Comparative Statics in Principal-Agent Problems

Richard T. Holden*

September 15, 2008

Abstract

Principal-agent problems are widespread in economics. Since it is usually believed little can be said in general settings (Mirrlees (1975), Grossman and Hart (1983)) applied work typically uses the first-order approach, justified by strong assumptions on the distribution of shocks, or restricts attention to linear contracts, justified by the assumptions under which Holmström and Milgrom (1987) have shown that this is the optimal contract. In this paper we dispense with these assumptions and develop a general approach to principal-agent problems. We do so by applying results from the theory of monotone comparative statics to the Grossman-Hart approach. We offer general results on the intensity of incentives, the shape of the optimal contract, and the value and use of information. The paper thus: (i) provides a technique for analyzing moral hazard problems and (ii) highlights which assumptions of the linear contracts or first-order approach are *critical* for drawing various conclusions. We show that the linear contracts model of Holmström and Milgrom (1987) is surprisingly robust and that all but one of its comparative statics hold under general assumptions in a static setting. Classic results on the value of information due to Holmström (1979) are also generalized. (JEL L13, L22).

*Massachusetts Institute of Technology, Sloan School of Management, 50 Memorial Drive, E52-410, Cambridge MA 02142. email: rholden@mit.edu. I am grateful to Daron Acemoglu, Robby Akerlof, Susan Athey, Drew Fudenberg, Bob Gibbons, Jerry Green, Emir Kamenica, Ian Jewitt, Jon Levin, Gregor Matvos, Paul Milgrom, Larry Samuelson and Chris Shannon for helpful discussions and comments. I owe special thanks to Philippe Aghion, Oliver Hart and Bengt Holmström.

"A very widespread economic situation is that of the relation between principal and agent. Even in ordinary and legal discourse, the principal-agent relation would be significant in scope and economic magnitude. But economic theory in recent years has recognized that analogous interactions are almost universal in the economy, at least as one significant component of almost all transactions." Kenneth Arrow (1986).

1 Introduction

Principal-agent problems are pervasive. Relationships between: shareholder and manager, manager and worker, doctor and patient, lawyer and client, even government and citizens all involve moral hazard. It is therefore not surprising that substantial attention has been paid to the formal methods for analyzing such problems. Beginning with Mirrlees (1975), it has been believed that general results are unavailable. Applied work typically uses the "first-order approach" (replacing the agent's incentive compatibility constraint with her first-order condition), or restricts attention to linear contracts.

The validity of the first-order approach depends on the agent's first-order condition and her incentive compatibility constraint being equivalent. This is true only under strong assumptions[1]. But this is something of a red-herring. Even when the first-order approach is valid it implies that the shape of the optimal incentive scheme depends crucially on the likelihood ratio. Since similar incentive contracts are observed in very different settings this is troubling. In a sense, the first-order approach is "the wrong model"[2].

[1] Sufficient conditions for first-order approach to be valid are twin assumptions of the Monotone Likelihood Ratio ("MLRP") and Convexity of the Distribution Function ("CDFC"). The Linear Distribution Function Condition (Hart and Holmström (1987)–previously called the Spanning Condition by Grossman and Hart (1983)) or certain restrictions on the utility function of the agent proposed by Jewitt (1988), also validate the first-order approach.

[2] This phrase is due to Bengt Holmstrom.

Realizing this, Holmström and Milgrom (1987) provide conditions under which *linear* contracts are optimal. This has been a remarkably useful model: the one which is typically used in applied research. There is a tension, however. The conditions which make linear contracts optimal are also strong. They include: the agent having an exponential utility function which is additively separable in action and reward, shocks being normally distributed, the agent controlling the mean (i.e. drift rate) and not the variance of a Brownian Motion, and there being no consumption by the agent until the end of the relationship. Moreover, researchers typically use the result obtained in a dynamic setting (indeed in continuous time) to justify a restriction to linear contracts in *static* models. At best this is internally inconsistent.

This paper provides a number of general results which do not involve the assumptions of either the first-order approach or the linear contracts approach. It does so by applying results from the theory of monotone comparative statics (Milgrom and Shannon (1994)) to the Grossman-Hart model of the principal-agent problem (Grossman and Hart (1983)). Often researchers are interested largely in comparative statics. For instance: input prices, tax rates and the nature of product market competition affect firm profits, the noisiness of the relationship between the agent's action and output affects the cost of insuring the (risk-averse) agent. We allow parameters of interest to affect the profits which accrue to the principal in a given state, and/or the probability distribution of which state occurs. Our approach is valid in situations where the either the first-order approach or restriction to linear contracts are valid, as well as those where one or both are not. Thus the technique not only provides a method for analyzing more general problems, but also highlights which assumptions of the more restrictive approaches are critical for the conclusions drawn.

The Grossman-Hart method explicitly decomposes the problem into the cost to the principal of implementing a certain action by the agent, and the (expected) benefits which accrue to her from this action. The approach consists of two steps. In step one, the principal determines the lowest cost way to implement a given action. Then–in step two–she chooses the action which maximizes the difference between the expected benefits and costs of implementation. They show how the first step can be solved, using a transformation which ensures that the principal solves a convex programming problem. They do not analyze the second step, however, because they show that it will generally not be a convex programming problem. Here we tackle the second step using monotone methods. In particular, we are not concerned about the potential non-convexity of the second step problem as we can do comparative statics on the *set of optima*. Utilizing the underlying monotonicity structure of the problem this technique requires only mild assumptions.

The remainder of the paper is organized as follows. In Section 2, we outline the our approach and provide necessary and sufficient conditions for a harder action to be implemented. Section 3 compares the comparative statics result in the linear model of Holmström and Milgrom (1987), with the general results obtained in the previous sections. Section 4 then generalizes Holmström's Sufficient Statistic Theorem. Section 5 contains some concluding remarks. Various extensions of the basic model, which are anticipated in earlier sections, are contained in Appendix A. Ommitted proofs are contained in Appendix B.

2 The Model

2.1 Statement of the Problem

There are two players, a risk-neutral principal and a risk-averse agent[3]. The principal hires the agent to perform an action. She does not observe the action the agent chooses, but does observe profits which are a noisy signal of the action.

Let ϕ be a parameter of interest which affects the profits which accrue to the principal. In Appendix A we generalize these results to the case where the parameter affects the probability of the states occurring and/or the profits in those states. Following Grossman and Hart (1983), suppose that there are a finite number of possible gross profit levels for

[3]The case where the principal is risk-averse is contained in Appendix A.

the firm. Denote these $q_1(\phi) < ... < q_n(\phi)$. These are profits before any payments to the agent. We assume that the principal is concerned only with net profit–i.e. gross profit less payments to the agent.

We say that a set X is a **product set** if there exist sets $X_1, ..., X_n$ such that $X = X_1 \times ... \times X_n$. X is a **product set in** $\mathbb{R}^n$ if $X_i \subseteq \mathbb{R}, i = 1, ..., n$.

The set of actions available to the agent is A, which is assumed to be a product set in $\mathbb{R}^n$ which is closed, bounded and non-empty[4]. Assume that there is a twice continuously differentiable function $\pi : A \to S$, where S is the standard probability simplex, i.e. $S = \{y \in \mathbb{R}^n | y \geq 0, \sum_{i=1}^n y_i = 1\}$. The probabilities of outcomes $q_1(\phi), ..., q_n(\phi)$ are therefore $\pi_1(a), ..., \pi_n(a)$.

Assumption A1. *The agent's von Neumann-Morgenstern utility function is of the form:*

$$U(a, I) = G(a) + K(a)V(I),$$

where I is a payment from the principal to the agent, and $a \in A$ is the action taken by the agent. V is a continuous, strictly increasing, real-valued, concave function on an open ray of the real line $\mathcal{I} = (\underline{I}, \infty)$. Let $\lim_{I \to \underline{I}} V(I) = -\infty$ and assume that G and K are continuous, real-valued functions and that K is strictly positive. Finally assume that for all $a_1, a_2 \in A$ and $I, \widehat{I} \in \mathcal{I}$ the following holds

$$G(a_1) + K(a_1)V(I) \geq G(a_2) + K(a_2)V(I)$$
$$\Rightarrow G(a_1) + K(a_1)V(\widehat{I}) \geq G(a_2) + K(a_2)V(\widehat{I})$$

As Grossman and Hart (1983) note, this assumption implies that each agent's preferences over income lotteries are independent of action[5], and that the agent's ranking over

[4]Thus, by the Heine-Borel Theorem it is compact. This is important for existence of an optimal second-best action.

[5]A result of Keeney (1973) implies the converse – that if the agent's preferences over income lotteries are independent of action then they can be represented by a utility function of the form $G(a) + K(a)V(I)$.

(only) perfectly certain actions is independent of income. This assumption is clearly not innocuous. It is worth noting, however, that if the agent's utility function is additively or multiplicatively separable in action and reward then the assumption is satisfied. Both the first-order approach and the linear contracts approach assume additive separability. In fact, in either of the separable cases, preferences over income lotteries are independent of actions *and* preferences over action lotteries are independent of income. This is stronger than A1. In Appendix A we consider relaxing A1.

Let the agent's reservation utility be $\overline{U}$ (i.e. the expected utility attainable if she rejects the contract offered by the principal), and let

$$\mathcal{U} = V(\mathcal{I}) = \{v | v = V(I) \text{ for some } I \in \mathcal{I}\}.$$

Grossman and Hart (1983) point out that to ensure that an optimal second best action and incentive scheme exists it is important that the reservation utility can be achieved with some I for any choice of action by the agent.

Assumption A2. $\left(\overline{U} - G(a)\right)/K(a) \in \mathcal{U}, \forall a \in A$.

A third assumption ensures that $\pi_i(a)$ is bounded away from zero and hence rules out the scheme proposed by Mirrlees (1975) through which the principal can approximate the first-best by imposing ever larger penalties for actions which occur with ever smaller probabilities if the desired action is not taken[6].

Assumption A3. $\pi_i(a) > 0$, $\forall a \in A$ *and* $i = 1, ..., n$.

The principal is assumed throughout to know the agent's utility function $U(a, I)$, the action set A, and the function π. The principal does not, of course, observe a. An **incentive scheme** is an n-dimensional vector $\mathbf{I} = (I_1, ..., I_n) \in \mathcal{I}^n$.

[6] This ensures that $\pi_i(a)$ is bounded away from zero because of the compactness of A, in conjunction with the fact that there are a finite number of states.

Given an incentive scheme the agent chooses $a \in A$ to maximize her expected utility $\sum_{i=1}^{n} \pi_i(a) U(a, I_i)$.

2.2 First-Best

In the first-best the principal observes the action chosen by the agent. $C_{FB} : A \rightarrow \mathbb{R}$ is the **first-best cost** of implementing action a given by:

$$C_{FB}(a) = h\left(\frac{\bar{U} - G(a)}{K(a)}\right),$$

where $h = V^{-1}$.

The contract involved in achieving the first-best is the following. The principal pays the agent $C_{FB}(a)$ if she chooses a and some $\tilde{I}$ otherwise, where $\tilde{I}$ is "close" to $\underline{I}$.

The **first-best action** is that which solves:

$$\max_{a \in A} \left\{ \sum_{i=1}^{n} \pi_i(a) q_i(\phi) - C_{FB}(a) \right\}.$$

Note that C_{FB} induces a complete ordering on A, which is independent of $\bar{U}$. We write $a \succeq a' \Leftrightarrow C_{FB}(a) \geq C_{FB}(a')$. When $a \succ a'$ we say that action a is **harder** than action a'.

2.3 Second-Best Step One: Lowest Cost Implementation

In the second-best case, the problem which the principal faces is to choose an action and a payment schedule to maximize expected output net of payments, subject to that action being optimal for the agent and subject to the agent receiving her reservation utility in expectation. i.e.

$$\max_{a,(I_1,\ldots,I_n)} \left\{ \sum_{i=1}^{n} \left(\pi_i(a)\left(q_i - I_i\right)\right) \right\} \tag{1}$$

subject to

$$a^* \in \arg\max_a \left\{ \sum_{i=1}^{n} \pi_i(a) U(a, I_i) \right\}$$

$$\sum_{i=1}^{n} \pi_i(a^*) U(a^*, I_i) \geq \overline{U}$$

Following Grossman and Hart (1983) we proceed in two stages. First we assume that the principal has an action $a^* \in A$ which she wishes to implement and consider what is the lowest cost way to implement this. We then consider, given this, what is the optimal a^* which she wishes to implement. Let us define $\upsilon_1 = V(I_1), \ldots, \upsilon_n = V(I_n)$ and $h \equiv V^{-1}$. These will be used as the control variables. The problem can now be stated as:

$$\min_{\upsilon_1,\ldots,\upsilon_n;\upsilon_i \in \mathcal{U},\forall i} \left\{ \sum_{i=1}^{n} \pi_i(a^*) h(\upsilon_i) \right\} \tag{2}$$

subject to

$$G(a^*) + K(a^*) \left(\sum_{i=1}^{n} \pi_i(a^*) \upsilon_i \right) \geq G(a) + K(a) \left(\sum_{i=1}^{n} \pi_i(a) \upsilon_i \right), \forall a \in A$$

$$G(a^*) + K(a^*) \left(\sum_{i=1}^{n} \pi_i(a^*) \upsilon_i \right) \geq \overline{U}$$

The constraints in (2) are linear in the υ_js and, since V is concave, h is convex. Consequently the problem in (2) is simply to minimize a convex function subject to a set of linear constraints.

A vector $(\upsilon_1, \ldots, \upsilon_n)$ which satisfies the constraints in (2) is said to **implement** action a^*.

We are now done with step one–finding the lowest cost way to implement a given action.

Let

$$C(a^*) = \inf \left\{ \sum_{i=1}^{n} \pi_i(a^*) h(v_i) | \boldsymbol{v} = (v_1, ..., v_n) \text{ implements } a^* \right\}$$

which implements a^* if the constraint set in (2) is non-empty. If the constraint set is empty then let $C(a^*) = \infty$.

2.4 Second-Best Step Two: Monotone Comparative Statics on the Optimal Action

The second-step of the Grossman-Hart approach is to choose which action should be implemented. That is, choose the action which maximizes the expected benefits minus the costs of implementation:

$$\max_{a \in A} \{B(a, \phi) - C(a)\}, \tag{3}$$

where $B(a, \phi) = \sum_{i=1}^{n} \pi_i(a) q_i(\phi)$. Grossman and Hart (1983) point out that this may well be a non-convex problem, for $C(a)$ will not generally be a convex function.

The theory of monotone comparative statics allows one to perform comparative statics on the *set of optima* in this second-step problem. Denote $a^{**}(\phi; C) = \arg\max_{a \in A} \{B(a, \phi) - C(a)\}$ as the solution to the problem. In order to make precise what it means for a set of optima to increase, we, following Athey, Milgrom, and Roberts (1998), state the following definitions of the Strong Set Order ("SSO") which makes precise what it means for a set to be higher than another, what it means for a set-valued function to be nondecreasing, and what it means for a function to exhibit increasing differences. We begin with the case where the action set, $A \in \mathbb{R}$. We show that this entails no loss of generality, and applies to $A \subseteq \mathbb{R}^n$, in Appendix A.

Let $A, B \subset \mathbb{R}^n$. Then A is **higher** than B in the **Strong Set Order** ($A \geq_s B$) iff for any $\mathbf{a} \in A$ and $\mathbf{b} \in B$, $\max\{\mathbf{a}, \mathbf{b}\} \in A$ and $\min\{\mathbf{a}, \mathbf{b}\} \in B$.[7] A set-valued function $H : \mathbb{R} \to 2^{\mathbb{R}}$

[7] In $\mathbb{R}^1$ the following is analogous. A set $S \subseteq \mathbb{R}$ is said to be as "High" as another set $T \subseteq \mathbb{R}$ ($S \geq_S T$), if and only if (i) each $x \in S \backslash T$ is greater than each $y \in T$, and (ii) each $x' \in T \backslash S$ is less than each $y' \in S$.

is said to be **nondecreasing** if for $x \geq y, H(x) \geq_S H(y)$. A function $f : X \to \mathbb{R}$ has **increasing differences** in $(x_n; x_m), n \neq m$ iff $\forall x'_n \in X_n$ and $x''_n \in X_n$ with $x'_n > x''_n$, and $\forall x_j, j \neq n, m$ we have

$$f(x_1, ..., x'_n, ..., x_N) - f(x_1, ..., x''_n, ..., x_N) \text{ is nondecreasing in } x_m$$

With these definitions in hand, we can now state the following result. It shows that a necessary and sufficient condition for the optimal action correspondence to be nondecreasing in the parameter ϕ is that the expected benefit function exhibit increasing differences.

Proposition 1. *$a^{**}(\phi; C)$ is nondecreasing in ϕ for all functions C iff B has increasing differences.*

The proof follows directly from the Milgrom-Shannon Monotonicity Thorem. This result deals with the possibility that *all* of the local optima are nondecreasing in ϕ but that the global optimum is actually decreasing in ϕ for some values[8].

We now make two assumptions for ease of exposition. The first requires the action set to be a subset of the real line, which makes the notion of "higher" and "lower" actions more straightforward. The second is a differentiability condition. Neither of these assumptions (A4 and A5) are required, even for strict comparative static conclusions (Edlin and Shannon (1998)). In Appendix A we show that they can easily be relaxed.

Assumption A4. $A \subseteq \mathbb{R}$

Assumption A5. *B is twice continuously differentiable in both its arguments.*

Lemma 1. *Assume A4-A5. Then B has increasing differences iff:*

$$\sum_{i=1}^{n} q'_i(\phi)\pi'_i(a) \geq 0, \forall a, \phi.$$

[8] See, for instance, AMR figure 2.1 and the accompanying discussion.

Proof. A function $f(x,\theta)$ which is twice continuously differentiable has increasing differences if and only if for all x, θ, $\frac{\partial^2}{\partial x \partial \theta} f(x,\theta) \geq 0$. Now note that $\frac{\partial^2}{\partial a \partial \phi} B$ is $\sum_{i=1}^{n} q_i'(\phi)\pi_i'(a)$. ■

We will sometimes be interested in a strict comparative static–$a^{**}(\phi; C)$ *strictly* increasing in ϕ (as opposed to merely nondecreasing). It is well known that strict comparative statics require slightly stronger assumptions (Edlin and Shannon (1998)). A function can have strictly increasing differences[9] but have the maximum not increase in the relevant parameter. For a^{**} to be strictly increasing the following is required.

Proposition 2. *Assume A4-A5, that $C(a)$ is continuously differentiable, and that $a^{**}(\phi; C) \in int\,(A)$ for all ϕ. Then $a^{**}(\phi; C)$ is strictly nondecreasing in ϕ for all functions C iff $\frac{\partial^2}{\partial a \partial \phi} B > 0$.*

Proof. See AMR Theorem 2.6 ■

Remark 1. *Edlin and Shannon (1998) show that A4 is not required for this result, and that A5 can be weakened to only require B to be continuously differentiable in a.*

By construction, an increase in effort affects the probabilities of different states occurring. We make a final assumption which implies that high effort increases the probabilities of high profit states occurring, and *vice versa.*

Assumption A6. *$\pi : A \to S$ satisfies First Order Stochastic Dominance ("FOSD") if $a_1 > a_2 \in A \Rightarrow \sum_{i=1}^{j} \pi_i(a_1) < \sum_{i=1}^{j} \pi_i(a_2), \forall j < n$.*

Before proceeding to analyze this for the case of n possible outcomes, consider the special case where there are only two outcomes. This highlights the key features of the analysis, without some of the complications which arise in the n outcome case.

[9] A function $f : \mathbb{R}^2 \to \mathbb{R}$ has "Strictly Increasing Differences" if for all $x'' > x'$, $f(x'', \theta) - f(x', \theta)$ is strictly increasing in θ.

2.4.1 Two Outcomes

Denote the two possible outcomes as H and L. Lemma 1 implies that for a^{**} to be nondecreasing in ϕ requires:

$$q_L'(\phi)\pi_L'(a) + q_H'(\phi)\pi_H'(a) \geq 0 \tag{4}$$

By definition $\pi_L(a) + \pi_H(a) = 1$. Differentiating this identity yields $\pi_L'(a) + \pi_H'(a) = 0$. Therefore $\pi_H'(a) = -\pi_L'(a)$ and one can write (4) as:

$$\pi_L'(a)\left[q_L'(\phi) - q_H'(\phi)\right] \geq 0$$

Since a harder action makes the low profit state less likely by FOSD, it must be that $\pi_L'(a) \leq 0$. Therefore we require $q_L'(\phi) - q_H'(\phi) \leq 0$, which amounts to $q_H'(\phi) \geq q_L'(\phi)$.

This condition has a particularly simple interpretation. If an increase in ϕ makes the high state relatively more profitable and the low state (i.e. $q_H'(\phi) > q_L'(\phi)$), the principal chooses to implement a harder action. We will show later that the reason for this is straightforward. If a higher value of ϕ makes the high profit state relatively more attractive to the principal, then she induces the agent to put more probability weight on that state by altering the incentive scheme.

2.4.2 n Outcomes

In the case where there are n possible outcomes there is a similar condition for a^{**} to be nondecreasing in ϕ. Direct application of Lemma 1 establishes the following result.

Proposition 3. *Assume A1-A5. Then a^{**} is nondecreasing in ϕ iff:*

$$\sum_{i=1}^{n} \pi_i'(a) q_i'(\phi) \geq 0.$$

Assume also that there exist interior maximizers for all values of ϕ. Then a^{**} is strictly *increasing in ϕ iff:*

$$\sum_{i=1}^{n} \pi_i'(a) q_i'(\phi) > 0.$$

Remark 2. *The same condition holds in the first-best case since the first-best problem is:*

$$\max_{a \in A} \left\{ \sum_{i=1}^{n} \pi_i(a) q_i(\phi) - C_{FB}(a) \right\},$$

and the condition holds for all functions C.

Often we are interested in cases where the likelihood ratio is monotonic[10]. Lemma 2, below, shows that, if one assumes that the likelihood ratio is monotonic, then an increase in effort decreases the probability of all events below a unique cutoff, and increases the probability of all events above that cutoff.

Condition 1 (Monotone Likelihood Ratio Property ("MLRP")). *(Strict) MLRP holds if, given $a, a' \in A$, $a' \preceq a \Rightarrow \pi_i(a')/\pi_i(a)$ is decreasing in i.*

It is well known that MLRP is a stronger condition than FOSD (in that MLRP $\Rightarrow$ FOSD, but FOSD $\not\Rightarrow$ MLRP).

Lemma 2. *Assume A1-A5 and that MLRP holds. Consider an action $a_1 \succeq a_2$. Then there exists j such that $\pi_i(a_1) > \pi_i(a_2)$ for all $i > j$ and $\pi_i(a_1) < \pi_i(a_2)$ for all $i < j$.*

We can now state the following result, which provides conditions for an increase in ϕ to lead to a harder action in the n outcome case when MLRP holds.

Proposition 4. *Assume A1-A5 and that MLRP holds. Then a^{**} is nondecreasing in ϕ iff:*

$$\sum_{i=j+1}^{n} \pi_i'(a) q_i'(\phi) \geq \sum_{i=1}^{j} \left| \pi_i'(a) \right| q_i'(\phi).$$

[10] See Milgrom (1981) and Karlin and Rubin (1956) for discussions of distributions which satisfy MLRP.

Assume also that C is differentiable and that there exist interior maximizers for all values of ϕ. Then a^{**} is strictly increasing in ϕ iff:

$$\sum_{i=j+1}^{n} \pi_i'(a) q_i'(\phi) > \sum_{i=1}^{j} \left|\pi_i'(a)\right| q_i'(\phi).$$

MLRP is a natural condition, but need not hold in general. Its usefulness in the above result was that it provided a unique cutoff point for increases and decreases in probabilities, given a harder action.

3 Comparison with the Linear Model

In an extremely influential paper, Holmström and Milgrom (1987) construct a dynamic principal-agent model and provide conditions under which a linear contract is the optimal contract. They obtain a solution for the slope of the incentive scheme which is

$$\beta = \frac{B'(a)}{1 + rc\sigma^2},$$

where $B'(a)$ is the derivative of the principal's benefit function with respect to agent effort (which is a real number), r is the agent's coefficient of absolute risk aversion, c is the second derivative of the agent's cost function, and σ^2 is the variance of the (mean zero) normally distributed shock to output. In the linear model there are clear and intuitive comparative statics. Incentives become more intense (and the equilibrium action is harder) when: the agent is less risk-averse, effort is less costly to her, when the shock has lower variance, and when effort is more valuable to the principal.

The basic question we ask here is which assumptions are critical for these comparative statics to hold *in a static setting*. Most applications of the linear model are in static settings, where we know that linear contracts are generically not optimal. We will show, however, that this does not affect most of the comparative static conclusions.

3.1 Benefit of Actions

Recall that in our setting the principal's benefit function is

$$B = \sum_{i=1}^{n} \pi_i(a)\, q_i(\phi).$$

First consider the comparative static in the linear model that there are more intense incentives when effort has a more positive impact on profits. The relevant condition in the general model is

$$\sum_{i=1}^{n} \pi_i'(a)\, q_i > \sum_{i=1}^{n} \pi_i'(a)\, \hat{q}_i \Longrightarrow a^{**} \geq_S \hat{a}^{**}, \tag{5}$$

which is equivalent to the condition from Lemma 1.

3.2 Cost of Effort

Second, in the linear model an increase in the second derivative of the agent's cost of effort function, c, leads to less intense incentives. Changes in c affect the cost of implementing an action, a^*, but do not affect the benefits to the principal. We thus need only focus on $C(a^*)$.

$$C(a^*) = \sum_{i=1}^{n} \pi_i(a^*)\, I_i.$$

Similarly, for action $a' \succ a^*$ we have

$$C(a') = \sum_{i=1}^{n} \pi_i(a')\, I_i',$$

where $\{I_i'\}_{i=1}^{n}$ is the optimal incentive scheme under action a'. Now consider an increase in the agent's cost of effort in the following sense. Transform the agent's utility function to $\hat{U}(a, I) = \alpha G(a) + K(a)V(I)$, with $\alpha > 1$. Denote the cost of implementing actions a^* and a' under this utility function $\hat{C}(a^*)$ and $\hat{C}(a')$ respectively. Recall that the principal's second-step problem is

$$\max_{a\in A}\left\{B(a,\phi)-C(a)\right\}.$$

A straightforward consequence of the Milgrom-Shannon Monotonicity Theorem is that a lower action is the SSO will be implemented if and only if

$$\hat{C}\left(a'\right)-C\left(a'\right)>\hat{C}\left(a^*\right)-C\left(a^*\right). \tag{6}$$

We can now establish

Theorem 1. *Suppose that either $G=0$ or K is constant on A. Then an increase in α leads the optimal action to decrease (in the SSO).*

Proof. Suppose K is constant on A, and w.l.o.g. let $K(\cdot)=1$. Now suppose, contrary to the conclusion of the theorem, that an increase in α leads the optimal action to increase. Then $\hat{C}\left(a'\right)-C\left(a'\right)<\hat{C}\left(a^*\right)-C\left(a^*\right)$ and hence by supposition $\hat{C}\left(a'\right)-\hat{C}\left(a^*\right)<C\left(a'\right)-C\left(a^*\right).$ $a'\succ a^*$ implies

$$\alpha G\left(a'\right)-\alpha G\left(a^*\right)<G\left(a'\right)-G\left(a^*\right).$$

The participation constraint (by Grossman-Hart Proposition 2) and we thus have

$$V\left(\hat{I}'\right)-V\left(\hat{I}^*\right)>V\left(I'\right)-V\left(I^*\right),$$

and thus

$$V\left(\hat{I}'\right)-V\left(I'\right)>V\left(\hat{I}^*\right)-V\left(I^*\right),$$

which implies

$$\hat{C}\left(a'\right)-C\left(a'\right)>\hat{C}\left(a^*\right)-C\left(a^*\right),$$

a contradiction.

Now consider the case where $G=0$. Again, suppose by way of contradiction that $\hat{C}\left(a'\right)-\hat{C}\left(a^*\right)<C\left(a'\right)-C\left(a^*\right).$ $a'\succ a^*$ and the binding participation constraint imply

that

$$K(a')V\left(\hat{I}'\right) - K(a^*)V\left(\hat{I}^*\right) > K(a')V(I') - K(a^*)V(I^*).$$

Rearranging we have

$$K(a')V\left(\hat{I}'\right) - K(a')V(I') > K(a^*)V\left(\hat{I}^*\right) - K(a^*)V(I^*),$$

which implies

$$\hat{C}(a') - C(a') > \hat{C}(a^*) - C(a^*),$$

a contradiction. ■

3.3 Risk-Aversion

Third, in the linear model an increase in the coefficient of absolute risk aversion, r, decrease the intensity of incentives. The risk aversion comparative static does not hold in general. Consider a transformation of the agent's utility function as follows. $\hat{U}(a, I) = G(a) + K(a)W(V(I))$, where W is a strictly concave function. This makes the agent more risk averse. It can be the case that more risk aversion allows a harder action to be implemented. If a harder action shifts the probabilities of different outcomes to be "more equal" (which is entirely consistent with distributions satisfying MLRP) then the agent gets more insurance if that action is implemented. The principal might not find it optimal to implement the harder action with the less risk averse agent because of the additional cost of effort. But with the more risk-averse agent she might, because that agent values the insurance effect from the harder action more. It is easy to provide counterexamples along these lines where the analogous condition to (6) is violated.

3.4 The Value of Information

Definition 1 (Lehmann, 1988)**.** *Consider two distribution functions F_θ and G_θ with density functions f_θ and g_θ with respect to a common σ-finite measure μ, and which satisfy MLRP in x. F_θ is* ***more accurate*** *than G_θ if and only if:*

$$h_\theta(x) = G_\theta^{-1}(F_\theta(x))$$

is a non-decreasing function of θ for all x.

We will be interested here in families of densities $f(\cdot|a), a \in A$ which give rise to the probabilities $\pi_1(a), ..., \pi_n(a)$. A key observation is that in representing the information available to the principal there is no loss of information in replacing any statistic with its associated likelihood ratio statistic. This is because in order to compute the posterior distribution the principal does not need to know the statistic itself, only the likelihood ratio statistic. As pointed out by Dynkin (1961) and Jewitt (1997), this is a consequence of the Halmos-Savage Factorization Theorem (Halmos and Savage (1949))[11]. In a given problem we will denote this sufficient likelihood ratio statistic as $T = f_a(q|a)/f(q|a)$ when A is infinite and

$$T = \left(\frac{f(q|a_1)}{f(q|a^*)}, ..., \frac{f(q|a_n)}{f(q|a^*)}\right)$$

in the case where A is a finite set $\{a_1, ..., a_n\}$.

Theorem 2. *Assume A1-A3 and consider two families of densities $f(\cdot|a)$ and $g(\cdot|a)$ which satisfy MLRP. Then the payoff to the principal is greater under f than under g if and only if f is more accurate than g.*

[11]Dynkin's formulation as presented in Jewitt (1997) makes this very clear. Let $f(x|a)$ be a density with $x \in \mathbb{R}^n$ and $a \in A$. Now consider the function T from the support of f to a function space defined by:

$$T(x) = g_x,\ g_x : A \to \mathbb{R},\ g_x(a) = f(x|a)/f(x|\bar{a}).$$

Since $f(x|a) = \Theta(T(x), a) f(x|\bar{a})$, the Halmos-Savage Factorization Theorem implies that T is a sufficient statistic for a.

Proof. The problem in (3) is

$$\max_{a\in A}\left\{B\left(a\right)-C\left(a,T\right)\right\},$$

and let $R(T)$ be the expected payoff to the principal. Since $R(T)$ depends only on C and hence we can substitute the likelihood ratio statistic into C. Now note that by MLRP the principal's second-step problem is in the class of *monotone decision problems* (see Lehmann (1988); Athey and Levin (2001)). Therefore we can appeal to Lehmann's Theorem 5.1. ■

Remark 3. *Grossman and Hart (1983) use a "garbling" in the sense of Blackwell (Blackwell and Girshick (1954)) to analyze a less informative signal. That is, if $\pi' = R\pi$, where R is an $n\times n$ stochastic matrix, then the incentive problem under π' is said to be noisier than under π. Because the Blackwell criterion is a partial order, this can only be a sufficient condition for a noisier problem*[12]*. For instance, the Blackwell criterion does not order natural increases in noise such as a uniform distribution with a wider support.*

Remark 4. *Jewitt (1997) establishes a remarkable result in a general principal-agent model which goes beyond the first-order approach, although he does not utilize the two-step approach that we have here. Jewitt shows that a necessary and sufficient condition for the principal to have better information is that the experiment be Blackwell More Informative for Mixture Dichotomies*[13]*. Jewitt's result is thus more general than the one we have just reported, as it*

[12] Except in the case of two outcomes where it is necessary and sufficient.

[13] Suppose that A is finite. Then f is **Blackwell more informative** than g if, for all $\psi\in\mathcal{C}$ (where $\mathcal{C}$ is the class of convex functions on $\mathbb{R}^{n-1}$), we have:

$$\int_{\psi}\left(\frac{f(q|a_1)}{f\left(q|a^*\right)},...,\frac{f(q|a_n)}{f\left(q|a^*\right)}\right)f\left(q|a^*\right)dq\geq\int_{\psi}\left(\frac{g(z|a_1)}{g\left(z|a^*\right)},...,\frac{g(z|a_n)}{g\left(z|a^*\right)}\right)g\left(q|a^*\right)dz$$

If for each two element subset $\{a_1,a_2\}\subset A$, the dichotomy

$$(f(q|a_1),f\left(q|a_2\right))$$

is more informative than the dichotomy

$$(g(q|a_1),g\left(q|a_2\right))$$

then we say that f is **Blackwell more informative** than g for **dichotomies**.

Let $f\left(q|\mu\right)$ and $g\left(z|\mu\right)$ be the mixture distributions obtained by selecting each $a_j\in A$ with probability $\mu_j, j=1,...,n$. If for each mixture μ and $a\in A$ the dichotomy $\left(f\left(q|\lambda\right),f\left(q|a\right)\right)$ is Blackwell more Informative than the dichotomy $\left(g\left(q|\lambda\right),g\left(q|a\right)\right)$ then we say that f is **Blackwell more informative** than g for **mixture dichotomies**.

extends to cases where MLRP does not hold (and hence Lehmann's Theorem does not apply).

4 Holmström's Sufficient Statistic Theorem

In analyzing the value of information Holmström (1979) asks when the Principal should condition the optimal incentive scheme on an additional signal which is available (see also Shavell (1979)). This is one of the most fundamental results of agency theory. A deep implication of the theorem is that the performance of one agent can be used to provide incentives for other agents when the performance of agents are correlated. This is the key insight behind the literatures on yardstick competition and tournaments (Lazear and Rosen (1981); Green and Stokey (1983); Nalebuff and Stiglitz (1983), among others). It also implies that, in an optimal contract, agents are not rewarded for luck. This has important implications for testing whether CEO contracts are actually optimal, or are the product of capture of the principal (see, for example, Bertrand and Mullainathan (2001)).

Holmström's result is obtained in the context where the first-order approach is valid (he assumes MLRP and CDFC). It is natural to try to extend this result to the more general setting considered in this paper.

The basic question is the following. Suppose that the principal observes not only q, but another signal x about the agent's action. Under what conditions will the optimal action/incentive scheme depend on x?

In the context of the model in this paper recall that in the second step the principal solves:

$$\max_{a \in A} \left\{ \sum_{i=1}^{n} \pi_i(a) q_i - C(a, T_x) \right\}, \tag{7}$$

where we again make use of the fact that there is no loss of generality in replacing any statistic with its likelihood ratio statistic, T. Here $T_x = f_a(q, x|a) / f(q, x|a)$. Let the likelihood ratio statistic when the principal does not observe x be $T = g_a(q|a) / g(q|a)$.

The main result of this section is that the optimal action depends on the signal x if and only if x provides additional information about the agent's action than that which is contained in q. That is, if q is not a sufficient statistic for (q, x) with respect to a.

Let the posterior distribution of A be $\xi(\cdot|x)$ and let $x_1 = (q_1, x_1)$ and $x_2 = (q_2, x_2)$. We say that q is a **sufficient statistic** for the family of *pdf*s $\{f(q, x|a), a \in A\}$ if $\xi(q, x|x_1) = \xi(q, x|x_2)$ for any prior *pdf* ξ and two pairs $x_1 \in \mathbb{R}^2, x_2 \in \mathbb{R}^2$ such that $q_1 = q_2$.

That is, the posterior distribution of the agent's action is only affected by x through q. If q is a sufficient statistic for (q, x) with respect to a then we say that the signal is **uninformative**, and **informative** otherwise.

Holmström's result and our generalization are the following.

Theorem 3 (Holmström, 1979)**.** *Assume MLRP and CDFC, $A \subseteq \mathbb{R}$ and $U = u(I) - c(a)$. Then the optimal action depends on a signal, x, if and only if the signal is informative.*

Theorem 4. *Assume A1-A3. Then the optimal action, a^{**}, depends on x if and only if q is not a sufficient statistic for (q, x) with respect to a.*

Proof. The optimization problem in (3) implies that the maximized value of the objective function does not depend on x if T does not depend on x. That is, if

$$\frac{f_a(q, x|a)}{f(q, x|a)} = \frac{g_a(q|a)}{g(q|a)}.$$

Integrating this with respect to x is equivalent to the existence two functions m and n such that

$$f(q, x|a) = m(q, a)\, n(q, x).$$

By the Halmos-Savage Factorization Theorem this representation is equivalent to the statement that q is a sufficient statistic for (q, x) with respect to a. ■

5 Discussion and Conclusion

It has been known since Mirrlees (1975) that the principal-agent problem is a complex one. Grossman and Hart (1983) provide an approach to the problem which is very general, and does not carry with it the strong assumptions that the first-order approach does. A contribution of this paper has been to show that the Grossman-Hart approach can be "operationalized" as a method for analyzing even complex principal-agent problems. If the goal of the analysis is to understand comparative statics, then the method used in this paper is substantially more general than recourse to the first-order approach, or the imposition of linear contracts. Much of the applied literature does indeed seek only comparative static conclusions. The closed form solutions made possible by assuming linearity are often used simply to derive comparative statics.

We have shown that all but one of the comparative static conclusions of the linear model do not depend critically on the relatively strong assumptions made (the risk-aversion result being the exception). Remarkably, the comparative static conclusions of the linear model hold under quite general assumptions *in a static setting*. Although strong assumptions are necessary to make the dynamic problem stationary, they are not critical for most of the comparative static conclusions of the linear model. This is surprising when one considers the crucial role they play in delivering a linear contract as the optimal contract.

A reasonable criticism of a researcher restricting attention to linear contracts in a static setting is that such contracts are generically not optimal. This paper shows that a reasonable rejoinder is that it does not matter for comparative statics.

References

Athey, S., and J. Levin (2001): "The Value of Information in Monotone Decision Problems," *Working paper.*

Athey, S., P. Milgrom, and J. Roberts (1998): *Robust Comparative Statics.* Unpublished manuscript.

Bertrand, M., and S. Mullainathan (2001): "Are CEOs Rewarded for Luck? The Ones Without Principals Are," *Quarterly Journal of Economics*, 116, 901–932.

Blackwell, D. A., and M. Girshick (1954): *Theory of Games and Statistical Decisions.* John Whiley, New York: NY.

Dynkin, E. (1961): "Necessary and Sufficient Statistics for a Family of Probability Distributions," *Selected Translations in Mathematical Statistics and Probability*, 1, 17–40.

Edlin, A. S., and C. Shannon (1998): "Strict Monotonicity in Comparative Statics," *Journal of Economic Theory*, 81, 201–219.

Green, J. R., and N. L. Stokey (1983): "A Comparison of Tournaments and Contracts," *Journal of Political Economy*, 91, 349–364.

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

Halmos, P. R., and L. Savage (1949): "Application of the Radon-Nikodym Theorem to the Theory of Sufficient Statistics," *Annals of Mathematics and Statistics*, 20, 225–241.

Hart, O. D., and B. Holmström (1987): "The Economics of Contracts," in *Advances in Economic Theory: Proceedings of the Fifth World Congress of the Econometric Society.*

Holmström, B. (1979): "Moral Hazard and Observability," *Bell Journal of Economics*, 10, 74–91.

Holmström, B., and P. Milgrom (1987): "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 55, 303–328.

Jewitt, I. (1988): "Justifying the First-Order Approach to Principal-Agent Problems," *Econometrica*, 56, 1177–1190.

——— (1997): "Information and Principal Agent Problems," *Mimeo, University of Bristol.*

Karlin, S., and H. Rubin (1956): "The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio," *Annals of Mathematical Statistics*, 27, 272–299.

Keeney, R. L. (1973): "Risk Independence and Multiattributed Utility Functions," *Econometrica*, 41, 27–34.

Lazear, E. P., and S. Rosen (1981): "Rank-Order Tournaments as Optimal Labor Contracts," *Journal of Political Economy*, 89, 841–864.

Lehmann, E. (1988): "Comparing Location Experiments," *Annals of Statistics*, 16, 521–533.

Milgrom, P., and C. Shannon (1994): "Monotone Comparative Statics," *Econometrica*, 62, 157–180.

Milgrom, P. R. (1981): "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12, 380–391.

Mirrlees, J. A. (1975): "The Theory of Moral Hazard and Unobservable Behavior: Part," *Mimeo, Nuffield College Oxford.*

Nalebuff, B. J., and J. E. Stiglitz (1983): "Information, Competition and Markets," *American Economic Review*, 73, 278–283.

Shavell, S. (1979): "Risk Sharing and Incentives in the Principal and Agent Relationship," *Bell Journal of Economics*, 10, 55–73.

6 Appendix A: Extensions

Here we extend the basic model in a number of directions. First we allow $A \subseteq \mathbb{R}^n$, and the functions B and C to be non differentiable. Second we allow another parameter of interest to affect the probability of certain states of nature occurring, π, in addition to the parameter which affects the profits in each state. Third,we consider the case where the principal is risk-averse. Fourth, we consider relaxing A1.

6.1 Relaxing A4 and A5

The more general version of Proposition 3 is as follows.

Proposition 5. *Assume A1-A3. Then a^{**} is nondecreasing in ϕ iff:*

$$\sum_{i=1}^{n} \pi_i(a)\, q_i(\phi) \text{ has increasing differences in } (a, \phi).$$

Proof. Follows from Lemma 1 and AMR Theorem 4.3. ■

6.2 Alternative Parameter Impacts

In sections 2 and 3 we were concerned with the parameter of interest, ϕ, having an impact on the profit levels in each of the states of nature. That is, the profits in the various states are $q_1(\phi), ..., q_n(\phi)$. While this is a natural case to consider, it is not of exclusive interest. The parameter of interest could affect the probabilities of the states occurring and/or the profit levels in the states. We now extend the results to the general case where there are two parameters of interest: ϕ and η. ϕ affects the profits in each state and $\eta \in \mathbb{R}$ affects the probabilities of those state occurring. Now the cost of implementing an action depends also on the value of η, since it provides information about a. Formally, the π function is now $\pi : A \times \mathbb{R} \to S$.

The principal's step-two problem is therefore:

$$\max_{a\in A}\left\{\sum_{i=1}^{n}\pi_i(a,\eta)q_i(\phi)-C(a,\eta)\right\} \tag{8}$$

Again, for the sake of exposition, we will deal with the case where the action set is a subset of the real line and B is differentiable. As before, the results extend straightforwardly to the non-differentiable, multidimensional action set case.

Proposition 6. *a^{**} is nondecreasing in η iff:*

$$\sum_{i=1}^{n}q_i(\phi)\frac{\partial^2\pi_i(a,\eta)}{\partial a\partial\eta}-\frac{\partial^2 C(a,\eta)}{\partial a\partial\eta}\geq 0,\qquad \forall a,\eta$$

Proof. Follows directly from AMR Theorem 2.2. ■

6.3 Risk-Averse Principal

The above analysis generalizes easily to the case where the principal is risk-averse.

Let the principal's utility function be given by U_P, which is, by construction, strictly concave. Following Grossman and Hart (1983), the principal's problem is then as follows:

$$\max_{\upsilon_1,\ldots,\upsilon_n}\left\{\sum_{i=1}^{n}\pi_i(a^*)U_P\left(q_i(\phi)-h(\upsilon_i)\right)\right\} \tag{9}$$

subject to

$$G(a^*)+K(a^*)\left(\sum_{i=1}^{n}\pi_i(a^*)\upsilon_i\right)\geq G(a)+K(a)\left(\sum_{i=1}^{n}\pi_i(a)\upsilon_i\right),\forall a\in A$$

$$G(a^*)+K(a^*)\left(\sum_{i=1}^{n}\pi_i(a^*)\upsilon_i\right)\geq\overline{U}$$

As Grossman and Hart (1983) point out, this is still a convex programming problem.

The second-step now entails:

$$\max_{a\in A}\left\{\sup_{\{v_1,...,v_n\}}\sum_{i=1}^{n}\pi_i(a^*)U_P\left(q_i(\phi)-h(v_i)\right)\right\} \tag{10}$$

Note that this is the case because the costs and benefits of selecting a particular action can no longer be analyzed separately. In this non-separable context, where there are multiple choice variables, $v_1, ..., v_n$,

A function $X : \Theta \rightarrow 2^S, S \subseteq \mathbb{R}^n$ from an ordered set Θ into the set of all subsets of $\mathbb{R}^n$ is **nondecreasing** if and only if $\forall\theta, \theta' \in \Theta$ such that $\theta > \theta', X(\theta) \geq_S X(\theta')$. If f has increasing differences in $(x_n; x_m)$, for all $n \neq m$, then f is **supermodular**.

Theorem 5 (AMR 4.9)**.** *Suppose $f : X \times \Theta \rightarrow \mathbb{R}$, where $\Theta \subset \mathbb{R}$ is a product set in $\mathbb{R}^n$. Let $X^*(\theta) \equiv \arg\max_{x\in X} f(\mathbf{x}, \theta)$. If f is supermodular then: (i) $X^*(\theta)$ is nondecreasing in θ, (ii) if $X^*(\theta)$ is nonempty and compact for each $\theta \in \Theta$, then $\mathbf{x}^L(\theta)$ and $\mathbf{x}^H(\theta)$ exist and are nondecreasing in θ.*

Proof. See AMR. ∎

We can now can state the following result which extends the conditions for an increase in ϕ to increase agent effort to the risk-averse principal case.

Proposition 7. *Assume A1-A6 and that the principal is risk-averse. Then an increase in ϕ weakly increases agent effort if:*

$$\frac{\partial^2}{\partial x \partial y}\left(\sum_{i=1}^{n}\pi_i(a)U_P\left(q_i(\phi)-h(v_i)\right)\right) \geq 0, \forall x \neq y, \text{ where } x, y = a, \phi, v_1, ..., v_n. \tag{11}$$

Proof. We require a^{**} to be nondecreasing in ϕ. By Theorem 5 this is the case if $\sum\limits_{i=1}^{n}\pi_i(\widehat{a})U\left(q_i(\phi)-h(v_i)\right)$ is supermodular. This requires (11). ∎

Note that where the principal is risk-neutral the v_is cancel out and we are left with the condition which was previously derived.

6.4 Non-Independent Utility Functions

We have already noted that Assumption 1 is rather strong. The difficulty with relaxing it, as Grossman and Hart (1983) point out, is that we have used $\boldsymbol{v} = (V(I_1), ..., V(I_n))$ as control variables independently from the action, a. The approach we take to this is to consider the impact of non-independence on the cost to the principal of implementing a given action. We build to a result which characterizes the way in which $C(a)$ depends on how agent's preferences for income lotteries vary with action.

First, recall the principal's problem, before using the simplification made possible by A1. She solves the following program:

$$\min_{I_1,\dots,I_n} \left\{ \sum_{i=1}^{n} \pi_i(a^*) I_i \right\} \tag{12}$$

subject to

$$a^* \in \arg\max_a \left\{ \sum_{i=1}^{n} \pi_i(a) U(a, I_i) \right\}$$

$$\sum_{i=1}^{n} \pi_i(a^*) U(a^*, I_i) \geq \overline{U}$$

This is, in general, a non-convex programming problem. We will be interested in movements in $C(a^*)$ as the agent's preferences for income lotteries changes with action. When A1 is invoked (12) can be converted into a convex programming problem using the Grossman-Hart transformation. Without A1, however, (12) is not a convex problem. To make explicit the agent's preferences for income lotteries as actions change, we introduce the parameter ω. A high value of ω corresponds to a *decreasing* preference for income lotteries as actions become harder. The following definition makes this precise.

Definition 2. *Consider an action $\overline{a} >_S \underline{a} \in A$ and an agent with utility function $U(a, I, \omega)$.*

Then for $\overline{\omega} > \underline{\omega}$ it must be that:

$$\sum_{i=1}^{n} \pi_i(\underline{a})U(\underline{a}, I_i, \underline{\omega}) - \sum_{i=1}^{n} \pi_i(\underline{a})U(\underline{a}, I_i, \overline{\omega}) <$$
$$\sum_{i=1}^{n} \pi_i(\overline{a})U(\overline{a}, I_i, \underline{\omega}) - \sum_{i=1}^{n} \pi_i(\overline{a})U(\overline{a}, I_i, \overline{\omega})$$

Now write the principal's program as:

$$\min_{I_1,\dots,I_n} \left\{ \sum_{i=1}^{n} \pi_i(a^*)I_i \right\} \tag{13}$$

subject to

$$a^* \in \arg\max_{a} \left\{ \sum_{i=1}^{n} \pi_i(a)U(a, I_i, \omega) \right\}$$
$$\sum_{i=1}^{n} \pi_i(a^*)U(a^*, I_i, \omega) \geq \overline{U}$$

We are now able to show how Proposition 3 extends to the case where the agent's preferences for income lotteries are not independent of action. First define the function $f : A \times \mathbb{R}^2 \to \mathbb{R}$ as the net return to the principal from the triple (a, ϕ, ω).

Proposition 8. *Assume A1-A5. Then a^{**} is nondecreasing in ϕ, if $f(a, \phi, \omega)$ is supermodular.*

Proof. Follows from the Monotonicity Theorem of Milgrom and Shannon (1994). ■

Thus, when the agent has a decreasing preference for income lotteries as actions become harder a sufficient condition for a^{**} to be nondecreasing in ϕ is that the payoff to the principal has increasing difference in (a, ϕ).

7 Appendix B: Omitted Proofs

Proof of Lemma 2. $a_1 \succ a_2 \Rightarrow$ (by FOSD) $\exists i$ such that $\pi_i(a_1) > \pi_i(a_2)$ and

$$\exists j \equiv \max\{g \text{ s.t.} \pi_g(a_1) < \pi_g(a_2)\}$$

such that $\pi_j(a_1) < \pi_j(a_2)$, since probabilities must sum to one. Suppose, by way of contradiction, that $j > i$. Then, by MLRP

$$\frac{\pi_i(a_1)}{\pi_i(a_2)} < \frac{\pi_j(a_1)}{\pi_j(a_2)} < 1,$$

but $\pi_i(a_1)/\pi_i(a_2) > 1$, a contradiction. Now, by MLRP $\pi_{i'}(a_1) > \pi_{i'}(a_2)$ for all $i' > i$ and $\pi_{j'}(a_1) < \pi_{j'}(a_2)$ for all $j' < j$. Now suppose $\exists k, k'$ such that $k, k' > j$ and $k, k' < i$. Without loss of generality assume that $k > k'$. By way of contradiction assume that $\pi_k(a_1) = \pi_k(a_2)$ and $\pi_{k'}(a_1) = \pi_{k'}(a_2)$. Then

$$\frac{\pi_k(a_1)}{\pi_k(a_2)} > \frac{\pi_{k'}(a_1)}{\pi_{k'}(a_2)},$$

in contradiction of MLRP. ■

Proof of Proposition 4. Part 1: Recall that we require $a^{**}(\phi; C)$ to be increasing in ϕ and hence that B has increasing differences. By Lemma 1 we require $\sum_{i=1}^{n} q_i'(\phi)\pi_i'(a) \geq 0, \forall a, \phi$. Now apply Lemma 2 and write the condition as

$$\sum_{i=1}^{j} q_i'(\phi)\pi_i'(a) + \sum_{i=j+1}^{n} q_i'(\phi)\pi_i'(a) \geq 0.$$

This requires

$$\sum_{i=j+1}^{n} q_i'(\phi)\pi_i'(a) \geq -\sum_{i=1}^{j} q_i'(\phi)\pi_i'(a).$$

Since $\pi_i'(a) \leq 0$ for all $i \leq j$ and $\pi_i'(a) > 0$ for all $i > j$, by Lemma 2:

$$-\sum_{i=1}^{j} q_i'(\phi)\pi_i'(a) = \sum_{i=1}^{j} q_i'(\phi) \left|\pi_i'(a)\right|.$$

This completes the proof of this part. Part 2 follows by identical arguments. ■