On the Optimality of *One-size-fits-all* Contracts: The Limited Liability Case

Felipe Balmaceda[1]

Economics Department

Diego Portales University[2]

April 8, 2013

[1]I would like to thank participants in the regular seminar at the Economics Department of University of Chile, the Latin American Meeting of the Econometric Society in Bs. As-Argentina, the Latin American Theory workshop at IMPA, Rio de Janeiro and the Latin American Econometric Society Theory workshop at PUC, Rio de Janeiro. Financial support was provided by Fondecyt through research grant #1100267.Any error however remains my own responsibility.

[2]Avda. Santa Clara 797, Santiago, 8580000, Chile felipe.balmaceda@udp.cl

Abstract

This paper studies a principal-agent model in which the principal and agent are risk-neutral, there are two actions, adverse selection, moral hazard and limited liability. When the two actions are subject to moral hazard, there is no distortion at the top, the optimal action profile is downward distorted for everyone else and the optimal menu of contract exhibits the *one-size-fits-all* property; that is, each ability type receives the same contract. The optimal contract pays a bonus when the outcome with the highest likelihood ratio is observed and the limited liability otherwise. When one of the actions is contractible and the other is subject to moral hazard, there is no distortion at the top, the non-contractible action is downward distorted for everyone else, the contractible action can be either upward or downward distorted. The optimal contract no longer exhibits the *one-size-fits-all* property. The *one-size-fits-all* property sheds light why we rarely observe menus of contracts in market that use franchising, credit and labor markets, and in regulated industries.

Keywords: Moral Hazard, Adverse Selection, Multiple Tasks, Limited Liability.

JEL-Classification: D82, D86, J33.

1 Introduction

The theory of incentives predicts that optimal contracts are highly complex and their terms should respond to differences in agents' characteristics. However, empirical and casual observation (see, for instance, Baker et al. (1994) and Sugato and Lafontaine (1995)) show that this is not usually the case. For instance, in markets that use franchising (for details see Lafontaine and Slade (1997, 2001)), franchisors instead of offering contracts tailored to the characteristics of each franchisee, most of them 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, different royalty rates and franchise fees, but a given franchisor offers the same terms to all potential franchisees at a given point in time. Salanié (2005, p. 474) concludes the following: "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 observed prevalence of simple contracts remains a challenge to incentive theory when one does not resort to things such as legal restrictions, transaction costs and/or fairness considerations.

This paper rationalizes the optimality of simple contracts in the sense that contracts are not tailored to agents' characteristics and the optimal allocation is implemented via a simple bonus-type contract. Crucial to the result is the fact that agents are subject to limited liability, which limits a principal's ability to punish agents for bad performance. This is usually the case in financial and labor market contracts; an employer is not free to punish poor performance with negative wages and an entrepreneur cannot be asked to re-paid more than what his venture returns. Mainly, the paper considers a model where two (weakly) complementary actions must be undertaken, there is an aggregated performance

[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, firms are risk neutral and unaware of workers' abilities. He shows that under certain parameterization the second-best menu has more than one contract, and under others the menu contains one contract only.

measure that can take more than two outcomes and the probability distribution satisfies the monotone likelihood ratio property (**MLRP**) in each input (i.e., actions and ability). The principal and the agent are risk neutral, the agent privately learns her type/ability before signing the contract. In addition, the cost of effort function as well as the reservation utility are type independent.

Within this model two different scenarios are considered: (i) moral hazard with regard to both actions, incomplete information with respect to the agent's ability type and limited liability; and (ii) moral hazard with regard to one of the two actions (the other is contractible), incomplete information with respect to the agent's ability type and limited liability. The results are as follows: (i) when both actions are subject to moral hazard, there is no distortion at the top and the optimal action profile is downward distorted for everyone else. The optimal action profile is implemented by a bonus-type contract that pays a bonus when the outcome with the highest likelihood ratio is observed and pays the limited liability otherwise. The optimal menu of contracts exhibits what I call the *one-size-fits-all* property; that is, all workers are offered the same bonus-type contract regardless of their ability type. In spite of the optimality of the *one-size-fits-all* property, agents' equilibrium behavior is such that more talented agents choose a higher action profile, are more productive and have a higher average compensation than less talented agents; and (ii) when there is moral hazard with regard to one of the two actions only, there is no distortion at the top and for all other ability types, the optimal non-observable action is downward distorted, while the contractible action could be either upward or downward distorted. The optimal contract no longer exhibits the *one-size-fits-all* property, yet the bonus type structure of the optimal compensation contract holds. As before, workers' equilibrium behavior is such that more talented agents are in equilibrium more productive and have a higher average compensation than less talented agents.

The main incentive problem is that an agent may wish to understate his type to convince the principal that higher outcomes are harder to achieve. If payments were not restricted by limited liability, the principal would mitigate the agent's underreporting incentives in the standard fashion; by promising a stepper payment scheme as the agent reports a higher type. This would work because **MLRP** with respect to the agent's ability type implies that higher ability types benefit more from a steeper contract. Thus, higher ability agents would bear more responsibility for their actions. If higher payments are more sensitive to the agent's ability type, it does not pay to an agent with low ability to claim to be more skillful than he really is since he is penalized with a contract that is more sensitive to higher outputs and these are less likely to be produced by a low-ability agent. Thus, his output will be lower and therefore his compensation will be lower too. The same logic explains why high-ability agents do not claim to have a lower ability than the one they really have. However, limited liability precludes low payments when low outcomes are observed. This implies that the incentive compatibility constraint requiring that the power of incentives increases with the agent's ability type cannot be satisfied without given sizeable informational rents. In order to induce truthful revelation and to minimize informational rents, the optimal contract exhibits the *one-size-fits-all* property; that is, the optimal menu of contracts has one contract only. This pays a bonus when the outcome with highest likelihood ratio (here the highest outcome) is realized and pays the limited liability L otherwise. The reason is that limited liability imposes a lower bound on punishments and therefore each ability type faced with a menu of contracts with at least two different contracts that pay the limited liability in some state would prefer the one with a steeper wage profile. To reduce the informational rent, the principal sets the payment in each state different from that for which the 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 bonus to each ability type. This means that the moral hazard problem is more important than the adverse selection problem in the sense that the principal designs optimal contracts to deal only with the moral hazard issue and relinquishes the right to customize the contract to each ability type. However, the action profile chosen is in general 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.

When one of the two actions is contractible and the other action is subject to moral hazard, the principal no longer needs to give up the right to customize the contract to each ability type in order to minimize the informational rent. The reason stands for the fact that the principal can use the contractible action as an instrument to achieve truthful revelation of private information when bonuses are allowed to vary with the agent's ability type. A varying bonus with the agent's ability type allows the principal to provide different incentives to different ability types, holding the limited liability as well as the informational rent as low as possible. In order to counteract the agent's incentives to misreport his type created by a menu of contracts with type-dependant bonuses, the principal asks different ability types to choose different contractible action levels. This action is downward distorted when the relative informativeness of the signal (output) with regard to the agent's ability and the non-contractible action is increasing with the contractible action, while the opposite happens when the relative informativeness is decreasing. Thus, the use of the contractible action as a screening device is determined by its impact on the principal's quality of the relative inference process with regard to the agent's ability type and the unobservable action. The *one-size-fits-all* property is recovered under fairly restrictive conditions. Mainly, when the relative informativeness of the signal with regard to the agent's ability and the non-contractible action is independent of the contractible action, and actions and ability are neither substitutes nor complements.

The provision of incentives in organizations is at the crux of the organizational design problem. A long standing literature studies the question of how optimal incentive contracts look like in different situations. The literature on optimal contracting under pure moral hazard is vast and that under pure adverse selection is also huge. For the sake of brevity I will not review them here. I will focus on the most closely related strand that deals with optimal contracting under both, moral hazard and adverse selection, risk neutrality and unlimited liability. Demougin (1989), Guesnerie et al. (1989), Caillaud et al. (1992), McAfee and McMillan (1986, 1987), Laffont and Tirole (1986) and Melumad and Reichelstein (1989) 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; that is, the models are just 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. One can, however, easily come up with examples of economically interesting situations where the production technology is such that the uncertainty and the type interact with each other.

Lewis and Sappington (2000, 2001) consider some situations of this sort. They 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) do so when wealth is private knowledge. In 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 to achieve his goal 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 of the ability level.

The rest of the paper is organized as follows. In Section 2, I introduce the model and the main assumptions. Section 3 presents the full information case. In Section 4, the optimal contract with adverse selection, moral hazard and limited liability is derived. In the next Section, I study the case in which one action is contractible and the other remains unobservable. Section 6 concludes with some remarks. All proofs can be found in the appendix as well as the case of pure moral hazard with observable ability. There the validity of the first-order approach is demonstrated.

2 The Model

Consider a relationship between a risk-neutral principal (P) and a risk-neutral agent (A). The agent is characterized by his type (e.g., ability level) indexed by θ. This is privately known by the agent before signing a contract. There is a continuum of types $\theta \in \Theta \equiv [\underline{\theta}, \overline{\theta}]$, and the principal knows that types are distritbuted according to the continuously differentiable distribution function $F(\theta)$ with associated density function $f(\theta)$ and full support. The full support assumption ensures that all outcomes can be realized no matter what the agent's type is. This is meant to avoid the inference that for some outputs, the principal will be able to rule out some types after observing a given outcome. As is standard in the literature, the inverse of the hazard function $H(\theta) \equiv \frac{1-F(\theta)}{f(\theta)}$ is assumed to be bounded and monotonically decreasing. The principal hires the agent to work in a project that requires two actions: a non-observable strategy, denoted by $s \in S \equiv [0, \bar{s}]$, and a non-observable effort, denoted by $e \in E \equiv [0, \bar{e}]$. Later I consider the case in which one of the two action is contractible.

The project's return is denoted by y_h for $h = 1, \ldots, n$ with $y_{h+1} > y_h, \forall h$. The probability that outcome y_h occurs depends on the agent's action profile $a \equiv (e, s) \in A \equiv E \times S$ and the agent's type θ in the following way: $\Pr(y = y_h|\theta, a) \equiv p_h(\theta, a)$ for $h \in \{1, \ldots, n\}$, with $\sum_{h=1}^{n} p_h(\theta, a) = 1, \ \forall\ (a, \theta) \in A \times \Theta$.

The agent's private cost of action profile $a \in A$ is $c(a) = c \cdot (e + s)$. Thus, the agent's private cost of actions is the same regardless of the agent's type and actions are neither substitutes nor complements in the cost function.[2]

I shall define the following likelihood ratios: (i) $\ell_{hs}(\theta) \equiv \frac{p_{hs}(\theta,a)}{p_h(\theta,a)}$; (ii) $\ell_{he}(\theta) \equiv \frac{p_{he}(\theta,a)}{p_h(\theta,a)}$; and (iii) $\ell_{h\theta}(a) \equiv \frac{p_{h\theta}(\theta,a)}{p_h(\theta,a)}$.

Hereinafter I assume the following:

[MLRP] (i) $\ell_{hs}(\theta, a)$ is increasing in h for all $(a, \theta) \in A \times \Theta$ and bounded above; (ii) $\ell_{he}(\theta, a)$ is increasing in h for all $(a, \theta) \in A \times \Theta$ and bounded above; and (iii) $\ell_{h\theta}(\theta, a)$ is increasing in h for all $(a, \theta) \in A \times \Theta$ and bounded above.

[2]The results holds if: (i) $c(\cdot)$ is a twice continuously differentiable and strictly convex function; and (ii) $c_a(\mathbf{0}) = c(\mathbf{0}) = \mathbf{0}$ and $c_{es}(a) \leq 0$ for any $a \in A$. Yet, the expression for the optimal contract becomes more cumbersome and the algebra messier without further gain in intuition.

Parts (i) and (ii) are the standard **MLRP** for one dimension moral hazard problems. Part (iii) is the **MLRP** with respect to the agent's type guaranteeing that higher outcomes are more likely to come from distributions parameterized by higher ability.[3] Assuming **MLRP** with respect to the agent's type implies that for any given effort level, the distribution of output for a more talented agent first-order stochastic dominates (FOSD) that for a less talented agent.[4] Thus, **MLRP** ensures that any principal who has a payoff function that is increasing in the project's outcome prefers the stochastic distribution of returns induced by higher actions and ability.

The next assumption, together with **MLRP**, validates the use of the first-order approach and ensures that the optimal action profile belongs to the interior of A.

[CP] (i) The upper cumulative probability distributions $\sum_{h>h'} p_h(\theta, a)$ are increasing and strictly concave in $(a, \theta) \in A \times \Theta$, $\forall h'$; (ii) $p_h(\theta, a)$ is thrice continuously differentiable for all $(a, \theta) \in A \times \Theta$, and $p_h(\theta, a) > 0$, $\forall (a, \theta) \in A \times \Theta$ and $\forall h \in \{1, \ldots, n\}$; and (iii) $\lim_{s \to 0} \sum_{h>h'} p_{hs}(\theta, a) y_h \to \infty$ and $\lim_{e \to 0} \sum_{h>h'} p_{he}(\theta, a) y_h \to \infty$;

Finally, I will assume the following complementarity property of the cumulative distribution functions in $A \times \Theta$.

[SC] (i) $\sum_{h>h'} p_{h\theta a}(\theta, a) \geq 0$ for all h'; and (ii) $\sum_{h>h'} p_{hes}(\theta, a) \geq 0$ for all $(a, \theta) \in A \times \Theta$ and for all h'.

Part (i) says that the distribution of the marginal return to actions rise in the sense of FOSD with the agent's ability. In short, ability and actions are complements. Part (ii) says that the distribution of marginal return to e rises in the sense of FOSD with s. In short, effort and strategy are complements.

Since there are n possible contractible outcomes to the principal, a contract is given by

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

[4]The expectation of $\ell_{h\theta}(\theta, a)$ is zero, and because $\ell_{h\theta}(\theta, a)$ is increasing in h, the covariance between $\ell_{h\theta}(\theta, a)$ and y_h is positive; hence

$$\sum_h p_h(\theta, a) \ell_{h\theta} y_h > 0 \text{ for all } a \in A.$$

.

a wage schedule: $W = \{w_h\}_{h=1}^n$, where w_h is the wage paid when y_h is realized. A direct consequence of the **SC** property is that for any monotone increasing wage schedule, greater actions lead to a higher expected wage the greater the agent's ability. A menu of contracts is defined as a function $(W(\cdot), a(\cdot)) : \Theta \rightarrow \Re^n \times A$ that considers a contract for each ability type.

Payments are restricted by a limited liability constraint that prevents the principal from paying the agent a wage in any state lower than $L \geq 0$; that is,[5]

$$\textbf{[LL]}\ w_h \geq L, \quad \forall h. \tag{1}$$

Finally, each ability type has an outside option given by U_θ, with $U_\theta = U \leq L$ for all $\theta \in \Theta$. That is, the outside option is type independent. Furthermore, $U \leq L$ implies that I am focusing on the case in which under limited liability the first-best cannot be implemented without paying an extra cost.

3 Benchmark: The Complete Information Case

In this section I study the case in which actions are contractible and ability is common knowledge. In appendix B, I consider the pure moral hazard case in which the action profile is subject to moral hazard, the ability type is common knowledge and the agent is subject to the limited liability constraint in equation (1).[6]

The action profile a that maximizes joint welfare is given by the solution to the following problem

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

[5] When $L = 0$, this constraint can be thought of as the case in which the agent owns no assets at the time of contracting.

[6] There is another possible benchmark which is the one where actions are contractible, ability is private information and there is limited liability. This case however has a trivial solution which is to offer a fixed wage contract that pays $w_h(\theta) = L + c(a^*(\theta))$ in each state, where $a^*(\theta)$ is the first-best efficient action profile. It is easy to check that this contract implements the first-best and induce truthful revelation of information (each type is indifferent between his type and any other type). The reason stands for the fact that neither the cost function nor the outside utility depend on the agent's type θ.

where

$$S(\theta, a) \equiv \sum_h p_h(\theta, a) y_h - c(a).$$

The first-order conditions that determine the efficient action profile, denoted by $a^*(\theta)$, are:

$$S_e(\theta, a) = 0$$

and

$$S_s(\theta, a) = 0.$$

Because of assumption **CP**, the first-order conditions are necessary and sufficient and the first-best efficient action profile belongs to the interior. Thus, the first-best surplus is given by $S(\theta, a^*(\theta))$.

Assumption **SC** implies that for any $\theta' > \theta$, $a^*(\theta') \geq a^*(\theta)$ and $S(\theta', a^*(\theta')) \geq S(\theta, a^*(\theta))$. Thus, the first-best efficient action profile is non-decreasing with the agent's type and total welfare rises with the agent's type.

The implementation is simple. The principal promises to pay the agent $c(a^*(\theta)) + L$ if he delivers $a^*(\theta)$ and pays L otherwise. The agent weakly prefers to deliver $a^*(\theta)$ in this setting, and so the principal's ideal outcome is ensured.

4 Moral Hazard, Adverse Selection and Limited Liability

In this section I study the environment in which the two actions are subject to moral hazard. The principal's goal here is twofold: on the one hand, she has to provide the agent with incentives to choose the desired action profile and, on the other hand, 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 he chooses the action profile prescribed by the principal for an agent of type θ, $a(\theta)$.

Lets suppose that an agent of ability $\theta \in \Theta$ is faced with a wage schedule $W \equiv \{w_h\}_{h=1}^n \in \Re^n$ and chooses the action profile $a \in A$, then his expected utility is given by:

$$U(W, a, \theta) \equiv \sum_h p_h(\theta, a) w_h - c(a). \tag{2}$$

By the extended revelation principle, the direct mechanism $(W(\theta), a(\theta))$ is incentive compatible if and only if

$$U(W(\theta), a(\theta), \theta) \geq U(W(\theta'), a(\theta, \theta'), \theta), \ \forall \theta, \theta' \in \Theta, \tag{3}$$

where

$$a(\theta, \theta') \equiv \arg\max_{a \in A} U(W(\theta'), a, \theta), \ \forall \theta, \theta' \in \Theta. \tag{4}$$

and

$$a(\theta) \equiv \arg\max_{a \in A} U(W(\theta), a, \theta), \ \forall \theta \in \Theta.^{7} \tag{5}$$

The mechanism satisfies individual rationality if and only if

$$U(W(\theta), a(\theta), \theta) \geq U, \ \forall \theta \in \Theta. \tag{6}$$

Constraint (3) is the incentive-compatibility constraint and states that the agent is better off announcing his true type, receiving the allocation $W(\theta)$ and choosing $a(\theta)$ than announcing a different type θ' when his true type is θ and receiving the allocation $W(\theta)$ and choosing $a(\theta, \theta')$. Constraint (6) is the standard individual rationality constraint establishing that each ability type prefers to participate to staying out.[8]

The problem faced by the principal is as follows

$$\max_{(W(\theta), a(\theta)) \in \Re^n \times A} \int \sum\nolimits_h p_h(\theta, a(\theta))(y_h - w_h(\theta)) f(\theta) d\theta$$

subject to

(1), (3) and (6).

Let $\frac{da(\theta,\theta')}{d\theta'} \equiv \sum_h \frac{\partial a(\theta,\theta')}{\partial w_h(\theta')} \frac{dw_h(\theta')}{d\theta'}$. The following is formally proven in appendix A as well as all the forthcoming results.

[7] I use the notation $a(\theta, \theta')$ introduced in equation (4) as refereing to $a(W(\theta'), \theta)$ and $a(\theta)$ introduced in equation (5) as refereing to $a(W(\theta), \theta)$.

[8] Later I will informally discuss the case in which the principal may shut down some ability types.

Proposition 1 *Suppose the first-order approach is valid. The following conditions are sufficient for a differentiable contract $W(\theta)$ to be implementable:*[9]

$$\sum\nolimits_h p_h(\theta, a(\theta,\theta'))\frac{dw_h(\theta')}{d\theta'} = 0,\ \forall\theta \in \Theta \tag{7}$$

and

$$\sum\nolimits_h p_{h\theta}(\theta, a(\theta,\theta'))\frac{dw_h(\theta')}{d\theta'} + \sum\nolimits_a \sum\nolimits_h p_{ha\theta}(\theta, a(\theta,\theta'))w_h(\theta')\frac{da(\theta,\theta')}{d\theta'} \geq 0,\ \forall\theta,\theta' \in \Theta, \tag{8}$$

where $a(\theta,\theta')$ is defined by (4)*.*

Equations (7) *and* (8) *holding for $\theta' = \theta$ constitute necessary conditions for implementability.*

The proof is based on the fact that one can eliminate the agency relationship from the problem by solving for the agent's action profile as a function of the mechanism. This procedure depends on the validity of the first-order approach. Equation (7) is the first-order condition for the agent's revelation of information problem taking into account that the agent chooses the action profile that maximizes his expected utility, and (8) is the equivalent to the second-order condition to the agent's revelation problem. This explains why this is required to hold for all $\theta, \theta' \in \Theta$. This constraint restricts the kind of mechanisms that the principal can use. This is equivalent to what the literature calls the monotonicity constraint. For instance, in Laffont and Tirole's (1986) model this entails a rising quantity and transfer with the agent's type. In fact, if actions and types are neither complements nor substitutes (that is, $p_{ha\theta}(\theta, a) = 0, \forall h$), sufficiency boils down to payoff monotonicity. However, the condition here is in general weaker since it allows decreasing actions with the agent's type as long as the compensation rises with the agent's type at a sufficiently fast rate. Thus, imposing monotonicity while it yields sufficiency, it rules out a large class of implementable mechanisms.

[9]The restriction to differentiable mechanisms is not without loss of generality. The statement should be said that is differentiable for almost all $\theta \in \Theta$; i.e., all point were the mechanism is differentiable. The required condition is that $U(\theta)$ is non-decreasing and convex in θ. Furthermore, the sufficient condition does not require the mechanism to be monotonic and thus almost everywhere differentiability is not implied by the sufficient condition.

Using the sufficient condition in equation (7) and the envelope theorem to obtain $U'(\theta)$ and ignoring, as usually done in the literature, the global incentive compatibility constraint stated in equation (8), the principal's problem can be written as the following optimal control problem

$$\max_{(W(\theta),U(\theta))\in\Re^{n+1}} \int (S(\theta, a(\theta)) - U(\theta))f(\theta)d\theta \qquad \text{P-AML}$$

subject to

$$U'(\theta) = \sum\nolimits_h p_{h\theta}(\theta, a(\theta))w_h(\theta) \tag{9}$$

$$U(\theta) \geq L, \ \forall\theta \in \Theta \tag{10}$$

$$w_h(\theta) \geq L, \ \forall h \in \{1, \ldots, n\}, \forall\theta \in \Theta. \tag{11}$$

In this optimization problem the control variables are the payments, the state variable is U, the co-state variable is μ, the Hamiltonian is

$$\mathcal{H}(U, W, a, \mu, \theta) = \big(S(\theta, a(\theta)) - U(\theta)\big)f(\theta) + \mu(\theta)\sum\nolimits_h p_{h\theta}(\theta, a(\theta))w_h(\theta) \tag{12}$$

and the Lagrangean is

$$\mathcal{L}(U, W, a, \mu, \theta, \lambda) = \mathcal{H}(U, W, a, \mu, \theta) + \gamma(\theta)(U(\theta) - L) + \sum\nolimits_h \lambda_h(\theta)(w_h(\theta) - L), \tag{13}$$

where $\gamma(\theta)$ is the multiplier for the participation constraint and $\lambda_h(\theta) \geq 0$ is the multiplier corresponding to the limited liability constraint for w_h.

The following is proven in the appendix.

Proposition 2

i) The principal's relaxed problem P-AML has a unique solution given by: $w_h^{AML}(\theta) = L$ *for all* $h < n$ *and* $\theta \in \Theta$ *and*

$$w_n^{AML}(\theta) = \frac{c(a(\underline{\theta}))}{p_n(\underline{\theta}, a(\underline{\theta}))} + L, \ \forall\theta \in \Theta. \tag{14}$$

ii) The optimal action profile $a^{AML}(\theta)$ *satisfies the following first-order conditions:*

$$p_{ne}(\theta, a^{AML}(\theta))(w_n^{AML}(\theta) - L) - c = 0,$$

and

$$p_{ns}(\theta, a^{AML}(\theta))(w_n^{AML}(\theta) - L) - c = 0.$$

iii) $a^{AML}(\theta)$ *is non-decreasing with* θ*.*

This shows that the optimal menu of contracts exhibits what I call the *one-size-fits-all* property; that is, the menu contains one and only one contract. The optimal contract is a bonus contract of the pass/fail type. It can be described by a non-contingent transfer $w^{AML}(\theta)$ and a bonus equal to $w_n^{AML}(\theta) - w^{AML}(\theta)$ paid to the agent when y_n is observed. The value of the non-contingent transfer $w^{AML}(\theta)$ is fully determined by the limited liability constraint and the bonus is determined by the lowest ability type's limited-liability rent. Thus, the bonus is independent of the agent's type. However, this does not imply that the expected output is the same for each ability type. The second-best optimal action profile rises with the agent's type θ and therefore the greater the type, the greater the expected output and the agent's expected compensation.

Because $H(\overline{\theta}) = 0$, there is no distortion at the top, actions are downward distorted for all other types and the lowest ability type gets a utility level $U(\underline{\theta}) = L$. Thus, the lowest type gets no informational rent and a positive limited-liability rent equal to $L - U$, while all other types get an informational rent. The reason for the inefficient action profile is that the principal's cost-the cost of the incentive compatibility constraint-, which is given by $H(\theta)\sum_h p_{h\theta}(\theta, a^{AML}(\theta))w_h^{AML}(\theta)$, is an increasing function of the action level whenever the optimal contract is non-decreasing. Thus, by downward distorting the optimal action profile, the principal captures part of the informational rent.

To better understand this result it is useful to take a closer look at how the informational rent changes as $w_h(\theta)$ varies. This is the principal's marginal cost of varying $w_h(\theta)$. Observe that the informational rent changes with $w_h(\theta)$ as follows

$$\sum\nolimits_h p_{h\theta e}(\theta, a(\theta))w_h(\theta)\frac{\partial e(\theta)}{\partial w_h(\theta)} + \sum\nolimits_h p_{h\theta s}(\theta, a(\theta))w_h(\theta)\frac{\partial s(\theta)}{\partial w_h(\theta)} + p_{h\theta}(\theta, a(\theta)).$$

The first two terms comprise the variation on the informational rent due to the change in the chosen action profile as the result of an increase in $w_h(\theta)$, and the third term is the direct effect of a change in the payment $w_h(\theta)$. Because of **MLRP** with respect to θ, the direct effect is negative for small values of h and positive for large values of it. It also follows from **MLRP** with respect to a, contract monotonicity (i.e., $w_h(\theta) \geq w_{h'}(\theta)$, $\forall h > h'$) and the complementarity between actions and θ, that the informational rent falls with $w_h(\theta)$ when

h is small and rises with it when this is large. Hence, in order to reduce the informational rent, the principal should pay little when the outcome is high and pay a lot when this is low, yet this will destroy the agent's incentives to choose a positive action profile.

The principal's marginal benefit of varying the wage $w_h(\theta)$ is given by

$$S_e(\theta, a(\theta))\frac{\partial e(\theta)}{\partial w_h(\theta)} + S_s(\theta, a(\theta))\frac{\partial s(\theta)}{\partial w_h(\theta)}.$$

It follows from MLRP with respect to a and the assumption that the contract is monotonic on the outcome and that actions are lower than the first-best efficient action profile that the surplus falls with $w_h(\theta)$ when h is small and rises with it when this is large. Hence, in order to get a higher surplus, the principal should pay little in low state and pay a lot in high states. This is exactly the opposite of what is required to reduce the informational rent. Hence, the principal must balance the positive incentive effect of paying more in good states against the increased informational rent that results from that. As a result of this when the wage is not restricted to be higher than a given level, the principal will downward distort the action profile in order to minimize the informational rent provided that the agent reveals his type truthfully. There is no need to provide the agent with a limited-liability rent since the "punishments" are unbounded below. The optimal contract that implements the principal's preferred action profile is monotonic with the output and is such that it provides more powerful incentives to higher types.

When a limited liability is imposed, the solution to the trade-off described above is no longer feasible since the principal is restricted to pay after any output a wage greater than or equal to L. The principal could increase the payment in each state in order to satisfy the limited liability constraint, however this will result in a sizeable limited-liability rent. Thus, the principal is better-off giving up on this and choosing a contract that provides no informational rent and downward distorting the choice of the action profile in order to reduce the informational rent.

The payment in the highest state cannot increase with the agent's ability type so as to provide stronger incentives to better types. The reason is that all agents will have an incentive to claim to be the highest type since this will provide them with the highest reward in the case that the highest outcome is observed and the same punishment when any other outcome is observed. This strategy yields a higher expected payoff to all types. The reason is twofold: first, each type can adjust his effort according to the optimal contract for the highest type; and second, an agent that chooses the action profile that maximizes his utility when he truthfully reveals his type, but claims to be the highest type possible, obtain a higher expected payoff since the bonus when the highest outcome is realized is higher when he lies than when he is honest. The principal, aware of this and restricted in the use of penalties for low outputs by the limited liability constraint, is forced to offer the same contract to each type so that it can induce the agent to truthfully revel his type at no extra cost (each ability type is indifferent between reveling his type truthfully and lying about it).

It is interesting to notice that despite the fact that the moral hazard problem overcomes the adverse selection problem, the action profile chosen under both, moral and adverse selection is different from that under pure moral hazard (derived in appendix B). The reason stands for the incentive compatibility with regard to the agent's type.

Let me finish this section by briefly discussing the consequences of considering different assumptions. In particular, I will focus on three of them which are: (i) type-independent outside utility; (ii) fixed support vs. moving support; (iii) task complementarities and **MLRP** in each dimension.

First, lets 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 in any given firm should be higher for an agent with a higher ability type. One can easily show that this would not change the result as long as the growth rate of U with respect to θ does not exceed $U'(\theta) = p_{n\theta}(\theta, a^{AML}(\theta))(w_n^{AML}(\theta) - L)$. The reason stands for the fact that the optimal contract provides each type with a utility higher than his outside option. If this is not the case, things get more complicated. On the one hand, the principal could not offer higher types a contract that promises a higher expected compensation since this will induce lower types to claim to have a higher ability level. So the solution will be to shutdown part of the type distribution. Which types will be shut down depend on the details of the model and it is hard to predict what would happen. However, complementarity between actions and types suggests that there is a force towards shutting down lower ability types unless the growth rate of U with θ is too high. 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, how strong are the countervailing forces coming from the outside option.

Second, lets consider the case of moving support. We know from the 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 coupled with unbounded utility almost solve the implementation problem. With limited liability, the principal is restricted in her capacity to punish the agent for outcomes that cannot come from other combinations of ability and actions. Hence, despite the fact that there are outcomes that are observable only if certain actions are taken, which will facilitate the inference problem, the principal would not be able to punish the agent in such a way to deter the agent's misbehavior when outcomes that are not consistent with the desired action profile are observed.

Third, lets consider the case of task complementarity. This assumption as well as that of **MLRP** are needed to ensure that the local approach with respect to action profile as well as types can be applied. It is my conjecture that **MLRP** with respect to one action only will do the job as long as **MLRP** with respect to θ holds since this ensures that the single-crossing property holds in the $(w_h, w_{h'})$ space. However, showing that the first-order approach is valid is harder to accomplish since I need to show that the optimal wage scheme is non-decreasing and this will require that a linear combination between ℓ_{he} and ℓ_{hs}, where the weights are the Langrange multipliers for each of the incentive compatible constraints, is increasing with h.

Finally, the fact that actions are complements is needed to satisfy the global incentive compatible constraint ensuring that

$$\sum\nolimits_a p_{na\theta}(\theta, a(\theta'))\big(w_n(\theta') - L\big)\frac{da(\theta')}{d\theta'} \geq 0, \ \forall\theta, \theta' \in \Theta.$$

If $\sum_{h>h'} p_{ha\theta}(\theta, a(\theta')) < 0$, the assumptions with regard to the relationship between $\sum_{h>h'} p_h(\theta, a(\theta'))$ and a should be such that $a(\theta)$ falls with θ.

It is known by now that the assumptions under which the results here are derived are demanding, one could dispense of some of them either by using the approach proposed by Poblete and Spulber (2012) or by using the less stringent conditions based on Monotone

Ratio assumptions proposed by Wang et al. (2011), which would allow me to characterize the optimal contracts without using the first-order approach. Yet this is not as straightforward since their techniques are developed only for the one action pure moral hazard case.

5 Limited Liability and Contractible Actions

In this section I study the same environment as in the preceding section with the modification that one of the two actions is contractible, while the other remains unobserved. Without loss of generality, I will assume that e is contractible and s is subject to moral hazard.

Lets define $s(\theta,\theta') \in \arg\max_{s\in S} U(W(\theta'), s, e(\theta'), \theta)$, $\forall \theta, \theta' \in \Theta$. Hence, $\frac{\partial s(\theta,\theta')}{\partial \theta'} \equiv \sum_h \frac{ds(\theta,\theta')}{dw_h(\theta')}\frac{dw_h(\theta')}{d\theta'} + \frac{\partial s(\theta,\theta')}{\partial e(\theta')}\frac{de(\theta')}{d\theta'}$. Then, the following is shown in the appendix.

Proposition 3 *Suppose the first-order approach is valid. The following conditions are sufficient for a differentiable contract $(W(\theta), e(\theta))$ to be implementable:*

$$\sum\nolimits_h p_h(\theta, e(\theta'), s(\theta,\theta'))\frac{dw_h(\theta')}{d\theta'} + \left(\sum\nolimits_h p_{he}(\theta, e(\theta'), s(\theta,\theta'))w_h(\theta') - c\right)\frac{de(\theta')}{d\theta'} = 0, \ \forall \theta \in \Theta, \tag{15}$$

and

$$\sum\nolimits_h p_{h\theta}(\theta, e(\theta'), s(\theta,\theta'))\frac{dw_h(\theta')}{d\theta'} + \sum\nolimits_h p_{he\theta}(\theta, e(\theta'), s(\theta,\theta'))w_h(\theta')\frac{de(\theta')}{d\theta'} + \sum\nolimits_h p_{hs\theta}(\theta, e(\theta'), s(\theta,\theta'))w_h(\theta')\frac{ds(\theta,\theta')}{d\theta'} \geq 0, \ \forall \theta, \theta' \in \Theta. \tag{16}$$

Equations (15) *and* (16) *holding for $\theta' = \theta$ constitute necessary conditions for implementability.*

Equation (15) is the first-order condition for the agent's revelation of information problem, while (16) is the equivalent to the second-order condition to the agent's revelation problem. This explains why this is required to hold for all $\theta, \theta' \in \Theta$.

In contrast to proposition (1), the necessary condition in equation (15) does not require $\frac{dw_n(\theta')}{d\theta'} = 0$ when $w_h(\theta') = L$. The reason stands for the fact that the principal can use the contractible action $e(\theta')$ as another instrument to induce truth-telling at a lower informational cost and to deter low ability types from claiming to be high types when the power of incentives is higher for contracts tailored to high-ability types.

Using the sufficient condition in equation (15) and the envelope theorem to obtain $U'(\theta)$ and ignoring, as usually done in the literature, the global incentive compatibility constraint stated in equation (16), the principal's problem is as follows

$$\max_{(W(\theta),U(\theta),e(\theta))\in\Re^{n+1}\times E}\int (S(\theta,a(\theta))-U(\theta))f(\theta)d\theta \qquad \text{P-AMC}$$

subject to

$$U'(\theta)=\sum\nolimits_h p_{h\theta}(\theta,a(\theta))w_h(\theta) \tag{17}$$

$$U(\theta)\geq L,\ \forall\theta\in\Theta \tag{18}$$

$$w_h(\theta)\geq L,\ \forall h\in\{1,\ldots,n\},\forall\theta\in\Theta. \tag{19}$$

The control variables are the payments and the action $e(\theta)$, the state variable is U, the co-state variable is μ, the Hamiltonian is

$$\mathcal{H}(U,W,a,\mu,\theta)=\big(S(\theta,a(\theta))-U(\theta)\big)f(\theta)+\mu(\theta)\sum\nolimits_h p_{h\theta}(\theta,a(\theta))w_h(\theta)$$

and the Lagrangean is

$$\mathcal{L}(U,W,a,\mu,\theta,\lambda)=\mathcal{H}(U,W,a,\mu,\theta)+\gamma(\theta)(U(\theta)-L)+\sum\nolimits_h\lambda_h(\theta)(w_h(\theta)-L), \tag{20}$$

where $\gamma(\theta)$ is the multiplier for the participation constraint and $\lambda_h(\theta)\geq 0$ is the multiplier corresponding to the limited liability constraint for the wage when outcome y_h is observed.

The following is proven in the appendix.

Proposition 4 (Limited Liability and Contractible Action) *Suppose* $\mathcal{H}(U,W,a,\mu,\theta)$ *is concave in* a*. Then,*

i) The principal's relaxed problem P-AMC has a unique solution given by: $w_h^{AMC}(\theta)=L$ *for all* $h<n$ *and* $\theta\in\Theta$ *and*

$$w_n^{AMC}(\theta)=\frac{c}{p_{ns}(\theta,a^{AMC}(\theta))}+L,\ \forall\theta\in\Theta. \tag{21}$$

ii) The optimal action profile $a^{AMC}(\theta)$ *is given by the solution to the following equations:*

$$S_s(\theta,a^{AMC}(\theta))=H(\theta)\frac{\partial\big(p_{n\theta}(\theta,a^{AMC}(\theta))/p_{ns}(\theta,a^{AMC}(\theta))\big)}{\partial s(\theta)}c$$

and

$$S_e(\theta, a^{AMC}(\theta)) = H(\theta)\frac{\partial\big(p_{n\theta}(\theta, a^{AMC}(\theta))/p_{ns}(\theta, a^{AMC}(\theta))\big)}{\partial e(\theta)}c.$$

This proposition shows that the optimal action profile is such that there is no distortion at the top, the non-contractible action is downward distorted and the contractible action could be either upward or downward distorted. The former holds if and only if

$$\frac{\partial\big(p_{n\theta}(\theta, a)/p_{ns}(\theta, a)\big)}{\partial e} > 0.$$

Notice that if the probability distribution is such that the ratio $p_{n\theta}(\theta, a)/p_{ns}(\theta, a)$, which can be interpreted as the rate of information substitution, is independent of action e, then the optimal contractible effort is the first-best efficient effort level. The reason is that for any given outcome different effort levels result in the same relative informativeness with regard to θ and s. In contrast when this ratio is increasing with e; i.e., the relative informativeness with regard to θ and s of the highest outcome increases with e, the effort is downward distorted. The reason is that more effort worsens the informativeness of the signal y_n with respect to the action s, while if the ratio is decreasing with e, the informativeness of the signal y_n with respect to the action s improves. In either case, the effort is distorted in order to keep the informational rent as low as possible.

This is somewhat reminiscent of Holmström and Milgrom (1991). They show in a multitasking model in which there is one performance measure per task that if the noise in the performance measure for a particular task is high, it might be optimal to make the agent's pay less sensitive to performance in other tasks. This is to prevent the agent from allocating too much effort in the well-measured tasks, which may harm the principal. Here, it is not the sensitivity of payment to performance what is adjusted to avoid the misallocation of effort across tasks, but the amount of contractible effort requested. This is to keep the relative informativeness of the aggregate performance measure between the non-contractible action and ability type as informative as possible, which helps to keep the informational rent low.

This result also highlights that the *one-size-fits-all* property of optimal contracts is due to the fact that when there is moral-hazard in each task, limited-liability restricts the set of incentive compatible contracts in such a way that the principal only focus on the moral hazard issue. In contrast, when the principal has an instrument different from payments to induce the agent to truthfully reveal his type, limited liability restrictions to the space of incentive compatible contracts has much less byte since the principal can use the contractible action to escape from the need to make payments identical for each ability type.

In fact, the necessary condition for incentive compatibility in equation (15) requires that

$$\frac{dw_n(\theta)}{d\theta} = -\frac{p_{ne}(\theta, e(\theta), s(\theta))w_n(\theta) - c}{p_n(\theta, e(\theta), s(\theta))}\frac{de(\theta)}{d\theta}, \ \forall\theta \in \Theta. \tag{22}$$

Thus, if $\frac{de(\theta)}{d\theta} \neq 0$, the *one-size-fits-all* property does not hold. This leads to the following corollary.

Corollary 5 *Suppose that*

$$\frac{\partial\big(p_{n\theta}(\theta,a)/p_{ns}(\theta,a)\big)}{\partial e} = 0, \ \frac{\partial^2\big(p_{n\theta}(\theta,a)/p_{ns}(\theta,a)\big)}{\partial s \partial\theta} = 0 \ \textit{and} \ S_{a\theta} = 0 \ \textit{for} \ a \in \{e, s\}.$$

Then $\frac{dw_n(\theta)}{d\theta} = 0, \ \forall\theta \in \Theta.$

This corollary establishes that if (i) actions and ability are not complements; and (ii) the relative informativeness of the performance with regard to θ and s is independent of the contractible action e, the optimal contract exhibits the *one-size-fits-all* property. The reason stands for the fact that the inference problem faced by the principal is not affected by e and thus this action is set to the efficient level, which is independent of the ability type since effort and ability are neither complements nor substitutes, and the interaction of the moral hazard and adverse selection problem remains unchanged.

One can interpret the contractibility of an action as the principal investing in monitoring to make some actions contractible. This shows that if the cost of doing so is small, the principal will do so. This is in contrast to the result in Zhao (2008) who shows that despite the fact that a principal can observe effort in all but one task and knows which action is optimal for each task, the principal would prefer to use only-output based incentive contracts. In their model the result follows from the fact that using a single bonus, the principal can motivate the agent to work hard in each task. This saves on the limited-liability rent compared to the case in which he contracts on effort in each task but one, which is non-contractible. This result is consistent with those encounter on the post-contractual private information such as

Baker (1992), Raith (2008) and Balmaceda (2010). The latter shows that it is optimal to use both output and input based measure of performance when there are contractible inputs. However, the reason is rather different, here the use of the input and output measure is due to the fact that it allows the principal to offer each ability type a lower informational rent and escape to the curse of limited liability.

The result here will be the same as that in Zhao (2008) if actions were contractible in each task since in this case it is optimal to contract only on actions and ignore the output measure. Here one can show the same, in fact, it is optimal to ask each type to exert the first-best efficient action in each task, pay each type a fixed wage equal to the cost $c(a^*(\theta)) + L$. It is easy to check that this will induce the agent to truthfully reveal his type.

Although proposition (4) has significantly simplified the problem, the ignored constraint in equation (16) is still potentially binding. One could take this constraint into account explicitly, for instance, using the ironing procedure in Maskin and Riley (1984) and Mussa and Rosen (1978). I have adopted the standard approach in the literature of ignoring this constraint. However, this might be important here since in principle this constraint allows for non-monotonic mechanisms. In fact, if $e(\theta)$ raises sufficiently fast with θ, the optimal bonus is allowed to fall with it and vice-versa. Furthermore, as shown in the appendix, a non-decreasing mechanism is sufficient for (16) to hold.[10] Thus allowing the principal to contract on e enlarges the set of mechanisms that are implementable in relation to the standard model and to the model with no contractible actions. Thus, since the main results here are statements of existence or lack of it of certain properties, I do not gain in generality by dealing explicitly with constraint (16).

The main difference between the contractible action case and the non-contractible action case is that in the former the principal uses two instruments, the effort and the bonus, to screen agents, while under non-contractibility he uses only the bonus. Thus, in the later

[10] García (2005) provides necessary conditions for implementability that closely resemble the ones in proposition 3. However, the comparison is not as straightforward since Garcia considers only adverse selection while the model here consider both adverse selection and moral hazard. The main difference stands for the fact that in Garcia's paper the main constraint is on $ds(\theta')/d\theta'$ whereas here is on $ds(\theta, \theta')/d\theta'$. Hence, the necessary condition here places restriction on the agent's on- and off-the-path behavior whereas Garcia requires restrictions on-the-path only.

case, the agent has more room for opportunistic behavior than in the former. This limits the principal's ability to choose the menu of contracts as he has to take into account the on- and off-path deviations as captured by $da(\theta')/d\theta'$ and $da(\theta,\theta')/d\theta'$ for all $\theta',\theta \in \Theta$ instead of only the on-path deviations as captured by $da(\theta')/d\theta'$ for all $\theta' \in \Theta$. This constraint result in the *one-size-fits all* feature of optimal contracts.

Let me end this section by a sufficient condition on the information technology under which the sufficient condition in equation (16) holds. This says that for any $(\theta, e, s) \in \Theta \times A$, $p_{ne\theta}(\theta,e,s)p_n(\theta,e,s) - p_{n\theta}(\theta,e,s)p_{ne}(\theta,e,s) \geq 0$ and $p_{nes}(\theta,e,s)p_n(\theta,e,s) - p_{ns}(\theta,e,s)p_{ne}(\theta,e,s) \geq 0$.

A well known technology used in all papers I am aware off satisfying this sufficient conditions is the weakly separable in $(\theta, a) \in \Theta \times A$ technology; that is, the probability distribution satisfies the following: take $\sum_{h=1}^{H} x_h$ and $\sum_{h=1}^{H} z_h$, with $H \in \{1, \ldots, n\}$, to be two different cumulative distribution functions, where the former dominates the latter in the sense of first-order stochastic dominance. Then the probability of outcome h occurring must be given by: $p_h(\theta, a) = p(\theta, a)x_h + (1 - p(\theta, a))z_h$, where $p(\theta, a) \equiv x(\theta)b(e)d(s)$ is chosen so that all the assumptions made are satisfied. The following example $p(\theta, a) = \theta e^{\gamma} s^{\beta}$, with $\gamma, \beta \in [0,1]^2$ and $\gamma + \beta < 1$, satisfies separability in each dimension and all the assumptions already made. This formulation was used by Lewis and Sappington (2000, 2001) to obtain their results for the particular case in which $\beta = 0$ and $n = 2$. Separability always holds in the case of two outcomes as well as for the standard additively separable technology.[11] Also as discussed at length by Faynzilberg and Praveen (1997, 2000), the optimal contracting problem with moral hazard, adverse selection and risk averse agents is extremely difficult when the technology is non-separable. In fact, Faynzilberg and Praveen (1997) and Faynzilberg and Praveen's (2000) existence result requires separability. Hence, asking for separability while restrictive does not a priori seems to be asking too much.[12]

[11]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).

[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.

6 Conclusions

This paper shows that the optimal menu of contracts in the presence of limited liability, risk neutrality, moral hazard and adverse selection has three highly empirically observed properties: (i) each contract lying in the optimal menu is of the bonus type; (ii) the optimal menu of contracts exhibits the *one-size-fits-all* property; that is, contracts are not customized to the agent's privately know ability; and (iii) better agents work harder and have a higher productivity (see, for instance, Lazear (2000), Paarsch and Shearer (2000) and Seiler (1984)).

The fact that the optimal menu of contracts is such that contract are type independent has an important practical consequence. Mainly, it relieves 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 controlling adequately for unobserved heterogeneity. 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, specially in those where risk-aversion plays a role, my result shows that this is not necessarily the main concern in a setting with risk-neutrality and limited liability such as financial contracting. 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 less importance than suggested by Chiappori and Salanié (2002). Furthermore, the *one-size-fits-all* property provides some theoretical justification for studies that are concerned with the consequences of optimal contracting on performance and make use of aggregate data.

Finally, the results in this paper can shed light on why many financial institutions offer *one-size-fits-all* debt contracts, tax systems do not offer menus of tax schedules where agents are free to choose the schedule that best suits them, and regulatory agencies do not use menu of contracts of the form predicted by the optimal regulation theory as developed by Laffont and Tirole (1986). This however must be taken with caution since the *one-size-fits-all* property of optimal contracts depends on several assumptions; one of particular importance is the co-existence of contractible and non-contractible actions. Hence, the sole observation of the *one-size-fits-all* property cannot be taken as evidence backing the model here without further investigation on the details of the particular environment where this particular feature is observed.

Appendix

A Proofs of the Results in the Main Text

Proof of Proposition 1. The agent's problem in terms of his revelation of type θ', or equivalent in terms of his choice from the menu of contracts offered by the principal, is

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

The first-order condition is given by:

$$\sum_h p_h(\theta, a(\theta, \theta')) \frac{dw_h(\theta')}{d\theta'} + \sum_a \left(\sum_h p_{ha}(\theta, a(\theta, \theta')) w_h(\theta') - c \right) \frac{da(\theta, \theta')}{d\theta'} = 0, \ \forall \theta \in \Theta. \tag{A.1}$$

Using the first-order condition for the action profile this can be re-writen as follows

$$\sum_h p_h(\theta, a(\theta, \theta')) \frac{dw_h(\theta')}{d\theta'} = 0, \ \forall \theta \in \Theta. \tag{A.2}$$

This shows that equation (A.2) evaluated at $\theta' = \theta$ is necessary for implementability.

Because equation (A.2) holds as an identity in θ, one can total differentiate to find

$$\frac{\partial^2}{\partial \theta' \partial \theta'} U(W(\theta'), a(\theta, \theta'), \theta) + \frac{\partial^2}{\partial \theta' \partial \theta} U(W(\theta'), a(\theta, \theta'), \theta) = 0. \tag{A.3}$$

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

$$\sum_h p_{h\theta}(\theta, a(\theta, \theta')) \frac{dw_h(\theta')}{d\theta'} + \sum_a \sum_h p_{ha\theta}(\theta, a(\theta, \theta')) w_h(\theta') \frac{da(\theta, \theta')}{d\theta'} \geq 0,$$

evaluated at $\theta' = \theta$ is necessary for the second-order condition to be satisfied.

The same arguments show that equations

$$\sum_h p_h(\theta, a(\theta, \theta')) \frac{dw_h(\theta')}{d\theta'} = 0 \tag{A.4}$$

$$\sum_h p_{h\theta}(\theta, a(\theta, \theta')) \frac{dw_h(\theta')}{d\theta'} + \sum_a \sum_h p_{ha\theta}(\theta, a(\theta, \theta')) w_h(\theta') \frac{da(\theta, \theta')}{d\theta'} \geq 0 \tag{A.5}$$

holding for all $\theta, \theta' \in \Theta$ are sufficient conditions for implementability. ■

Proof of Proposition 2. The first-order conditions for the Hamiltonian in equation (12) are as follows

$$\mu'(\theta) = -\frac{\partial \mathcal{L}}{\partial U} = f(\theta) - \gamma(\theta), \ \forall \theta \in \Theta \tag{A.6}$$

$$\sum\nolimits_a \big(S_a(\theta, a(\theta))f(\theta) + \mu(\theta)\sum\nolimits_h p_{h\theta a}(\theta, a(\theta))w_h(\theta)\big)\frac{\partial a(\theta)}{\partial w_h(\theta)} +$$

$$\mu(\theta)p_{h\theta}(\theta, a(\theta)) + \lambda_h(\theta) = 0 \tag{A.7}$$

$$\gamma(\theta)(U(\theta) - U) = 0, \ \gamma(\theta) \geq 0, \ U(\theta) \geq U, \ \forall \theta \in \Theta \tag{A.8}$$

$$\mu(\underline{\theta})(U(\underline{\theta}) - L) = 0, \ \mu(\underline{\theta}) \leq 0, \ \mu(\overline{\theta})(U(\overline{\theta}) - L) = 0, \ \mu(\overline{\theta}) \geq 0 \tag{A.9}$$

$$\lambda_h(\theta)(w_h(\theta) - L) = 0, \ \lambda_h(\theta) \geq 0, \ w_h(\theta) \geq L, \ \forall h = \{1, \ldots, n\}, \ \forall \theta \in \Theta \tag{A.10}$$

Solving the differential equation for $\mu'(\theta)$ in equation (A.6) as done before one can show as before that

$$\mu(\theta) = -(1 - F(\theta)).$$

Let me guess a non-decreasing wage profile as the solution to the optimal control problem; that is, for all $\forall\, \theta$ and any pair (h', h) with $h' > h$, $w_{h'}(\theta) \geq w_h(\theta)$.

Recall that

$$\sum\nolimits_h p_{he}(\theta, a(\theta))w_h(\theta) = c \tag{A.11}$$

$$\sum\nolimits_h p_{hs}(\theta, a(\theta))w_h(\theta) = c \tag{A.12}$$

It readily follows from these first-order conditions that

$$\frac{\partial a(\theta)}{\partial w_h(\theta)} = \frac{-p_{ha}(\theta, a(\theta))U_{a'a'}(W, a, \theta) + p_{ha'}(\theta, a(\theta))U_{es}(W, a, \theta)}{U_{ee}(W, a, \theta)U_{ss}(W, a, \theta) - U_{es}^2(W, a, \theta)}, \forall a, a' \in \{e, s\}, a \neq a'$$

It readily follows from this, **MLRP** with respect to a and the fact that the wage profile is monotone in h that $\frac{1}{p_h(\theta, a(\theta))}\frac{\partial a(\theta)}{\partial w_h(\theta)}$ increases with h.

Multiplying both sides of the first-order condition for $w_h(\theta)$ by $w_h(\theta) - L$ and using the complementary-slackness condition for the limited liability constraint, I get that

$$(w_h(\theta) - L)\Big(\sum\nolimits_a \big(S_a(\theta, a(\theta)) - H(\theta)\sum\nolimits_h p_{h\theta a}(\theta, a(\theta))w_h(\theta)\big)\frac{\partial a(\theta)}{\partial w_h(\theta)} -$$

$$H(\theta)p_{h\theta}(\theta, a(\theta))\Big) = 0, \ \forall h \in \{1, \ldots, n\}.$$

Summing over all these constraints and using the first-order conditions for actions, one gets that

$$\sum_a \left(S_a(\theta, a(\theta)) - H(\theta) \sum_h p_{h\theta a}(\theta, a(\theta)) w_h(\theta)\right) \frac{U_{es} - U_{a'a'}}{U_{ee}U_{ss} - U_{es}^2} c = H(\theta) \sum_h p_{h\theta}(\theta, a(\theta))(w_h(\theta) - L).$$

Observe that the RHS can be written as

$$H(\theta) \sum_h \ell_{h\theta}(\theta, a(\theta)) p_h(\theta, a(\theta)) w_h(\theta),$$

which is the covariance between $w_h(\theta)$ and $\ell_{h\theta}(\theta, a(\theta))$.

Because **MLRP** implies that $\ell_{h'\theta}(\theta, a) > \ell_{h\theta}(\theta, a),\ \forall h' > h$, the mean of $\ell_{h\theta}(\theta, a)$ is zero and I am guessing that the optimal wage profile is monotone in h, the RHS is positive. Thus, the term on the LHS is positive.

It readily follows from this, **MLRP** with respect to θ and equation (A.7) that if $\lambda_h = 0$ for any $h \in \{2, \ldots, n-1\}$, then $\lambda_{h'} < 0$ for all $h' > h$, which is a contradiction since $\lambda_h \geq 0,\ \forall h$. If $\lambda_1 \geq 0$, then either $\lambda_h > 0$ for all $h > 1$, which implies that only a zero effort can be implemented, or there exists a threshold $\hat{h} < n$ such that $\lambda_h > 0,\ \forall h < \hat{h}$ and $\lambda_h < 0,\ \forall h > \hat{h}$. Thus, the only possible solution entails $\lambda_n = 0$ and $\lambda_h > 0,\ \forall h \in \{1, \ldots, n-1\}$. This implies that $w_h(\theta) = L,\ \forall h \in \{1, \ldots, n-1\},\ \forall \theta \in \Theta$ and $w_n(\theta) > L,\ \forall \theta \in \Theta$.

Recall that a necessary condition for implementability requires that

$$\sum_h p_h(\theta, a^{AML}(\theta)) \frac{dw_h(\theta)}{d\theta} = 0$$

Because $w_h(\theta) = L,\ \forall h \in \{1, \ldots, n-1\},\ \forall \theta \in \Theta$ and $p_n(\theta, a^{AML}(\theta)) > 0$, this becomes

$$\frac{dw_n(\theta)}{d\theta} = 0.$$

Thus, integrating both sides one gets that $w_n(\theta) = w_n(\underline{\theta})$

Using the transversality condition for $\underline{\theta}$ in equation (A.9), one gets that $U(\underline{\theta}) = L$ (since $L \geq U$ and $\mu(\underline{\theta}) < 0$). Thus,

$$w_n(\underline{\theta})) = \frac{c(a(\underline{\theta}))}{p_n(\underline{\theta}, a(\underline{\theta}))} + L. \tag{A.13}$$

Because $\frac{dw_n(\theta)}{d\theta} = 0$, this together with $w_n(\underline{\theta})$ yields the result in equation (21). Hence, the optimal contract as guessed is monotonic in h and type independent.

Before ending the proof I need to check that this solution satisfies the second-order condition in equation (A.5), which requires the following

$$\sum\nolimits_h p_{h\theta}(\theta, a(\theta'))\frac{dw_h(\theta')}{d\theta'} + \sum\nolimits_a p_{na\theta}(\theta, a(\theta'))\big(w_n(\theta') - L\big)\frac{da(\theta')}{d\theta'} \geq 0,\ \forall \theta, \theta' \in \Theta.$$

Because $\frac{dw_h(\theta')}{d\theta'} = 0$ for all $\theta' \in \Theta$, then $a(\theta, \theta') = a(\theta)$ for all $\theta' \in \Theta$. This together with the fact that $w_n(\theta') > L$ for all $\theta' \in \Theta$ implies that the global condition becomes

$$\sum\nolimits_a p_{na\theta}(\theta, a(\theta'))\big(w_n(\theta') - L\big)\frac{da(\theta')}{d\theta'} \geq 0,\ \forall \theta, \theta' \in \Theta.$$

Observe that this holds since $p_{na\theta}(\theta, a(\theta')) \geq 0$ for all $a \in A$, $w_n(\theta') > L$ and $\frac{da(\theta')}{d\theta'} \geq 0$ for $a = e, s$. The latter follows from the fact that $\frac{dw_n(\theta)}{d\theta} = 0$ and from partially differentiating the agent's first-order conditions

$$p_{ne}(\theta, a(\theta))\frac{c(a(\underline{\theta}))}{p_n(\underline{\theta}, a(\underline{\theta}))} - c = 0,$$
$$p_{ns}(\theta, a(\theta))\frac{c(a(\underline{\theta}))}{p_n(\underline{\theta}, a(\underline{\theta}))} - c = 0.$$

This leads to

$$\frac{de(\theta)}{d\theta} = \frac{-p_{nss}(\theta, a)p_{ne\theta}(\theta, a) + p_{nes}(\theta, a)p_{ns\theta}(\theta, a)}{p_{nss}(\theta, a)p_{nee}(\theta, a) - p_{nes}(\theta, a)^2} \geq 0$$

and

$$\frac{ds(\theta)}{d\theta} = \frac{-p_{nee}(\theta, a)p_{ns\theta}(\theta, a) + p_{nes}(\theta, a)p_{ne\theta}(\theta, a)}{p_{nss}(\theta, a)p_{nee}(\theta, a) - p_{nes}(\theta, a)^2} \geq 0$$

Finally, I need to show that the first-order approach. To this end notice recall that

$$\sum\nolimits_h p_{he}(\theta, a^{AMU}(\theta))w_h(\theta) = c > 0,$$
$$\sum\nolimits_h p_{hs}(\theta, a^{AMU}(\theta))w_h(\theta) = c > 0.$$

Let $P_h(\theta, a)$ be $\sum_{h' \geq h} p_{h'}(\theta, a)$ and $\triangle w_h$ be $w_h - w_{h-1}$ for $h \geq 1$ and $\triangle w_h = w_1$ for $h = 1$. Then note that applying the summation by parts formula to the first-order conditions for actions one gets that this re-writes as follows

$$\sum\nolimits_h P_{he}(\theta, a^{AMU}(\theta))\triangle w_h(\theta) = c \tag{A.14}$$

$$\sum\nolimits_h P_{hs}(\theta, a^{AMU}(\theta))\triangle w_h(\theta) = c \tag{A.15}$$

Hence,

$$U_{aa}(W,a,\theta) = \sum\nolimits_h \triangle w_h P_{haa}(\theta,a) \text{ for } a \in \{e,s\}.$$

and

$$U_{es}(W,a,\theta) = \sum\nolimits_h \triangle w_h P_{hes}(\theta,a).$$

Note that $U_{aa}(W,a,\theta) \leq 0$ by **CP** and $\triangle w_h \geq 0$ for all h. Thus, the agent's second order condition is satisfied if and only if for all $a \in A$

$$\Big(\sum\nolimits_h \triangle w_h P_{hee}(\theta,a)\Big)\Big(\sum\nolimits_h \triangle w_h P_{hss}(\theta,a)\Big) - \Big(\sum\nolimits_h \triangle w_h P_{hes}(\theta,a)\Big)^2 \geq 0.$$

Assumption **CP** together with the fact $\triangle w_h \geq 0$ guarantees that this term is positive. To see this note that

$$\begin{aligned}&\sum\nolimits_h \triangle w_h P_{hee}(\theta,a) \sum\nolimits_h \triangle w_h P_{hss}(\theta,a) - \Big(\sum\nolimits_h \triangle w_h P_{hes}(\theta,a)\Big)^2 \\ \geq &\Big(\sum\nolimits_h \triangle w_h \sqrt{P_{hee}(\theta,a)}\sqrt{P_{hss}(\theta,a)}\Big)^2 - \Big(\sum\nolimits_h \triangle w_h P_{hes}(\theta,a)\Big)^2\end{aligned}$$

where the inequality follows from the Cauchy-Schwartz inequality.

It follows then that the agent's utility function is concave in a for any $a \in A$ if

$$\sum\nolimits_h \triangle w_h \Big(\sqrt{P_{hee}(\theta,a)}\sqrt{P_{hss}(\theta,a)} - P_{hes}(\theta,a)\Big) \geq 0.$$

Observe that assumption **CP** ensures that the term in square brackets is positive for all h. This together with the fact that $\triangle w_h \geq 0$ for all h yields the desired result. Thus, $U(W,a,\theta)$ is concave in a whenever $\triangle w_h \geq 0$ holds.

Because concavity of the agent's problem requires that $w_{h'}(\theta) \geq w_h(\theta),\ \forall h, h' \in \{1,\ldots,n\},\ h' > h$. This is a necessary condition for optimality.

■

Proof of Proposition 3. The agent's problem in terms of his revelation of type θ', or 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:

$$\sum_h p_h(\theta, e(\theta'), s(\theta,\theta'))\frac{dw_h(\theta')}{d\theta'} + \big(\sum_h p_{hs}(\theta, e(\theta'), s(\theta,\theta'))w_h(\theta') - c\big)\frac{ds(\theta,\theta')}{d\theta'} + \big(\sum_h p_{he}(\theta, e(\theta'), s(\theta,\theta'))w_h(\theta') - c\big)\frac{de(\theta')}{d\theta'} = 0,\ \forall\theta\in\Theta. \tag{A.16}$$

Using the first-order condition for the action profile this can be re-written as follows

$$\sum_h p_h(\theta, e(\theta'), s(\theta,\theta'))\frac{dw_h(\theta')}{d\theta'} + \big(\sum_h p_{he}(\theta, e(\theta'), s(\theta,\theta'))w_h(\theta') - c\big)\frac{de(\theta')}{d\theta'} = 0,\ \forall\theta\in\Theta. \tag{A.17}$$

This shows that equation (A.17) evaluated at $\theta' = \theta$ is necessary for implementability.

Because this holds as an identity in θ, I can total differentiate to find

$$\frac{\partial U(W(\theta'), a(\theta'), \theta)}{\partial\theta'\partial\theta'} + \frac{\partial U(W(\theta'), a(\theta'), \theta)}{\partial\theta'\partial\theta} = 0 \tag{A.18}$$

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

$$\sum_h p_{h\theta}(\theta, e(\theta'), s(\theta,\theta'))\frac{dw_h(\theta')}{d\theta'} + \sum_h p_{he\theta}(\theta, e(\theta'), s(\theta,\theta'))w_h(\theta')\frac{de(\theta')}{d\theta'} + \sum_h p_{hs\theta}(\theta, e(\theta'), s(\theta,\theta'))w_h(\theta')\frac{ds(\theta,\theta')}{d\theta'} \geq 0, \tag{A.19}$$

evaluated at $\theta' = \theta$ is necessary for the second-order condition to be satisfied.

The same arguments show that equations (A.16) and (A.19) holding for all $\theta, \theta' \in \Theta$ are sufficient conditions for implementability.

■

Proof of Proposition 4. The first-order conditions for the Hamiltonian in equation

(13) are as follows

$$\mu'(\theta) = -\frac{\partial \mathcal{L}}{\partial U} = f(\theta) - \gamma(\theta), \ \forall \theta \in \Theta \tag{A.20}$$

$$\big(S_s(\theta, a(\theta))f(\theta) + \mu(\theta) \textstyle\sum_h p_{h\theta s}(\theta, a(\theta)) w_h(\theta)\big) \frac{\partial s(\theta)}{\partial w_h(\theta)} +$$

$$\mu(\theta) p_{h\theta}(\theta, a(\theta)) + \lambda_h(\theta) = 0 \tag{A.21}$$

$$S_e(\theta, a(\theta))f(\theta) + \mu(\theta) \textstyle\sum_h p_{h\theta e}(\theta, a(\theta)) w_h(\theta) + \tag{A.22}$$

$$\big(S_s(\theta, a(\theta))f(\theta) + \mu(\theta) \textstyle\sum_h p_{h\theta s}(\theta, a(\theta)) w_h(\theta)\big) \frac{\partial s(\theta)}{\partial e(\theta)} = 0$$

$$\gamma(\theta)(U(\theta) - U) = 0, \ \gamma(\theta) \geq 0, \ U(\theta) \geq U, \ \forall \theta \in \Theta \tag{A.23}$$

$$\mu(\underline{\theta})(U(\underline{\theta}) - L) = 0, \ \mu(\underline{\theta}) \leq 0, \ \mu(\overline{\theta})(U(\overline{\theta}) - L) = 0, \ \mu(\overline{\theta}) \geq 0 \tag{A.24}$$

$$\lambda_h(\theta)(w_h(\theta) - L) = 0, \ \lambda_h(\theta) \geq 0, \ w_h(\theta) \geq L, \ \forall h = \{1, \dots, n\}, \ \forall \theta \in \Theta \tag{A.25}$$

Solving the differential equation for $\mu'(\theta)$ in equation (A.6) as done before one can show as before that

$$\mu(\theta) = -(1 - F(\theta)).$$

Recall that $s(\theta)$ satisfies

$$\textstyle\sum_h p_{hs}(\theta, a(\theta)) w_h(\theta) = c$$

It readily follows from this that

$$\frac{\partial s(\theta)}{\partial w_h(\theta)} = -\frac{p_{hs}(\theta, a(\theta))}{U_{ss}(W, a, \theta)},$$

and

$$\frac{\partial s(\theta)}{\partial e(\theta)} = -\frac{U_{es}(W, a, \theta)}{U_{ss}(W, a, \theta)} \geq 0.$$

Using the same argument as that in proposition 2, one can show that $w_h(\theta) = L, \ \forall h \in \{1, \dots, n-1\}$ and $w_n(\theta) > L$.

Multiplying both sides of the first-order condition for $w_h(\theta)$ by $w_h(\theta) - L$ and using the complementary-slackness condition for the limited liability constraint, summing over all these equations and using the first-order conditions for actions, one gets that

$$\big(S_s(\theta, a(\theta)) - H(\theta) \textstyle\sum_h p_{h\theta s}(\theta, a(\theta)) w_h(\theta)\big) \frac{c}{U_{ss}} = H(\theta) \textstyle\sum_h p_{h\theta}(\theta, a(\theta)) w_h(\theta).$$

Substituting this into the first-order condition for $e(\theta)$ one gets that

$$S_e(\theta, a(\theta)) - H(\theta) \sum_h \Big(p_{h\theta e}(\theta, a(\theta)) - p_{h\theta}(\theta, a(\theta)) \frac{U_{es}(W(\theta), a(\theta), \theta)}{c}\Big) w_h(\theta) = 0$$

Using the fact that $w_h(\theta) = L,\ \forall h \in \{1, \dots, n-1\}$ and the first-order condition for action s, this rewrites as follows

$$S_e(\theta, a(\theta)) - H(\theta) \Big(p_{n\theta e}(\theta, a(\theta)) - p_{n\theta}(\theta, a(\theta)) \frac{p_{nes}(\theta, a(\theta))}{p_{ns}(\theta, a(\theta))}\Big) (w_n(\theta) - L) = 0. \tag{A.26}$$

This re-writes as follows

$$S_e(\theta, a(\theta)) - H(\theta) \frac{\partial \big(p_{n\theta}(\theta, a(\theta))/p_{ns}(\theta, a(\theta))\big)}{\partial e(\theta)} c = 0.$$

Thus, there is no effort distortion at the top since $H(\overline{\theta}) = 0$ and the optimal effort is downward distorted when $\frac{\partial \left(p_{n\theta}(\theta, a(\theta))/p_{ns}(\theta, a(\theta))\right)}{\partial e(\theta)} > 0$ and upward distorted otherwise.

It readily follows from equation (A.26) and $w_h(\theta) = L,\ \forall h \in \{1, \dots, n-1\}$ that the optimal strategy $s^{AMC}(\theta)$ and the optimal wage are determined by the following equations

$$p_{ns}(\theta, a^{AMC}(\theta)) \hat{w}_n(\theta) = c$$

and

$$S_s(\theta, a^{AMC}(\theta)) = H(\theta) \frac{\partial \big(p_{n\theta}(\theta, a^{AMC}(\theta))/p_{ns}(\theta, a^{AMC}(\theta))\big)}{\partial s(\theta)} c > 0.$$

The global condition in equation (A.19) requires the following

$$\begin{aligned} &p_{ne\theta}(\theta, e(\theta'), s(\theta, \theta'))(w_n(\theta') - L) \frac{de(\theta')}{d\theta'} + p_{n\theta}(\theta, e(\theta'), s(\theta, \theta')) \frac{dw_n(\theta')}{d\theta'} + \\ &p_{ns\theta}(\theta, e(\theta'), s(\theta, \theta'))(w_n(\theta') - L) \frac{ds(\theta, \theta')}{d\theta'} \geq 0. \end{aligned}$$

Using the fact that $\frac{ds(\theta,\theta')}{d\theta'} \equiv \sum_h \frac{\partial s(\theta,\theta')}{\partial w_h(\theta')} \frac{dw_h(\theta')}{d\theta'} + \frac{\partial s(\theta,\theta')}{\partial e(\theta')} \frac{de(\theta')}{d\theta'}$ and $p_{nss}(\theta, e(\theta'), s(\theta, \theta')) < 0$, the sufficient condition re-writes as follows $\forall\ \theta, \theta' \in \Theta$,

$$\begin{aligned} &\Big(p_{ne\theta}(\theta, e(\theta'), s(\theta, \theta')) p_{nss}(\theta, e(\theta'), s(\theta, \theta')) - p_{ns\theta}(\theta, e(\theta'), s(\theta, \theta')) p_{nes}(\theta, e(\theta'), s(\theta, \theta'))\Big) \hat{w}_n(\theta') \frac{de(\theta')}{d\theta'} + \\ &\Big(p_{n\theta}(\theta, e(\theta'), s(\theta, \theta')) p_{nss}(\theta, e(\theta'), s(\theta, \theta')) - p_{ns}(\theta, e(\theta'), s(\theta, \theta')) p_{ns\theta}(\theta, e(\theta'), s(\theta, \theta'))\Big) \frac{dw_n(\theta')}{d\theta'} \leq 0. \end{aligned}$$

Notice that a sufficient condition but not a necessary condition for this to hold is that both, $e(\theta)$ and $w_n(\theta)$ are both non-decreasing with θ. However, as we will show below, this is restrict the set of implementable mechanisms too much.

Lets define $J(\theta, a(\theta))$ as $S(\theta, a(\theta)) - H(\theta)\Big(p_{n\theta}(\theta, a(\theta))/p_{ns}(\theta, a(\theta))\Big)$. It is easy to check that

$$\frac{de(\theta)}{d\theta} = -\Big(J_{ee}(\theta, a(\theta))J_{ss}(\theta, a(\theta)) - J_{es}^2(\theta, e(\theta))\Big)^{-1} \tag{A.27}$$
$$\Big(J_{e\theta}(\theta, a(\theta))J_{ss}(\theta, a(\theta)) - J_{es}(\theta, a(\theta))J_{s\theta}(\theta, a(\theta))\Big),$$

$$\frac{ds(\theta)}{d\theta} = -\Big(J_{ee}(\theta, a(\theta))J_{ss}(\theta, a(\theta)) - J_{es}^2(\theta, a(\theta))\Big)^{-1} \tag{A.28}$$
$$\Big(J_{s\theta}(\theta, a(\theta))J_{ee}(\theta, a(\theta)) - J_{es}(\theta, a(\theta))J_{e\theta}(\theta, a(\theta))\Big)$$

and

$$\frac{dw_n(\theta)}{d\theta} = -\frac{\hat{w}(\theta)}{p_{ns}(\theta, a(\theta))}\Big(p_{nss}(\theta, a(\theta))\frac{ds(\theta)}{d\theta} + p_{nse}(\theta, a(\theta))\frac{de(\theta)}{d\theta} + p_{ns\theta}(\theta, a(\theta))\Big) \tag{A.29}$$

Observe that for any $a, a' \in \{e, s\}$

$$J_{aa'}(\theta, e(\theta)) = S_{aa'}(\theta, a(\theta)) - H(\theta)\frac{\partial}{\partial a \partial a'}\Big(p_{n\theta}(\theta, a(\theta))/p_{ns}(\theta, a(\theta))\Big)$$

and for any $a \in \{e, s\}$

$$J_{a\theta}(\theta, e(\theta)) = S_{a\theta}(\theta, a(\theta)) - H'(\theta)\frac{\partial}{\partial a}\Big(p_{n\theta}(\theta, a(\theta))/p_{ns}(\theta, a(\theta))\Big) -$$
$$H(\theta)\frac{\partial}{\partial a \partial \theta}\Big(p_{n\theta}(\theta, a(\theta))/p_{ns}(\theta, a(\theta))\Big).$$

Using the necessary condition in equation (A.17) and the first-order condition for $s(\theta, \theta')$, this re-writes as follows

$$\Bigg(p_{ne\theta}(\theta, e(\theta'), s(\theta, \theta'))p_{nss}(\theta, e(\theta'), s(\theta, \theta')) - p_{ns\theta}(\theta, e(\theta'), s(\theta, \theta'))p_{nes}(\theta, e(\theta'), s(\theta, \theta')) +$$
$$\Big(p_{ns}(\theta, e(\theta'), s(\theta, \theta'))p_{ns\theta}(\theta, e(\theta'), s(\theta, \theta')) - p_{n\theta}(\theta, e(\theta'), s(\theta, \theta'))p_{nss}(\theta, e(\theta'), s(\theta, \theta'))\Big)$$
$$\frac{p_{ne}(\theta, e(\theta'), s(\theta, \theta')) - p_{ns}(\theta, e(\theta'), s(\theta, \theta'))}{p_n(\theta, e(\theta'), s(\theta, \theta'))}\Bigg)\frac{de(\theta')}{d\theta'} \leq 0, \ \forall \theta, \theta' \in \Theta.$$

Rearing terms this can be written as

$$\frac{1}{p_n(\theta, e(\theta'), s(\theta, \theta'))}\Bigg(p_{nss}(\theta, e(\theta'), s(\theta, \theta'))\Big(p_{ne\theta}(\theta, e(\theta'), s(\theta, \theta'))p_n(\theta, e(\theta'), s(\theta, \theta')) -$$
$$p_{n\theta}(\theta, e(\theta'), s(\theta, \theta'))p_{ne}(\theta, e(\theta'), s(\theta, \theta')) + p_{n\theta}(\theta, e(\theta'), s(\theta, \theta'))p_{ns}(\theta, e(\theta'), s(\theta, \theta'))\Big) -$$
$$p_{ns\theta}(\theta, e(\theta'), s(\theta, \theta'))\Big(p_{nes}(\theta, e(\theta'), s(\theta, \theta'))p_n(\theta, e(\theta'), s(\theta, \theta')) - p_{ns}(\theta, e(\theta'), s(\theta, \theta'))p_{ne}(\theta, e(\theta'), s(\theta, \theta')) +$$
$$p_{ns}^2(\theta, e(\theta'), s(\theta, \theta'))\Big)\Bigg)\frac{de(\theta')}{d\theta'} \leq 0, \ \forall \theta, \theta' \in \Theta.$$

A sufficient condition for this to hold is the following: for any $(\theta, e, s) \in \Theta \times A$, $p_{ne\theta}(\theta, e, s)p_n(\theta, e, s) - p_{n\theta}(\theta, e, s)p_{ne}(\theta, e, s) \geq 0$ and $p_{nes}(\theta, e, s)p_n(\theta, e, s) - p_{ns}(\theta, e, s)p_{ne}(\theta, e, s) \geq 0$. Furthermore, this also ensures that $\frac{de(\theta')}{d\theta'} \geq 0$.

In particular, separability with respect to each input ensures that this sufficient condition holds. Separability requires that the probability distribution satisfies the following: take $\sum_{h=1}^{H} x_h$ and $\sum_{h=1}^{H} z_h$, with $H \in \{1, \ldots, n\}$, to be two different cumulative distribution functions, where the former dominates the latter in the sense of first-order stochastic dominance. Then the probability of outcome h occurring must be given by: $p_h(\theta, a) = p(\theta, a)x_h + (1 - p(\theta, a))z_h$, where $p(\theta, a) \equiv x(\theta)b(e)d(s)$ is chosen so that all the assumptions made are satisfied. The following example $p(\theta, a) = \theta e^{\gamma} s^{\beta}$, with $\gamma, \beta \in [0, 1]^2$ and $\gamma + \beta < 1$, satisfies separability in each dimension and all the assumptions already made. The reason is that for any $(\theta, e, s) \in \Theta \times A$, $p_{ne\theta}(\theta, e, s(\theta,)p_n(\theta, e, s) - p_{n\theta}(\theta, e, s)p_{ne}(\theta, e, s) = 0$ and $p_{nes}(\theta, e, s(\theta,)p_n(\theta, e, s) - p_{ns}(\theta, e, s)p_{ne}(\theta, e, s) = 0$. Furthermore, this also ensures that $\frac{de(\theta')}{d\theta'} \geq 0$.

Finally, one can show following the same steps as above that the first-order approach is valid. ■

B Moral Hazard, Multiple Tasks and Limited Liability

In this section, I analyze the case in which the agent's ability is known to the principal and the agent before signing the contract, but the action profile is subject to moral hazard.

Lets suppose that an agent of ability $\theta \in \Theta$ is faced with wage schedule $W \equiv \{w_h\}_{h=1}^{n}$ and chooses the action profile $a \in A$, then his expected utility is given by:

$$U(W, a, \theta) \equiv \sum\nolimits_h p_h(\theta, a)w_h - c(a). \tag{B.1}$$

Observe that the linearity of the agent's payoff function in equation (B.1) with respect to wages implies that if the principal could offer a contract in which payments are unbounded, the first-best efficient action profile could be implemented by offering a contract with a sufficiently harsh punishment to the agent. This however is prevented by the limited-liability constraint.

The literature on optimal contracting under moral hazard is vast. In the multitasking case, this article is more closely related to the contributions of Ratto and Schnedler (2008) and Dewatripont and Tirole (1999). Ratto and Schnedler (2008) study a situation where production requires two non-conflicting tasks, and the manager wants to direct production to achieve a preferred allocation of effort across tasks. However, aggregated production is the only indicator of agent activity. The main result is that the principal cannot implement the preferred allocation with a single agent, yet he is able to do so by inducing a game among two agents. These results complement those in Dewatripont and Tirole (1999), who also found that under multitasking there are implementation problems, but based on direct conflicts between tasks. They show that it always better to split the task of finding evidence in favor and against a decision between two agents. The reason stands for the fact that the optimal compensation can be based only on an aggregated measure of the task and this is increasing in the outcome of one task and decreasing in the outcome of the other task. This implies that it is impossible to induce one agent to exert more effort in both tasks and thus it is optimal to split the task between two different agents to avoid conflict of interest in job design.[13]

There is a less closely related strand which deals with moral hazard only under risk aversion.[14] While this is less related to the paper here, I will mentioned for the sake of completeness. One strand of this literature focuses on noisy performance measures and risk averse workers. The cornerstone of this literature is the existence of a negative trade-off between risk and incentives; that is, the more noisy the performance measure, the less power the incentives. However, as Prendergast (2002b) and Prendergast (2002a) point out, the data do not confirm the existence of such trade-off. Another strand focuses on the issue of how to reward a given task without harming other tasks and how the number of tasks affects the risk and incentives trade-off within the effort substitution approach.

[13]Conflicting tasks also provides incentives for sabotage and collusion is unavoidable, which is not the case with non-conflicting tasks.

[14]See Gibbons (1998) for an extensive review of the incentive contracting literature.

Understanding this dimension of incentive contracting is at the core of the multi-tasking literature as developed by Holmstrom and Milgrom (1987), Holmström and Milgrom (1991), Baker (2002), Itoh (1991, 1992, 1994), Dewatripont et al. (1999) and MacDonald and Marx (2001). The main result of this literature is that firms reduce the power of incentives to avoid rewarding the wrong behavior. The driving force behind this is the existence of noisy or distorted performance measures coupled with effort substitution that result in distorted incentives.

It readily follows from the agent's payoff in equation (B.1) that a contract W satisfying the limited liability constraint **LL** induces an agent with ability θ to choose the action profile $a \in A$ if and only if the following incentive-compatibility constraint holds

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

and it induces the agent to participate if the following individually rationality constraint holds,

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

Thus the principal's problem consists on choosing a contract (W, a) that maximizes his expected payoff subject to these constraints and the **LL** constraint. Thus, the principal must solve the following unrelaxed program

$$\max_{(W,a)\in\Re^n\times A} \sum_h p_h(\theta, a)(y_h - w_h)$$

subject to

(1), (B.2) and (B.3).

Let the solution to this program be (W^m, a^m).

Assumption 1

i) A solution to the unrelaxed program exists with $e \in (0, \bar{e})$ *and* $s \in (0, \bar{s})$.

ii) $\sum_h p_h(\theta, a^m)(y_h - w_h^m) > 0$.

Making use of the first-order approach, which entails to enlarging the constraint set, the incentive compatibility constraints in equation (B.2) can be replaced by the first- and second-order conditions for the agent's problem, which are:

$$\sum_h p_{he}(\theta, a)w_h - c \begin{cases} \leq 0 & \text{if } e = 0 \\ = 0 & \text{if } e \in (0, \bar{e}) \\ \geq 0 & \text{if } e = \bar{e} \end{cases} \tag{B.4}$$

$$\sum_h p_{hs}(\theta, a)w_h - c \begin{cases} \leq 0 & \text{if } s = 0 \\ = 0 & \text{if } s \in (0, \bar{s}) \\ \geq 0 & \text{if } s = \bar{s} \end{cases} \tag{B.5}$$

and

$$\sum_h p_{hss}(\theta, a)w_h \sum_h p_{hee}(\theta, a)w_h - \left(\sum_h p_{hes}(\theta, a)w_h\right)^2 \geq 0 \text{ for any } a \in A. \tag{B.6}$$

Then the principal's relaxed problem becomes

$$\max_{(W,a)\in\Re^n \times A} \sum_h p_h(\theta, a)(y_h - w_h) \tag{P-MH}$$

subject to

$$(1),\ (\text{B.3}),\ (\text{B.4}),\ (\text{B.5}) \text{ and } (\text{B.6}).$$

The Lagrangean for this problem when the second-order condition for the agent's problme in equation (B.6) is ignored is as follows

$$\mathcal{L}(W, \mu_e, \mu_s, \gamma, \lambda) = \sum_h p_h(\theta, a)(y_h - w_h) + \mu_e\Big(\sum_h p_{he}(\theta, a)w_h - c\Big) + \tag{B.7}$$
$$\mu_s\Big(\sum_h p_{hs}(\theta, a)w_h - c\Big) + \gamma\Big(\sum_h p_h(\theta, a)w_h - c(a) - U\Big) + \sum_h \lambda_h\big(w_h - L\big),$$

where μ_a is the Lagrange multiplier for the first-order condition with regard to action $a \in \{e, s\}$, γ is the multiplier for the agent's participation constraint and λ_h is the multiplier for the limited liability constraint with regard to the compensation in state h.

The following is proven in the appendix.

Proposition B.1 *Suppose **MLRP**, **CP** and **SC** hold and* (W,a) *is a solution to the relaxed program. Then,* (W,a) *is a solution to the unrelaxed program.*

Proof of Proposition B.1.

Lemma 1 *If* (W,a) *solves the relaxed program, there exists real numbers* γ, μ_e, μ_s *and* $\lambda_h, h=1,\dots,n$ *such that*

$$w_h: \; -p_h(\theta,a)+\mu_e p_{he}(\theta,a)+\mu_s p_{hs}(\theta,a)+\gamma p_h(\theta,a)+\lambda_h=0, \tag{B.8}$$

$$\begin{aligned} e: &\sum\nolimits_h p_{he}(\theta,a)(y_h-w_h)+\mu_e\sum\nolimits_h p_{hee}(\theta,a)w_h+ \\ &\mu_s\sum\nolimits_h p_{hse}(\theta,a)w_h+\gamma\Big(\sum\nolimits_h p_{he}(\theta,a)w_h-c\Big)\begin{cases}\leq 0 & \text{if } e=0\\ =0 & \text{if } e\in(0,\bar{e})\\ \geq 0 & \text{if } e=\bar{e}\end{cases}\end{aligned} \tag{B.9}$$

$$\begin{aligned} s: &\sum\nolimits_h p_{hs}(\theta,a)(y_h-w_h)+\mu_e\sum\nolimits_h p_{hes}(\theta,a)w_h+ \\ &\mu_s\sum\nolimits_h p_{hss}(\theta,a)w_h+\gamma\Big(\sum\nolimits_h p_{hs}(\theta,a)w_h-c\Big)\begin{cases}\leq 0 & \text{if } s=0\\ =0 & \text{if } s\in(0,\bar{s})\\ \geq 0 & \text{if } s=\bar{s}\end{cases}\end{aligned} \tag{B.10}$$

and

$$\mu_e \gtreqless 0,\; \mu_s \gtreqless 0,\; \lambda_h\geq 0 \;\; \text{and}\; \gamma\geq 0. \tag{B.11}$$

Proof. These are the Khun-Tucker necessary conditions for optimality. ■

Lemma 2 *If* (W,a) *solves the relaxed program and* $a\in(0,\bar{a})$ *for* $a=e,s$, *then* $w_h = L$, $\forall h=1,\dots,n-1$ *and* $w_n>L$. *If* $a=0$, *then* $w_n=L$.

Proof. Summing up the first-order conditions in equation (B) over h, one gets that $\sum_h \lambda_h = 1-\gamma\geq 0$, since $\sum_h p_{ha}(\theta,a)=0$ for all a. Thus, if the participation constraint does not bind, the **LL** constraint binds for at least one outcome. Multiplying both sides of the FOC by $w_h(\theta) - L$ and summing over h one gets that

$$-\sum_h p_h(\theta,a)(w_h - L) + \mu_e \sum_h p_{he}(\theta,a)(w_h - L) + \mu_s \sum_h p_{hs}(\theta,a)(w_h - L) + \gamma \sum_h p_h(\theta,a)(w_h - L) + \sum_h \lambda_h (w_h - L) = 0.$$

Using the incentive constraint for actions, the complementarity slackness constraint for the agent's participation decision and complementarity slackness constraint for w_h, this can be written as follows

$$\sum_h p_h(\theta,a)(w_h - L) = \mu_e \sum_h p_{he}(\theta,a)(w_h - L) + \mu_s \sum_h p_{hs}(\theta,a)(w_h - L) + \gamma\big(U + c(a) - L\big). \tag{B.12}$$

In what follows I will assume that the participation constraint does not bind and then I will show that at the optimal contract this is always satisfied.

Suppose that μ_e and μ_s are such that $\mu_e p_{he}(\theta,a) + \mu_s p_{hs}(\theta,a)$ is strictly decreasing in h. Then it follows from the FOC in equation (B) that if the wage is greater than L for two different outcomes, say y_h and $y_{h'}$, then $\lambda_h = \lambda_{h'} = 0$. Because $\mu_e p_{he}(\theta,a) + \mu_s p_{hs}(\theta,a)$ falls with h, it follows from the FOC in equation (B) that

$$\gamma - 1 + \ell_e p_{he}(\theta,a) + \ell_s p_{hs}(\theta,a) = \gamma - 1 + \ell_e p_{h'e}(\theta,a) + \ell_s p_{h's}(\theta,a) \tag{B.13}$$

which contradicts the hypothesis that $\mu_e p_{he}(\theta,a) + \mu_s p_{hs}(\theta,a)$ falls with h. It follows from this that $\lambda_h \geq 0$ for $h = 1$ and $\lambda_h > 0$ for all $h > 1$.

This implies that $w_h \geq w_{h+1}$ for all $h = 1, \ldots, n-1$ and thus the agent's first order condition for effort e and that for s in equations (B.4) and (B.5) are respectively such taht $U_e(W,a,\theta) < 0$ and $U_s(W,a,\theta) < 0$. This implies that $e = 0$ and $s = 0$ and thus the optimal contract cannot implement a positive effort in either task.

Suppose now $\mu_e = 0$ and $\mu_s = 0$ and therefore $\mu_e p_{he}(\theta,a) + \mu_s p_{hs}(\theta,a) = 0$ for all $h = 1, \ldots, n$. Then it follows from that $\lambda_h > 0$, $\forall h$ and therefore $w_h = L$, $\forall h$. This implies that $U_e(W,a,\theta) < 0$ and $U_s(W,a,\theta) < 0$. This implies that $e = 0$ and $s = 0$ and thus the optimal contract cannot implement a positive effort in either task. Thus, μ_e or μ_s must be such that $\mu_e p_{he}(\theta,a) + \mu_s p_{hs}(\theta,a)$ rises with h.

Let define the set Ψ as the set of outcomes for which the **LL** constraint is non-binding; i.e., $\Psi = \{h \in \{1, ..., n\} \mid \lambda_h = 0\}$. This implies that $w_h = L$ for all $h \in \Psi^c$, where Ψ^c is the complement of Ψ.

The first-order condition for all $h \in \Psi$ becomes

$$-p_h(\theta, a) + \mu_e p_{he}(\theta, a) + \mu_s p_{hs}(\theta, a) + \gamma p_h(\theta, a) = 0. \tag{B.14}$$

Let the wage be greater than L for two different outcomes, say y_h and $y_{h'}$. This implies that $\lambda_h = \lambda_{h'} = 0$ and therefore the following must hold,

$$\mu_e \ell_{he} + \mu_s \ell_{hs} + \gamma - 1 = \mu_e \ell_{h'e} + \mu_s \ell_{h's} + \gamma - 1. \tag{B.15}$$

Then because $\mu_e p_{he}(\theta, a) + \mu_s p_{hs}(\theta, a)$ rises with h, this results in a contradiction. Thus, the payment after only one outcome, say outcome h, is greater than L. In particular, for this outcome $\lambda_h = 0$ and thus the following holds

$$\gamma + \mu_e \ell_{he} + \mu_s \ell_{hs} = 1.$$

Because $\mu_e p_{he}(\theta, a) + \mu_s p_{hs}(\theta, a)$ rises with h, if this holds for any $h' < n$, then $\lambda_h < 0$ for all $h > h'$, which is a contradiction, since $\lambda_h \geq 0$ for all h. Thus, only the wage when the highest outcome is realized is greater than L.

It follows now from the fact that $w_h = L$ for all $h \leq n-1$ and $w_n > L$, $U_e(W, a, \theta) \geq 0$ and $U_s(W, a, \theta) \geq 0$ and that $U < L$ that the participation constraint is always satisfied. ■

Lemma 3 *If (W, a) solve the relaxed program, then the agent's expected utility at W is concave in the action profile a for any $a > 0$.*

Proof.

Recall that

$$\sum_{h=0}^{n} f_h g_h = f_0 \sum_{h=0}^{n} g_h + \sum_{j=0}^{n-1} (f_{j+1} - f_j) \sum_{h=j+1}^{n} g_h. \tag{B.16}$$

Let $P_h(\theta,a)$ be $\sum_{h'\geq h} p_{h'}(\theta,a)$ and $\triangle w_h$ be $w_h - w_{h-1}$ for $h \geq 1$ and $\triangle w_h = w_1^M$ for $h=1$. Then note that applying the summation by parts formula in equation (B.16)

$$U(W,a,\theta) \equiv \sum\nolimits_h \triangle w_h P_h(\theta,a) - c(a).$$

Hence,

$$U_{aa}(W,a,\theta) = \sum\nolimits_h \triangle w_h P_{haa}(\theta,a) \text{ for } a \in \{e,s\}.$$

and

$$U_{es}(W,a,\theta) = \sum\nolimits_h \triangle w_h P_{hes}(\theta,a).$$

Note that the first equation is negative by **CP** and $\triangle w_h \geq 0$ for all h. Thus, the agent's second order condition is satisfied if and only if for all $a \in A$

$$\Big(\sum\nolimits_h \triangle w_h P_{hee}(\theta,a)\Big)\Big(\sum\nolimits_h \triangle w_h P_{hss}(\theta,a)\Big) - \Big(\sum\nolimits_h \triangle w_h P_{hes}(\theta,a)\Big)^2 \geq 0.$$

Assumption **CP** together with the fact $\triangle w_h \geq 0$ guarantees that this term is positive. To see this note that

$$\begin{aligned}&\sum\nolimits_h \triangle w_h P_{hee}(\theta,a) \sum\nolimits_h \triangle w_h P_{hss}(\theta,a) - \Big(\sum\nolimits_h \triangle w_h P_{hes}(\theta,a)\Big)^2\\ \geq&\Big(\sum\nolimits_h \triangle w_h \sqrt{P_{hee}(\theta,a)}\sqrt{P_{hss}(\theta,a)}\Big)^2 - \Big(\sum\nolimits_h \triangle w_h P_{hes}(\theta,a)\Big)^2\end{aligned}$$

where the inequality follows from the Cauchy-Schwartz inequality. It follows then that the agent's utility function is concave in a for any $a \in A$ if

$$\sum\nolimits_h \triangle w_h \Big(\sqrt{P_{hee}(\theta,a)}\sqrt{P_{hss}(\theta,a)} - P_{hes}(\theta,a)\Big) \geq 0.$$

Observe that assumption **CP** ensures that the term in square brackets is positive for all h. This together with the fact that $\triangle w_h \geq 0$ for all h yields the desired result. Thus, $U(W,a,\theta)$ is concave in a for any $a \in A$ and this proves that condition (B.6) holds. ■

Part (i) follows from 3. If (W,a) solves the relaxed program, then the following holds

$$U_{aa}(W,a,\theta) \leq 0,\ \forall a \in \big(0,\bar{a}\big) \text{ for } a = e,s,$$

$$U_a(W,a,\theta) \geq 0,\ \forall a \in \big(0,\bar{a}\big) \text{ for } a = e,s$$

These conditions imply that the agent's action profile is global maximum and (W, a) is an element of the unrelaxed constraint set. Since (W, a) maximize the principal's utility over the relaxed constraint set, it maximizes the principal's utility over the unrelaxed constraint set since it was assume that the unrelaxed program has a positive solution (assumption 1).

■

The next result characterizes the optimal contract and the optimal effort.

Proposition B.2 (Pure Moral Hazard) *Suppose* $p_{naaa'}(\theta, a^M(\theta) \leq 0$, $\forall a, a' \in \{e, s\}$, $\forall \theta \in \Theta$. *Then*

i) The optimal contract is as follows: for all $h < n$,

$$w_h^M(\theta) = L, \ \forall \theta \in \Theta \tag{B.17}$$

and

$$w_n^M(\theta) = S_e(\theta, a^M(\theta))\ell_{ne}(\theta, a^M(\theta)) \frac{2p_{nse}(\theta, a^M(\theta)) - p_{nee}(\theta, a^M(\theta)) - p_{nss}(\theta, a^M(\theta))}{p_{nee}(\theta, a^M(\theta))p_{nss}(\theta, a^M(\theta)) - p_{nes}^2(\theta, a^M(\theta))} + L, \ \forall \theta \in \Theta.$$

ii) The optimal action profile for a type θ*-agent, denoted by* $a^M(\theta)$*, is determined by the unique solution to the following equations*

$$p_{ne}(\theta, a^M(\theta))(w_n^M(\theta) - L) - c = 0,$$

and

$$p_{ns}(\theta, a^M(\theta))(w_n^M(\theta) - L) - c = 0.$$

Proof of Proposition B.2. Let $\hat{w}_h$ be $w_h - L$. Then, using the first-order conditions for action a, the agent's incentive constraint, the FOC for e can be written as follows

$$e : \sum_h p_{he}(\theta, a)y_h - c + \mu_e p_{nee}(\theta, a)\hat{w}_n + \mu_s p_{nse}(\theta, a)\hat{w}_n = 0 \tag{B.18}$$

and that for s

$$s : \sum_h p_{hs}(\theta, a)y_h - c + \mu_e p_{nes}(\theta, a)\hat{w}_n + \mu_s p_{nss}(\theta, a)\hat{w}_n = 0. \tag{B.19}$$

Furthermore, the first-order condition for w_n is given by:

$$\mu_e p_{ne} + \mu_s p_{ns} = p_n \tag{B.20}$$

Solving for μ_e and μ_s from equations (B.18) and (B.19), one gets that

$$\mu_s = \frac{S_e(\theta,a)p_{nse}(\theta,a) - S_s(\theta,a)p_{nee}(\theta,a)}{\hat{w}_n(p_{nee}(\theta,a)p_{nss}(\theta,a) - p_{nes}^2(\theta,a))} \tag{B.21}$$

and

$$\mu_e = \frac{S_s(\theta,a)p_{nse}(\theta,a) - S_e(\theta,a)p_{nss}(\theta,a)}{\hat{w}_n(p_{nee}(\theta,a)p_{nss}(\theta,a) - p_{nes}^2(\theta,a))}. \tag{B.22}$$

Substituting this into equation (B.20) and solving for w_n^M one gets that the optimal wage is given by:

$$\hat{w}_n^M(\theta) = \frac{p_{ne}(\theta,a)(S_s(\theta,a)p_{nse}(\theta,a) - S_e(\theta,a)p_{nss}(\theta,a)) + p_{ns}(\theta,a)(S_e(\theta,a)p_{nse}(\theta,a) - S_s(\theta,a)p_{nee}(\theta,a))}{p_n(\theta,a)(p_{nee}(\theta,a)p_{nss}(\theta,a) - p_{nes}^2(\theta,a))} \tag{B.23}$$

Using the FOC for the agent's effort choice, one gets the optimal effort profile $a^M(\theta)$ is determined as the unique solution to the following equations

$$p_{ne}\big(\theta, a^M(\theta)\big)(w_n^M(\theta) - L) - c = 0, \ \forall\theta \in \Theta, \tag{B.24}$$

and

$$p_{ns}\big(\theta, a^M(\theta)\big)(w_n^M(\theta) - L) - c = 0, \ \forall\theta \in \Theta. \tag{B.25}$$

It follows from this that the equilibrium effort profile must be such that: $\ell_{ne}(\theta, a^M(\theta)) = \ell_{ns}(\theta, a^M(\theta))$. Substituting this into the optimal wage one gets that

$$\hat{w}_n^M(\theta) = \frac{S_s(\theta,a)(p_{nse}(\theta,a)p_{ne}(\theta,a) - p_{nee}(\theta,a)p_{ns}(\theta,a)) + S_e(\theta,a)(p_{nse}(\theta,a)p_{ns}(\theta,a) - p_{nss}(\theta,a)p_{ne}(\theta,a))}{p_n(\theta,a)(p_{nee}(\theta,a)p_{nss}(\theta,a) - p_{nes}^2(\theta,a))}. \tag{B.26}$$

Substituting this into the first-order conditions in equations (B.18) and (B.20) one gets that

$$\sum_h p_{he}(\theta,a)y_h - c = \frac{p_n\big(\theta,a\big)}{p_{ne}^2\big(\theta,a\big)}\left(\frac{p_{nee}(\theta,a)p_{nss}(\theta,a) - p_{nes}^2(\theta,a)}{2p_{nse}(\theta,a) - p_{nee}(\theta,a) - p_{nss}(\theta,a)}\right)c$$

and

$$\sum_h p_{hs}(\theta,a)y_h - c = \frac{p_n\big(\theta,a\big)}{p_{ns}^2\big(\theta,a\big)}\left(\frac{p_{nee}(\theta,a)p_{nss}(\theta,a)-p_{nes}^2(\theta,a)}{2p_{nse}(\theta,a)-p_{nee}(\theta,a)-p_{nss}(\theta,a)}\right)c$$

It readily follows from this and the fact that $p_{ne}^2\big(\theta,a(\theta)\big) = p_{ne}^2\big(\theta,a(\theta)\big)$ that $\sum_h p_{hs}y_h - c = \sum_h p_{he}y_h - c > 0$, $\mu_e > 0$, $\mu_s > 0$. This together with the fact that $p_{nse}(\theta,a) \geq 0$ implies that $a^M(\theta) \leq a^*(\theta)$ and $a^M(\theta) \neq a^*(\theta)$, $\forall\theta \in \Theta$.

Substituting the fact that $S_e(\theta,a^M(\theta)) = S_s(\theta,a^M(\theta))$ into the (B.26) one gets that

$$\hat{w}_n^M(\theta) = S_e(\theta,a^M(\theta))\ell_{ne}(\theta,a^M(\theta))\frac{2p_{nse}(\theta,a^M(\theta)) - p_{nee}(\theta,a^M(\theta)) - p_{nss}(\theta,a^M(\theta))}{p_{nee}(\theta,a^M(\theta)p_{nss}(\theta,a^M(\theta)) - p_{nes}^2(\theta,a^M(\theta)}. \tag{B.27}$$

Notice that in order for the following FOCs

$$p_{ne}(\theta,a^M(\theta))\hat{w}_n^M(\theta) - c = 0$$
$$p_{ns}(\theta,a^M(\theta))\hat{w}_n^M(\theta) - c = 0$$

to be concave and therefore to have a unique solution, since the objective function is concave requires that $p_{aa'a''}(\theta,a^M(\theta)) \leq 0$, $\forall(a,a',a'') \in \{e,s\}^3$. It is easy to check that this ensures that

$$p_{nssa}p_{nee} + p_{neea}p_{nss} - 2p_{nes}p_{nesa} \geq 0.$$

Furthermore it is easy to check that This leads to

$$\frac{\partial e(\theta)}{\partial\theta} = \frac{-p_{nss}(\theta,a)p_{ne\theta}(\theta,a) + p_{nes}(\theta,a)p_{ns\theta}(\theta,a)}{p_{nss}(\theta,a)p_{nee}(\theta,a) - p_{nes}(\theta,a)^2}\frac{c(a(\underline{\theta}))}{p_n(\underline{\theta},a(\underline{\theta}))} > 0$$

and

$$\frac{\partial s(\theta)}{\partial\theta} = \frac{-p_{nee}(\theta,a)p_{ns\theta}(\theta,a) + p_{nes}(\theta,a)p_{ne\theta}(\theta,a)}{p_{nss}(\theta,a)p_{nee}(\theta,a) - p_{nes}(\theta,a)^2}\frac{c(a(\underline{\theta}))}{p_n(\underline{\theta},a(\underline{\theta}))} > 0$$

■

The optimal contract is a bonus contract of the pass/fail type. It can be described by a non-contingent transfer $w^M(\theta)$ and a bonus equal to $w_n^M(\theta) - w^M(\theta)$ paid to the agent when

$y_h = y_n$ is observed. The value of the non-contingent transfer $w^M(\theta)$ is fully determined by the limited liability constraint and therefore $w^M(\theta) = L$.[15] The cost to the principal is that the agent earns a limited-liability rent equal to $L + p_n(\theta, a(\theta))(w_n^M(\theta) - L) - c(a^M(\theta)) - U$

First, notice that **MLRP** implies that the partition of the set of signals on which payments are conditioned does not depend on the action implemented (nor does it depend on the agent's liability limit or on his cost function). This says that from the principal's viewpoint most of the relevant information provided by the performance measure is irrelevant in the sense that the optimal contract would be feasible even if the principal were able to observe only a binary performance measure.[16]

Second, observe that there is a multitasking problem which is of a different nature from that in Holmström and Milgrom (1991). Mainly, the multitasking problem here is one of implementation. That is, the set of effort pairs that is implementable is restricted to effort pairs that yield the same marginal utility of effort. In fact, it readily follows from the agent's incentive compatibility constraints that the only action profiles that are implementable under a contract $W(\theta)$ are those satisfying the following

$$\sum\nolimits_h \big(\ell_{he}(\theta, a(\theta)) - \ell_{hs}(\theta, a(\theta))\big) w_h(\theta) p_h(\theta, a(\theta)) = 0.$$

The reason for this result is twofold: first, the performance measure is aggregated and confounds the efforts of two non-conflicting tasks; and second, the performance measure and the agents' costs are such that for any incentive intensity, the agent has no preferences over either task. Thus, if a contract induces the agent to choose a positive action in one task, he will choose a positive action in the other task. Furthermore, the stronger the incentives in one task, the higher are those in the other task. This means that the principal cannot implement very different actions in each task. In fact, he can never implement a positive

[15] 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.

[16] Demougin and Fluet (1998) shows that having multiple signals has no added value in the single-task case. However, given the results here and those in Holmström and Milgrom (1991), I conjecture that result does not hold for the multi-task case since the principal could use one signal to provide more incentives in one task and the other to do the same in the other task. Thus, having more signals would in general increase the set of implementable action profiles.

action in one task and zero action in the other task. Hence, the incentive compatibility constraint for each task implies that if it is optimal for the principal to induce the agent to exert effort in one task, then it is optimal to induce him to exert effort in both tasks.

It is easy to see that the optimal action profile $a^M(\theta)$ satisfies the following

$$S_e(\theta, a^M(\theta)) = S_s(\theta, a^M(\theta)).$$

The reason for this result is twofold: first, the performance measure is aggregated and confounds the efforts of two non-conflicting tasks; and second, the performance measure and the agent's cost are such that for any incentive intensity, the agent has no preferences over either task. Because the least costly contract compensates the agent only after the highest outcome is realized and actions are complements, the incentive compatibility constraint for each task implies that if it is optimal for the principal to induce the agent to exert higher effort in one task, then it is optimal to induce him to exert higher effort in both tasks. To better understand this observe that the ratio $p_{he}(\theta, a)/p_{hs}(\theta, a)$ can be interpreted as the rate of information substitution. Suppose this is large for larger outputs. Then, a large output may turn out to be so informative about the agent's action e that the principal, upon observing this output, can confidently attribute it to e rather than s. Given this, the principal may not want to pay a bonus only when the highest outcome is realized (due to **MLRP**) since this may induce too much action s. When the ratio is independent of output, any given output results in the same relative informativeness with regard to e and s and thus the principal's inference process is the same regardless of the output. This would lead to a less distorted effort profile in the sense that the ratio $e^M(\theta)/s^M(\theta) = e^*(\theta)/s^*(\theta)$. This however makes the inference problem more difficult since output does not allow the principal to discriminate well ex-post between e and s. When $p_{he}(\theta, a)/p_{hs}(\theta, a)$ varies with output, the inference problem is easier, but in order to reduce the agent's limited liability and use this extra information, the principal distorts more the effort allocation in the sense that in general $e^M(\theta)/s^M(\theta) \neq e^*(\theta)/s^*(\theta)$.

Third, notice that the pure moral hazard action profile is downward distorted for all $\theta \in \Theta$.

Corollary B.3 *Suppose the agent's type is common knowledge. Then,* $a^M(\theta) \leq a^*(\theta)$, $\forall \theta \in$

Θ.

It is interesting to note that the distortion of the action profile with respect to the first-best is not due to incongruence of the information system with the contribution of the agent's action to the expected firm value $\sum_h p_h(\theta, a(\theta))y_h$. It is the result of the following things: (i) the use of an aggregated performance measure (i.e,. the one-signal-per-task assumption made by Holmstrom and Milgrom (1987) is not satisfied); (ii) the principal's incentive to reduce the agent's limited-liability rent by mean of making an efficient use of the information system; (iii) the effort complementarity across tasks; and (iv) **MLRP** holds for both actions and therefore the wage profile implementing any positive action profile must be increasing.

Finally, it is instructive to study how the limited-liability rent changes with the agent's ability type. It is easy to check that incentive compatibility with respect to actions implies that the rent varies with θ as follows

$$p_{n\theta}(\theta, a(\theta))(w_n^M(\theta) - L) + p_n(\theta, a(\theta))\frac{\partial w_n^M(\theta)}{\partial \theta}.$$

The first term comprises the increases in the limited-liability rent due to an increase in the agent's type when the contract is held constant. This effect is positive since the return distribution satisfies **MLRP** with respect to the agent's ability parameter θ. The second term consist of the effect of θ on the optimal contract. This cannot be signed without further assumptions. The reason stands for the following. On the one hand, an increase in θ rises the marginal productivity of actions and therefore, holding the wage constant, actions increase with the ability type. This induces the principal to offer a higher bonus to agents with a higher ability type since the marginal benefit of bonus from the principal's viewpoint increases (recall that actions are downward distorted). On the other hand, a higher θ, holding the bonus and actions constant, increases the limited-liability rent and this provides the principal with incentives to lower the wage. Thus, when the first effect dominates it is optimal to increase the wage as θ rises.[17]

This result is similar to the one derived by Demougin and Fluet (1998) in the one action moral hazard model with a risk-neutral principal and agent and limited liability and to

[17]Which effects dominates is difficult to determine since it depends on the sign of the following derivatives $p_{aa'\theta}(\theta, a)$, $\forall a, a' \in \{e, s\}$).

the one obtained by Laux (2001) in the multiple action moral hazard model with limited liability.[18]

[18] Laux (2001) assumes that the agent's reservation utility is non-binding, there is an observable outcome for each possible task and the effort level in each task is binary.

References

Baker, G. (2002). Distortion and risk in optimal incentive contracts. *Journal of Human Resources*, 37(4):728–751.

Baker, G., Gibbs, M., and Holmstrom, B. (1994). The wage policy of a firm. *The Quarterly Journal of Economics*, 109(4):921–55.

Baker, G. P. (1992). Incentive contracts and performance measurement. *Journal of Political Economy*, 100(3):598–614.

Balmaceda, F. (2009). Uncertainty, pay for performance, and asymmetric information. *Journal of Law, Economics and Organization*, 58(1):211–236.

Balmaceda, F. (2010). Job design: Complexity, incentives and discretion. Working Paper Series CEA-134, CEA-University of Chile.

Baron, D. P. and Besanko, D. (1984). Regulation, asymmetric information, and auditing. *RAND Journal of Economics*, 15(4):447–470.

Baron, D. P. and Besanko, D. (1987). Monitoring, moral hazard, asymmetric information, and risk sharing in procurement contracting. *RAND Journal of Economics*, 18(4):509–532.

Caillaud, B., Guesnerie, R., and Rey, P. (1992). Noisy observation in adverse selection models. *Review of Economic Studies*, 59(3):595–615.

Chiappori, P. A. and Salanié, B. (2002). Testing contract theory: A survey of some recent work. The MIT Press.

Demougin, D. (1989). A renegotiation-proof mechanism for a principal-agent model with moral hazard and adverse selection. *RAND Journal of Economics*, 20(2):256–267.

Demougin, D. and Fluet, C. (1998). Mechanism sufficient statistic in the risk-neutral agency problem. *Journal of Institutional and Theoretical Economics (JITE)*, 154(4):622–638.

Dewatripont, M., Jewitt, I., and Tirole, J. (1999). The economics of career concerns, part ii: Application to missions and accountability of government agencies. *The Review of Economic Studies*, 66(1):199–217.

Dewatripont, M. and Tirole, J. (1999). Advocates. *Journal of Political Economy*, 107(1):1–39.

Faynzilberg, P. S. and Praveen, K. (1997). Optimal contracting of separable production technologies. *Games and Economic Behavior*, 21(1-2):15 – 39.

Faynzilberg, P. S. and Praveen, K. (2000). On the generalized principal-agent problem: Decomposition and existence results. *Review of Economic Design*, 5:23–58. 10.1007/s100580050046.

García, D. (2005). Monotonicity in direct revelation mechanisms. *Economics Letters*, 88(1):21 – 26.

Gibbons, R. (1998). Incentives in organizations. NBER Working Papers 6695, National Bureau of Economic Research, Inc.

Guesnerie, R., Picard, P., and Rey, P. (1989). Adverse selection and moral hazard with risk neutral agents. *European Economic Review*, 33(4):807–823.

Holmstrom, B. and Milgrom, P. (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica*, 55:303–328.

Holmström, B. and Milgrom, P. (1991). Multi-task principal-agent analyses: Incentive contracts, asset ownership and job desing. *Journal of Law, Economics and Organization*, 7:24–52.

Itoh, H. (1991). Incentive to help in mutli-agent situations. *Econometrica*, 59:611–636.

Itoh, H. (1992). Cooperation in hierarchical organizations: An incentive perspective. *Journal of Law, Economics and Organization*, 8:321–345.

Itoh, H. (1994). Job design, delegation and cooperation: A principal-agent analysis. *European Economic Review*, 38:691–700.

Laffont, J.-J. and Tirole, J. (1986). Using cost observation to regulate firms. *Journal of Political Economy*, 94(3):614–41.

Lafontaine, F. and Slade, M. (2001). Incentive contracting and the franchise decision. *K. Chatterjee and W. Samuelson, eds., Advances in Business Applications of Game Theory, Kluwer Academic Press.*

Lafontaine, F. and Slade, M. E. (1997). Retail contracting: Theory and practice. *The Journal of Industrial Economics*, 45(1):1–25.

Laux, C. (2001). Limited liability and incentive contracting with multiple projects. *RAND Journal of Economics*, 32:514–526.

Lazear, E. (2000). Performance pay and productivity. *American Economic Review*, 90:1346–1361.

Lewis, T. R. and Sappington, D. E. M. (1989). Countervailing incentives in agency problems. *Journal of Economic Theory*, 49(2):294–313.

Lewis, T. R. and Sappington, D. E. M. (2000). Motivating wealth-constrained actors. *American Economic Review*, 90(4):944–960.

Lewis, T. R. and Sappington, D. E. M. (2001). Optimal contracting with private knowledge of wealth and ability. *Review of Economic Studies*, 68(1):21–44.

MacDonald, G. and Marx, L. (2001). Adverse specialization. *Journal of Political Economy*, 109(4):864–899.

Maskin, E. and Riley, J. (1984). Monopoly with incomplete information. *RAND Journal of Economics*, 15(2):171–196.

McAfee, R. P. and McMillan, J. (1986). Bidding for contracts: A principal-agent analysis. *RAND Journal of Economics*, 17(3):326–338.

McAfee, R. P. and McMillan, J. (1987). Competition for agency contracts. *RAND Journal of Economics*, 18(2):296–307.

Melumad, N. D. and Reichelstein, S. (1989). Value of communication in agencies. *Journal of Economic Theory*, 47(2):334–368.

Mirrlees, J. (1975). The theory of moral hazard and unobservable behavior. Working paper, Nuffield College, Oxford.

Mussa, M. and Rosen, S. (1978). Monopoly and product quality. *Journal of Economic Theory*, 18(2):301–317.

Paarsch, H. J. and Shearer, B. (2000). Piece rates, fixed wages, and incentive effects: Statistical evidence from payroll records. *International Economic Review*, 41(1):59–92.

Poblete, J. J. and Spulber, D. F. (2012). The form of incentive contracts: agency with moral hazard, risk neutrality and limited liability. rand journal of economics. *Rand Journal of Economics*, forthcoming.

Prendergast, C. (2002a). The tenuous trade-off between risk and incentives. *Journal of Political Economy*, 110:1071–1102.

Prendergast, C. (2002b). Uncertainty and incentives. *Journal of Labor Economics*, 20:115 – 137.

Raith, M. (2008). Specific knowledge and performance measurement. *RAND Journal of Economics*, 39(4):1059–1079.

Ratto, M. and Schnedler, W. (2008). Too few cooks spoil the broth: Division of labor and directed production. *The B.E. Journal of Economic Analysis & Policy*, 8(1):27.

Salanié, B. (2005). *The Economics of Contracts: A Primer, 2nd Edition*, volume 1 of *MIT Press Books*. The MIT Press.

Seiler, E. (1984). Piece rate vs. time rate: The effect of incentives on earnings. *The Review of Economics and Statistics*, 66(3):363–76.

Sugato, B. and Lafontaine, F. (1995). Double-sided moral hazard and the nature of share contracts. *RAND Journal of Economics*, 26(4):761–781.

Wang, W., Hu, S., and Hu, X. (2011). Moral hazard with limited liability: Incentive contracting under monotone ratio assumptions.

Zhao, R. R. (2008). All-or-nothing monitoring. *American Economic Review*, 98(4):1619–28.