International Journal of Game Theory (2020) 49:601–637
https://doi.org/10.1007/s00182-019-00700-5

REVIEW

Contracting with moral hazard, adverse selection and risk neutrality: when does one size fit all?

Felipe Balmaceda[1]

Accepted: 12 December 2019 / Published online: 23 December 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

This paper studies a principal-agent relationship when both are risk-neutral and in the presence of adverse selection and moral hazard. Contracts must satisfy the limited-liability and monotonicity conditions. We provide sufficient conditions under which the optimal contract is simple, in the sense that each type is offered the same contract. These are: the action and the agent's type are complements, and the output's cumulative distribution function is such that the marginal rate of substitution between the action and the agent's type is the same for each possible output realization. Furthermore, under the average monotone likelihood ratio property, the optimal contract is a call-option contract as in Innes (J Econ Theory 52(1):45–67, 1990). The results shed light on the fact that sometimes contracts are not highly dependent on individual characteristics as predicted in most pure moral hazard and pure adverse selection settings.

Keywords Moral hazard · Adverse selection · Limited liability · Monotonicity · Optimal mechanism

JEL Classification D82 · D86 · J33

I would like to thank the participants at the regular seminar series at the Economics Department of the University of Chile, the Latin American Theory workshop at IMPA 2006, the Latin American Econometric Society Theory workshop 2008, the Latin American Meeting of the Econometric Society 2009, and IIOC-Boston 2013. This is a different version of the paper presented at these conferences. The main difference stems from the fact that in this version monotonicity constraints are imposed. I am grateful for the financial support of the Instituto de Sistemas Complejos de Ingeniería (ISCI), grant CONICYT PIA/BASAL AFB180003 and FONDECYT FB016, and the Institute for Research in Market Imperfections and Public Policy, MIPP, grant ICM IS130002.

✉ Felipe Balmaceda
fbalmace@gmail.com

1 Economics Department, Diego Portales University, Avda. Santa Clara 797, Santiago 8580000, Chile

1 Introduction

The theory of incentives shows that providing incentives to agents with different characteristics calls for highly complex contracts. However, empirical and casual observations (see, for instance, Baker et al. 1994; Bhattacharyya and Lafontaine 1995) show that, in many circumstances, contracts are simple in the sense that they are not tailored to agents' characteristics. For instance, in markets that use franchising (for details see Lafontaine and Slade 1997, 2001), instead of offering contracts tailored to the characteristics of each franchisee, most of the franchisors employ a limited set of contracts, often just two: a business-format franchising and an integrated contract.[1] Within the first category, different franchisors choose different contract terms, as well as different royalty rates and franchise fees, but any given franchisor offers the same terms to all potential franchisees at a given point in time. Paarsch and Shearer (1999, 2000) study the British Columbia tree-planting industry and find that piece-rate contracts are identical within each tract and, through a structural model that controls for the endogeneity of the piece rate, that the elasticity of worker effort with respect to changes in the piece rate across tracts is positive and large. Using data on young drivers, Chiappori and Salanié (2002) cannot reject the hypothesis that coverage and accident frequency are statistically independent. Finkelstein and Poterba (2004) test for adverse selection in annuity markets in the UK, and do not find adverse selection in coverage, but they do find adverse selection in other dimensions. Finkelstein and McGarry (2006) find little or no evidence of positive correlation between risk types and policy choices. In fact, Salanié (2005, p 474) concludes: "The recent literature provides very strong evidence that contractual forms have large effects on behavior. As the notion that "incentive matters" 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?" Thus, the simplicity of real-life contracts in many situations remains a challenge to incentive theory, despite the fact that important progress has been made on the last decade or so based on the idea that contracts/mechanisms must be robust (see, for instance, Lopomo 2001; Bergemann and Morris 2005; Chassang 2013; Carroll 2015).

Under a set of fairly natural assumptions, we show that we can rationalize simple contracts such that payment schemes are not tailored to agents' characteristics. In order to study our notion of simplicity, this paper proposes a standard principal-agent model where the agent must choose an action that determines, together with his type (e.g. ability), the probability distribution of a contractible output that has more than two realizations. The principal and the agent are risk neutral; the agent privately learns his type before signing the contract and the principal observes neither the agent's action nor his type. The agent is subject to limited liability, which limits the principal's ability to punish the agent for bad performance. Limited liability is prevalent in financial and labor market contracts; an employer is not free to punish poor performance with

[1] Balmaceda (2009) provides a rationale for the emergence of pay-for-performance contracts and self-selection in a competitive labor market setting, where workers are risk averse, while firms are risk neutral and unaware of workers' abilities. He shows that, under certain parameterizations, the second-best menu has more than one contract, and under others, the menu contains only one contract.

negative wages, and an entrepreneur cannot be asked to re-pay more than the returns on his venture. Following Innes (1990), we restrict the contract space to monotone contracts; that is, the payment to the agent is non-decreasing with realized output, and the return to the principal is also non-decreasing with it.[2]

A full characterization of the optimal mechanism is difficult, yet we provide sufficient conditions under which this consists of a single contract, and as a result the optimal mechanism exhibits what we call the *one-size-fits-all* property. This result requires that the action and the agent's type improve the output distribution in the sense of first-order stochastic dominance (**FOSD**), and that the agent's payoff function satisfies the strict complementarity between the action and the agent's type (**SC**), as well as the weak-separability property (**WSEP**). This was first introduced in the literature by Faynzilberg and Kumar (1997, 2000). Furthermore, if the output distribution satisfies the average (over types) monotone likelihood ratio property (**EMLRP**), the optimal contract is a call-option contract; that is, the contract pays the limited liability for low output realizations, and it pays the output minus a positive face value for high output realizations. This is the same contract as the one derived by Innes (1990). Thus, under **EMLRP**, the contract is simple not only in the sense defined above, but also in how payments vary with realized output.

WSEP imposes that the marginal rate of substitution between the action and the agent's type is the same for each possible output realization. In the case of pure moral hazard, Grossman and Hart (1983) argue that a similar condition, known as the linear distribution function, is sufficient for the validity of the first-order approach. Faynzilberg and Kumar (1997, 2000) extend this to the case of moral hazard and adverse selection, and partially characterize the optimal contract with moral hazard, adverse selection and risk averse agents. **WSEP** allows agents to order different contracts according to the power of incentives in such a way that all types agree on which contracts provide more powerful incentives. Thus, **WSEP** allows an ordering on the contract space.

To better understand the intuition, assume for the time being that there are only two outcomes, and that the contract offers a fixed payment and a bonus upon success. In order to provide incentives for effort, the principal must provide high-powered incentives. To induce information revelation, higher-power incentives are offered to higher types, since, ceteris paribus, they have an incentive to understate their ability level in order to convince the principal that success is unlikely. Because effort is unobservable, each type benefits from high-powered incentives, hence the fixed payment must decrease with ability. But, under limited liability, the fixed payment has a lower bound on how much it could be decreased. Because the principal wants to minimize the informational and limited-liability rent without sacrificing efficiency too much, she has incentives to lower the fixed payment as much as possible, but then the lower bound forces the same fixed payment to everyone. This makes the contract with highest incentive power more attractive to everyone. Thus, the principal has to choose between a pooling contract that leaves no rent to the least able type, and a menu that pays a fixed payment decreasing in ability and a bonus increasing in it, which leaves positive

[2] This is a revised version of Balmaceda (2011) where we study the same problem without imposing monotonicity constraints. This makes the result more intuitive and plausible in real-world situations.

rents to everyone and sacrifices efficiency. Because effort increases with ability, and a dollar of fixed payment costs the principal more than a dollar in a higher bonus, ultimately the principal prefers pooling types, since more able types choose higher efforts. The presence of multiple outcomes expands the scope for profitable screening; yet, under **WSEP**, each type evaluates each contract as consisting of a fixed payment plus a bonus that aggregates all possible outcomes and is independent of the agent's action.

When there are multiple outcomes, the intuition for *one-size-fits-all* is based on the fact that, for any incentive-compatible menu of contracts, there exists a degenerate contract that is preferred by the principal and chosen by each possible agent's type. To see this, consider an incentive-compatible menu of contracts and a degenerate menu that has one contract, namely the one with the most powerful incentives among the incentive-compatible contracts, denoted by w^*. Because of **WSEP**, i.e. the cumulative distribution function being such that the marginal rate of substitution between the action and the agent's type is the same for each possible output realization, everyone agrees on which contract provides the most powerful incentives. Incentive compatibility implies that, conditional on the incentive-compatible action, all agents are weakly better-off under the incentive-compatible contract than under w^*. Because the action chosen is assumed to be constant and agents are risk neutral, this means that, for each ability type, expected compensation is higher under the incentive-compatible contract. However, when an agent chooses contract w^*, he chooses an action that is at least as large as the incentive-compatible action, since actions and types are complements (condition **SC**), and w^* provides more powerful incentives. Because higher types and higher actions improve the output distribution in the sense of **FOSD**, and the contract is such that the return to the principal is non-decreasing with realized output, the fact that agents choose a (weakly) higher action under w^* implies that profits are higher under the degenerate mechanism than under the incentive-compatible mechanism. Furthermore, we show that the optimal action is identical to the one that maximizes the virtual surplus, which entails a downward distortion for each possible type but the highest. The limited-liability constraint impairs the principal's ability to reduce the informational rent any further. In its absence, the principal could lower the payments in bad states so as to make contracts with different power of incentives more appealing to agents with higher types. This will leave room for separation to be optimal for at least a subset of types. Hence, under limited liability, **SC** and **WSEP** offering a degenerate mechanism achieves the dual role of reducing informational rents and increasing efficiency; that is, each type chooses the action that maximizes the virtual surplus.

When we further impose **EMLRP**, the optimal mechanism offers a single call-option contract to all types; that is, a contract that makes the principal the full residual claimant when the output is low, by paying the agent the lowest possible value allowed by the limited-liability constraint, and it makes the agent the full residual claimant when the output is high, by paying him the realized output minus a positive face value. The face value is chosen so that the participation constraint of the lowest type binds and therefore all other types get a positive informational rent. To reduce the informational rent, the principal would like to set the payment in each state different from that for which the average likelihood ratio is the highest equal to the limited liability and, in order to stop agents from overstating their ability type, the principal is forced to

offer the same wage profile to each ability type. However, the monotonicity constraint imposing that the principal's payoff does not decrease with realized output makes this solution infeasible. Hence, the principal is forced to offer the steepest wage profile satisfying the monotonicity constraints. Because of **EMLRP** and **SC**, this contract induces higher types to choose higher actions and, therefore, minimizes informational rents and maximizes efficiency.

These results highlight the fact that, under **WSEP**, the moral hazard problem is more important than the adverse selection problem, in the sense that the optimal mechanism is such that contract terms are not customized to the agent's type, yet they are designed to provide incentives for effort. However, the action chosen is different from that under pure moral hazard, since incentive compatibility with respect to ability types imposes constraints on the contract through both the first- and second-order conditions of the agent's revelation of information problem.

Related Literature. A long-standing literature studies the question of how optimal incentive contracts should look like in different situations. The literature on optimal contracting under pure moral hazard is vast, and that under pure adverse selection is also extensive. For the sake of brevity, we will focus on the most closely related strands, which are: the one dealing with contract simplicity and robustness, the one dealing with moral hazard under limited liability with and without monotonicity constraints, and that dealing with moral hazard and adverse selection.

The literature regarding contract simplicity and robustness has made considerable progress in the last decade or so, and it has mainly focused on the optimality of linear contracts and ex-post implementation. Carroll (2015) shows that the optimal contract under risk neutrality and limited liability is linear when the principal's objective is to maximize the worst-case expected payoff in a setting where the principal knows only a subset of the actions available to the agent. Under uncertainty about the agent's technology, the principal knows a lower bound on the agent's expected payoff, based on the fact that she knows a subset of the actions available to the agent. The way she translates the agent's lower bound into a lower bound for his payoff is by a linear sharing relationship, since this is tight in each instance. Holmstrom and Milgrom (1987) demonstrate that linear contracts are robust to aggregation over time–the optimal contract is a linear function of the endpoint. Carroll and Meng (2016) show that if the principal has knowledge of a lower bound, but not of an upper bound on the shocks, linear contracts are reliable because they give the same incentives for effort at every point along the contract. Balmaceda et al. (2016) also study robustness by quantifying welfare loss as the ratio between the first-best social welfare and that arising from the principal's optimal pay-for-performance contract. They look at the performance of linear contracts in terms of this metric and show that linear contracts attain the lower bound, which is consistent with the Carroll's (2015) result. Chassang (2013) also studies a principal-agent model with limited liability, moral hazard and adverse selection in a dynamic context. He calibrates the dynamic contract to a simple benchmark that satisfies three properties: the principal is guaranteed a positive expected payoff, is renegotiation-proof, and the contract satisfies an efficiency bound which is sufficiently tight to imply the max-min optimality of linear contracts. However, this benchmark is not feasible because it does not satisfy limited liability. He then pro-

vides a class of dynamic limited-liability contracts that satisfy efficiency properties for a very broad class of stochastic environments and are free from details. By contrast, in our setting, limited liability is crucial for the simplicity, since it is the lower bound on payments that limited liability imposes that which makes pooling optimal under either a two-outcome setting or the assumptions adopted in this paper. In a sense, this robustness literature deals with a different kind of simplicity related to linearity, a not type-independent contracts. Yet, it is related to the fact that we also find that the optimal type-independent contract is simple when **EMLRP** is assumed. Namely, a single call-option contract is optimal.

Lopomo (1998, 2001) ask why is it that the English auction is often used to determine the terms of trade when the owner of a single object faces a number of potential buyers. He argues in favor of it due to its simplicity and robustness. Lopomo (1998) shows that the perfect Bayesian equilibrium with undominated strategies of the English auction maximizes the seller's expected revenue among all open bidding procedures in which the buyers at any stage during the auction have only three options: they can purchase the object at the current asking price, remain in the auction by declaring their willingness to pay the current bid, or drop out irrevocably. Lopomo (2001) argues that English auctions are optimal when considering all posterior-implementable outcome functions among all equilibrium outcomes of selling procedures, which satisfy a no-regret condition: each buyer has no incentive to revise his decision after observing his opponents' behavior. Bergemann and Morris (2005) study robust mechanisms in the sense that ex-post implementation is equivalent to interim (or Bayesian) implementation for all possible types. The robustness here regards the fact that Bayesian implementation requires too much common knowledge, while ex-post implementation does not. They show that this kind of robustness requires quasi-linear preferences without restrictions on transfers, and that the principal implements a function and not a correspondence. Because our model considers limited liability, their result does not apply to our setting. By contrast, our simplicity is derived under Bayesian implementation and common knowledge, yet in order to do so, our problem needs more structure than theirs, and as such it is less general. In addition, the notion of simplicity in our setting is different and it relates to type-independence more than implementation issues.

With regard to the optimal contracting under moral hazard, adverse selection and risk neutrality, the most closely related papers are: Demougin (1989), Guesnerie et al. (1989), Caillaud et al. (1992), Guesnerie and Laffont (1984), McAfee and McMillan (1986, 1987), Laffont and Tirole (1986) and Melumad and Reichelstein (1989). They study the value of information in agency when there is unlimited liability. These papers study under what conditions an incentive-compatible and individually-rational allocation can be implemented by a contract that is type-independent and there is no efficiency loss from adding noise to the production technology. The first three papers consider the moving support case, while the fourth studies the case of fixed support. The main assumption in these papers is that the noise in the production technology is independent of the agent's type; they study noisy hidden information models. The main result of this literature is that, in most such models, the principal can reach the same utility as in the absence of noise. Guesnerie and Laffont (1984) show that the optimal mechanism is to offer the same contract to everyone when the first-best allocation is

decreasing in the agent's type, and incentive compatibility requires the opposite. This is known as non-responsiveness. However, the reason why the optimal mechanism exhibits the *one-size-fits-all* property here has nothing to do with non-responsiveness. Laffont and Tirole (1986) also show that bunching may occur when contracts are not restricted and the standard increasing hazard rate assumption is violated.

Escobar and Pulgar (2016) show, in the absence of limited-liability and when the outcome space is binary, that given certain conditions, the optimal contract exhibits the *one-size-fits-all* property for a subset of the agent's type set. This stems from the fact that their outcome space is binary and, therefore, all types agree that a contract which pays a higher wage when the high outcome is realized and a lower wage when the low outcome is observed provides stronger incentives for effort and truth-telling. This must not be the case when the outcome space has more than two outcomes. Gottlieb and Moreira (2013) study a principal-agent model with moral hazard and adverse selection. They focus on the case in which types are multidimensional, and show that a positive mass of types with low conditional probabilities of success gets a constant payment and zero rents and there is distortion everywhere.[3] They also find, under certain assumptions, that the optimal mechanism offers only finitely many contracts. They, as Balmaceda (2011) does, find that when the agent is risk neutral and has limited liability, all agents are offered a single contract.[4] The main difference stems from the fact that they, like Escobar and Pulgar (2016), focus on the case in which there are only two outcomes. Ollier and Thomas (2013) also study moral hazard and adverse selection with ex-post constraints and, as in Escobar and Pulgar (2016), the outcome space is binary. They, as we do, show that the optimal contract pool types, since there are countervailing forces due to the ex-post constraints. However, considering several outcomes increases the scope for profitable screening, which makes our result more surprising. The countervailing forces in our setting arise from limited liability.

Lewis and Sappington (2000, 2001) study a principal-agent model with adverse selection, moral hazard and limited liability. They use this setting to study how wealth constraints affect the optimal mechanism for selling a project to different bidders. Lewis and Sappington (2000) do so when wealth is publicly observed and Lewis and Sappington (2001) consider the case of private wealth. In the Lewis and Sappington's (2001) model, which is closest to our setting, the outcome is dichotomic and they allow the agent to post a bond before output is realized. Hence, the principal has two instruments, which are: the payment after success (failure leads to zero outcome in their model) and the bond posted by the agent that results from the agent's announced initial wealth. This payment cannot exceed the agent's wealth. When ability and wealth are private information, they show that an agent requires both higher ability and greater wealth to secure a more powerful compensation scheme: more of either of them does not suffice. When wealth is observable they show that. for low-ability agents, the payment after success rises with the ability level, while for high-ability agents, this is independent on the ability level.

[3] We just became aware that Gottlieb and Moreira (2017) extend their analysis to the case of a continuum of outputs. Therefore, their result is like ours but for the continuous output case.

[4] We became aware of this result just recently.

Faynzilberg and Kumar (1997, 2000) show that, when the agent is risk averse and there are neither monotonicity nor limited-liability constraints, **WSEP** is needed to ensure the existence of an optimal mechanism. Furthermore, Faynzilberg and Kumar (1997) show that **WSEP** and **MLRP** together result in a monotone mechanism, while Faynzilberg and Kumar (2000) propose a duality method in the spirit of Grossman and Hart (1983) for finding the optimal mechanism and argue that this is type-independent if the technology satisfies strong separability; that is, the marginal impact of the action and the agent's type are independent of each other, and these are independent of the randomness in the production function. This stems from the fact that the action chosen is type-independent and therefore all types agree on which contract provides stronger incentives for effort. Hence they need a much stronger condition so that agents can order the contract space.

Regarding the pure moral hazard literature, Innes (1990) and Matthews (2001) show the optimality of debt contracts for the principal (call-option contract for the agent) under the assumption of monotone likelihood ratio (**MLRP**) and contract monotonicity constraints. Poblete and Spulber (2012) show the optimality of the call-option contract under a less stringent condition called monotonicity with respect to the state of the critical ratio, defined as the product of the hazard rate of the state and the ratio of the marginal product of effort to the marginal product of the state. There is another strand that imposes limited liability but not monotonicity constraints. Kim (1997) and Demougin and Fluet (1998) show that the optimal contract entails a bonus only when the outcome is higher than a given threshold, and a fixed wage equal to the limited liability for all outcomes lower than the threshold. Thus, in both cases, the principal leaves a rent to the agent, known as the limited-liability rent. The effort required is downward distorted with respect to the first-best effort in order to reduce the limited-liability rent. We add to this literature the adverse selection dimension, and focus on monotone contracts as defined by Innes (1990).[5] Under the assumptions made here, we obtain the Innes's (1990) result in the case when both moral hazard and adverse selection are taken into account.

The rest of the paper is organized as follows. Section 2 presents the model and the main assumptions. Section 3 derives the optimal mechanism with adverse selection and moral hazard. Section 4 discusses the robustness with regard to some of the assumptions. Section 5 concludes with some remarks. All proofs can be found in the Appendix.

2 The model

2.1 Setup

Consider a relationship between a risk-neutral principal and a risk-neutral agent. The agent chooses an action $a \in A \equiv [0, \bar{a}]$ and is characterized by a privately known type (e.g., ability level) indexed by $\theta \in \Theta \equiv [\underline{\theta}, \overline{\theta}]$. Type θ has a cumulative distribution

[5] See, Balmaceda (2011) for the case in which monotonicity constraints are disregarded and there exist multiple tasks with an aggregated across-tasks performance measure.

function $F(\theta)$, which is common knowledge, twice continuously differentiable with the associated density function $f(\theta)$ and full support. The inverse of the hazard function $H(\theta) \equiv \frac{1-F(\theta)}{f(\theta)}$ is assumed to be monotonically decreasing, as is the standard in the literature.

The agent's action and type jointly determine the output distribution, which can take $n+1$ possible realizations, $0 \leq y^0 < y^1 < \cdots < y^n$. Given action a and type θ, output y^i, $i \in I \equiv \{0, \ldots, n\}$, occurs with probability $p^i(\theta, a) \in [0, 1]$, with $\sum_{i=0}^{n} p^i(\theta, a) = 1$.

The agent's cost of action $a \in A$ is $c(a)$, with: (i) $c(a') \geq c(a)$ for all $a' \geq a$; (ii) $c(\cdot)$ is a twice continuously differentiable and strictly convex function for all $a \in A$; and (iii) $c_a(0) = c(0) = 0$.[6]

A contract is given by the payment scheme: $w \equiv (w^0, \ldots, w^n)$, where w^i is the payment when y^i is realized. Payments are restricted by a limited-liability constraint that prevents the principal from paying the agent a wage lower than $L \geq 0$ in any state, with $L \leq y_0$. This constraint can be thought of as the case in which the agent owns no assets at the time of contracting. Hence,

$$w^i \geq L \quad \forall i \in I. \tag{\textbf{LL}}$$

In addition, contracts are restricted to satisfy Innes's (1990) monotonicity constraints, as follows:

$$w^{i+1} \geq w^i, \quad \forall i \in I \setminus \{n\} \tag{\textbf{MONA}}$$

and

$$y^{i+1} - w^{i+1} \geq y^i - w^i, \quad \forall i \in I \setminus \{n\}. \tag{\textbf{MONP}}$$

Restriction **MONA** must be satisfied if the agent can costlessly reduce the output or if the principal can secretly borrow money from an outsider to inflate output. **MONP** must be satisfied if the principal can costlessly reduce profits. In the case with two possible outcomes **MONP** can be discarded since optimality requires this to be the case. Otherwise, the principal will offer a flat contract that induces action zero. Finally, contracts are restricted to be piecewise continuously differentiable in θ.

Define the contract space as:

$$\mathcal{C} = \{w \in \Re^n | w \text{ satisfies } \textbf{LL}, \textbf{MONA} \text{ and } \textbf{MONP}\}.$$

Finally, the agent has a reservation utility equal to zero.

2.2 Main assumptions

[FOSD] $\sum_{i>i'} p_a^i(\theta, a) \geq 0$ and $\sum_{i>i'} p_\theta^i(\theta, a) \geq 0$ for all $i' \in I$ and all $(a, \theta) \in A \times \Theta$.

[6] The sub-index a denotes the partial derivative with respect to a. Similarly, with respect to θ.

This implies that higher actions as well as higher types increase, ceteris paribus, the agent's expected output, his expected compensation when the mechanism satisfies **MONA**, as well as his expected profits when the mechanism satisfies **MONP**.

[CP] (i) The upper cumulative probability distributions $\sum_{i>i'} p^i(\theta, a)$ are strictly concave in $(a, \theta) \in A \times \Theta$, $\forall i' \in I \setminus \{n\}$; (ii) $p^i(\theta, a)$ is thrice continuously differentiable for all $(a, \theta) \in A \times \Theta$, and $p^i(\theta, a) > 0$, $\forall (a, \theta) \in A \times \Theta$ and $\forall i \in I$; and (iii) $\lim_{a \to 0} \sum_i p^i_a(\theta, a) y^i - c_a(a) > 0$ and $\lim_{a \to \bar{a}} \sum_i p^i_a(\theta, a) y^i - c_a(a) < 0, \forall \theta \in \Theta$.

This is meant to ensure that the first-best action is unique and belongs to the interior. This, together with the fact that contracts satisfy **MONA**, ensures that the locally optimal action is unique and that this is globally optimal for all $(\theta, w) \in \Theta \times \mathcal{C}$.

[SC] $\sum_{i>i'} p^i_{\theta a}(\theta, a) \geq 0$ for all $i' \in I$ and all $(a, \theta) \in A \times \Theta$.[7]

This assumption implies that, when a increases, the marginal decrease in the cumulative probability distribution increases with type. This assumption ensures that the optimal action increases with ability; therefore, even when each type is offered the same contract, the more able types choose a higher action.

Following Faynzilberg and Kumar (1997), we will also consider the possibility that the technology satisfies weak separability between (θ, a) and the random component of the output, as follows:

[WSEP] For any $a, a' \in A$ with $a \neq a'$ and $\theta, \theta' \in \Theta$ with $\theta \neq \theta'$, the following holds:

$$\frac{p^i(\theta, a') - p^i(\theta, a)}{p^i(\theta', a) - p^i(\theta, a)} \text{ is independent of } i.$$

Faynzilberg and Kumar (1997) show that this implies that the probability that outcome y^i occurs can be written as follows:

$$p^i(\theta, a) = (q^i - r^i) g(\theta, a) + r^i$$

where $q^i \in [0, 1]$, $r^i \in [0, 1]$, $\sum_i^n r^i = 1$ and $\sum_{i=1}^n q^i = 1$. This assumption and the **CP**, **SC** and **FOSD** will be sufficient for the optimal mechanism to be type-independent or to satisfy the one-size-fits-all property.

[EMLRP] For any $a' > a$, $\frac{\int_{\theta \in \Theta} p^i(\theta, a') dF(\theta)}{\int_{\theta \in \Theta} p^i(\theta, a) dF(\theta)}$ increases with i.

This assumption implies that any principal who has an aggregated-over-types payoff function that is increasing in the project's outcome prefers the stochastic distribution of returns induced by higher actions. This ensures that the optimal type-independent mechanism is option-like, as is the case in the pure moral hazard contract under standard **MLRP** (see Appendix B for details).[8]

[7] Athey (1996) call this property l-increasing differences.

[8] See, Balmaceda (2009) for a use of the **MLRP** with respect to θ to derive the optimal contract under pure adverse selection, risk aversion and labor market competition.

2.3 The First Best

In this sub-section we study the case in which action a is contractible and the agent's type θ is common knowledge.

Let $y \equiv (y^0, \ldots, y^n)$ and $p(\theta, a) \equiv (p^0(\theta, a), \ldots, p^n(\theta, a))$. The total surplus from a type-θ agent when he chooses action a as inner product is given by: $S(\theta, a) \equiv p(\theta, a)y - c(a)$. Then, the first-best action, denoted by $a^{**}(\theta)$, is the solution to the following problem:

$$\max_{a' \in A} S(\theta, a'),$$

which entails the following first-order condition:

$$\sum_i p_a^i(\theta, a')y^i - c_a(a') = 0.$$

Because of assumptions **FOSD** and **CP**, the first-order condition is necessary and sufficient and the first-best action profile belongs to the interior. Thus, the first-best surplus is given by: $S(\theta, a^{**}(\theta))$. Because $y > 0$, $S(\theta, a^{**}(\theta)) \geq 0$.

Assumption **SC** implies that $S(\theta, a)$ is supermodular in (θ, a) and therefore it readily follows from theorem 3 in Edlin and Shannon (1998) that for any $\theta' > \theta$, $a^{**}(\theta') > a^{**}(\theta)$ and $S(\theta', a^{**}(\theta')) > S(\theta, a^{**}(\theta))$.[9] Thus, the first-best action is increasing with the agent's type and total surplus rises with the agent's type when **FOSD** holds.

One contract that implements this action promises to pay the agent $c(a^{**}(\theta)) + L$ if he delivers $a^{**}(\theta)$, and it pays L otherwise. The agent weakly prefers to deliver $a^{**}(\theta)$ in this setting, and so the principal's ideal outcome ensues. Because the reservation utility is zero and $L \geq 0$, this ensures that the agent is willing to participate.

This contract is not incentive compatible–every agent will choose to deliver zero effort since this will ensure a payment equal to L ,and because compensation is not contingent on realized output, the agent has neither incentive to exert effort nor to truthfully reveal his type.

3 Moral hazard and adverse selection

The principal's goal here is twofold: on the one hand, she has to provide the agent with incentives to choose the desired action, and on the other, she has to induce the agent to truthfully reveal his type. In short, any offer $(w(\theta), a(\theta))$ must be incentive compatible; i.e. a θ-type agent prefers contract $w(\theta)$ to any other contract, and is obedient in the sense that he chooses the action prescribed by the principal $a(\theta)$.

An agent of type $\theta \in \Theta$, when faced with contract $w \in \mathcal{C}$ and chooses action $a \in A$, has an expected utility given by:

$$U(\theta, w, a) \equiv p(\theta, a)w - c(a), \tag{1}$$

[9] This requires differentiability and that $S(\theta, a)$ has an increasing marginal return to a as θ increases.

and the principal has an expected utility given by:

$$V(\theta, w, a) \equiv p(\theta, a)(y - w).$$

By the revelation principle, a direct mechanism $(w(\theta), a(\theta)) : \Theta \to \mathcal{C} \times A$ is incentive-compatible if and only if, for all $\theta \in \Theta$,

$$(\theta, a(\theta)) \in \underset{(\theta', a') \in \Theta \times A}{\operatorname{argmax}} \; U(\theta, w(\theta'), a'). \tag{IC}$$

The mechanism satisfies individual rationality if and only if, for all $\theta \in \Theta$,

$$U(\theta, w(\theta), a(\theta)) \geq 0. \tag{IR}$$

Constraint (IC) is the incentive-compatibility constraint. This states that a type-θ agent is better off announcing his true type, receiving the allocation $w(\theta)$ and choosing action $a(\theta)$, than announcing a different type θ', receiving the allocation $w(\theta')$ and choosing any other action profile $a' \in A$. Because the action and the type are not observable, the agent can deviate by choosing any action/type he wishes and not be detected. Constraint (IR) is the standard individual rationality constraint establishing that, on the equilibrium path, each ability type prefers participating to staying out.

We say that action a is implementable if there exists a contract $w \in \mathcal{C}$, such that (w, a) is incentive-compatible. If w is such that $w(\theta) = w(\underline{\theta})$ for all $\theta \in \Theta$, the contract satisfies the one-size-fits-all property at Θ, and a is implementable by a one-size-fits-all contract at Θ. Define $\alpha(\theta, \theta') \in \operatorname{argmax}_{a' \in A} U(\theta, w(\theta'), a')$ and denote $\alpha(\theta, \theta)$ by $a(\theta)$.

Proposition 1 *Let $(w(\theta), a(\theta)) : \Theta \to \mathcal{C} \times A$ with $a(\theta) > 0$. Then $(w(\theta), a(\theta))$ is incentive compatible if and only if:*

(i) *For all $\theta \in \Theta$, $a(\theta)$ is differentiable almost everywhere, and satisfies the follows first-order condition:*

$$U_a(\theta, w(\theta), a(\theta)) = 0. \tag{2}$$

(ii) *For all $\theta, \theta' \in \Theta$,*

$$U_{aa}(\theta, w(\theta), a(\theta)) \leq 0. \tag{3}$$

(iii) *For almost all $\theta \in \Theta$,*

$$\sum\nolimits_i p^i(\theta, a(\theta)) \frac{dw^i(\theta)}{d\theta} = 0 \tag{4}$$

(iv) *For almost all* $\theta \in \Theta$,

$$\sum_i p^i_\theta(\theta, a(\theta)) \frac{dw^i(\theta)}{d\theta} \Big(\sum_i p^i_{aa}(\theta, a(\theta)) w^i(\theta) - c_{aa}(a(\theta)) \Big)$$
$$- \sum_i p^i_{a\theta}(\theta, a(\theta)) w^i(\theta) \sum_i p^i_a(\theta, a(\theta)) \frac{dw^i(\theta)}{d\theta} \leq 0. \tag{5}$$

Equation (2) is the first-order condition for the action when the agent truthfully reveals his type by choosing $w(\theta)$. Equation (3) ensures that the agent's utility is concave in a, for all $\theta \in \Theta$. Equation (4) is the first-order condition for the agent's revelation of information problem, taking into account that, upon a deviation, the agent will choose the action that maximizes his expected utility. This is also evaluated at the local optimal for (θ', a'). Equation (5) is the local second-order condition to the agent's revelation problem. This ensures the concavity of the agent's utility in $(\theta', w(\theta'), a')$, evaluated at the local optimal. This, together with (2) and (3), is sufficient for global incentive compatibility. This constraint restricts the kind of incentive-compatible mechanisms that the principal can use, and is equivalent to what the literature generally calls the monotonicity constraint. For instance, in the Laffont and Tirole's (1986) model, this entails a rising quantity and transfer with the agent's type. However, the condition here is weaker in general, since it allows decreasing as well as increasing payments with the agent's type. Thus, while imposing that payments are non-decreasing with respect to θ does yield sufficiency, it does so with loss of generality.[10] Observe that this holds when the contract is type-independent or satisfies the one-size-fits-all property. In addition, observe that conditions **MONA** and **CP** ensure that the agent's utility function is strictly concave in a for all $\theta, \theta' \in \Theta$, and that $\alpha(\theta, \theta')$ belongs to the interior of A. This implies that the locally optimal action for any $\theta, \theta' \in \Theta$ is globally optimal.

The key in the implementability result above is Eq. (4), which states that, in any incentive-compatible mechanism, conditional on the incentive-compatible action being chosen, the expected compensation could not vary with the agent's reported type. This stems from the fact that both the principal and the agent are risk-neutral. Therefore, conditional on any given action, they both only care about the expected compensation and, ceteris paribus, each agent type prefers the contract with the highest expected compensation, and the principal the contract with the lowest expected compensation.[11] However, this does not rule out menus where different contracts are customized to intervals of different types, since $w(\theta)$ is piecewise continuously differentiable. For instance, the principal could offer two contracts, one with low-powered incentives for an interval of low types, and one with high-powered incentives for an interval of high types, where high-powered means that payments increase more with

[10] If the outcome space were binary, that is, $i \in \{0, 1\}$, then what this second-order condition boils down to is that the difference between a higher and lower payoff increases with the agent's type, since in that case $p^1_\theta(\theta, a) > 0$.

[11] If the agent were to be risk averse, not only should he care about the expected compensation, but also about the payment variation across outputs. Payment variation will entail losses, since it would result in an inefficient risk allocation between the principal and the agent, but it will provide another instrument for increasing efficiency at a lower cost.

output realizations than in the low-powered contract. However, we will show that, when the probability function satisfies **WSEP**, jumps in $w(\theta)$ cannot be optimal.

Before characterizing the optimal mechanism, it is worthwhile to briefly discuss the dimensionality of the output space. If we were to restrict the output space to only two outputs, the only way to provide incentives will be by means of paying a high payment when the higher output is observed and a low payment when the lower output is observed. When an increase in (θ, a) increases the probability that the high output is observed, which is a very natural ordering, a greater high payment and a lower low payment will, due to risk neutrality, be preferred by each possible type. Hence, every possible type agrees on which contract provides stronger incentives. When the output space is greater than two, this is not sufficient,since different (θ, a) give rise to distribution functions in which the payment difference for different output levels $w^{i+1} - w^i$ will provide more powerful incentives to different types and, therefore, the types will evaluate the appeal of different contracts differently. We will show that, together, **FOSD** and **WSEP** provide a natural ordering of the contracts such that different types agree on which contracts provide stronger incentives, which is at the crux of the one-size-fits-all property.

In what follows, we will denote $dw^i(\theta)/d\theta$ by $\dot{w}^i(\theta)$ and $dU(\theta)/d\theta$ by $\dot{U}(\theta)$. Using the necessary condition in Eq. (4) and the envelope theorem to obtain $\dot{U}(\theta)$, while noting that $\dot{U}(\theta) \geq 0$, which means that only the lowest type's participation constraint matters, and integrating by parts, the principal's problem can be written as follows:

$$\max_{(w,a):\Theta\to\mathcal{C}\times A} \int_{\theta\in\Theta} \Big(S(\theta, a(\theta)) - H(\theta) \sum\nolimits_i p^i_\theta(\theta, a(\theta))w^i(\theta)\Big) f(\theta)d\theta - U(\underline{\theta})$$

subject to (6)

$$\sum\nolimits_i p^i_\theta(\theta, a(\theta))\dot{w}^i(\theta)\Big(\sum\nolimits_i p^i_{aa}(\theta, a(\theta))w^i(\theta) - c_{aa}(a(\theta))\Big) -$$

$$\sum\nolimits_i p^i_{a\theta}(\theta, a(\theta))w^i(\theta) \sum\nolimits_i p^i_a(\theta, a(\theta))\dot{w}^i(\theta) \leq 0, \text{ for almost all } \theta \in \Theta \tag{7}$$

$$\sum\nolimits_i p^i_a(\theta, a(\theta))w^i(\theta) - c_a(a(\theta)) = 0, \ \forall\theta \in \Theta, \tag{8}$$

$$U(\underline{\theta}) \geq 0, \tag{9}$$

$$w^i(\theta) \geq L, \ \forall\theta \in \Theta \text{ and } \forall\, i \in I. \tag{10}$$

This optimization problem bears some resemblance to the standard mechanism design problem (Myerson (1982)). The objective function is the virtual surplus; that is, total welfare minus information rents, which are given by $H(\theta) \sum_i p^i(\theta, a(\theta))w^i(\theta)$. However, constraint (7) is different from the standard monotonicity constraint due to the multidimensionality of the contract space. This constraint does not entail monotonicity of either the action or the payments. Nonetheless, this does not preclude us from studying the problem of virtual surplus maximization ignoring constraint (7) and then checking if the unconstrained solution satisfies this constraint. If the solution to the relaxed problem satisfies constraint (7), then it is the solution to the principal-agent problem. We will analyze this problem first and then will come back to the

virtual surplus maximization, taking into account the local second-order constraint in Eq. (7).

We first derive the optimal mechanism when the local second-order condition is ignored and **WSEP** is not imposed. Define $P^i(\theta, a) \equiv \sum_{j \geq i+1} p^i(\theta, a)$, $\xi^i(\theta, a) \equiv \frac{P^i_a(\theta,a)}{P^i_\theta(\theta,a)}$ and $\triangle w^i(\theta) \equiv w^{i+1}(\theta) - w^i(\theta)$.

Proposition 2 *Suppose that **CP**, **FOSD** and **SC** hold, and that $P^i_{a\theta}(\theta, a)/P^i_a(\theta, a)$ is non-increasing with a and non-decreasing with θ. Then, the mechanism that maximizes the virtual surplus, denoted by $(\hat{w}(\theta), \hat{a}(\theta))$, where $\hat{w}(\theta)$ satisfies **MONA**, **MONP** and **LL**, satisfies the following:*

(i) *There exists a threshold $\xi(\theta) \geq 0$, such that: $\triangle \hat{w}^i(\theta) = \triangle y^i$ if $\xi^i(\theta, \hat{a}(\theta)) \geq \xi(\theta)$, and $\triangle \hat{w}^i(\theta) = 0$ otherwise.*
(ii) *For all $\theta \in \Theta \setminus \bar{\theta}$, $\hat{a}(\theta) < a^{**}(\theta)$ and $\hat{a}(\bar{\theta}) = a^{**}(\bar{\theta})$*
(iii) *$\hat{a}(\theta)$ rises with θ.*

This result establishes that the mechanism that maximizes the aggregated virtual surplus is such that either $\triangle \hat{w}^i(\theta) = 0$ or $\triangle \hat{w}^i(\theta) = \triangle y^i$. This is due to the linearity of the virtual surplus with respect to $\triangle w^i(\theta)$. However, characterizing the optimal payment scheme is not possible without further assumptions. The reason stems from the fact that different types value different payments differently. This means that a certain type may prefer a higher $\hat{w}^i(\theta)$ and a lower $\hat{w}^{i'}(\theta)$, while a different type may prefer the opposite. Thus, not all types value increases in a given payment in the same way and they do not agree on which contract provides more powerful incentives. In fact, the condition that determines whether $\triangle \hat{w}^i(\theta)$ is zero or positive depends on whether $\xi^i(\theta, \hat{a}(\theta))$ is greater or lower than the threshold $\xi(\theta)$ and both functions are type-dependent.

This also establishes that the mechanism that maximizes the aggregated virtual surplus is such that the optimal action is downward distorted for all types but the highest one due to limited liability. In addition, the higher the type, the higher the action. The reason for the inefficient action is that the principal's cost of incentive compatibility, which is given by $H(\theta) \sum_i p^i_\theta(\theta, a(\theta)) w^i(\theta)$, is an increasing function of the action chosen due to: the complementarity between the action and the type, the fact that **MONA** implies that $w^i(\theta)$ is non-decreasing in i, as well as the increasing hazard rate assumption. Thus, the principal faces a trade-off between efficiency and informational rents, which is solved by a downward distortion of the optimal action so as to reduce the informational rent. Because both the rent and the optimal action are increasing in the agent's type θ, the principal requires the first-best action from the highest type.

The non-distortion at the top result seems counter-intuitive at first glance, since it contradicts the fact that, under moral hazard and limited-liability, the effort is downward distorted. This stems from the fact that here incentive compatibility with respect to private information requires higher effort from higher types. Thus, if the principal requires lower effort for the highest type, she has to lower the effort required to all other types. In other words, the principal faces a decreasing marginal cost of effort that trades off efficiency against the information rent; by lowering the effort requested to

the highest type the principal lowers the information rent for that type, but also lowers the efficiency of all other types.

Observe that if the mechanism that maximizes the aggregated virtual surplus when constraint (7) is ignored is such that, for some θ:

$$\sum_i \Big(p^i_\theta(\theta, \hat{a}(\theta))U_{aa}(\theta, w(\theta), \hat{a}(\theta)) \\ -p^i_a(\theta, \hat{a}(\theta)) \sum_i p^i_{a\theta}(\theta, \hat{a}(\theta))w^i(\theta)\Big)\dot{w}^i(\theta) \geq 0, \tag{11}$$

then constraint (7) must be binding when evaluated at the optimal mechanism for that particular type, and therefore the pointwise optimal mechanism for that type cannot be an optimum.

Observe that, assuming that $p(\theta, a)$ satisfies **WSEP** (i.e., $p^i(\theta, a) = (q^i - r^i)g(\theta, a) + r^i$ for all i), and after summation-by-parts, the local-second order condition can be written s follows:

$$\Big(U_{aa}(\theta, w(\theta), a(\theta))g_\theta(\theta, a(\theta)) - g_a(\theta, a(\theta)) \sum_{i=0}^{n-1} P^i_{a\theta}(\theta, a(\theta))\triangle w^i(\theta)\Big)\times \\ \sum_{i=0}^{n-1} \sum_{j\geq i+1} (q^j - r^j)\triangle \dot{w}^i(\theta) \leq 0.$$

The term in parenthesis is negative when we assume **MONA**, **CP** and **SC**. Hence, the sufficient condition reduces to:

$$\sum_{i=0}^{n-1} \sum_{j\geq i+1} (q^j - r^j)\triangle \dot{w}^i(\theta) \geq 0.$$

Observe that, in this case, the sufficient condition is independent of the action and the agent's type other than through the contract chosen. This greatly simplifies the principal's problem since all types rank the different contracts in the menu in terms of the power of incentives exactly in the same way. This is exactly the same with what happens when there are two outcomes.

Proposition 3 *Suppose that $g_{aaa}(\theta, a) \leq 0$, $g_{aa\theta}(\theta, a) \geq 0$, and **CP**, **FOSD**, **SC** and **WSEP** hold. Then the optimal incentive compatible mechanism, denoted by $(w^*(\theta), a^*(\theta))$, where $w^*(\theta)$ satisfies **MONA**, **MONP** and **LL**, is such that:*

(i) *For all $\theta \in \Theta$, $\dot{w}^{i*}(\theta) = 0$, $\forall i \in I$.*
(ii) *For all $\theta \in \Theta$, $\triangle w^i(\theta) = \triangle w^i \in [0, \triangle y^i]$ for all $\theta \in \Theta$.*
(iii) *For all $\theta \in \Theta \setminus \bar{\theta}$, $a^*(\theta) = \hat{a}(\theta) < a^{**}(\theta)$ and $a^*(\bar{\theta}) = \hat{a}(\bar{\theta}) = a^{**}(\bar{\theta})$.*
(iv) *For all $\theta \in \Theta \setminus \{\underline{\theta}\}$, $U(\theta) > L$ and $U(\underline{\theta}) = L$.*

This result shows that, under **WSEP**, the optimal mechanism satisfies the *one-size-fits all* property; that is, each type is offered the same contract.[12] This is at odds with the case in which there is either pure moral hazard or pure adverse selection.

[12] Observe that separability makes the principal's inference about the agent's private information (θ, a) harder since the informational content of each output is the same.

To grasp the intuition, consider any incentive-compatible non-degenerate menu of contracts $\{w(\theta)\}_{\theta\in\Theta}$, and assume that, instead, the principal offers the degenerate menu with the contract that has the most powerful incentives among all incentive-compatible contracts in the menu $\{w(\theta)\}_{\theta\in\Theta}$. Denote this contract by w^*. Because of **WSEP**, all types rank contracts in terms of the power of incentives in the same way, since the expected utility can be written as follows:

$$U(\theta, w, a) = w_0 + g(\theta, a) \sum_{i=0}^{n-1} Q^i \triangle w^i(\theta) - c(a),$$

where $Q^i = \sum_{j\geq i+1}(q^j - r^j)$.

Holding the incentive-compatible action constant, each type is at least weakly better-off choosing the incentive-compatible contract over w^*. Due to risk neutrality, this implies that the expected compensation for each type under the incentive compatible contract is higher than, or equal to that under w^*. Thus, holding the action constant, the principal's payoff is higher under contract w^* than under the corresponding incentive-compatible contract. However, when a type-θ agent chooses w^* instead of $w(\theta)$, he deviates and chooses an action accordingly. Due to the fact that the power of incentives is higher under w^* for each type and that $p(\theta, a)$ satisfies **SC**, a type-θ agent chooses an action, under contract w^*, that is at least as large as the one he would have chosen under $w(\theta)$. Because of this, the fact that **MONP** implies that $\triangle w^i \leq \triangle y^i$ for all $i \in I/\{n\}$ and **FOSD**, the principal's payoff is higher under w^* than under $w(\theta)$. In order to minimize the informational rent, the optimal action is downward distorted for everyone but the highest-ability agent. Thus, offering a degenerate menu decreases informational rents and improves efficiency, since no one chooses a strictly lower action than the one they would have chosen in the presence of a non-degenerate incentive-compatible menu.[13] In fact, the optimal action is identical to the one that maximizes the virtual surplus when **WSEP** is assumed, since a degenerate menu satisfies the local second-order condition.

Faynzilberg and Kumar (2000) argue that a sufficient condition for a type-independent mechanism in the case of risk aversion without limited liability is that the distribution function satisfies Grossman and Hart's (1983) linear cumulative distribution (**LDFC**) in actions and types.[14] The reason is that, in this case, the incentive-compatible action is type-independent. Here, we show that we do not need such an extreme form of separability in order to obtain that the optimal mechanism satisfies the one-size-fits-all property. Also, as discussed at length by Faynzilberg and Kumar (1997, 2000), the optimal contracting problem with moral hazard, adverse selection and risk averse agents is extremely difficult when the technology is non-

[13] It is interesting to note that, despite the fact that the moral hazard problem overcomes the adverse selection problem, the action chosen by each type under both moral and adverse selection is different from that under pure moral hazard. The reason stems from the incentive compatibility with regard to the agent's type. See Appendix B for the optimal action and contract under pure moral hazard.

[14] Using the same notation as in assumption **WSEP**, this means that $p^i(\theta, a) = (q^i - r^i)(\theta + g(a)) + r^i$.

separable. In fact, the Faynzilberg and Kumar (1997, 2000) existence result requires separability.[15]

As a last step, we derive the optimal type-independent contract when **EMLRP** is imposed.

Proposition 4 *Suppose that $g_{aaa}(\theta, a) \leq 0$, $g_{aa\theta}(\theta, a) \geq 0$, and **CP**, **FOSD**, **SC**, **WSEP** and **EMLRP** hold. Then, there exists an index i, denoted by i^*, such that:*

(i) *The optimal contract is an option-call contract of the form $w^{i*}(\theta) = L$ for all $i \leq i^*$ and $w^{i*}(\theta) = y^i - y^*$ for all $i > i^*$, where y^* solves the following equation:*

$$g(\underline{\theta}, a^*(\underline{\theta}))\Big(L \sum\nolimits_{i \leq i^*} Q^i + \sum\nolimits_{i > i^*} Q^i (y^i - y^*)\Big) - c(a^*(\underline{\theta})) = 0.$$

(ii) *$a^*(\theta)$ is non-decreasing with θ.*

This shows that the unique contract within the optimal mechanism is a call-option contract that pays L if the output is lower than or equal to y^{i*}, and pays $y^i - \bar{y}^*$ whenever the output exceeds y^{i*}. That is, the principal is the full residual claimant for small outputs and the agent is the full residual claimant for high outputs. The value of $\bar{y}^*$ is set to leave no informational rent to the lowest type; meaning that $\bar{y}^*$ is set so that $U(\underline{\theta}) = L$ and the rest receive a positive informational rent. The fact that the optimal contract is type-independent does not imply that the expected output is the same for each ability type. The second-best optimal action grows with the agent's type θ, and therefore the greater the type, the greater the expected output and the expected compensation.

This result is due to **EMLRP**, since the largest output provides more information in favor of a higher action. As a result, in the absence of constraint (7), the optimal payment scheme would be to pay a bonus when the highest output is observed and the limited liability otherwise.[16] While this solution satisfies **MONA**, it does not satisfy **MONP**. Hence, **MONP** makes this solution infeasible and forces the principal to spread payments through outputs so as to satisfy this monotonicity constraint. Because implementation requires offering the same contract to each type, condition **SC** ensures that the incentives of all types are aligned in the sense that each type's action increases as any of the prizes (i.e., $\triangle w^i$) rises and, conditional on any action, an increase in one of the prizes makes the contract more attractive to each ability type. If this is lacking, the principal could attempt sorting by mean of providing bigger prizes for intermediate outputs, since these will provide agents with less powerful incentives.

[15] Separability in each dimension was also assumed by McAfee and McMillan (1986, 1987), Baron and Besanko (1984, 1987) and Laffont and Tirole (1986) to name a few, and separability with respect to the uncertainty was assumed by Melumad and Reichelstein (1989). A well-known weakly separable technology satisfying **SC** and **CP** used in most papers is given by $p^i(\theta, a) = g(\theta, a)q^i + (1 - g(\theta, a))r^i$, where $g(\theta, a) = \theta^\beta a^\alpha$. This formulation was used by Lewis and Sappington (2000, 2001) to obtain their results for the particular case in which $\beta = 0$ and $n = 1$. Separability always holds in the case of two outcomes as well as for the standard additively separable technology.

[16] See Balmaceda (2011) for a formal proof of this claim.

4 Discussion

This section briefly discusses the consequences of considering different assumptions.

First, consider the case of type-dependent outside utility. The most plausible case is the one in which U increases with θ, since the opportunity cost of working elsewhere should be higher for an agent with a higher type, despite the fact that this is not observable. It can be easily shown that this would not change the result as long as the growth rate of U with respect to θ does not exceed the informational rent increase with θ (i.e., $\dot{U}(\theta)$). The reason stems from the fact that the optimal contract provides each type with a utility higher than their outside option. Otherwise, the principal may not be able to offer higher types a contract that promises, conditional on a given action, a higher expected compensation so as to satisfy his participation constraint without inducing lower types to claim they are a higher type. Hence, incentive compatibility may require to shut down a subset of types. The subset that will be shut down depends on the details of the model and it is hard to predict what would happen. However, complementarity between the action and the type suggests that there is a force towards shutting down low types. However, as shown by Lewis and Sappington (1989), this would also depend on whether the outside option is convex or concave in θ. In other words, it would also depend on how strong the countervailing forces coming from the outside option are.

Second, consider the case of moving support. We know from the moral hazard literature without limited liability that moving support facilitates implementation. However, under limited liability it is hard to see how this would help, since, as shown by Mirrlees (1975), the moving support assumption together with unbounded utility almost solve the implementation problem. With limited liability, the principal is restricted in her capacity to punish the agent for outputs that cannot come from a combination of a given type with a given action. Hence, despite the fact that there are outputs that are observable only if certain actions are chosen, which facilitates the inference problem, the principal would not be able to punish the agent in such a way that would deter the agent's misbehavior when outcomes not consistent with the desired actions are observed.

Third, in the absence of monotonicity constraints, as shown in Balmaceda (2011), **WSEP** is not needed. The reason is that **MLRP** with respect to both the action and the type, **SC** , and limited liability imply that it is optimal to pay a bonus only when the highest output is observed. Given that limited liability precludes negative payments, it is optimal to set all other payments equal to the limited liability. Because there is only one optimal instrument that provides incentives –the payment when the highest outcome is observed– all types prefer the contract with the highest payment. This implies that, provided it is optimal to pay a bonus only when the highest outcome is observed, it is not incentive-compatible to offer a menu with more than one contract. In other words, the model behaves as if there were only two outcomes.

Finally, observe that moral hazard is crucial for the *one-size-fits-all* result. To see this, lets discuss the local implementability conditions when there is adverse selection and no moral hazard; that is, actions are contractible.

Proposition 5 *Let* $(w(\theta), a(\theta)) : \Theta \to \mathcal{C} \times A$ *with* $a(\theta) > 0$ *and* $w(\theta)$ *piecewise continuously differentiable for all* $\theta \in \Theta$. *Then* $(w(\theta), a(\theta))$ *is incentive compatible only if:*

$$\sum_i p^i(\theta, a(\theta)) \frac{dw^i(\theta)}{d\theta} + \frac{da(\theta)}{d\theta} \Big(\sum_i p^i_a(\theta, a(\theta)) w^i(\theta) - c_a(a(\theta)) \Big) = 0 \quad (12)$$

and

$$\sum_i p^i_\theta(\theta, a(\theta)) \frac{dw^i(\theta)}{d\theta} + \frac{da(\theta)}{d\theta} \sum_i p^i_{a\theta}(\theta, a(\theta)) w^i(\theta) \geq 0. \quad (13)$$

In contrast to the necessary condition in Eq. (4) in proposition (1), the necessary condition in Eq. (12) does not require that:

$$\sum_i p^i(\theta, a(\theta)) \frac{dw^i(\theta)}{d\theta} = 0.$$

This results from the fact that the principal can use the action profile $a(\theta)$ as another instrument to induce truth-telling at a lower informational rent, and to deter low types from claiming to be high types when the power of incentives is higher for contracts tailored to high-ability types. In fact, given that the cost of the action is independent of the agent's type, the principal can implement the first-best action by offering a compensation contract that pays the cost of effort in each state. The agent would be indifferent between all actions and will choose the principal's preferred action profile. If the cost of the action profile were to depend on the agent's type, this solution would no longer be possible since, if she were to compensate each type according to the cost of the action taken, each type would claim to be the one with the highest cost.

5 Conclusions

This paper shows that the optimal menu of contracts in the presence of limited liability, monotonicity constraints, risk neutrality, moral hazard and adverse selection has three highly empirically observed properties: (i) when **WSEP** is satisfied, the menu of contracts exhibits the *one-size-fits-all* property; that is, contracts are not customized to the agent's privately know ability; (ii) better agents choose higher actions, are more productive, and have a higher expected compensation (see, for instance, Lazear 2000; Paarsch and Shearer 2000; Seiler 1984 for empirical evidence backing this result); and (ii) when **EMLRP** is imposed, the unique contract is a call-option contract that makes the agent the residual claimant when output is high, and the principal the residual claimant when output is low.

The fact that the optimal mechanism is such that only one contract is offered has an important practical consequence. Namely, it unburdens an econometrician studying the consequence of optimal contracting from controlling for unobserved heterogeneity. In fact, Chiappori and Salanié (2002) argue that the main concern about testing contract theory is the necessity of adequately controlling for unobserved heterogeneity.

Mainly, they argue that if this "is not done properly, then the combination of unobserved heterogeneity and of endogenous matching of agents to contracts is bound to create selection biases on the parameters of interest." While this concern is valid in many settings, especially in those where risk-aversion plays a role, the result here shows that this is not necessarily the main concern in a setting with risk-neutrality and limited liability, such as financial contracting, procurement, and regulation. Thus, when studying whether observed contracts have the properties predicted by contract theory in a setting likely to satisfy the assumptions above, the need of estimating fixed effect models is of a lesser importance than suggested by Chiappori and Salanié (2002). Furthermore, the *one-size-fits-all* property provides some theoretical justification for studies concerned with the consequences of optimal contracting on performance that use aggregated data.

Finally, the results presented in this paper provide a rationale for why many financial institutions offer *one-size-fits-all* credit card interest rates and payday loans, why tax systems do not offer menus of tax schedules where agents are free to choose the schedule that best suits them, and why regulatory agencies do not use menus of contracts that adopt the form predicted by the optimal regulation theory as developed by Laffont and Tirole (1986). In addition, it offers a rationale for taxi-driver contracts when **EMLRP** is imposed, since taxicab owners are full residual claimants up to a certain amount, which is the same for most taxis, and after that amount is reached, the drivers become full residual claimants.

Appendix

A Proofs of moral hazard, adverse selection and limited liability

Observe that, after summation by parts, a type-θ agent's utility rewrites as follows:

$$U(\theta, w, a) = w_0 + \sum_{i=0}^{n-1} \triangle w^i P^i(\theta, a) - c(a)$$

where $\triangle w^i \equiv w^{i+1} - w^i \geq 0$ and $P^i(\theta, a) \equiv \sum_{j \geq i+1}^{n} p^j(\theta, a)$ and the Principal's payoff from type θ is given by:

$$V(\theta, w, a) = y_0 - w_0 + \sum_{i=0}^{n-1} (\triangle y^i - \triangle w^i) P^i(\theta, a)$$

where $\triangle y^i \equiv y^{i+1} - y^i$.

Observe that FOSD with respect to a implies that, for any $a' \geq a$, $V(\theta, w, a') \geq V(\theta, w, a)$, since $\triangle y^i \geq \triangle w^i$ for all $i \in I/\{n\}$ by **MONP** and $U(\theta, w, a') \geq U(\theta, w, a)$ since $\triangle w^i \geq$ for all $i \in I/\{n\}$.

Lemma 1 *$U(\theta, w, a)$ is supermodular in $(\theta, \triangle w, a)$.*

Proof Recall that $P(\theta, a) \equiv (P_1(\theta, a), \ldots, P_I(\theta, a))$ and $P^i(\theta, a) \equiv \sum_{h' \geq h} p^i(\theta, a)$. Suppose that $a' > a$, then lets define $h(i)$ as the lowest h such that $P^i(\theta, a', a_{-i}) -$

$P^i(\theta, a, a_{-i}) \geq 0$; the existence of $h(i)$ is ensured by the fact that condition **SC** implies $P^i(\theta, a') - P^i(\theta, a)$ rises with h whenever $a' \geq a$. Suppose that $\theta^{'} > \theta$, then lets define $h(\theta)$ as the lowest h such that $P^i(\theta', a) - P^i(\theta, a) \geq 0$; the existence of $h(\theta)$ is ensured by the fact that condition **SC** implies that $P^i(\theta', a) - P^i(\theta, a)$ rises with h whenever $a' \geq a$.

Suppose that $\theta' > \theta$ and $a' \geq a$, then supermodularity requires that $U(\theta', a', \triangle w) - U(\theta', a, \triangle w) \geq U(\theta, a', \triangle w) - U(\theta, a, \triangle w)$. Observe that:

$$\begin{aligned}
&U(\theta', a', \triangle w) - U(\theta', a, \triangle w) \geq U(\theta, a', \triangle w) - U(\theta, a, \triangle w) \\
&\quad = \triangle w P(\theta', a') - c(a') - (\triangle w P(\theta', a) - c(a)) - (\triangle w P(\theta, a') - c(a') \\
&\qquad - (\triangle w P(\theta, a) - c(a))) \\
&\quad = \triangle w(P(\theta', a') - P(\theta', a)) - \triangle w(P(\theta, a') - P(\theta, a)) \\
&\quad = \sum\nolimits_i \triangle w^i \big(P^i(\theta', a') - P^i(\theta', a) - (P^i(\theta, a') - P^i(\theta, a))\big) \\
&\quad \geq 0,
\end{aligned}$$

where the inequality follows from the fact that condition **SC** ensures that $P^i(\theta, a') - P^i(\theta, a)$ rises with θ for all h, and the fact that the monotonicity constraint **MONA** implies that $\triangle w^i \geq 0$.

Suppose that $\theta' > \theta$ and $\triangle w^{i'} > \triangle w^i$. Then supermodularity requires that $U(\theta', a, \triangle w^{i'}, \triangle w_{-h}) - U(\theta', a, \triangle w^i, \triangle w_{-h}) \geq U(\theta, a, \triangle w^{i'}, \triangle w_{-h}) - U(\theta, a, \triangle w^i, \triangle w_{-h})$. Observe that this inequality is equal to:

$$(\triangle w^{i'} - \triangle w^i) P^i(\theta', a) \geq (\triangle w^{i'} - \triangle w^i) P^i(\theta, a).$$

Because $P^i(\theta', a) \geq P^i(\theta, a)$ and $\triangle w^{i'} > \triangle w^i$, this holds for all $i \in I$. This proves the claim.

The proof for a is identical to that for θ and therefore omitted for the sake of brevity.

Suppose that $\triangle w^{'}_{h'} > \triangle w_{h'}$ and $\triangle w^{i'} > \triangle w^i$. Then supermodularity requires:

$$\begin{aligned}
&U(\theta, a, \triangle w^{'}_{h'}, \triangle w^{i'}, \triangle w_{-hh'}) - U(\theta', a, \triangle w^{'}_{h'}, \triangle w^{i}, \triangle w_{-hh'}) \geq \\
&U(\theta, a, \triangle w_{h'}, \triangle w^{i'}, \triangle w_{-hh'}) - U(\theta, a, \triangle w_{h'}, \triangle w^{i}, \triangle w_{-hh'}).
\end{aligned}$$

Observe that this inequality is equal to:

$$(\triangle w^{'}_{h'} - \triangle w_{h'}) P^i(\theta, a) \geq (\triangle w^{'}_{h'} - \triangle w_{h'}) P^i(\theta, a).$$

Because $P^i(\theta, a) > 0$ and $\triangle w^{i'} > \triangle w^i$, this holds for all $i \in I$. This proves the claim. □

Proof of Proposition 1 We restrict attention to actions and contracts that are piecewise continuously differentiable. The agent's problem in terms of his revelation of type θ' and choice of action profile a', or the equivalent in terms of his choice from the menu of contracts offered by the principal, is:

$$\max_{\theta' \in \Theta, a' \in A} U(\theta, w(\theta'), a').$$

For any reported type $\theta' \in \Theta$, the first-order condition for a for any type-θ is given by:

$$\sum\nolimits_i p_a^i(\theta, a') w^i(\theta') - c_a(a') \leq 0, \;\; a' \geq 0 \; \forall \theta \in \Theta \text{ and}$$
$$\left(\sum\nolimits_i p_a^i(\theta, a') w^i(\theta') - c_a(a')\right) a = 0. \tag{A.1}$$

Let the solution to this problem be $\alpha(\theta, \theta')$, and when $\theta' = \theta$ the solution is denoted by $a(\theta)$. Supermodularity of $U(\theta, w, a)$ in (θ, w, a) ensures its existence. If $\alpha(\theta, \theta') > 0$, because the objective function is jointly continuous in (θ, a), this implies that $\alpha(\theta, \theta')$ is continuous in (θ, θ'). Conditions **MONA** and **CP** ensure that, for any contract for which $\triangle w^i(\theta') > 0$ for some i, $U(\theta, w, a)$ is strictly concave in a and $\alpha(\theta, \theta') > 0$ for all $\theta, \theta' \in \Theta$. Thus, local conditions are necessary and sufficient for $\alpha(\theta, \theta')$ to be the optimal effort for all $\theta, \theta' \in \Theta$. To see this result, observe that:

$$U(\theta, w(\theta'), a) = w_0 + \sum\nolimits_{i=0}^{n-1} \triangle w^i(\theta') P^i(\theta, a) - c(a)$$

Thus,

$$U_a(\theta, w(\theta'), a) = \sum\nolimits_{i=0}^{n-1} \triangle w^i(\theta') P_a^i(\theta, a) - c_a(a)$$

and

$$U_{aa}(\theta, w(\theta'), a) = \sum\nolimits_{i=0}^{n-1} \triangle w^i(\theta') P_{aa}^i(\theta, a) - c_{aa}(a)$$

Because assumption **CP** ensures that $P^i(\theta, a)$ is strictly concave in a and $c(a)$ is convex in a, this, together with **MONA**, implies that $U(\theta, w(\theta'), a)$ is strictly concave in a for all $\theta, \theta' \in \Theta$. Together with the Inada's type conditions assumed in assumption **CP**, this ensures that $\alpha(\theta, \theta') > 0$ for all $\theta, \theta' \in \Theta$.

Equation (A.1) is a necessary condition for optimality of $\alpha(\theta, \theta')$. We also need the following to hold for all $\theta \in \Theta$,

$$U_{aa}(\theta, w(\theta'), \alpha(\theta, \theta'))|_{\theta'=\theta} \leq 0. \tag{A.2}$$

It follows from the above that:

$$U_{aa}(\theta, w(\theta'), \alpha(\theta, \theta')) \frac{\partial \alpha(\theta, \theta')}{\partial \theta} + \sum\nolimits_i p_{a\theta}^i(\theta, \alpha(\theta, \theta')) w^i(\theta') = 0 \tag{A.3}$$

and

$$U_{aa}(\theta, w(\theta'), \alpha(\theta, \theta')) \frac{\partial \alpha(\theta, \theta')}{\partial \theta'} + \sum\nolimits_i p_a^i(\theta, \alpha(\theta, \theta')) \frac{d w^i(\theta')}{d\theta'} = 0 \tag{A.4}$$

Maximizing at a point of differentiability yields the first-order condition with respect to θ'; for almost all $\theta \in \Theta$,

$$\sum_i p^i(\theta, \alpha(\theta, \theta')) \frac{dw^i(\theta')}{d\theta'} + \Big(\sum_i p^i_a(\theta, \alpha(\theta, \theta')) w^i(\theta') - c_a(\alpha(\theta, \theta')) \Big) \frac{\partial \alpha(\theta, \theta')}{\partial \theta'} \Big|_{\theta'=\theta} = 0 \quad \text{(A.5)}$$

Using the first-order condition for the action profile, and assuming that $\alpha(\theta, \theta') > 0$, this can be re-written as follows: for almost all $\theta \in \Theta$,

$$\sum_i p^i(\theta, \alpha(\theta, \theta')) \frac{dw^i(\theta')}{d\theta'} \Big|_{\theta'=\theta} = 0. \quad \text{(A.6)}$$

This shows that Eq. (A.6) is necessary for implementability.

Next, we need to show that the indirect utility is concave in θ' when evaluated at $\theta' = \theta$. That is,

$$\frac{\partial^2}{\partial \theta' \partial \theta'} U(\theta, w(\theta'), \alpha(\theta, \theta')) \Big|_{\theta'=\theta} \leq 0.$$

Because Eq. (A.6) holds as an identity in θ when evaluated at the optimal action $\alpha(\theta, \theta')$, one can total differentiate (A.6) to find:

$$\frac{\partial^2}{\partial \theta' \partial \theta'} U(\theta, w(\theta'), \alpha(\theta, \theta')) + \frac{\partial^2}{\partial \theta' \partial \theta} U(\theta, w(\theta'), \alpha(\theta, \theta')) \Big|_{\theta'=\theta} = 0. \quad \text{(A.7)}$$

This implies that, for almost all $\theta \in \Theta$, the local second-order condition can be written as:

$$\frac{\partial^2}{\partial \theta' \partial \theta} U(\theta, w(\theta'), \alpha(\theta, \theta')) \Big|_{\theta'=\theta} \geq 0 \quad \text{(A.8)}$$

Its derivative is given by:

$$\Bigg(\sum_i p^i_\theta(\theta, \alpha(\theta, \theta')) \frac{dw^i(\theta')}{d\theta'} + \sum_i p^i_{a\theta}(\theta, \alpha(\theta, \theta')) w^i(\theta') \frac{\partial \alpha(\theta, \theta')}{\partial \theta'} + \Big(U_{aa}(\theta, w(\theta'), \alpha(\theta, \theta')) \frac{\partial \alpha(\theta, \theta')}{\partial \theta'} + \sum_i p^i_a(\theta, a') \frac{dw^i(\theta')}{d\theta'} \Big) \frac{\partial \alpha(\theta, \theta')}{\partial \theta} + U_a(\theta, w(\theta'), \alpha(\theta, \theta')) \frac{\partial^2 \alpha(\theta, \theta')}{\partial \theta' \partial \theta} \Bigg) \Big|_{\theta'=\theta} \geq 0.$$

Noticing that $U_a(\theta, w(\theta'), \alpha(\theta, \theta')) = 0$, and that the second term multiplying $\frac{\partial \alpha(\theta, \theta')}{\partial \theta}$ is zero due to Eq. (A.4) and substituting Eq. (A.3) into this, we deduce that the local second-order condition is as follows:

$$\frac{1}{U_{aa}(\theta, w(\theta'), \alpha(\theta, \theta'))}\Bigg(\sum\nolimits_i p_\theta^i(\theta, \alpha(\theta, \theta'))\frac{dw^i(\theta')}{d\theta'}U_{aa}(\theta, w(\theta'), \alpha(\theta, \theta'))$$
$$-\sum\nolimits_i p_{a\theta}^i(\theta, \alpha(\theta, \theta'))w^i(\theta')\sum\nolimits_i p_a^i(\theta, \alpha(\theta, \theta'))\frac{dw^i(\theta')}{d\theta'}\Bigg)\Bigg|_{\theta'=\theta} \geq 0. \qquad \text{(A.9)}$$

Because $U_{aa}(\theta, w(\theta'), \alpha(\theta, \theta')) \leq 0$, a necessary condition for implementability is that the term inside the parenthesis in Eq. (A.9) is negative. This gives rise to the equation in the main text.

Global Incentive Constraints: Suppose next that (A.1), (A.2) (A.6) and (A.9) hold. Then, it must be the case that all agent's incentive-compatibility conditions hold. To see this result, suppose by contradiction that for at least one type θ the agent's incentive compatibility is violated. Then there exists a type $\theta' \neq \theta$, such that the following holds:

$$\sum\nolimits_i p^i(\theta, \alpha(\theta, \theta'))w^i(\theta') - c(\alpha(\theta, \theta')) > \sum\nolimits_i p^i(\theta, a(\theta))w^i(\theta) - c(a(\theta)).$$

Integrating this re-writes as follows:

$$\int_{\theta'}^{\theta}\left(\sum\nolimits_i p^i(\theta, a(s))\frac{dw^i(s)}{ds} + \left(\sum\nolimits_i p_a^i(\theta, a(s))w^i(s) - c_a(a(s))\right)\frac{da(s)}{ds}\right)ds < 0.$$

The local incentive constraints, together with (A.9), implies that, if $\theta > \theta'$:

$$\int_{\theta'}^{\theta}\left(\sum\nolimits_i p^i(\theta, a(s))\frac{dw^i(s)}{ds} + \left(\sum\nolimits_i p_a^i(\theta, a(s))w^i(s) - c_a(a(s))\right)\frac{da(s)}{ds}\right)ds \geq 0,$$

which leads to a contradiction. If $\theta < \theta'$, the same reasoning leads us to a similar contradiction. This shows that Eqs. (A.1), (A.6), (A.9) are sufficient conditions for incentive compatibility. □

Proof of Proposition 2 It readily follows from Theorem 2 in Milgrom and Segal (2002) that:

$$\dot{U}(\theta) = U(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} P_\theta^i(x, a(x))\triangle w^i(x)f(x)dx,$$

where $P_\theta^i(\theta, a) \equiv \sum_{j\geq i+1} p_\theta^j(\theta, a) \geq 0,\ \forall i \in I \setminus \{n\}$ and the inequality is due to **FOSD**.

After summation by parts the virtual surplus can be re-written as follows:

$$\Phi(\theta, w(\theta), a(\theta)) = (S(\theta, a(\theta)) - H(\theta)\sum\nolimits_{i\neq n} P_\theta^i(\theta, a(\theta))\triangle w^i(\theta))f(\theta) - U(\underline{\theta}).$$

Observe also that the agent's first-order condition for the action re-writes as follows:

$$\sum\nolimits_{i\neq n} P_a^i(\theta, a(\theta))\triangle w^i(\theta) - c_a(a(\theta)) = 0,$$

where $P_a^i(\theta, a) \equiv \sum_{j \geq i+1} p_a^j(\theta, a) \geq 0,\ \forall i \in I \setminus \{n\}$ due to **FOSD**.

Let $\gamma(\theta)$ be the Lagrange multiplier for the first-order condition for the action, $\delta^i(\theta)$ be the multiplier corresponding to **MONA**, and $\eta^i(\theta)$ be the multiplier corresponding to **MONP**. Then, for each $\theta \in \Theta$, the first-order condition with respect to $\triangle w^i$ evaluated at the optimal mechanism, denoted by $(\hat{w}(\theta), \hat{a}(\theta))$, is as follows:

$$-H(\theta)P_\theta^i(\theta, \hat{a}(\theta))f(\theta) + \gamma(\theta)P_a^i(\theta, \hat{a}(\theta)) + \delta^i(\theta) - \eta^i(\theta) = 0, \tag{A.10}$$

$$\delta^i(\theta) \geq 0,\ \delta^i(\theta)\triangle\hat{w}^i(\theta) = 0,\ \triangle\hat{w}^i(\theta) \geq 0, \forall\, i \in I \setminus \{n\}, \tag{A.11}$$

$$\eta^i(\theta) \geq 0,\ \eta^i(\theta)(\triangle y^i - \triangle\hat{w}^i(\theta)) = 0,\ \triangle\hat{w}^i(\theta) \leq \triangle y^i, \forall\, i \in I \setminus \{n\}, \tag{A.12}$$

and the one with respect to a is given by:

$$\Big(S_a(\theta, \hat{a}(\theta)) - H(\theta)\sum\nolimits_{i \neq n} P_{\theta a}^i(\theta, \hat{a}(\theta))\triangle\hat{w}^i(\theta)\Big)f(\theta) + \gamma(\theta)\Big(\sum\nolimits_{i=0}^{n-1} P_{aa}^i(\theta, \hat{a}(\theta))\triangle\hat{w}^i(\theta) - c_{aa}(\hat{a}(\theta))\Big) = 0. \tag{A.13}$$

It readily follows from Eq. (A.10) that, for all $\theta \in \Theta$, $\triangle\hat{w}^i(\theta) = \triangle y^i$ if

$$\gamma(\theta)\frac{P_a^i(\theta, \hat{a}(\theta))}{P_\theta^i(\theta, \hat{a}(\theta))} \geq H(\theta)f(\theta),$$

and $\triangle\hat{w}^i(\theta) = 0$ otherwise. To save on notation, define $\xi^i(\theta, a) \equiv \gamma(\theta)\frac{P_a^i(\theta,a)}{P_\theta^i(\theta,a)}$ and $\xi(\theta) = H(\theta)f(\theta)$. Then, $\triangle\hat{w}^i(\theta) = \triangle y^i$ if $\xi^i(\theta, a) \geq \xi(\theta)$ and $\triangle\hat{w}^i(\theta) = 0$ otherwise.

It readily follows from (A.13) that:

$$\gamma(\theta) = -\frac{S_a(\theta, \hat{a}(\theta)) - H(\theta)\sum_{i \neq n} P_{\theta a}^i(\theta, \hat{a}(\theta))\triangle\hat{w}^i(\theta)}{\sum_{i \neq n} P_{aa}^i(\theta, \hat{a}(\theta))\triangle\hat{w}^i(\theta) - c_{aa}(\hat{a}(\theta))}f(\theta). \tag{A.14}$$

Multiplying the first-order condition in Eq. (A.10) by $\triangle\hat{w}^i(\theta)$, summing over all i and substituting into the first-order condition for the action, one gets that:

$$\gamma(\theta)c_a(\hat{a}(\theta)) = H(\theta)\sum\nolimits_{i=0}^{n-1} P_\theta^i(\theta, \hat{a}(\theta))\triangle\hat{w}^i(\theta)f(\theta) + \sum\nolimits_{i=0}^{n-1} \eta^i(\theta)\triangle\hat{w}^i(\theta) \geq 0.$$

It readily follows from this that $\gamma(\theta) > 0$ for all $\theta \in \Theta \setminus \bar{\theta}$, since there exists an $i \in I \setminus \{n\}$, such that $\triangle\hat{w}^i(\theta) > 0$; otherwise $\hat{a}(\theta) = 0$. We deduce from the FOC in Eq. (A.13), that $S_a(\theta, a(\theta)) > 0$ for all $\theta \in \Theta \setminus \bar{\theta}$ and, therefore, $\hat{a}(\theta) < a^{**}(\theta)$ for all $\theta \in \Theta \setminus \bar{\theta}$.

Setting $H(\bar{\theta}) = 0$ in Eq. (A.10), we get that: $\gamma(\bar{\theta})P_a^i(\bar{\theta}, \hat{a}(\bar{\theta})) + \delta^i(\bar{\theta}) = \eta^i(\bar{\theta})$. Suppose that $\gamma(\bar{\theta}) > 0$, then because $P_a^i(\bar{\theta}, \hat{a}(\bar{\theta})) > 0$, $\eta^i(\bar{\theta}) > 0$ for all $i \in I \setminus \{n\}$ and therefore $\triangle\hat{w}^i(\bar{\theta}) = \triangle y^i$ for all $i \in I \setminus \{n\}$. We deduce from the FOC in Eq.

(A.13) that $\hat{a}(\bar{\theta}) = a^{**}(\bar{\theta})$ and, therefore, $S_a(\bar{\theta}, a(\bar{\theta})) = 0$. It follows from Eq. (A.14), that this implies that $\gamma(\bar{\theta}) = 0$, which is a contradiction. Suppose that $\gamma(\bar{\theta}) = 0$, then $\delta^i(\bar{\theta}) = \eta^i(\bar{\theta}) = 0$ for all $i \in I \setminus \{n\}$. It follows from Eq. (A.14) that $S_a(\bar{\theta}, a(\bar{\theta})) = 0$, and therefore $\hat{a}(\bar{\theta}) = a^{**}(\bar{\theta})$.

Observe that condition **CP**, together with the following assumption that $\frac{P^i_{a\theta}(\theta,a(\theta))}{P^i_a(\theta,a(\theta))}$ is non-decreasing with a, ensures concavity of $\Phi(\cdot)$ with respect to a. A sufficient condition for this is that $P^i_{aa\theta}(\theta, a) \geq 0$ for all $i \in I \setminus \{n\}$. In addition, notice that the optimal action increases with θ only if conditions **SC** and **CP**, together with the following assumption that $\frac{P^i_{a\theta}(\theta,a(\theta))}{P^i_a(\theta,a(\theta))}$ is non-increasing with θ hold. A sufficient condition for this is that $P^i_{a\theta\theta}(\theta, a) \leq 0$ for all $i \in I \setminus \{n\}$, since $H(\theta)$ is non-increasing with θ. Also, observe that the determinant of the Hessian regarding $\Phi(\theta, w, a)$ is equal to zero due to the linearity of the objective function with respect to $\triangle w^i$. □

Proof of Proposition 3 Assume that $p(\theta, a)$ satisfies **WSEP**, and define $Q^i = \sum_{j \geq i+1}(q^j - r^j)$. Then, it is easy to see that the principal's problem can be written as follows:

$$\max_{(w,a):\Theta \to \mathcal{C} \times A} \int_{\theta \in \Theta} \Big(S(\theta, a(\theta)) - H(\theta) g_\theta(\theta, a(\theta)) \sum\nolimits_{i \neq n} Q^i \triangle w^i(\theta) \Big) f(\theta) d\theta - U(\underline{\theta})$$

subject to

$$\sum\nolimits_{i \neq n} Q^i \triangle \dot{w}^i(\theta) \geq 0,$$

$$g_a(\theta, a(\theta)) \sum\nolimits_{i \neq n} Q^i \triangle w^i(\theta) - c_a(a(\theta)) = 0,$$

$$\triangle w^i(\theta) \in [0, \triangle y^i], \ \forall\, i \in I \setminus \{n\},$$

$$U(\underline{\theta}) \geq L.$$

Suppose that there exists an incentive-compatible mechanism $(w(\theta), a(\theta))$; we will then show that there is another mechanism that pays the same to everyone, denoted by w^*, and asks each agent to exert the effort $a^*(\theta) \equiv \text{argmax}_{a \in A}\, U(\theta, w^*, a)$ which increases the principal's payoff pointwise and is incentive compatible.

Let $(w(\theta), a(\theta))$ be an incentive-compatible mechanism and let contract w^* be the contract with $w^{0*} = \inf\{w(\hat{\theta}) : \hat{\theta} \in \Theta\}$ and $\sum_{i \neq n} Q^i \triangle w^{i*} = \sup\big\{ \sum_{i \neq n} Q^i \triangle w^{i*}(\hat{\theta}) : \hat{\theta} \in \Theta \big\}$. If $w^{0*} > L$, reducing all payments uniformly will keep all constraints satisfied and will increase the principal's expected payoff. Thus, we can assume that $w^{0*} = L$. In addition, either there is a contract in the menu such that the sup and inf are attained, or w^* is a limit point of $\{w(\hat{\theta}) : \hat{\theta} \in \Theta\}$ due to the piecewise continuity of $w(\hat{\theta})$.

Because $w(\theta)$ is incentive compatible, holding the chosen effort constant, the following holds:

$$\begin{aligned} & w_0(\theta) + \sum\nolimits_{i=0}^{n-1} \triangle w^i(\theta) g(\theta, a(\theta))) Q^i - c(a(\theta)) \\ & \geq L + \sum\nolimits_{i=0}^{n-1} \triangle w^{i*} g(\theta, a(\theta)) Q^i - c(a(\theta)), \ \forall \theta \in \Theta. \end{aligned}$$

It readily follows from this that, for all $\theta \in \Theta$,

$$w_0(\theta) - L + \sum_{i=0}^{n-1} (\triangle w^i(\theta) - \triangle w^{i*}) g(\theta, a(\theta))) Q^i \geq 0 \tag{A.15}$$

We deduce from the above that expected compensation is higher under the incentive-compatible mechanism that under the alternative menu w^* whenever the agent chooses the incentive-compatible action $a(\theta)$. When the agent chooses the action $a^*(\theta) \in \text{argmax}_{a \in A} U(\theta, w^*, a)$, we have the following: for all $\theta \in \Theta$,

$$\begin{aligned}
&y^0 - w^0(\theta) + \sum_{i=0}^{n-1} (\triangle y - \triangle w^i(\theta)) g(\theta, a(\theta))) Q^i \\
&\quad - \Big(y^0 - L + \sum_{i=0}^{n-1} (\triangle y - \triangle w^{i*}) g(\theta, a^*(\theta)) Q^i \Big) \\
&= L - w^0(\theta) + \sum_{i=0}^{n-1} (\triangle y - \triangle w^{i*})(g(\theta, a(\theta)) - g(\theta, a^*(\theta))) Q^i \\
&\quad + g(\theta, a(\theta)) \sum_{i=0}^{n-1} (\triangle w^{i*} - \triangle w^i(\theta)) Q^i \\
&\leq 0,
\end{aligned}$$

where the inequality follows from Eq. (A.15), the fact that **FOSD**, **CP** and **SC** imply that $a^*(\theta)$ exists and $a^*(\theta) \geq a(\theta)$ and **MONP** implies that $\triangle y - \triangle w^{i*} \geq 0$ for all $i \in I$. Thus, the principal's payoff is higher when menu w^* is offered than when any other incentive-compatible menu is offered.

Taking into account the fact that the incentive-compatible contract that yields the largest payoff to the principal is type-independent, the principal's problem becomes the following:

$$\max_{(w,a):\Theta \to \mathcal{C} \times A} \int_{\theta \in \Theta} \Big(S(\theta, a(\theta)) - H(\theta) g_\theta(\theta, a(\theta)) \sum_{i \neq n} Q^i \triangle w^i \Big) f(\theta) d\theta - U(\underline{\theta})$$

subject to

$$g_a(\theta, a(\theta)) \sum_{i \neq n} Q^i \triangle w^i - c_a(a(\theta)) = 0,$$

$$\triangle w^i \in [0, \triangle y^i], \ \forall\, i \in I \setminus \{n\},$$

$$U(\underline{\theta}) \geq L,$$

since for a type-independent contract trivially satisfies the ignored constraint: $\sum_{i \neq n} Q^i \triangle \dot{w}^i(\theta) \geq 0$.

It is easy to see that it is optimal to set $U(\underline{\theta}) = L$.

The first-order condition for w^i is given by:

$$\big(- H(\theta) g_\theta(\theta, a(\theta)) + \gamma(\theta) g_a(\theta, a(\theta)) \big) Q^i + \delta^i(\theta) - \eta^i(\theta) = 0 \tag{A.16}$$

and that for a is given by

$$S_a(\theta, a) f(\theta) - H(\theta) g_{a\theta}(\theta, a(\theta)) f(\theta) \sum_{i \neq n} Q^i \triangle w^i$$
$$+ \gamma(\theta)\big(g_{aa}(\theta, a(\theta)) \sum_{i \neq n} Q^i \triangle w^i - c_{aa}(a(\theta))\big) = 0. \tag{A.17}$$

It readily follows from Eq. (A.16) that, if $H(\theta) g_\theta(\theta, a(\theta)) f(\theta) - \gamma(\theta) g_a(\theta, a(\theta)) > 0$, then $\delta^i(\theta) > 0$ for all $i \in I \setminus \{n\}$. Therefore, $\triangle w^i = 0$ for all $i \in I \setminus \{n\}$, which implies that $a(\theta) = 0$ for all $\theta \in \Theta$. This can never be optimal. Thus, $H(\theta) g_\theta(\theta, a(\theta)) f(\theta) - \gamma(\theta) g_a(\theta, a(\theta)) \leq 0$. This requires that $\gamma(\theta) \geq 0$ for all $\theta \in \Theta$.

It readily follows from this and Eq. (A.17) that:

$$\gamma(\theta) = -\frac{S_a(\theta, a(\theta)) - H(\theta) g_{a\theta}(\theta, a(\theta)) \sum_{i \neq n} Q^i \triangle w^i}{g_{aa}(\theta, a(\theta)) \sum_{i \neq n} Q^i \triangle w^i - c_{aa}(a(\theta))} f(\theta), \ \forall \theta \in \Theta, \tag{A.18}$$

since $g_{aa}(\theta, a(\theta)) \sum_{i \neq n} Q^i \triangle w^i - c_{aa}(a(\theta)) < 0$.

This implies that $S_a(\theta, a(\theta)) > 0, \ \forall \theta \in \Theta \setminus \{\bar{\theta}\}$. Hence, the optimal action, denoted by $a^*(\theta)$, satisfies the following: $a^*(\theta) < a^{**}(\theta), \ \forall \theta \in \Theta \setminus \{\bar{\theta}\}$.

Suppose that $H(\theta) g_\theta(\theta, a(\theta)) f(\theta) - \gamma(\theta) g_a(\theta, a(\theta)) < 0$, then it readily follows from Eq. (A.16) that $\eta^i(\theta) > 0$ for all $i \in \setminus\{n\}$, which implies that $\triangle w^i = \triangle y^i$ for all $i \in I \setminus \{n\}$. This, together with the agent's first-order condition in Eq. (A.17), implies that $a^*(\theta) = a^{**}(\theta), \ \forall \theta \in \Theta$. This contradicts the fact that if $\gamma(\theta) > 0$, then $a^*(\theta) < a^{**}(\theta)$.

Thus, $H(\theta) g_\theta(\theta, a(\theta)) f(\theta) - \gamma(\theta) g_a(\theta, a(\theta)) = 0$, then $\eta^i(\theta) = \delta^i(\theta) = 0$ and $\triangle w^i \in [0, \triangle y^i]$. This implies that $\gamma(\theta) > 0$ for all $\theta \in \Theta \setminus \{\bar{\theta}\}$ and $\gamma(\bar{\theta}) = 0$. Thus, $a^*(\theta) < a^{**}(\theta), \ \forall \theta \in \Theta \setminus \{\bar{\theta}\}$ and $a^*(\bar{\theta}) = a^{**}(\bar{\theta})$.

Observe that condition **CP**, together with the following assumption that $\frac{g_{a\theta}(\theta, a(\theta))}{g_a(\theta, a(\theta))}$ is non-decreasing with a, ensures concavity of $\Phi(\cdot)$ with respect to a. A sufficient condition for this is that $g_{aa\theta}(\theta, a) \geq 0$ for all $i \in I \setminus \{n\}$. In addition, notice that the optimal action increases with θ only if conditions **SC** and **CP**, together with the following assumption that $\frac{P^i_{a\theta}(\theta, a(\theta))}{P^i_a(\theta, a(\theta))}$ is non-increasing with θ hold.

A sufficient condition for this is that $P^i_{a\theta\theta}(\theta, a) \leq 0$ for all $i \in I \setminus \{n\}$, since $H(\theta)$ is non-increasing with θ. Also, observe that the determinant of the Hessian regarding $\Phi(\theta, w, a)$ is equal to zero due to the linearity of the objective function with respect to $\triangle w^i$. □

Proof of Proposition 4 Let $a(\theta)$ be the incentive-compatible action and w be a type-independent mechanism that implements $a(\theta)$. Because under **WSEP** the agent's incentive-compatible action satisfies the following:

$$-g_a(\theta, a(\theta)) \sum_{i=0}^{n-1} \sum_{j=0}^{i} (q^j - r^j) \triangle w^i - c_a(a(\theta)) = 0,$$

then, any other type-independent mechanism $\tilde{w}$ that satisfies the following:

$$\sum_{i=0}^{n-1} \sum_{j=0}^{i} (q^j - r^j) \triangle \tilde{w}^i = W \equiv \sum_{i=0}^{n-1} \sum_{j=0}^{i} (q^j - r^j) \triangle w^i \tag{A.19}$$

is also incentive-compatible. Hence, any incentive-compatible, type-independent mechanism must solve the following:

$$\max_{w \in \mathcal{C}} \int_{\theta \in \Theta} V(\theta, a(\theta), w) d\theta \text{ subject to (A.19).}$$

Let λ be the Lagrange multiplier for constraint (A.19), δ^i the Lagrange multiplier for **MONA**, and η^i the Lagrange multiplier for **MONP**. Then, the first-order condition is given by:

$$-\int_{\theta \in \Theta} (1 - P^i(\theta, a(\theta)) dF(\theta) + \lambda \sum\nolimits_{j=0}^{i} (q^j - r^j) + \delta^i - \eta^i = 0 \quad \text{(A.20)}$$

Because λ is independent of i, and the first-order condition is independent of w, the solution is bang-bang. Observe that, if $\delta^i \geq 0$, then $\eta^i = 0$ and vice versa. It is easy to see that **FOSD** implies that:

$$\delta^i \geq 0 \Leftrightarrow \frac{\int_{\theta \in \Theta} (1 - P^i(\theta, a(\theta)) dF(\theta)}{\sum_{j=0}^{i} (q^j - r^j)} - \lambda \leq 0,$$

and $\eta^i > 0$ otherwise. Hence, if the ratio on the RHS is increasing in i, the optimal contract is option-like. Observe that, for any h,

$$\frac{\int_{\theta \in \Theta} (1 - P^{i+1}(\theta, a(\theta)) dF(\theta)}{\sum_{j=0}^{i+1} (q^j - r^j)} \geq \frac{\int_{\theta \in \Theta} (1 - P^i(\theta, a(\theta)) dF(\theta)}{\sum_{j=0}^{i} (q^j - r^j)}.$$

After a few steps of simple algebra, it can be easily shown that this is equivalent to:

$$(q_{i+1} - r_{i+1})(1 - \sum\nolimits_{j=0}^{i} r^j) + r_{i+1} \sum\nolimits_{j=0}^{i} (q^j - r^j) \leq 0.$$

Define $l^i \equiv q^i / r^i$. Notice that **EMLRP** implies that, for any $i \in I \setminus \{n\}$, $l^{i+1} \geq l^i$. Using this notation, the last equation re-writes as follows:

$$r^{i+1}\big((l^{i+1} - 1)(1 - \sum\nolimits_{j=0}^{i} r^j) + \sum\nolimits_{j=0}^{i} r^j (l^j - 1)\big) \leq 0$$
$$\Leftrightarrow l^{i+1} \leq 1 + \sum\nolimits_{j=0}^{i} r^j (l^{i+1} - l^j)$$

Because of **EMLRP**, the RHS and the LHS increase with i. In addition, it is easy to check that, at $i = n - 1$, the RHS is equal to the LHS and is lower than the RHS at $i = 0$, and the LHS increases at a rate $l^{i+2} - l^{i+1}$, while the RHS does so at a rate $r^{i+1}(l^{i+2} - l^{i+1})$. Hence, putting this together we get that the inequality above holds for all $i \in I \setminus \{n\}$. □

Proof of Proposition 5 The agent's problem in terms of his revelation of type θ', or the equivalent in terms of his choice from the menu of contracts offered by the principal, is:

$$\max_{\theta' \in \Theta} U(w(\theta'), a(\theta'), \theta).$$

The first-order condition is given by:

$$\Big(\sum\nolimits_i p^i(\theta, a(\theta'))\frac{dw^i(\theta')}{d\theta'} + \Big(\sum\nolimits_i p^i_a(\theta, a(\theta'))w^i(\theta') - c_a(a(\theta'))\Big)\frac{da(\theta')}{d\theta'}\Big)\Big|_{\theta'=\theta} = 0, \ \forall\theta \in \Theta. \tag{A.21}$$

Because this holds as an identity in θ, we can do a total diferentiation to find:

$$\Big(\frac{\partial U(\theta, w(\theta'), a(\theta'))}{\partial\theta'\partial\theta'} + \frac{\partial U(\theta, w(\theta'), a(\theta'))}{\partial\theta'\partial\theta}\Big)\Big|_{\theta'=\theta} = 0 \tag{A.22}$$

Because the second-order condition requires the first term to be non-positive, this implies that:

$$\Big(\sum\nolimits_i p^i_\theta(\theta, a(\theta'))\frac{dw^i(\theta')}{d\theta'} + \sum\nolimits_i p^i_{a\theta}(\theta, e(\theta'))w^i(\theta')\frac{da(\theta')}{d\theta'}\Big)\Big|_{\theta'=\theta} \geq 0. \tag{A.23}$$

The missing details of the proof resemble that of proposition 1 and, for the sake of brevity, are omitted. □

B Pure moral hazard case: not to be published

In this subsection, we analyze the case in which the agent's ability is known to both the principal and the agent before signing the contract, but the action is not observed by the principal.[17]

Let the likelihood ratio be $\frac{p^i(\theta,a')}{p^i(\theta,a)}$ for any $a' > a$. In this section we assume the following:

[MLRP] $\frac{p^i(\theta,a')}{p^i(\theta,a)}$ increases with i for all $\theta \in \Theta$.

This ensures that any principal who has a payoff function increasing in output prefers the stochastic distribution of returns induced by higher actions. We will show here

[17] There is another possible benchmark in which there is no moral hazard but there is adverse selection. However, this case has a trivial solution, which is to offer a fixed wage contract that pays $w^i(\theta) = c(a^*(\theta))$ in each state, where $a^*(\theta)$ is the first-best efficient action. It is easy to check that this contract implements the first-best and induces truthful revelation of information (each type is indifferent between revealing his true type or any other type). The reason stems from the fact that neither the cost function nor the outside utility depends on the agent's type θ.

that the optimal contract under this assumption is option-like, as in Innes (1990) and Matthews (2001).

Suppose that an agent of type $\theta \in \Theta$ is faced with contract $w \in \mathcal{C}$ and chooses action $a \in A$; his expected utility is then given by:

$$U(\theta, w, a) \equiv p(\theta, a)w - c(a). \tag{B.1}$$

Observe that the linearity of the agent's payoff function in Eq. (B.1) with respect to payments implies that if the principal could offer a contract in which payments are unbounded, the first-best action could be implemented by offering a contract that pays the output minus a face value after each output realization. However, this is prevented by the limited-liability constraint.

Contract w induces an agent of type θ to choose action $a \in A$ if and only if the following incentive-compatibility constraint holds:

$$U(\theta, w, a) \geq U(\theta, w, a'), \quad \forall a' \in A, \tag{B.2}$$

and induces him to participate if the following individually-rationality constraint holds:

$$U(\theta, w, a) \geq 0. \tag{B.3}$$

Let the principal's expected utility be $V(\theta, w, a) \equiv p(\theta, a)(y - w)$. Then, the principal's problem when faced with an agent of type $\theta \in \Theta$ is the following:

$$\max_{(w,a)\in\mathcal{C}\times A} V(\theta, w, a) \text{ subject to (B.2) and (B.3)}$$

Let $a(\theta, w)$ be largest element satisfying $a(\theta, w) \in \text{argmax}_{a\in A}\, U(\theta, w, a)$. Then, the principal's problem is given by:

$$\max_{w\in\mathcal{C}} V(\theta, w, a(\theta, w)) \text{ subject to } U(\theta, w, a(\theta, w)) \geq 0. \qquad \text{P-MH}$$

Proposition 6 (Pure Moral Hazard) *Suppose **CP** and **MLRP** hold. Then,*

(i) *The optimal contract is a call-option contract of the form $w^i(\theta) = \max\{0, y^i - \bar{y}(\theta)\}$, and the optimal action satisfies $a^m(\theta) \leq a^*(\theta)$; and*
(ii) *If **SC** holds, $a^m(\theta)$ rises with θ and $\bar{y}(\theta)$ falls with θ.*

Proof Define the call-option contract with face value $\bar{y}$ by $o^i(\bar{y}) = \max\{y^i - \bar{y}, 0\}$, $\forall i \in I$. Under this contract, the principal's payoff is given by $d^i(\bar{y}) = y^i - o^i(\bar{y}) = \min\{y^i, \bar{y}\}$. In addition, consider any other contract w inducing a return to the principal equal to $z^i = y^i - w^i$. Then, the following lemma, which is closely related to Lemma 1 in Matthews (2001), shows that $p(\theta, a)d'$ and $p(\theta, a)z'$ single-cross at the action a satisfying $dp(\theta, a) = zp(\theta, a)$, because the payments specified by d are at a maximum for low-output realizations and at a minimum for high-output realizations, given monotonicity constraints.

Lemma 2 *Suppose that* ***MLRP*** *holds. Then, for any* $(\theta, a) \in \Theta \times A$ *and* $w \in \mathcal{C}$, *and a call-option contract* $o \neq w$, *if* $p(\theta, a)o = p(\theta, a)w$, *then:*

$$p(\theta, a')o \lesseqgtr p(\theta, a')w \text{ for } a' \gtreqless a$$

Proof Let $v = z - d$. Routine algebra proves the result if, for all $a' \in A$, the equality $vp(\theta, a) = 0$ implies that $v(p(\theta, a') - p(\theta, a)) > 0, \forall a' > a$, and $v(p(\theta, a') - p(\theta, a)) < 0, \forall a' < a$.[18] Let $vp(\theta, a) = 0$. Since $v \neq 0$ and $p(\theta, a)$ has full support, v takes positive and negative values. Let k be the largest i such that $v_i < 0$. Because o is a call-option contract and w satisfies **LL**, **MONA** and **MONP**, $v_i \leq 0$ for all $i \leq k$ and $v_i > 0$ for all $i > k$. Thus, because v has a positive component (since $vp(\theta, a) = 0$), **MLRP** and $vp(\theta, a) = 0$ implies the following for all $a' > a$:

$$\begin{aligned} v(p(\theta, a') - p(\theta, a)) &= \sum\nolimits_{i=0}^{n} \left(\frac{p^i(\theta, a')}{p^i(\theta, a)} - 1\right) p^i(\theta, a) v^i \\ &> \sum\nolimits_{i=0}^{k} \left(\frac{p^k(\theta, a')}{p^k(\theta, a)} - 1\right) p^i(\theta, a) v^i \\ &\quad + \sum\nolimits_{i=k+1}^{n} \left(\frac{p^k(\theta, a')}{p^k(\theta, a)} - 1\right) p^i(\theta, a) v^i \\ &= \left(\frac{p^k(\theta, a')}{p^k(\theta, a)} - 1\right) \sum\nolimits_{i=0}^{n} p^i(\theta, a) v^i \\ &= \left(\frac{p^k(\theta, a')}{p^k(\theta, a)} - 1\right) vp(\theta, a) = 0. \end{aligned}$$

The argument is identical for all $a' < a$. □

Lemma 3 *Suppose that* $a(\theta, w) \in argmax_{a \in A} U(\theta, w, a)$, *and that* $a(\theta', w) \in argmax_{a \in A} U(\theta', w, a)$, *with* $a(\theta, w)$ *in the interior of A. Then, (i)* $a(\theta', w) > a(\theta, w)$ *if* $\theta' > \theta$, *and* $a(\theta', w) < a(\theta, w)$ *if* $\theta' < \theta$; *and (ii)* $a(\theta, w) > a(\theta, w)$ *if* $\triangle w^{i'} > \triangle w^i$, *and* $a(\theta, w) < a(\theta, w)$ *if* $\triangle w^{i'} < \triangle w^i$.

Proof The proof follows directly from Lemma 1 and theorem 3 in Edlin and Shannon (1998). □

This lemma establishes that the agent's chosen action rises with the agent's type as well as with any bonus $\triangle w^i$.

Lemma 4 *A non-trivial optimal contract exists.*

Proof The proof here follows closely the one in Poblete and Spulber (2012). Let $a(o) \in \text{argmax}_{a \in A} U(\theta, a, o)$, $a(w) \in \text{argmax}_{a \in A} U(\theta, w, a)$, and the call-option contract o be such that $U(\theta, a(w), o) = U(\theta, a(w), w)$. Then, the single-crossing

[18] As $z \in \mathcal{C}$, $zp(\theta, a) \leq yp(\theta, a)$. In addition, $dp(\theta, a)$ increases continuously with $\bar{y}$ from $-\infty$ to $yp(\theta, a)$ as $\bar{y}$ ranges from $-\infty$ to y_I. Thus, for each $(w, a) \in \mathcal{C} \times A$, a unique debt contract is defined by $vp(\theta, a) = 0$.

property of $U(\theta, w, a)$ in Lemma 2 implies that: (i) $a(o) > a(w)$; (ii) Letting $U(\theta, a(w), w) = \max_{a \in A} U(\theta, w, a')$, single-crossing at $a(w)$ implies that $U(\theta, a(o), o) \geq U(\theta, a(w), w)$; and (iii) $V(\theta, a(o), o) > V(\theta, a(w), w)$. This follows from the fact that $a(o) > a(w)$ implies that $V(\theta, a(w), w) = V(\theta, a(w), o) < V(\theta, a(o), o)$, where the last inequality follows from **MLRP** and the fact that $y^i - o^i$ is increasing with h. From (ii) and (iii) contract o yields greater net benefits to the principal than contract w, and it satisfies the agent's participation constraint. Hence, if an optimal contract exists, it must be a call-option contract.

Observe that the agent's utility under a call-option contract $U(\theta, a(\bar{y}), o(\bar{y}))$ is continuous and decreasing in $\bar{y}$ and supermodular in $(\theta, a, -\bar{y})$, as shown in Lemma 1. Let $A(\bar{y}) = \{a \in A | a \in \text{argmax}_{x \in A} U(\theta, x, o(\bar{y}))\}$. It is easy to check that the set $A(\bar{y})$ is closed and bounded. Furthermore, let $a(\bar{y})$ be the largest element in $A(\bar{y})$. Then monotone comparative statics show that $a(\bar{y})$ is decreasing with $\bar{y}$, because boundedness and monotonicity $a(\bar{y})$ is continuous except at countable number points.

The expected payoff of the principal is given by $V(\theta, a, o(\bar{y})) = \sum_i p^i(\theta, a) \min\{y^i, \bar{y}\}$. Observe that $V(\theta, a(\bar{y}), o(\bar{y}))$ is discontinuous only if $a(\bar{y})$ is discontinuous. Because $a(\bar{y})$ is decreasing, $V(\theta, a(\bar{y}), o(\bar{y}))$ is decreasing at any discontinuity point. Let $\tilde{V}(\theta, a(\bar{y}), o(\bar{y})) = \sup_{\bar{y}' \leq \bar{y}} V(\theta, a(\bar{y}'), o(\bar{y}))$. Hence, $\tilde{V}(\theta, a(\bar{y}), o(\bar{y}))$ is non-decreasing and continuous in $\bar{y}$ because $V(\theta, a(\bar{y}), o(\bar{y}))$ is decreasing at any discontinuity point. Consider now the following problem:

$$\max_{\bar{y}} \tilde{V}(\theta, a(\bar{y}), o(\bar{y})) \text{ subject to } U(\theta, a(\bar{y}), o(\bar{y})) \geq 0$$

Clearly, this has a solution, since we are maximizing a continuous function on a closed set and we can restrict $\bar{y}$ to be lower than a finite number which makes the set compact, since $U(\theta, a(\bar{y}), o(\bar{y}))$ converges to 0 as $\bar{y}$ grows. Let the solution to this problem be $\tilde{V}^*$, and $\bar{y}^*$ the lowest value $\bar{y}$ such that $\tilde{V}(\theta, a(\bar{y}), o(\bar{y})) = \tilde{V}^*(\theta, a(\bar{y}), o(\bar{y}))$. Notice that $\bar{y}^*$ is well defined since $\tilde{V}(\theta, a(\bar{y}), o(\bar{y}))$ is continuous in $\bar{y}$.

The claim to be proven is that a call-option contract with face value $\bar{y}^*$ implements $\tilde{V}^*$. This means that there exists a sequence $a^n \in A(\bar{y}^n)$ with $\lim_{n \to \infty} \bar{y}^n \to y^*$ and $\lim_{n \to \infty} a^n \to a^*$, such that $\lim_{n \to \infty} \tilde{V}(\theta, a^n, o) \to \tilde{V}^*$. If $\bar{y}^*$ cannot implement $\tilde{V}^*$, then $a^* \notin A(\bar{y}^*)$. This means that there exists $\tilde{a} \in A(\bar{y}^*)$, such that $U(\theta, \tilde{a}, o(\bar{y}^*)) > U(\theta, a^*, o(\bar{y}^*))$. Since $U(\theta, a, o(\bar{y}))$ is continuous in both arguments $(a, \bar{y})$; by picking n sufficiently large this also means that $U(\theta, \tilde{a}, o(\bar{y}^n)) > U(\theta, a^n, o(\bar{y}^n))$, which is contradiction. Finally, suppose that the contract with face value $\bar{y}^*$ is not optimal. Then, there exists $\bar{y}$ such that $V(\theta, a(\bar{y}), o(\bar{y})) > \tilde{V}^*$ and $U(\theta, a(\bar{y}), o(\bar{y})) \geq 0$. But this implies that $\tilde{V}(\theta, a(\bar{y}), o(\bar{y})) \geq V(\theta, a(\bar{y}), o(\bar{y})) > \tilde{V}(\theta, a(\bar{y}^*), o(\bar{y}^*))$, which contradicts the optimality of $\bar{y}^*$ in the modified problem. □

□

The optimal contract is a call-option contract that makes the principal the full residual claimant for output realizations lower than the face value $\bar{y}(\theta)$ by paying the agent nothing, and makes the agent the full residual claimant for high output

realizations by paying him the output minus the face value equal to $\bar{y}(\theta)$.[19] This yields the result in Innes (1990).

Because the agent is risk neutral, a call-option contract solves the standard principal-agent problem in equation (P-MH). The reason has to do with the fact that a risk-neutral principal and agent only care about the expected payment. Because conditional on a given action a, a call-option contract always yields to the principal a lower expected cost than a non-call-option contract for action a, a call-option contract either yields a greater expected payoff conditional on action a being chosen, or it induces an action $a' \geq a$ at the same expected cost for the principal as a, which rises his expected payoff due to that **MLRP** implies **FOSD**. Hence, a call-option contract is optimal for the principal.

The optimal contract is different for each type and better types are offered more powerful incentives and therefore choose higher actions. However, the optimal action is downward distorted with respect to the first-best, since the principal wants to lower the agent's limited-liability rent by means of making an efficient use of the information or monitoring system.

It is interesting to study how the limited-liability rent changes with the agent's ability type. Using the Envelope theorem one can show that the limited-liability rent varies with θ as follows:

$$\sum_i p^i_\theta(\theta, a^m(\theta)) \max\{y^i - \bar{y}(\theta), 0\} - \sum_i p^i(\theta, a^m(\theta)) \frac{\partial \max\{y^i - \bar{y}(\theta), 0\}}{\partial \theta}.$$

The first term comprises the increase in the limited-liability rent due to an increase in the agent's type when the contract is held constant. This is due to the fact that the return distribution satisfies **MLRP** with respect to the agent's ability parameter θ, and the contract satisfies **MONA**. The second term consists of the increase in the limited-liability rent due to the effect that a change in θ has on the optimal face value of the call-option contract. Supermodularity in (θ, a) implies that $\bar{y}(\theta)$ decreases with θ. The reason stems from the following: on the one hand, an increase in θ improves the output distribution in the sense of **FOSD**, which, ceteris paribus, rises the principal's payoff; on the other hand, an increase in the face value $\bar{y}$ reduces the optimal action and, since the marginal return to the action increases with the agent's type, the principal chooses a lower $\bar{y}$ as θ rises. Hence, the limited-liability rent rises with θ. This implies that if types were unobservable, then every type will claim to be the highest type. Thus, the optimal contracts under pure moral hazard are not implementable under moral hazard and adverse selection.

References

Athey S (1996) Characterizing properties of stochastic objective functions. Working papers 96-1, Massachusetts Institute of Technology (MIT), Department of Economics

Baker G, Gibbs M, Holmstrom B (1994) The wage policy of a firm. Q J Econ 109(4):921–55

[19] If U were such that the limited liability does not bind, then $w^m(\theta)$ would be set to satisfy the participation constraint. In this case, the contract leaves no rent to the agent.

Balmaceda F (2009) Uncertainty, pay for performance, and asymmetric information. J Law Econ Organ 58(1):211–236
Balmaceda F (2011) On the Optimality of One-size-fits-all Contracts: the Limited liability case. Tech Rep
Balmaceda F, Balseiro JSR (2016) Bounds on the welfare loss from moral hazard with limited liability. Games Econ Behav 95:137–155
Baron DP, Besanko D (1984) Regulation, asymmetric information, and auditing. RAND J Econ 15(4):447–470
Baron DP, Besanko D (1987) Monitoring, moral hazard, asymmetric information, and risk sharing in procurement contracting. RAND J Econ 18(4):509–532
Bergemann D, Morris S (2005) Robust mechanism design. Econometrica 73(6):1771–1813
Bhattacharyya S, Lafontaine F (1995) Double-sided moral hazard and the nature of share contracts. RAND J Econ 26(4):761–781
Caillaud B, Guesnerie R, Rey P (1992) Noisy observation in adverse selection models. Rev Econ Stud 59(3):595–615
Carroll G (2015) Robustness and linear contracts. Am Econ Rev 105(2):536–563
Carroll G, Meng D (2016) Robust contracting with additive noise. J Econ Theory 166:586–604
Chassang S (2013) Calibrated incentive contracts. Econometrica 81(5):1935–1971
Chiappori PA, Salanié B (2002) Testing contract theory: a survey of some recent work
Demougin D (1989) A renegotiation-proof mechanism for a principal-agent model with moral hazard and adverse selection. RAND J Econ 20(2):256–267
Demougin D, Fluet C (1998) Mechanism sufficient statistic in the risk-neutral agency problem. J Inst Theor Econ JITE 154(4):622–638
Edlin AS, Shannon C (1998) Strict monotonicity in comparative statics. Journal of Economic Theory 81(1):201–219
Escobar JF, Pulgar C (2016) Motivating with simple contracts. Tech Rep
Faynzilberg PS, Kumar P (1997) Optimal contracting of separable production technologies. Games Econ Behav 21(1–2):15–39
Faynzilberg PS, Kumar P (2000) On the generalized principal-agent problem: decomposition and existence results. Rev Econ Des 5:23–58. https://doi.org/10.1007/s100580050046
Finkelstein A, McGarry K (2006) Multiple dimensions of private information: Evidence from the long-term care insurance market. Am Econ Rev 96(4):938–958
Finkelstein A, Poterba J (2004) Adverse selection in insurance markets: policyholder evidence from the u.k. annuity market. J Polit Econ 112(1):183–208
Gottlieb D, Moreira H (2013) Simultaneous adverse selection and moral hazard
Gottlieb D, Moreira H (2017) Simple contracts with adverse selection and moral hazard
Grossman S, Hart O (1983) An analysis of the principal-agent problem. Econometrica 51:7–45
Guesnerie R, Laffont J-J (1984) A complete solution to a class of principal-agent problems with an application to the control of a self-managed firm. J Public Econ 25(3):329–369
Guesnerie R, Picard P, Rey P (1989) Adverse selection and moral hazard with risk neutral agents. Eur Econ Rev 33(4):807–823
Holmstrom B, Milgrom P (1987) Aggregation and linearity in the provision of intertemporal incentives. Econometrica 55:303–328
Innes RD (1990) Limited liability and incentive contracting with ex-ante action choices. J Econ Theory 52(1):45–67
Kim SK (1997) Limited liability and bonus contracts. J Econ Manag Strategy 6(4):899–913
Laffont J-J, Tirole J (1986) Using cost observation to regulate firms. J Polit Econ 94(3):614–41
Lafontaine F, Slade M (2001) Incentive contracting and the franchise decision. In: Chatterjee K, Samuelson W (eds) Advances in business applications of game theory. Kluwer Academic Press, Dordrecht
Lafontaine F, Slade ME (1997) Retail contracting: theory and practice. J Ind Econ 45(1):1–25
Lazear E (2000) Performance pay and productivity. Am Econ Rev 90:1346–1361
Lewis TR, Sappington DEM (1989) Countervailing incentives in agency problems. J Econ Theory 49(2):294–313
Lewis TR, Sappington DEM (2000) Motivating wealth-constrained actors. Am Econ Rev 90(4):944–960
Lewis TR, Sappington DEM (2001) Optimal contracting with private knowledge of wealth and ability. Rev Econ Stud 68(1):21–44
Lopomo G (2001) Optimality and robustness of the english auction. Games Econ Behav 36(2):219–240

Lopomo GP (1998) The english auction is optimal among simple sequential auctions. J Econ Theory 82(1):144–166
Matthews SA (2001) Renegotiating moral hazard contracts under limited liability and monotonicity. J Econ Theory 97(1):1–29
McAfee RP, McMillan J (1986) Bidding for contracts: a principal-agent analysis. RAND J Econ 17(3):326–338
McAfee RP, McMillan J (1987) Competition for agency contracts. RAND J Econ 18(2):296–307
Melumad ND, Reichelstein S (1989) Value of communication in agencies. J Econ Theory 47(2):334–368
Milgrom P, Segal I (2002) Envelope theorems for arbitrary choice sets. Econometrica 70(2):583–601
Mirrlees J (1975) The theory of moral hazard and unobservable behavior. Working paper, Nuffield College, Oxford
Myerson RB (1982) Optimal coordination mechanisms in generalized principal-agent problems. J Math Econ 10(1):67–81
Ollier S, Thomas L (2013) Ex post participation constraint in a principal-agent model with adverse selection and moral hazard. J Econ Theory 148(6):2383–2403
Paarsch HJ, Shearer B (2000) Piece rates, fixed wages, and incentive effects: statistical evidence from payroll records. Int Econ Rev 41(1):59–92
Paarsch HJ, Shearer BS (1999) The response of worker effort to piece rates: evidence from the british columbia tree-planting industry. J Hum Resour 34(4):643–667
Poblete JJ, Spulber DF (2012) The form of incentive contracts: agency with moral hazard, risk neutrality and limited liability. rand journal of economics. Rand J Econ, forthcoming
Salanié B (2005) The economics of contracts: a primer, vol 1, 2nd edn. MIT Press, Cambridge
Seiler E (1984) Piece rate vs. time rate: the effect of incentives on earnings. Rev Econ Stat 66(3):363–76

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.