequality (52) is automatically satisfied whenever (50) holds and can be ignored throughout all of our analysis.

We want to obtain conditions for the optimal mechanism to be: \( s^A = s^B = 0 \), \( b^A = b^B = b \). The first-order conditions, already substituting this mechanism and ignoring (18), are:

\[
[s^A] : -\mu \left[ (1 - p_H^A) v'(0) + p_H^A v'(b) \right] + \lambda_2 + \lambda_3 - \lambda_5 - \lambda_6 \leq 0
\]

\[
[s^B] : -(1 - \mu) \left[ (1 - p_H^B) v'(0) + p_H^B v'(b) \right] - \lambda_2 - \lambda_3 + \lambda_5 + \lambda_6 \leq 0
\]

\[
[b^A] : -\mu p_H^A v'(b) + \lambda_1 + \lambda_2 p_H^A + \lambda_3 p_H^A - \lambda_5 p_H^B - \lambda_6 p_H^B = 0
\]

\[
[b^B] : -(1 - \mu) p_H^B v'(b) + \lambda_5 p_H^B + \lambda_6 p_H^B - \lambda_2 p_H^A - \lambda_3 p_L^A = 0
\]

There are two cases to consider: (i) \( p_H^B > p_H^A \) and (ii) \( p_H^B \leq p_H^A \).

Case (i) Take the following values for the multipliers:

\[
\lambda_1 = \left[ \mu p_H^A + (1 - \mu) p_H^B \right] v' \left( \frac{c_H^A}{p_H^A - p_L^A} \right) > 0,
\]

\[
\lambda_2 = (1 - \mu) \left[ v' \left( \frac{c_H^A}{p_H^A - p_L^A} \right) - v'(0) \right] \frac{p_H^B (1 - p_H^B)}{p_H^B - p_H^A} > 0,
\]

\[
\lambda_3 = 0
\]

\[
\lambda_5 = (1 - \mu) \left\{ v' \left( \frac{c_H^A}{p_H^A - p_L^A} \right) + \left[ v' \left( \frac{c_H^A}{p_H^A - p_L^A} \right) - v'(0) \right] \frac{p_H^B (1 - p_H^B)}{p_H^B - p_H^A} \right\} > 0,
\]

\[
\lambda_6 = 0.
\]

Substituting these multipliers in the first-order conditions – and using the convexity of the program – establishes that, for any (weakly concave) utility function, the optimal mechanism is \( s^A = s^B = 0 \) and \( b^A = b^B = \frac{c_H^A}{p_H^A - p_L^A} \).
Case (ii) Take the following values for the multipliers:

\[
\begin{align*}
\lambda_1 &= \mu p_H^A v'(b) + (1 - \mu) p_B^B v'(b) - \frac{(1 - \mu)p_H^A(1 - p_B^B)(p_H^A - p_L^A)}{p_H^A - p_L^B} [v'(b) - v'(0)], \\
\lambda_2 &= 0, \\
\lambda_3 &= \frac{(1 - \mu)p_H^A(1 - p_B^B)}{p_H^A - p_L^B} [v'(b) - v'(0)] > 0, \\
\lambda_5 &= (1 - \mu) \left\{ v'(b) + \frac{p_A^B(1 - p_B^B)}{p_H^A - p_L^B} [v'(b) - v'(0)] \right\} > 0, \\
\lambda_6 &= 0,
\end{align*}
\]

where \( b := \frac{c_A}{p_H^A - p_L^A} \). We will verify that there exists \( \epsilon > 0 \) such that whenever \( v' \left( \frac{c_A}{p_H^A - p_L^A} \right) - v'(0) < \epsilon \), the mechanism above is optimal.

For notational simplicity, let \( \bar{\epsilon} := v' \left( \frac{c_A}{p_H^A - p_L^A} \right) - v'(0) \). We first verify that \( \lambda_1 \geq 0 \), so all the multipliers defined above are non-negative. Rearranging the expression for \( \lambda_1 \) above, gives:

\[
\lambda_1 = \left[ \mu p_H^A + (1 - \mu) p_B^B \right] v'(b) - \frac{(1 - \mu)p_H^A(1 - p_B^B)(p_H^A - p_L^A)}{p_H^A - p_L^B} \bar{\epsilon}
\]

for \( \bar{\epsilon} \approx 0 \).

Next, we verify that the first-order conditions are satisfied. Substituting the multipliers in the first-order condition with respect to \( s^A \) gives

\[
\left\{ -\mu \left[ (1 - p_H^A) v'(0) + p_H^A v'(b) \right] + \frac{(1 - \mu)p_H^A(1 - p_B^B)}{p_H^A - p_L^B} [v'(b) - v'(0)] \\
- (1 - \mu) \left[ v'(b) + \frac{p_A^B(1 - p_B^B)}{p_H^A - p_L^B} (v'(b) - v'(0)) \right] \right\} \leq 0,
\]

which, using the definition of \( \bar{\epsilon} \), becomes

\[-[\mu v'(0) + (1 - \mu)v'(b)] + \left[ \frac{(1 - \mu)(1 - p_B^B)}{p_H^B - p_L^A} (p_H^B - p_L^A) - \mu p_H^A \right] \bar{\epsilon} \leq 0.\]

Since the first term in the left hand side is strictly negative whereas the second term is linear in \( \bar{\epsilon} \), there exists \( \epsilon > 0 \) such that the left hand side is negative whenever \( \bar{\epsilon} < \epsilon \), so that the first-order conditions with respect to \( s^A \) holds (with inequality). Substitution establishes that the first-order condition with respect to \( s^B \), \( b^A \), and \( b^B \) all hold with equality for any \( \bar{\epsilon} \). To summarize, there exists \( \epsilon > 0 \) such that offering the contract above is optimal whenever \( v' \left( \frac{c_A}{p_H^A - p_L^A} \right) - v'(0) < \epsilon \). \( \square \)

Implementing high effort from one type and low effort from the other

Suppose now that the principal wants to implement high effort from one type and low effort from the other with uniform contract, i.e., \( s^A = s^B = 0 \) and \( b^A = b^B = b > 0 \). We divide the analysis in two cases: (i) \( e(A) = L \) and \( e(B) = H \); (ii) \( e(A) = H \) and \( e(B) = L \).
Case (i): \( e(A) = L \) and \( e(B) = H \)

Let us denote

\[
    b_B^* = \frac{c_B^H}{p_B^H - p_B^L}
\]

the uniform bonus for types A and B.

**Lemma 12.** If \( p_A^L > p_B^L \), then there exists \( \epsilon > 0 \) such that, whenever \( v'(b_B^*) - v'(0) < \epsilon \), offering the contract \((0, b_B^*)\) to all types is optimal.

**Proof.** The proof is analogous to the proof of Lemma 11.

Case (ii): \( e(A) = H \) and \( e(B) = L \)

We first claim that the effort profile \( e^A = 1 \) and \( e^B = 0 \) cannot be optimal for sufficiently low risk aversion. For this, let us first show the following lemma.

**Lemma 13.** The optimal mechanism must offer a different contract to each type.

**Proof.** Notice that for the uniform contract \((0, b)\) to be incentive compatibility we need that

\[
    b \geq \frac{c_A^H}{p_A^H - p_A^L} > \frac{c_B^H}{p_B^H - p_B^L} \geq b,
\]

where the strict inequality is (8). This is a contradiction.

Suppose, by contradiction, that the above claim is not true. Then, we can find a sequence of utility function \( v_n \) converging to the identity function with the respective non-uniform optimal contracts \((s_A^n, s_B^n, b_A^n, b_B^n)\). By Lemma 10, this sequence of contracts converge to an optimal contract \((s_A^0, s_B^0, b_A^0, b_B^0)\) of the model with risk neutrality. By Lemma 13 and its proof, we have

\[
    b_A^0 \geq \frac{c_A^H}{p_A^H - p_A^L} > \frac{c_B^H}{p_B^H - p_B^L} \geq b_B^0,
\]

which is contradiction since this contract is strictly dominated by a uniform contract under risk neutrality.

**Implementing low effort from both types**

Suppose that the principal wants to implement high effort from both types, i.e., one wants to implement \( e(A) = e(B) = L \). It is trivial that optimal contract will the uniform null contract: \( s^A = s^B = 0 \) and \( b^A = b^B = 0 \).
Proof of (i) and (ii)

(i) If the optimal effort profile is $e(A) = e(B) = L$, offering the constant payment of zero to both types is optimal. If the optimal effort profile $e(A) = e(B) = H$ or $e(A) = L$ and $e(B) = H$, we only need to apply Lemmas 9, 10, 11, 12 and 13.

(ii) Follows from Lemma 10.

Proof of Proposition 4

We first introduce some notation. Let $P$ denote the space of distributions satisfying $p_{\theta}^{\theta}(x_i) > \underline{p}$ for some $\underline{p} > 0$ (consistently with Assumption 2), let $\# \Theta$ denote the the number of elements in $\Theta$ and, for notational clarity, let $\Delta_{\gamma,\hat{\gamma}} := c_{\gamma} - c_{\hat{\gamma}}$. For the (finite-dimensional) distribution $(p_{\theta,i}^{\theta})$ and cost function $(c_{\gamma}^{\theta})$, take one of the equivalent norms of the Euclidean space.

Let $\Psi : P \mapsto E^{\# \Theta} \times \mathbb{R}^{\# \Theta \times N}$ denote the feasibility correspondence:

$$
\Psi(p) := \left\{ (\hat{e}, \hat{w}) \in E^{\# \Theta} \times \mathbb{R}^{\# \Theta \times N} : \forall \hat{\gamma}, \hat{\theta} \in \Theta, \sum_{i=1}^{N} \left[ p_{\hat{\gamma},i}^{\theta}(\tau_{i}^{\hat{\gamma}}) \hat{w}_{i}^{\hat{\gamma}} - p_{\hat{\gamma},i}^{\hat{\theta}}(\tau_{i}^{\hat{\gamma}}) \hat{w}_{i}^{\hat{\theta}} \right] \geq \Delta_{\gamma,\hat{\gamma}} \right\},
$$

that is, the set of incentive compatible mechanisms under $p$. Let $\Gamma : P \mapsto E^{\# \Theta} \times \mathbb{R}^{\# \Theta \times N}$ denote the policy correspondence of the principal’s program:

$$
\Gamma(p) = \arg \max_{(\hat{e}, \hat{w}) \in \Psi(p)} \sum_{\theta} \mu^{\theta} \sum_{i=1}^{N} p_{\hat{\gamma},i}^{\theta}(x_i - \hat{w}_{i}^{\theta}),
$$

and let $V : P \rightarrow \mathbb{R}$ denote its optimal value:

$$
V(p) = \sup_{(\hat{e}, \hat{w}) \in \Psi(p)} \sum_{\theta} \mu^{\theta} \sum_{i=1}^{N} p_{\hat{\gamma},i}^{\theta}(x_i - \hat{w}_{i}^{\theta}).
$$

Throughout the proof, we take one of the equivalent norms of Euclidean space for contracts $w \in \mathbb{R}^{\# \Theta \times N}$ and the discrete topology for efforts $e \in E^{\# \Theta}$ (i.e., the topology that contains all subsets of $E^{\# \Theta}$).

**Lemma 14.** $V$ is a continuous function and $\Gamma$ is an upper semi-continuous correspondence at any $p$ for which the interior of $\Psi(p)$ is non-empty.

**Proof.** As shown in the existence part of the proof of Theorem 2, we can assume, without loss of generality, that all feasible contracts belong to $[0, L]^{\# \Theta \times N}$ for some $L > 0$. Notice that $\Psi$ is a nonempty, compact valued correspondence. The proof proceeds by verifying the conditions for the Maximum Theorem (Berge, 1963). For this, we need to show that $\Psi$ is a continuous correspondence.
(a) $\Psi$ is upper semi-continuous (u.s.c.): Let $(p^n)$ be a sequence of distributions in $\mathcal{P}$ converging to $p$. For any $(\bar{e}_n, \bar{w}_n) \in \Psi(p_n)$, the finiteness of $E$ and $\Theta$, the compactness of $[0,L]^N$ (and passing to a convergent subsequence if necessary), we can suppose that $(\bar{e}_n, \bar{w}_n)$ converges to $(\hat{e}, \hat{w}) \in E^\# \times [0,L]^\# \times N$. By the continuity of the objective function and the constraints of the maximization problem that defines $\Gamma$, we have that $(\hat{e}, \hat{w}) \in \Psi(p)$. Therefore, $\Psi$ is u.s.c.

(b) $\Psi$ is lower semi-continuous (l.s.c.): Let $(p^n)$ be a sequence of distributions in $\mathcal{P}$ that converges to $p$, and let $(\bar{e}_n, \bar{w}_n) \in \Psi(p_n)$ be such that

$$\sum_{i=1}^{N} \left[ p^n_{\bar{e}(\theta),i} w^n_i - p^n_{\bar{e}(\theta),i} \tilde{w}^n_i \right] > \Delta \epsilon_{\bar{e}(\theta),\hat{e}}, \text{ for all } \hat{e} \in E \\setminus \{ \bar{e}(\theta) \} \text{ and all } \hat{\theta} \neq \theta.$$ 

Then, for $n$ sufficiently large we have that the previous inequality is also true for $p^n$ instead of $p$. This implies that the constant sequence $(\bar{e}_n, \bar{w}_n) \in \Psi(p_n)$ converges to $(\bar{e}, \bar{w}) \in \Psi(p)$, which shows that $\Psi$ is l.s.c.

By the Maximum Theorem, since the objective function of the maximization problem in $\Gamma$ is continuous, $V$ is a continuous function and $\Gamma$ is u.s.c.

Lemma 15. For almost all distributions in $\mathcal{P}$ that satisfy MS for which the optimal contract is non-null, the state at which the optimal contract pays a positive amount is unique.

Proof. Fix $p \in \mathcal{P}$ that satisfies MS and for which the optimal contract is non-null. By Proposition 1, each solution of the principal-agent problem can be represented by a triple $(i, w, e)$ defined by the state $i$ at which the contract pays a positive amount $w \geq 0$ and by the recommended effort profile $e$. Denote $\mathcal{S}$ the set of all these feasible triples. Fix $(i^*, w^*, e^*) \in \mathcal{S}$ such that

$$w^* \in \text{argmax } \{ w; (i, w, e) \in \mathcal{S} \text{ is a solution of the principal-agent problem } \}.$$ 

By our assumption, $w^* > 0$. Suppose without loss of generality that $h_{i^*} > 0$ (if $h_{i^*} < 0$, the proof is analogous). The proof has several steps.

a) Constructing the distribution for which all optimal contracts pay a positive amount at a unique state: Let $\epsilon > 0$ be sufficiently small and define the distribution $\tilde{p} \in \mathcal{P}$ equals to $p$ except at $(\theta, e^*(\theta))$ where we have

$$\tilde{p}_{\theta,i} = \begin{cases} p^n_{\theta,i} - \epsilon & \text{if } i = i^* \\ p^n_{\theta,i} + \epsilon & \text{if } h_i \leq 0 \end{cases},$$

and $\tilde{p}$ satisfies MS with the same $h$ and $I$ that defines $p$, for all $\theta$, where $k \geq 1$ is the cardinality of $\{i; h_i \leq 0\}$.

b) $(i^*, w^*, e^*)$ is an optimal solution for the principal-agent problem with $\tilde{p}$: For any contract $w = (w_1, ..., w_N)$ and any effort recommendation profile $e(\theta)$, the incentive compatibility
The principal’s payoff at \((i^*, w^*, e^*)\) satisfies

\[
\sum_\theta \mu^\theta \left[ \sum_{i=1}^N p^\theta_{e^*(\theta),i} x_i - p^\theta_{e^*(\theta),i^*} w^* \right] \geq \sum_\theta \mu^\theta \left[ \sum_{i=1}^N p^\theta_{\hat{e}(\theta),i} x_i - p^\theta_{\hat{e}(\theta),i^*} \hat{w} \right],
\]

for each \((\hat{i}, \hat{w}, \hat{e}) \in S\). Using MS, this last condition is equivalent to

\[
\sum_\theta \mu^\theta \left[ I(\hat{e}(\theta), \theta) - I(e^*(\theta), \theta) \right] \sum_{i=1}^N h_i x_i + \sum_\theta \mu^\theta \left[ p^\theta_{\hat{e}(\theta),i} \hat{w} - p^\theta_{e^*(\theta),i^*} w^* \right] \geq 0. \tag{56}
\]

If \(\hat{i} \neq i^*\), substituting \(\hat{p}\) for \(p\), the inequality in (56) becomes strict, i.e.,

\[
\sum_\theta \mu^\theta \left[ I(\hat{e}(\theta), \theta) - I(e^*(\theta), \theta) \right] \sum_{i=1}^N h_i x_i + \sum_\theta \mu^\theta \left[ p^\theta_{\hat{e}(\theta),i} \hat{w} - p^\theta_{e^*(\theta),i^*} w^* \right] > 0,
\]

because if \(h_i \leq 0\)

\[
\hat{p}^\theta_{\hat{e}(\theta),i} \hat{w} - \hat{p}^\theta_{e^*(\theta),i^*} w^* = \left( \hat{p}^\theta_{e^*(\theta),i} + (I(e^*(\theta), \theta) - I(\hat{e}(\theta), \theta))h_i \right) \hat{w} - \hat{p}^\theta_{e^*(\theta),i^*} w^*
= \left( p^\theta_{e^*(\theta),i} + \frac{\epsilon}{k} + (I(e^*(\theta), \theta) - I(\hat{e}(\theta), \theta))h_i \right) \hat{w} - \left( p^\theta_{e^*(\theta),i^*} - \epsilon \right) w^*
> p^\theta_{\hat{e}(\theta),i} \hat{w} - p^\theta_{e^*(\theta),i^*} w^*,
\]

and if \(h_i > 0\)

\[
\hat{p}^\theta_{\hat{e}(\theta),i} \hat{w} - \hat{p}^\theta_{e^*(\theta),i^*} w^* = p^\theta_{\hat{e}(\theta),i} \hat{w} - \left( p^\theta_{e^*(\theta),i^*} - \epsilon \right) w^*
= p^\theta_{\hat{e}(\theta),i} \hat{w} - p^\theta_{e^*(\theta),i^*} w^* + \epsilon w^*
> p^\theta_{e^*(\theta),i} \hat{w} - p^\theta_{e^*(\theta),i^*} w^*,
\]

where we have used MS and the definition of \(\hat{p}\).

If \(\hat{i} = i^*\), substituting \(\hat{p}\) for \(p\), the inequality (56) is equivalent to

\[
\sum_\theta \mu^\theta \left[ I(\hat{e}(\theta), \theta) - I(e^*(\theta), \theta) \right] \sum_{i=1}^N h_i x_i + \sum_\theta \mu^\theta \left[ p^\theta_{\hat{e}(\theta),i} \hat{w} - p^\theta_{e^*(\theta),i^*} w^* \right] + \epsilon (w^* - \hat{w}) \geq 0,
\]

where we have used MS and the definitions of \(\hat{p}\) and \(w^*\).

Therefore, under \(\hat{p}\), \((i^*, w^*, e^*)\) provides strict higher payoff for the principal than \((\hat{i}, \hat{w}, \hat{e}),\)
for all \( i \neq i^* \).

c) \( i^* \) is the unique state at which the optimal contract pays a positive amount: From the proof of Proposition 1, state \( i^* \) is characterized by

\[
\sum_{\theta} \mu^\theta P_{e^*(\theta),i^*} \frac{h_i}{h_{i^*}} \leq \sum_{\theta} \mu^\theta P_{e^*(\theta),i} \frac{h_i}{h_{i^*}}, \text{ for all } i \neq i^*.
\]

Replacing \( p \) by \( \tilde{p} \), these inequalities become

\[
\sum_{\theta} \mu^\theta \tilde{P}_{e^*(\theta),i^*} \frac{h_i}{h_{i^*}} - \sum_{\theta} \mu^\theta \tilde{P}_{e^*(\theta),i} \frac{h_i}{h_{i^*}} \leq \sum_{\theta} \mu^\theta P_{e^*(\theta),i^*} \frac{h_i}{h_{i^*}} - \sum_{\theta} \mu^\theta P_{e^*(\theta),i} - \frac{h_i}{h_{i^*}} \varepsilon < 0, \tag{57}
\]

for all \( i \neq i^* \). Therefore, \( i^* \) is the unique state at which the optimal contract pays a positive amount at the optimal solution \((i^*, w^*, e^*)\) with distribution \( \tilde{p} \).

Let \((i^*, \tilde{w}, \tilde{e})\) be another solution of the principal-agent problem for \( \tilde{p} \). Again from Proposition 1, \( i^* \) is the unique state at which this optimal contract pays a positive amount if and only if

\[
\frac{1}{h_{i^*}} \sum_{\theta} \mu^\theta \tilde{P}_{e^*(\theta),i^*} < \frac{1}{h_{i}} \sum_{\theta} \mu^\theta P_{e^*(\theta),i}, \text{ for all } i \neq i^*.
\]

Adding \( \sum_{\theta} \mu^\theta (I(e^*(\theta), \theta) - I(\tilde{e}(\theta), \theta)) \) on both sides and using MS, this last inequality is equivalent to (57). Therefore, at any optimal solution with distribution \( \tilde{p} \), \( i^* \) is the unique state at which the optimal contract pays a positive amount.

d) The set of distributions in \( P \) that satisfy MS and the state at which any optimal contract pays a positive amount is unique is generic. Indeed, from steps (a)-(c), for every neighborhood of a distribution in \( P \) that satisfies MS, there exists a distribution in \( P \) that satisfies MS and for which this uniqueness property holds. Moreover, it is straightforward to see that the set of distributions satisfying the uniqueness property is open in the set of distributions in \( P \) that satisfy MS. This step concludes the proof.

Lemma 16. For almost all distributions in \( P \) that satisfy MS and for almost all cost functions, there is a neighborhood in \( P \) for which the optimal mechanism is implemented by only one contract that pays a positive amount in only one state.

Proof. Let \( p \in P \) satisfying MS.

Claim 1. There exists a neighborhood \( \mathcal{N} \) of \( p \) such that each optimal effort recommendation profile under the distribution \( \tilde{p} \in \mathcal{N} \) is also an optimal effort recommendation profile under the distribution \( p \). Otherwise, there exists a sequence \((\tilde{p}_n, \tilde{e}_n, \tilde{w}_n) \in \Gamma(\tilde{p}_n)\) such
that \( \tilde{e}_n \) is not an optimal effort profile under \( p \), for all \( n \). By Lemma 14, \((\tilde{e}_n)\) converges to some \( \tilde{e} \) which is an optimal effort recommendation profile under \( p \). Since \( E \) and \( \Theta \) are finite sets, for sufficiently large \( n \), \( \tilde{e}_n = \tilde{e} \), which is a contradiction.

Let \((i^*, w^*, e^*)\) be any solution of the principal-agent problem (we are using the same notation of the proof of Lemma 16). Notice that incentive compatibility at this optimal solution is equivalent to

\[
[I(\tilde{e}, \theta) - I(e^*(\theta), \theta)] h_{i^*} w^* \geq \Delta c_{e^*(\theta), \tilde{e}},
\]

for all \( \theta \) and \( \tilde{e} \).

Claim 2. For a generic set of cost functions, the inequality (58) is strict, i.e., \( \Psi(p) \) has non-empty interior. Indeed, let \( \epsilon > 0 \) be sufficiently small. Let us define the cost function \( \tilde{c}^\theta \) which is exactly the same as \( c^\theta \) except at \( \tilde{e} \) and \( \theta \) for which the constraint (58) binds where we define

\[
\tilde{c}^\theta_{\tilde{e}} = c^\theta_{\tilde{e}} \begin{cases} +\epsilon & \text{if } \Delta c^\theta_{e^*(\theta), \tilde{e}} \geq 0 \\ -\epsilon & \text{if } \Delta c^\theta_{e^*(\theta), \tilde{e}} < 0. \end{cases}
\]

For the model with cost function \( \tilde{c} \) and distribution \( p \), the constraint (58) is slack for all \( \tilde{e} \) and \( \theta \), i.e., \( \Psi(p) \) has non-empty interior.

There are two cases to be considered:

(i) The optimal contract is null, i.e., \( w^* = 0 \) and the inequality (58) is slack for all \( \theta \) and \( \tilde{e} \). Therefore, the interior of \( \Psi(p) \) is non-empty. By Lemma 14 and the claim 2, there exists a neighborhood \( \mathcal{N} \) of \( p \) such that \((0, e^*) \in \Gamma(\tilde{p})\), for all \( \tilde{p} \in \mathcal{N} \), where \( e^* \) is an optimal effort recommendation profile under \( p \).

(ii) The optimal contract is positive, i.e., \( w^* > 0 \). By Lemma 16, we can generically assume that state \( i^* \) at which the optimal contract pays a positive amount is unique among all possible solutions. Without loss of generality, let us assume that \( h_{i^*} > 0 \).

Under \( \tilde{p} \in \mathcal{N} \), define the relaxed version of the cost minimization program of implementing an optimal effort recommendation profile \( e^*(\theta) \):

\[
\min_{(w^\theta)} \sum_{\theta} \mu^\theta \sum_{i=1}^N \tilde{p}^\theta_{e^*(\theta), i} w_i^\theta \\
\text{s.t.} \quad \sum_{i=1}^N \tilde{p}^\theta_{e^*(\theta), i} w_i^\theta - \Delta c^\theta_{e^*(\theta), \tilde{e}} \geq 0, \forall \theta, \tilde{e} \\
\sum_{i=1}^N (w_i^\theta - w_i^\theta) \geq 0, \forall \theta \neq \theta \\
w^\theta \geq 0, \forall \theta.
\]

By the choice of the cost function and taking a smaller neighborhood \( \mathcal{N} \), there exists a mechanism formed by only one contract paying a positive amount at state \( i^* \) which is feasible for the above program. Taking an even smaller \( \mathcal{N} \) if necessary, since \( \Gamma \) is u.s.c. at \( p \) and the optimal contract
under $p$ is non-null, the optimal contract under $\tilde{p}$ is also non-null, for all $\tilde{p} \in \mathcal{N}$. Restricting to mechanisms with only one contract paying a positive amount at state $i^*$, let $\theta^*$ and $\tilde{e}^*$ be a type and an effort recommendation at which the first constraint of the above program binds at the optimal contract. Hence, we must have $\tilde{p}^{\theta^*,i^*}_c - \tilde{p}^{\theta^*,i^*}_{\tilde{e}^*} \neq 0$.

The necessary and sufficient first-order conditions of the linear Lagrangian are

$$\mu^\theta \tilde{p}^{\theta}_{e^*(\theta)} - \sum_{\tilde{e}} \tilde{\lambda}^{\theta,\tilde{e}} \left[ \tilde{p}^{\theta}_{e^*(\theta)} - \tilde{p}^{\theta}_{\tilde{e}} \right] + \sum_{\theta \neq \theta} \tilde{\sigma}^{\theta,\tilde{\theta}} \tilde{p}^{\theta}_{e^*(\theta)} - \left( \sum_{\theta \neq \theta} \tilde{\sigma}^{\theta,\tilde{\theta}} \right) \tilde{p}^{\theta}_{e^*(\theta)} - \tilde{\gamma}^\theta = 0.$$ 

Let us build multipliers $\lambda$, $\sigma$ and $\gamma$ for which the mechanism formed by one contract paying a positive amount at state $i^*$ is a critical point of the Lagrangian. For this, let us define the following continuous mapping on $\mathcal{N}$:

$$\tilde{\lambda}^{\theta,\tilde{e}}(\tilde{p}) = \begin{cases} \sum_{\theta} \mu^\theta \tilde{p}^{\theta}_{e^*(\theta),i^*} - \tilde{p}^{\theta}_{e^*(\theta),i^*} & \text{if } \theta = \theta^* \text{ and } \tilde{e} = \tilde{e}^* \\ 0 & \text{otherwise} \end{cases}$$

and

$$\tilde{\sigma}^{\theta,\tilde{\theta}} = \begin{cases} \mu^\theta & \text{if } \tilde{\theta} = \theta^* \text{ and } \theta \neq \theta^* \\ 0 & \text{otherwise} \end{cases}.$$ 

For $\theta = \theta^*$, the first-order condition at state $i$ becomes

$$\sum_{\tilde{e}} \tilde{\lambda}^{\theta^*,\tilde{e}}(\tilde{p}) \left[ \tilde{p}^{\theta^*}_{e^*(\theta^*)} - \tilde{p}^{\theta^*}_{\tilde{e},i^*} \right] = \tilde{\gamma}^{\theta^*}.$$ 

By the definition of $\tilde{\lambda}$, we have $\tilde{\gamma}^{\theta^*} = 0$.

For $\theta \neq \theta^*$, the first-order condition becomes

$$\tilde{\gamma}^\theta = \mu^\theta \tilde{p}^{\theta}_{e^*(\theta)} - \left( \sum_{\theta \neq \theta} \tilde{\sigma}^{\theta,\tilde{\theta}} \right) \tilde{p}^{\theta}_{e^*(\theta)} = 0,$$

where the last equality is a consequence of the definition of $\tilde{\sigma}$.

Notice that the only remaining slackness condition that must be checked is $\tilde{\gamma}^{\theta^*} \geq 0$. For this, define the mapping $T : \Delta^\#\Theta \times \#E \to \mathbb{R}^N$ by

$$T(\tilde{p}) = \left( \sum_{\theta} \mu^\theta \tilde{p}^{\theta}_{e^*(\theta),i} - \sum_{\tilde{e}} \tilde{\lambda}^{\theta^*,\tilde{e}}(\tilde{p}) \left[ \tilde{p}^{\theta^*}_{e^*(\theta^*)} - \tilde{p}^{\theta^*}_{\tilde{e},i} \right] \right),$$

where $\Delta$ is the $(N-1)$-dimensional simplex. We have that $T$ is a continuous mapping, $T(p)_{i^*} = 0$.
and
\[ T(p)_i = \sum_{\theta} \mu_i \hat{p}_{c^*(\theta),i} - \sum_{\theta} \mu_i \hat{p}_{c^*(\theta),i} \frac{h_{i_i}}{h_{i^*_i}} > 0, \]
for all \( i \neq i^* \). By the continuity of \( T \), we can find an even smaller \( \mathcal{N} \) such that \( T(\tilde{p})_{i^*} = 0 \) and \( T(\tilde{p})_i > 0 \), for all \( i \neq i^* \) and \( \tilde{p} \in \mathcal{N} \).

B. Examples

Screening with Pure Adverse Selection (Example 1 - continuation)

Let \( w^i_j \) denote type \( i \)'s payment in state \( j \). The optimal mechanism must solve the following program:

\[
\begin{align*}
\min_{w^i_H, w^i_L \geq 0} & \quad 2w^A_H + w^A_L + w^B_H + 2w^B_L \\
\text{subject to} & \quad 2w^A_H + w^A_L \geq 2w^B_H + w^B_L \\
& \quad w^B_H + 2w^B_L \geq w^A_H + 2w^A_L \\
& \quad 2w^A_H + w^A_L \geq 3 \\
& \quad w^B_H + 2w^B_L \geq 2.
\end{align*}
\]

The two first constraints require \( A \) and \( B \) to prefer to report their types truthfully (IC constraints). Since effort is observable, the principal does not need to worry about deviations on effort. Then, it is no longer the case that LL implies IR. The last two constraints are precisely the IR constraints. Disregarding the IC constraints, the IR constraints must bind for both types at optimal mechanism, which gives the one described in the text. It is straightforward to check that the IC constraints are satisfied for this mechanism.

Example 2 (continuation)

We will calculate the cheapest way of implementing high effort from all types. Consider the relaxed program:

\[
\begin{align*}
\min_{w^i_j \geq 0} & \quad \sum_{i=1}^{N-1} \frac{\mu_i}{2} \left( w^i_{i+1} + \frac{\sum_{j \neq i+1} w^i_j}{N-1} \right) \\
\text{subject to} & \quad \frac{1}{2} \left( w^i_{i+1} + \frac{\sum_{j \neq i+1} w^i_j}{N-1} \right) - \frac{N-2}{N-1} \geq \frac{1}{2} \left( w^i_j + \frac{\sum_{j \neq 1} w^i_j}{N-1} \right)
\end{align*}
\]

for \( i = 1, \ldots, N - 1 \). This is a relaxed program in that it only considers the “pure moral hazard” ICs. The unique solution is

\[
w^i_j = \begin{cases} 
2 & \text{if } j = i + 1 \\
0 & \text{if } j \neq i + 1
\end{cases}.
\]
The relaxed program omitted the ICs that require that each type does not benefit by picking a different contract while exerting high effort. However, those constraints are satisfied at the solution obtain above, since:

\[
0 = \frac{1}{2} \cdot 2 - \frac{N - 2}{N - 1} > \frac{1}{2(N - 1)} \cdot 2 - \frac{N - 2}{N - 1}.
\]

Thus, all omitted ICs are satisfied.

Sufficient conditions for it to be optimal to recommend high effort to all types are:

\[
x_{i+1} - 2 + \frac{1}{N - 1} \sum_{j \neq i+1} x_j > x_1 + \frac{1}{N - 1} \sum_{j \neq 1} x_j
\]

or

\[
x_{i+1} - 2 \frac{N - 1}{N - 2} > x_1
\]

for all \(i = 1, \ldots, N - 1\).

C. Multiplicative Separability

This appendix examines the multiplicative separability condition (MS). We provide two results. The first is a characterization of MS in terms of the ordering of incentives. The second one is the graphical interpretation given in the text regarding how distributions satisfying MS change with effort.

Recall the condition presented in the text, which allows contracts to be ranked by their incentives:

**Definition 3.** A distribution of outputs is *ordered in terms of incentives* if, for some \(\theta_0 \in \Theta\), \(e, \tilde{e} \in E\), \(w\) and \(\tilde{w}\) satisfying LL,

\[
\sum_{i=1}^{N} w_i \left[ p_{e,i}^{\theta_0} - p_{\tilde{e},i}^{\theta_0} \right] = \sum_{i=1}^{N} \tilde{w}_i \left[ p_{e,i}^{\theta_0} - p_{\tilde{e},i}^{\theta_0} \right],
\]

then

\[
\sum_{i=1}^{N} w_i \left[ p_{e,i}^{\theta} - p_{\tilde{e},i}^{\theta} \right] = \sum_{i=1}^{N} \tilde{w}_i \left[ p_{e,i}^{\theta} - p_{\tilde{e},i}^{\theta} \right]
\]

for all \(\theta \in \Theta\).

Substitution shows that any distribution that satisfies MS must be ordered in terms of incentives. The next proposition shows that the reverse is also true, so these are equivalent conditions:
Proposition 6. The following statements are equivalent:

1. The distribution of outputs $p^\theta_e(x)$ is ordered in terms of incentives.
2. The distribution of outputs $p^\theta_e(x)$ satisfies MS.
3. For all $i, j$, there exist constants $\phi_{i,j}$ (that depend on the outputs $x_i$ and $x_j$ but not on type or effort) such that

$$\frac{p^\theta_{e,i} - p^\theta_{e,j}}{p^\theta_{e,j} - p^\theta_{e,j}} = \phi_{i,j}$$

whenever $p^\theta_{e,j} - p^\theta_{e,j} \neq 0$.

Proof. We first show that (1) implies (2). Define the linear functional $\varphi^{\theta,e,\tilde{e}} : \mathbb{R}^N \to \mathbb{R}$ as

$$\varphi^{\theta,e,\tilde{e}}(w) = \sum_{i=1}^{N} w_i [p^\theta_{e,i} - p^\theta_{e,i}].$$

Since $w(x) = w_+(x) - w_-(x)$, where $w_+(x) = \max \{ w(x), 0 \}$ and $w_-(x) = \max \{ -w(x), 0 \}$, Definition 3 implies that for some $\theta_0 \in \Theta$ and some $e, \tilde{e} \in E$,

$$\varphi^{\theta_0,e,\tilde{e}}(w) = 0 \text{ if and only if } \varphi^{\theta,e,\tilde{e}}(w) = 0, \forall \theta \in \Theta.$$

Hence, functionals $\varphi^{\theta,e,\tilde{e}}$ are equivalent for all $\theta \in \Theta$ and $e, \tilde{e} \in E$, i.e., there exist constants $\lambda^{\theta,e,\tilde{e}} \in \mathbb{R}$ and a linear functional $\tilde{\varphi}$ in $\mathbb{R}^N$ such that $\varphi^{\theta,e,\tilde{e}} = \lambda^{\theta,e,\tilde{e}} \tilde{\varphi}$. Indeed, we have that the null spaces of $\varphi^{\theta,e,\tilde{e}}$ are all the same, which we denote by $\mathcal{N}$. By the Rank-Nullity Theorem, there exists $v \in \mathbb{R}^N \setminus \mathcal{N}$ such that $\mathbb{R}^N = [v] \oplus \mathcal{N}$, where $[v]$ is the subspace generated by vector $v$ and $\oplus$ represents the direct sum between vector spaces. Let $\tilde{\varphi}$ be the unique linear functional such that $\tilde{\varphi}(v) = 1$ and $\tilde{\varphi}(n) = 0$ for all $n \in \mathcal{N}$. Hence,

$$\lambda^{\theta,e,\tilde{e}} = \varphi^{\theta,e,\tilde{e}}(v) = \sum_{i=1}^{N} v_i [p^\theta_{e,i} - p^\theta_{e,i}].$$

Defining $I(e, \theta) = -\sum_{i=1}^{N} v_i p^\theta_{e,i}$ and $h = \tilde{\varphi}$, we have that

$$p^\theta_{e} - p^\theta_{e} = \varphi^{\theta,e,\tilde{e}} = \lambda^{\theta,e,\tilde{e}} \tilde{\varphi} = [I(\tilde{e}, \theta) - I(e, \theta)] h,$$

which implies MS.

Next, we show that (2) implies (3). From Definition 1, a distribution satisfies MS if and only if, for all $(x, \theta, e, \tilde{e})$,

$$p^\theta_{e}(x) - p^\theta_{e}(x) = [I(\tilde{e}, \theta) - I(e, \theta)] h(x).$$

Since (2) and (3) trivially hold if $p^\theta_{e,i} = p^\theta_{e,i}$ for all $i, \theta, e$, and $\tilde{e}$, it suffices to consider the case where $p^\theta_{e,j} \neq p^\theta_{e,j}$ for some $(j, \theta, e, \tilde{e})$. 

56
Suppose (2) holds. Then, we must have \( h_j \neq 0 \). Multiplying both sides of equation (60) by \( \frac{h_i}{h_j} \), we obtain:

\[
[p^\theta_{e,j} - p^\theta_{\tilde{e},j}] \frac{h_i}{h_j} = [I(\tilde{e}, \theta) - I(e, \theta)] h_i = p^\theta_{e,i} - p^\theta_{\tilde{e},i}.
\]

Setting \( \phi_{i,j} \equiv \frac{h_i}{h_j} \) establishes the equation in the statement.

Finally, we establish that (3) implies (1). To do so, suppose that (3) holds, and let \( w \) and \( \tilde{w} \) be two payment vectors satisfying LL such that for some \( \theta_0 \in \Theta \)

\[
\sum_{i=1}^{N} w_i [p^{\theta_0}_{e,i} - p^{\theta_0}_{\tilde{e},i}] = \sum_{i=1}^{N} \tilde{w}_i [p^{\theta_0}_{e,i} - p^{\theta_0}_{\tilde{e},i}] \]

(61) for all \( e, \tilde{e} \in E \). Without loss of generality, suppose that \( p^{\theta_0}_{e_0,j} \neq p^{\theta_0}_{\tilde{e}_0,j} \) for some \((j, \theta_0, e_0, \tilde{e}_0)\) (the distribution would immediately be ordered in terms of incentives if \( p^\theta_{e,i} = p^\theta_{\tilde{e},i} \) for all \((i, \theta, e, \tilde{e})\)). For notational simplicity (relabeling if needed), let \( j = 1 \).

Use equation (59) to write:

\[
\sum_{i=1}^{N} w_i [p^{\theta_0}_{e_0,i} - p^{\theta_0}_{\tilde{e}_0,i}] = [p^{\theta_0}_{e_0,1} - p^{\theta_0}_{\tilde{e}_0,1}] \sum_{i=1}^{N} w_i \phi_{i,1},
\]

and

\[
\sum_{i=1}^{N} \tilde{w}_i [p^{\theta_0}_{e_0,i} - p^{\theta_0}_{\tilde{e}_0,i}] = [p^{\theta_0}_{e_0,1} - p^{\theta_0}_{\tilde{e}_0,1}] \sum_{i=1}^{N} \tilde{w}_i \phi_{i,1}.
\]

Substitute these equations in (61):

\[
[p^{\theta_0}_{e_0,1} - p^{\theta_0}_{\tilde{e}_0,1}] \sum_{i=1}^{N} w_i \phi_{i,1} = [p^{\theta_0}_{e_0,1} - p^{\theta_0}_{\tilde{e}_0,1}] \sum_{i=1}^{N} \tilde{w}_i \phi_{i,1},
\]

which, since \( p^{\theta_0}_{e_0,1} \neq p^{\theta_0}_{\tilde{e}_0,1} \), implies:

\[
\sum_{i=1}^{N} w_i \phi_{i,1} = \sum_{i=1}^{N} \tilde{w}_i \phi_{i,1}.
\]

For each \( \theta, e, \) and \( \tilde{e} \), multiply both sides by \( p^\theta_{e,1} - p^\theta_{\tilde{e},1} \) to obtain:

\[
[p^\theta_{e,1} - p^\theta_{\tilde{e},1}] \sum_{i=1}^{N} w_i \phi_{i,1} = [p^\theta_{e,1} - p^\theta_{\tilde{e},1}] \sum_{i=1}^{N} \tilde{w}_i \phi_{i,1}.
\]
Then, using condition (59) again, gives:

\[ \sum_{i=1}^{N} w_i[p_{e,i}^\theta - p_{\tilde{e},i}^\theta] = \sum_{i=1}^{N} \tilde{w}_i[p_{e,i}^\theta - p_{\tilde{e},i}^\theta], \]

which establishes (1).

Proposition 6 shows the equivalence between three properties of a probability distribution. As discussed in the main text, property (1) states that if one type has the same incentives to exert two efforts under contracts \( w \) and \( \tilde{w} \), so do all other types. In other words, all contracts can be ranked in terms of the incentives that they provide. Property 2 is the MS condition used in the text.

Property 3 provides a straightforward graphic interpretation of what it means for a distribution to be multiplicatively separable as explained in the text.

**D. Generalized Multiplicative Separability**

**Definition 4.** Let \( \mathcal{T} \) be a finite set. The distribution of outputs satisfies generalized multiplicative separability (GMS) if there exist a function \( H : X \times \mathcal{T} \to \mathbb{R}_+ \) and measurable function \( I : E \times \Theta \times \mathcal{T} \to \mathbb{R} \) such that

\[
F_{\tilde{e}}^\theta(x) + \sum_{t=1}^{T} H(x, t) I(\tilde{e}, \theta, t) = F_{e}^\theta(x) + \sum_{t=1}^{T} H(x, t) I(e, \theta, t), \quad \forall e, \tilde{e}, \theta, x,
\]

(GMS)

and

\[
I(\tilde{e}, \theta, t) \geq I(e, \theta, t) \iff I(\tilde{e}, \theta, \tilde{t}) \geq I(e, \theta, \tilde{t})
\]

for all \( e, \tilde{e}, t, \tilde{t}, \theta \).

Notice that, although we adopted the term “generalized” in the definition of GMS, this condition is only a generalization of MS when distributions are ordered according to the first-order stochastic dominance.

**Example 3.** Let \( \mathcal{T} = \{1, ..., T\} \) and consider the following generalization of the Linearity of the Distribution Function Condition:

\[
F_{\tilde{e}}^\theta(x) = \sum_{t=0}^{T} I(e, \theta, t) F_t(x) = - \sum_{t=1}^{T} I(e, \theta, t) H(x, t) + F_0(x),
\]

(62)

where \( E = [0, 1], \{ \tilde{F}_t(x); t = 0, 1, ..., T \} \) is a family of cumulative distributions, \( I(e, \theta, t) \geq 0, \sum_{t=0}^{T} I(e, \theta, t) = 1 \) and \( H(x, t) = F_0(x) - F_t(x) \). If \( F_t \) first-order stochastically dominates \( F_0 \) and
I(\cdot, \theta, t) is non-decreasing for all \( t \in T \) and all \( \theta \in \Theta \), then the distribution of outputs satisfies GMS.

Notice that Example 2 is special case of Example 3. Indeed, define \( T = N \), \( \bar{f}_0 = p_0^\theta \), \( \bar{f}_\theta = p_1^\theta \),

\[
I(0, \theta, t) = \begin{cases} 
1 & \text{if } t = 0 \\
0 & \text{if } t \neq 0
\end{cases}
\quad \text{and} \quad
I(1, \theta, t) = \begin{cases} 
1 & \text{if } t = \theta + 1 \\
0 & \text{if otherwise}
\end{cases},
\]

where \( \bar{f}_t \) is the probability distribution of \( \bar{F}_t \). It is straightforward to check that GMS is satisfied.

**Theorem 4.** Suppose that GMS holds and \( \Delta x_i > 0 \), for all \( i = 1, \ldots, n \), where \( x_0 = 0 \). There exists an optimal mechanism that offers a number of contracts of at most the cardinality of the set \( T \). Moreover, for any optimal mechanism, there exists an equivalent mechanism that offers a number of contracts of at most the cardinality of the set \( T \).

**Proof.** By the Lemma 18, we can assume without loss of generality that free disposal (FD) holds. By Lemma 5, there is no loss of generality in assuming that all contracts are uniformly bounded. By GMS, a type-\( \theta \) agent who switches from effort \( \bar{e} \) to \( e \) while keeping the same contract \( w \) gains

\[
v^\theta_e(w) - v^\theta_{\bar{e}}(w) = \sum_{t=1}^{T} [I(\bar{e}, \theta, t) - I(e, \theta, t)] \sum_{i=1}^{N} w_i h_i(t) + \epsilon^\theta_{\bar{e}} - \epsilon^\theta_e. \tag{63}
\]

In turn, this switch changes the principal’s payoff by

\[
u^\theta_e(w) - u^\theta_{\bar{e}}(w) = \sum_{t=1}^{T} [I(\bar{e}, \theta, t) - I(e, \theta, t)] \sum_{i=1}^{N} (x_i - w_i) h_i(t). \tag{64}
\]

Let \((w, e)\) be a feasible mechanism. The set of all bonuses in this mechanism,

\[
\mathcal{M} := \{w^\theta : \theta \in \Theta\},
\]

is well defined and is composed of uniformly bounded functions. Therefore, its closure \( \overline{\mathcal{M}} \) is a compact set.

For each \( t \in T \), let \( w^{t*} \) be a solution of the following maximization program:

\[
\max_{w \in \mathcal{M}} \sum_{i=1}^{N} w_i h_i(t), \tag{65}
\]

where \( h_i(t) = H(x_i, t) - H(x_{i-1}, t) \) and \( H(x_0, t) = 0 \), for all \( i = 1, \ldots, N \) and all \( t \in T \).

Since \( \mathcal{M} \) is compact and \( N \) is finite, the objective function is a continuous linear functional and the constraint defines a closed set, (65) has a solution \( w^{t*} \in \mathcal{M} \). Let \( \Theta_t \) be the set of types.
that choose $w^{t*}$ when restricted to the menu $\mathcal{M}^* = \{w^{\tilde{t}}; \tilde{t} \in \mathcal{T}\}$. Without loss of generality, we can always define $\Theta_t$ such that the collection $\{\Theta_t; t \in \mathcal{T}\}$ forms a partition of $\Theta$.

Let $\theta \in \Theta_t$ and $e^*(\theta)$ be an effort that maximizes the agent’s payoff under contract $w^{t*}$:

$$v_{e^*(\theta)}(w^{t*}) \geq v_{e(\theta)}(w^{t*}),$$

which, by GMS, can be written as

$$\sum_{t=1}^{T} [I(e(\theta), \theta, t) - I(e^*(\theta), \theta, t)] \sum_{i=1}^{N} w^{t*}_{i} h_{i}(t) \geq c^{\theta}_{e^*(\theta)} - c^{\theta}_{e(\theta)}. \quad (66)$$

Similarly, because $e(\theta)$ is his effort choice with contract $w_0(x)$,

$$\sum_{t=1}^{T} [I(e(\theta), \theta, t) - I(e^*(\theta), \theta, t)] \sum_{i=1}^{N} w^{\theta}_{i} h_{i}(t) \leq c^{\theta}_{e^*(\theta)} - c^{\theta}_{e(\theta)}. \quad (67)$$

Combining (66) and (67), we obtain

$$\sum_{t=1}^{T} [I(e(\theta), \theta, t) - I(e^*(\theta), \theta, t)] \sum_{i=1}^{N} w^{\theta}_{i} h_{i}(t) \leq c^{\theta}_{e^*(\theta)} - c^{\theta}_{e(\theta)} \leq \sum_{t=1}^{T} [I(e(\theta), \theta, t) - I(e^*(\theta), \theta, t)] \sum_{i=1}^{N} w^{t*}_{i} h_{i}(t). \quad (68)$$

Since $w^{t*}$ solves program (65), we have

$$\sum_{i=1}^{N} w^{t*}_{i} h_{i}(t) \geq \sum_{i=1}^{N} w^{\theta}_{i} h_{i}(t), \text{ for all } t \in \mathcal{T}.$$ 

Therefore, it follows from GMS and (68) that $I(e(\theta), \theta, t) \geq I(e^*(\theta), \theta, t)$, for all $t \in \mathcal{T}$.

We now establish that replacing contract $w^{\theta}$ by $w^{t*}$ increases the principal’s payoff from type $\theta \in \Theta_t$. As in the proof of Theorem 2, we first show that, holding effort fixed, the principal is better off with the substitution of contracts. Since $w^*$ is the limit of sequence in $\mathcal{M}$, the agent’s utility is continuous, and the original mechanism is incentive compatible, it follows that

$$v^{\theta}_{e(\theta)}(w^{\theta}) \geq v^{\theta}_{e(\theta)}(w^{t*}). \quad (69)$$

Multiply both sides by $-1$ and add $\sum_{i=1}^{N} x_i p^{\theta}_{e(\theta), i}$ to both sides of this inequality to obtain:

$$\sum_{i=1}^{N} (x_i - w^{t*}_{i}) p^{\theta}_{e(\theta), i} \geq \sum_{i=1}^{N} (x_i - w^{\theta}_{i}) p^{\theta}_{e(\theta), i}, \quad (70)$$

which states that, holding effort $e(\theta)$ fixed, the principal gets a higher profit with contract $w^{t*}$.
than with $w^\theta$.

Next, we show that the change in effort also benefits the principal. Notice first that FD implies that the function $x \rightarrow x - w^*(x)$ is non-decreasing. Therefore, applying integration by parts and using GMS and FD, we conclude that

$$
\sum_{i=0}^{N} (x_i - w_{i+1}^*) h_i(t) = - \sum_{i=1}^{N+1} (\Delta x_i - \Delta w_{i+1}^*) H(x_{i-1}, t) \leq 0, \text{ for all } t \in \mathcal{T},
$$

where $h_0(t) = 0$, $x_0 = x_1 = w_0^*$ and $x_{N+1} = x_N = w_N^*$. Hence, since $I(e(\theta), \theta, t) \geq I(e^*(\theta), \theta, t)$, for all $t \in \mathcal{T}$ and $\theta \in \Theta_t$, we have that

$$
[I(e(\theta), \theta, t) - I(e^*(\theta), \theta, t)] \sum_{i=1}^{N} (x_i - w_{i+1}^*) h_i(t) \leq 0, \text{ for all } t \in \mathcal{T} \text{ and } \theta \in \Theta_t.
$$

Using again GMS, these last inequalities imply

$$
\sum_{i=1}^{N} (x_i - w_{i+1}^*) p^\theta_{e,i} \geq \sum_{i=1}^{N} (x_i - w_{i+1}^*) p^\theta_{e^*(\theta),i},
$$

which shows that the principal gains from the change in effort for each $\theta \in \Theta_t$.

Combining (70) and (71) establishes that the principal’s profit from $\theta$ with the new contract exceeds her profit with the original contract:

$$
u^\theta_{e^*(\theta)}(w^*) \geq u^\theta_{e(\theta)}(w^\theta).
$$

Consider the mechanism $(\bar{w}, \bar{e})$: $\bar{w}^\theta(x) = w^*(\theta)$ and $\bar{e}(\theta) = e^*(\theta)$, where $e^*(\theta) \in \arg \max e^v_{e}(w^*)$, for all $\theta \in \Theta_t$ and $t \in \mathcal{T}$. As previously shown, this mechanism raises the principal’s payoff pointwise and satisfies LL and FD. Notice that the expected utility of type $\theta$ at contract $w^*$ who chooses effort $e$ is

$$
v^\theta_{e}(w^*) = \sum_{i=1}^{N} w^*_{i} p^\theta_{e,i} - c^\theta_e.
$$

Since $w^* \geq 0$ and the agent must weakly prefer the recommended effort $e^*$ to the one with $c^\theta_e \leq 0$, the IR constraint must hold. By construction the mechanism $(\bar{w}, \bar{e})$ satisfies IC. The proof of the second part is analogous to the proof from Theorem 2.

The next two lemmas show that we can assume FD in the proof of Theorem 4 without loss of generality.

**Lemma 17.** Fix $\alpha > 0$. A mechanism $(w, e)$ satisfies IC and LL under probability and cost functions $(p, c)$ if and only if the mechanism $(\alpha w, e)$ satisfies IC and LL constraints under probability
and cost functions \((p, \alpha c)\).

**Proof.** This result follows from the fact that, for each \(\alpha > 0\), the IC constraint for the mechanism \((w, e)\) is equivalent to

\[
\sum_i p_{e(\theta),i}^\theta \cdot \alpha w_i^\theta - p_{\hat{e},i}^\theta \cdot \alpha w_i^\hat{\theta} \geq \alpha c_{e(\theta)}^\theta - \alpha c_{\hat{e}}^\hat{\theta}
\]

for all \(\theta, \hat{\theta} \in \Theta\) and all \(\hat{e} \in E\), and LL is equivalent to

\[
\alpha w_i^\theta \geq 0
\]

for all \(\theta \in \Theta\) and \(i = 1, ..., N\).

\(\square\)

**Lemma 18.** Suppose that \(\Delta x_i > 0\), for all \(i\) and \(x_0 = 0\). Then, assuming FD in the proof of Theorem 4 is without loss of generality.

**Proof.** For a given probability \(p\) satisfying Assumption A2 we know that all contracts in an optimal mechanism is uniformly bounded. In particular, there exists \(L > 0\) such that \(\Delta w_i \leq L\), for all \(i\) and all contract \(w\) belonging to any optimal mechanism. Take \(\alpha > 0\) sufficiently small such that \(\alpha L \leq \min \{\Delta x_i; i = 1, ..., n\}\). From Lemma 17, if \((w, e)\) is an optimal mechanism under probability and cost functions \((p, c)\), then \((\alpha w, e)\) is an optimal mechanism under probability and cost functions \((p, \alpha c)\). We conclude that all optimal mechanism in the economy with parameters \((x, p, \alpha c)\) satisfies FD and every optimal mechanism in this economy is a homogenous transformation (of degree one) of an optimal mechanism in the economy with parameters \((x, p, c)\). Therefore, the assumption that economy with parameters \((x, p, c)\) satisfies FD is without loss of generality. \(\square\)

**Ordering in terms of FOSD across types**

This subsection examines the generalized multiplicative separability (GMS) condition. We provide the graphical interpretation regarding how distributions satisfying GMS change with effort. For this, recall that a cone \(A\) in \(\mathbb{R}^N\) is finitely generated by a (finite) set of vectors \(\{a_1, ..., a_K\} \subset \mathbb{R}^N\) if for each \(a \in A\) there exist scalars \(\lambda_k \geq 0\) such that

\[
a = \sum_{k=1}^K \lambda_k a_k.
\]

**Definition 5.** The family of distributions \(F_{e}^\theta(x)\) of outputs is ordered in terms of FOSD across types if:

(i) \(\{F_{e}^\theta; e \in E\}\) is ordered according to the FOSD, for all \(\theta \in \Theta\);

(ii) \(\{F_{e}^\theta - F_{\hat{e}}^\theta \geq 0; \theta \in \Theta\) and \(e, \hat{e} \in E\}\) \(\subset A\), where \(A\) is a cone finitely generated in \(\mathbb{R}^N_+\).
It is straightforward to verify that any distribution that satisfies GMS must be ordered in terms of FOSD across types. The next proposition shows that the reverse is also true, so these are equivalent conditions.

**Proposition 7.** $F^\theta_e(x)$ is ordered in terms of FOSD across types if and only if $F^\theta_e(x)$ satisfies GMS.

**Proof.** By the definition of order in terms of FOSD across types, there exists a finite set $T$ and 
{\{H(t); t \in T\} \subset \mathbb{R}^N_+}$ such that 
$$F^\theta_e - F^\theta_{\tilde{e}} = \sum_{t \in T} \lambda^\theta_{t, e, \tilde{e}} H(t),$$
where $\lambda^\theta_{t, e, \tilde{e}} \in \mathbb{R}$, for all $\theta \in \Theta$. Hence,

$$F^\theta_e = F^\theta_{e_0} + \sum_{t \in T} \lambda^\theta_{t, e, e_0} H(t),$$
for some $e_0 \in E$. Letting $I(e, \theta, t) = -\lambda^\theta_{t, e, e_0}$, we have that

$$F^\theta_{\tilde{e}} - F^\theta_e = \sum_{t \in T} [I(\tilde{e}, \theta, t) - I(e, \theta, t)] H(t).$$

Again using the definition of order in terms of FOSD across types, $F^\theta_{\tilde{e}}$ FOSD $F^\theta_e$ if and only if $I(\tilde{e}, \theta, t) - I(e, \theta, t) \geq 0$, for all $t \in T$. Therefore, $F^\theta_{\tilde{e}}(x)$ satisfies GMS. \qed

**Uniform MLRP**

Whenever condition GMS holds, we can write

$$p^\theta_{e_L, i} + \sum_{i=1}^T I(e_L, \theta, t) h_i(t) = p^\theta_{e_H, i} + \sum_{i=1}^T I(e_H, \theta, t) h_i(t),$$

which can be rearranged as

$$\frac{p^\theta_{e_H, i} - p^\theta_{e_L, i}}{p^\theta_{e_L, i}} - 1 = -\sum_{i=1}^T \frac{[I(e_H, \theta, t) - I(e_L, \theta, t)] h_i(t)}{p^\theta_{e_L, i}}.$$

A distribution that satisfies GMS has the monotone likelihood ratio property (MLRP) if $\frac{p^\theta_{e_H, i}}{p^\theta_{e_L, i}}$ is increasing in $i$ for any $e_L, e_H, t$, and $\theta$ with $I(e_H, \theta, t) < I(e_L, \theta, t)$. Therefore, a sufficient condition for MLRP to hold is that

$$\frac{h_i(t)}{p^\theta_{e, i}}$$
is increasing in $i$ for any $e, \theta, t$. 63
We call this last condition *uniform MRLP*.

Our next result establishes that when the distribution satisfies uniform MLRP, not only is it optimal to offer a single contract, but this contract takes the form of debt. Since the proof of this result is similar to the proof of Theorem 3, we present it in the online appendix.

**Theorem 5.** Suppose GMS holds and the distributions satisfy uniform MLRP. There exists an optimal FD mechanism that gives the principal a single debt contract. Moreover, any optimal FD mechanism is equivalent to a mechanism that gives the principal a single debt contract.

**References**


Online Appendix

Existence

This general proof covers the model with two outputs as a particular case. By Lemma 5, there is no loss of generality in taking uniformly bounded contracts. Let $L > 0$ denote this uniform bound so that $\mathcal{W} := [0, L]^N \subset \mathbb{R}^N$ denotes the space of contracts. Let $\mathcal{B}(\mathcal{W})$ be the Borel $\sigma$-field in $\mathcal{W}$.

Let $v^\theta_e(w) := \sum_{i=1}^N w(x_i)p^\theta_e(x_i) - c^\theta_e$. By Assumption 2, $v^\theta(\cdot)$ is continuous on $\mathcal{W} \times E$, and, for each $(w, e) \in \mathcal{W} \times E$, $v^\theta_e(w)$ is $\mathcal{B}(\Theta)$-measurable. Since $E$ is compact and $v^\theta_e(w)$ continuous, the agent’s effort choice problem has a solution. Let $V^\theta(w) := \sup_{e \in E} v^\theta_e(w)$ denote the utility of type $\theta$. Let the non-empty, compact-valued mapping $(\theta, w) \rightarrow e^*(\theta, w)$ denote the “reaction function” of type $\theta$:

$$e^*(\theta, w) := \{ e \in E : v^\theta_e(w) \geq V^\theta(w) \}.$$  

By Berge’s Maximum Theorem, $w \rightarrow V^\theta(w)$ is continuous for each $\theta$ and $w \rightarrow e^*(\theta, w)$ is upper semicontinuous on $\mathcal{W}$ for each $\theta$.

Let $u^\theta_e(w) := \sum_{i=1}^N (x_i - w(x_i)) p^\theta_e(x_i)$ denote the principal’s payoff. By Assumption 2, for each $\theta \in \Theta$, $u^\theta(\cdot)$ is continuous on $\mathcal{W} \times E$, and, for each $(w, e) \in \mathcal{W} \times E$, $u^\theta_e(w)$ is $\mathcal{B}(\Theta)$-measurable. Notice that $p$ is $\mu$-integrable on $\Theta \times E$ because $|p^\theta_e| \leq 1$ on $\Theta \times E$.

Since $e^*(\theta, w)$ is compact and $u^\theta_e(w)$ continuous, for each $(\theta, w)$, the problem $\sup_{e \in e^*(\theta, w)} u^\theta_e(w)$ has a solution, say $e^\theta_w$. Thus, for each contract $w$, the principal can provide the agent with a list, $\{ e^\theta_w, \theta \in \Theta \}$, of recommended efforts. Since $e^\theta_w \in e^*(\theta, w)$ for each $\theta \in \Theta$, a type-$\theta$ agent has no incentive to take an effort other than the one requested by the principal $e^\theta_w$ (i.e., the agent is obedient). To choose the best list of requests, the principal must therefore solve the problem

$$\sup_{e \in e^*(\theta, w)} u^\theta_e(w)$$

for each $(\theta, w) \in \Theta \times \mathcal{W}$. Let $U^\theta(w) := \sup_{e \in e^*(\theta, w)} u^\theta_e(w)$. Since $u$ is $\mathcal{B}(\Theta) \times \mathcal{B}(\mathcal{W}) \times \mathcal{B}(E)$-measurable and continuous on $E$, and since $e^*(\cdot, \cdot)$ is $\mathcal{B}(\Theta) \times \mathcal{B}(\mathcal{W})$-measurable (see Nowak (1984), Lemma 1.10) and compact-valued, it follows that $U$ is $\mathcal{B}(\Theta) \times \mathcal{B}(\mathcal{W})$-measurable (see Himmelberg et.al. (1976), Theorem 2). Moreover, since $e^*(\theta, \cdot)$ is upper semicontinuous on $\mathcal{W}$, it follows from Theorem 2 of Berge (1963) that $U^\theta(\cdot)$ is upper semicontinuous on $\mathcal{W}$ for each $\theta$. Finally, since the principal’s payoff $u$ is integrably bounded, $U : \Theta \times \mathcal{W} \rightarrow \mathbb{R}$ is also integrably bounded on $\Theta \times \mathcal{W}$.
With these observations, we can write the principal’s program as
\[ \sup_{w \in \mathcal{W}} \int U^\theta(w) d\mu(\theta) \]
which has a solution since \( \mathcal{W} \) is compact.

Remark 1. Notice that Theorems 2 and 3 still hold with a continuum of outputs if we impose that the space of feasible contracts is uniformly bounded. The maximization and minimization problems defined in proof of Theorem 2 have solutions if we consider the \((L^\infty(X), L^1(X))\)-weak* topology on the space of feasible contracts. Indeed, by the Banach-Alaoglu theorem (see Rudin, 1991), the set \( \mathcal{M} \) defined for those problems has a weak*-compact closure, and their objective functions are continuous with respect to weak* topology. The rest of the proof is a simple adaptation of the arguments from Theorems 2 and 3.

Non-robustness with respect to risk neutrality

Notice that when the optimal mechanism recommends low effort to all types, it must pay zero in all states to all types. Therefore, any mechanism that offers different contracts to different types must recommend high effort from at least one type. The proposition below provides a partial converse to Proposition 5:

**Proposition 8.** Consider the model with a risk averse agent and binary outputs, efforts, and types. Suppose that either (i) \( p^B_H \leq p^A_L \) and \( e(A) = e(B) = H \) is the effort recommendation in any optimal mechanism, or (ii) \( p^B_L \geq p^A_L \) and \( e(A) = L \) and \( e(B) = H \) is the effort recommendation in any optimal mechanism. Then, for any strictly concave utility function \( u \), there is no optimal mechanism that offers a single contract to all types.

The proof is a consequence of Lemmas 10, 19, and 20.

**Lemma 19.** Suppose that the principal wants to implement \( e(A) = e(B) = H \). If \( p^B_H \leq p^A_L \), the optimal mechanism must offer a different contract to each type.

**Proof.** The proof verifies that the best mechanism among those that offer the same contract to all types does not satisfy the first-order conditions that are necessary for an optimal mechanism. The best uniform contract that implements high effort from both types is \( s^* = 0 \) and \( b^*_A = \frac{c^A_H}{p^A_H - p^A_L} \). To see this, notice this contract is the unique solution to the principal’s cost minimization program:

\[
\min_{s,b} \mu \left[ (1 - p^A_H)v(s) + p^A_H v(s + b) \right] + (1 - \mu) \left[ (1 - p^B_H)v(s) + p^B_H v(s + b) \right]
\]

subject to

\[ b \geq \frac{c^A_H}{p^A_H - p^A_L}, \quad b \geq \frac{c^B_H}{p^B_H - p^B_L}, \quad s \geq 0. \]
Evaluating the first-order conditions (excluding the one that defines $\lambda_1$) derived in the proof of Lemma 11 at this uniform contract, gives:

\[-(1 - \mu) \left[ (1 - p_H^B) v'(0) + p_H^B v'(b) \right] - \lambda_2 - \lambda_3 + \lambda_5 + \lambda_6 \leq 0 \]

\[\lambda_5 + \lambda_6 = (1 - \mu)v'(b) + \lambda_2 \frac{p_A^A}{p_H^A} + \lambda_3 \frac{p_A^B}{p_H^B} > 0.\]

Substituting out $\lambda_5 + \lambda_6$ and rearranging, we obtain:

\[(1 - \mu)(1 - p_H^B) \left[ v'(b) - v'(0) \right] \leq \lambda_2 \left( 1 - \frac{p_H^A}{p_H^B} \right) + \lambda_3 \left( 1 - \frac{p_H^A}{p_H^B} \right).\]

Since $p_H^A \geq p_L^A \geq p_H^B$ and the multipliers are non-negative, it follows that

\[(1 - \mu)(1 - p_H^B) \left[ v'(b) - v'(0) \right] \leq 0 \implies v'(b) \leq v'(0).\]

But this is not possible because $b > 0$ and $v$ is convex. Therefore, offering a single contract is not optimal. \qed

**Lemma 20.** Suppose that the principal wants to implement $e(A) = L$ and $e(B) = H$. If $p_L^B \geq p_L^A$, the optimal mechanism must offer a different contract to each type.

**Proof.** The proof is similar to the proof of Lemma 19. The best uniform contract is $s^* = 0$ and $b^* = \frac{c_B}{p_H^A - p_L^B}$. To see this, notice that this contract solves the principal’s cost minimization program:

\[
\min_{s,b} \mu \left[ (1 - p_L^A) v(s) + p_L^A v(s + b) \right] + (1 - \mu) \left[ (1 - p_H^B) v(s) + p_H^B v(s + b) \right]
\]

subject to

\[b \leq \frac{c_A}{p_H^A - p_L^A}, \quad b \geq \frac{c_B}{p_H^B - p_L^B}, \quad s \geq 0.\]

Evaluating the first-order conditions derived in the proof of Lemma 12 at this uniform contract, we obtain:

\[-\mu \left[ (1 - p_L^A) v'(0) + p_L^A v'(b) \right] - \lambda_2 - \lambda_3 + \lambda_5 + \lambda_6 \leq 0 \]

\[\lambda_5 + \lambda_6 = \mu v'(b) + \lambda_2 \frac{p_H^A}{p_L^A} + \lambda_3 \frac{p_H^B}{p_L^B} > 0.\]

Substituting out $\lambda_5 + \lambda_6$ and rearranging, gives

\[\mu(1 - p_L^A) \left[ v'(b) - v'(0) \right] \leq \lambda_2 \left( 1 - \frac{p_H^A}{p_L^A} \right) + \lambda_3 \left( 1 - \frac{p_H^B}{p_L^B} \right).\]
Since $p^B_H > p^B_L \geq p^A_L$, this inequality implies that

$$
\mu (1 - p^A_L) [v'(b) - v'(0)] \leq 0 \iff v'(b) \leq v'(0),
$$

which is not possible because $b > 0$ and $v$ is strictly convex. Therefore, offering a single contract is not optimal. \qed

**Proof of Theorem of 5**

Recall that, as shown in the proof of Theorem 4, if it is optimal for the agent $\theta \in \Theta_t$ to pick effort $e^*(\theta)$ when offered contract $w^t*$, it is also optimal to do so it when offered any contract $\bar{w}$ with $\sum_{i=1}^N w_i h_i(t) = \sum_{i=1}^N \bar{w}_i h_i(t)$, for all $t$. We are now ready to present the proof:

Let $w^t*$ be an optimal contract and let $e(\theta)$ denote the effort chosen by type $\theta$ when offered this contract. Then, for $K_t = \sum_{i=1}^N w^t_i h_i(t)$, this contract must solve the following program:

$$
\begin{align*}
\min_w & \sum_{i=1}^N \int_{\Theta_t} p^\theta_{e^*(\theta),i} \, d\mu(\theta) \\
\text{subject to} & \sum_{i=1}^N w_i h_i(t) = K_t, \forall t, \quad \text{(IC')} \\
& w_i \geq 0, \quad \text{(LL)} \\
& x_i - x_{i-1} \geq w_i - w_{i-1}. \quad \text{(M)}
\end{align*}
$$

As argued above, the first constraint ensures that effort $e(\theta)$ is still optimal for the agent. The second and third constraints are LL and FD. This is a restricted program: any contract that satisfies these constraints is feasible but not every feasible contract satisfies these constraints. Therefore, since $w^*$ is optimal among all feasible contracts, it must also solve this more restricted program that only includes a subset of feasible contracts.

Let $\bar{p}_{i,t} \equiv \int_{\Theta_t} p^\theta_{e^*(\theta),i} \, d\mu(\theta)$ denote the marginal distribution of outputs induced by effort $e(\cdot)$. It is convenient to rewrite the program above in terms of increments:

$$
\begin{align*}
\min_{\Delta w_i} & \sum_{j=0}^N \bar{p}_{j,t} \sum_{i=0}^j \Delta w_i & (72) \\
\text{subject to} & \sum_{j=0}^N h_j(t) \sum_{i=0}^j \Delta w_i = K_t, \quad \text{(IC')} 
\end{align*}
$$
\[
\sum_{i=0}^{j} \Delta w_i \geq 0, \quad \forall j \quad (LL_j)
\]
\[
\Delta x_j - \Delta w_j \geq 0, \quad \forall j. \quad (M_j)
\]

We claim that if \( LL_j \) holds with equality, then \( M_j \) holds with strict inequality and vice-versa. To see this, notice that if \( LL_j \) holds with equality, then
\[
\sum_{i=0}^{j} \Delta w_i = \sum_{i=0}^{j-1} \Delta w_i + \Delta w_j = 0 \quad \because \Delta w_j \leq 0 < \Delta x_j,
\]
\[
\leq 0 \text{ by } M_{j-1}
\]
showing that \( M_j \) holds with inequality. Conversely, if \( M_j \) holds with equality, then
\[
\Delta x_j = \Delta w_j + \sum_{i=0}^{j-1} \Delta w_i \geq \Delta x_j > 0,
\]
\[
\geq 0 \text{ by } LL_{j-1}
\]
so \( LL_j \) holds with inequality.

The necessary first-order conditions associated with program (72) are
\[
- \sum_{j=i}^{N} \bar{p}_{j,t} + \lambda_i^C \sum_{j=i}^{N} h_j(t) + \sum_{j=i}^{N} \mu_{LL}^j - \mu_{M}^j = 0, \quad \forall i \quad (73)
\]
along with the usual complementary slackness conditions.

Let \( \xi_j \equiv \bar{p}_{j,t} - \sum_{t=1}^{T} \lambda_i^C h_j(t) \) and notice that
\[
\xi_j > 0 \iff 1 > \frac{\lambda_i^C h_j(t)}{\bar{p}_{j,t}}.
\]
Notice that uniform MLRP implies that \( \frac{h_j(t)}{\bar{p}_{j,\theta,j}} \) is increasing in \( j \) for all \( \theta, t \), so that \( \frac{\lambda_i^C h_j(t)}{\int_{\Theta} \bar{p}_{j,\theta,j} d\mu(\theta)} = \frac{\lambda_i^C h_j(t)}{\bar{p}_{j,t}} \) is also increasing in \( j \). Hence, there exists a unique \( k \in \{1, \ldots, N\} \) such that \( \xi_j > (\leq) 0 \) for all \( j < (\geq) k \).

Substituting \( \xi_i \) in (73), we obtain
\[
\xi_i = \mu_{LL}^i - \mu_{M}^i + \mu_{i+1}^M, \quad i < N
\]
\[
\xi_N = \mu_{LL}^N - \mu_{N}^M.
\]
There are three cases: (1) \( \xi_N > 0 \), (2) \( \xi_N < 0 \), and (3) \( \xi_N = 0 \).

Case 1) \( \xi_N > 0 \).
In this case, $LL_N$ binds and $M_N$ does not. Because $\xi_i$ crosses 0 once from above, $\xi_i > 0$ for all $i$ and, by induction, none of the free disposal constraints bind. As a result, the solution is $w_i = 0$ for all $i$.

Case 2) $\xi_N < 0$.

In this case, $M_N$ binds. For $N - 1$, we have

$$\xi_{N-1} = \mu^{LL}_{N-1} - \mu^M_{N-1} - \xi_N \geq \mu^{LL}_{N-1} - \mu^M_{N-1}.$$ 

If $N - 1 \geq k$ so that $\xi_{N-1} \leq 0$, it then follows that $\mu^{LL}_{N-1} - \mu^M_{N-1} \leq 0$ so that $M_{N-1}$ binds (and $LL_{N-1}$ doesn't). Inductively, it follows that $M_i$ binds for all $i \geq k$.

Let $j$ be such that $LL_j$ binds (and, therefore $\mu^{LL}_j = 0$). By the previous argument, it must be the case that $j < k$ so that $\xi_j > 0$. We have that

$$0 < \xi_{j-1} = \mu^{LL}_{j-1} - \mu^M_{j-1} + \mu^M_j = \mu^{LL}_{j-1} - \mu^M_{j-1}.$$ 

Then, we must have that $LL_{j-1}$ binds and $M_{j-1}$ doesn’t. Hence, there exists $i^* \in \{0, ..., N\}$ such that $LL_i$ binds on all $i \leq i^*$ and $M_i$ binds on all $i > i^*$. The contract must therefore be an option with a price between $x_{i^*}$ and $x_{i^*+1}$.

Case 3) $\xi_N = 0$.

In this case, $\mu^{LL}_N = \mu^M_N$. Since we have previously shown that we cannot have both constraints holding with equality, we must have $\mu^{LL}_N = \mu^M_N = 0$. Substituting at the condition for the previous output gives

$$\xi_{N-1} = \mu^{LL}_{N-1} - \mu^M_{N-1} > 0,$$

where we used the fact that $\xi_i$ crosses 0 once from above and $\mu^M_N = 0$. Thus, $\mu^{LL}_{N-1} > 0 = \mu^M_{N-1}$ so that $LL_{N-1}$ binds. Substituting inductively shows that all other LL constraints also bind. Hence, the solution in this case is an option contract with a strike price $x^* \in \{x_{N-1}, x_N\}$.

Since $w^{t*}$ is a debt contract for each $t \in T$ and $\{w^{t*}; t \in T\}$ is incentive compatible, this set should singleton.

The last part of the proposition follows similar argument of the proof of Theorem 2.

References

