Holmström's Sufficient Statistic Theorem

Richard T. Holden*
Massachusetts Institute of Technology
50 Memorial Drive, E52-410
Cambridge, MA, 02142
rholden@mit.edu

June 6, 2006

Abstract

An important result, obtained by Holmström (1979), is that the optimal incentive scheme in a principal-agent problem should depend on a signal if and only if that signal independently affects the likelihood ratio. This has striking implications for the use of signals such as the performance of other agents, or other firms. The result was obtained in a setting where the first-order approach is valid. It has been known since Mirrlees (1975), that the first-order approach is only valid under strong assumptions, such as the Monotonic Likelihood Ratio Property combined with the Convexity of the Distribution Function Condition. These do not hold for most common families of distributions. In this paper, I prove Holmström's theorem in a general principal-agent setting due to Grossman and Hart (1983), using an approach developed by Holden (2005).

Keywords: Lattice, likelihood ratio, principal-agent problem, sufficient statistic

JEL Classification Codes: L13, L22.

*I am grateful to Philippe Aghion, Robert Akerlof, Oliver Hart and Bengt Holmström for valuable discussions; and to Shiyang Cao for excellent research assistance.

1 Introduction

Holmström (1979) showed that the optimal incentive scheme that a principal offers an agent should depend on a signal if and only if the signal independently affects the likelihood ratio. This is one of the most fundamental results of agency theory. It is essentially the agency theory equivalent of the famous Rao-Blackwell theorem of statistical inference. 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, Mullainathan and Shoar, 2000).

These conclusions are indicative of the fact that the princpal-agent problem, whilst not actually a statistical inference problem, behaves like one (Hart and Holmström (1987)). Although the principal does not observe the agent's action - only a noisy signal or set of signals of it - she knows the action in equilibrium. The incentive scheme offered by the principal, however, takes into account this inference problem in the consideration of out-of-equilibrium actions by the agent.

Holmström's theorem was proved using the first-order approach to the principal-agent problem (i.e. replacing the agent's incentive compatibility constraint with the first-order condition of her problem). It was first noted by Mirrlees (1975), that the first-order approach is only valid under strong assumptions. He showed that if the Monotonic Likelihood Ratio Property ("MLRP") and the Convexity of the Distribution Function Condition ("CDFC") hold then the first-order approach is valid because the agent's incentive compatibility constraint and her first-order condition are equivalent. Many authors have noted how restrictive this pair of assumptions are. As Jewitt (1988) points out, if shocks are additive to output then any *pdf* which satisfies both MLRP and the CDFC must have everywhere increasing density! Other assumptions such as the Spanning Condition ("SC") (Grossman and Hart, 1983), the Linearity of the Distribution Function Condition ("LDFC") Hart and Holmström (1987), and Jewitt's (1988) restrictions on the agent's utility function are all rather strong.

In light of the importance of Holmström's result, and the special setting in which it was obtained, it is seems important to understand whether it is a robust finding which is true in general. This paper tackles this question, and answers it in the affirmative.

To do so, I use an approach developed by Holden (2005), which is based on an important insight of Grossman and Hart (1983). They proposed a two-step procedure, in which the principal first determines the lowest cost way to implement any action, and then chooses the action which maximizes the difference between the expected benefits and the costs of implementation. This approach does not entail the strong assumptions that the first-order approach does.

The remainder of the paper is organized as follows. Section 2 provides the basis for the model by outlining the first step of the Grossman-Hart approach to the principal-agent problem. Section 3 contains the result, and Section 4 contains some concluding remarks.

2 The Grossman-Hart Approach

2.1 Statement of the Problem

This section is based heavily on Holden (2005). There are two players, a risk-neutral principal and a risk-averse agent. The agent takes an action, a, which is unobservable to the principal. Suppose that there are a finite number of possible profit levels for the firm. Denote these $q_1, ..., q_n$ with $q_1 < q_2 < ... < q_n$. The set of actions available to the agent is A, which we will take for the moment to be $\mathbb{R}$ in order to simplify the notation. In section 4 I discuss the fact that this can be relaxed so that A is a compact subset of $\mathbb{R}^n$. Define the ordinary probability simplex $S = \{y \in \mathbb{R}^n | y \geq 0, \sum_{i=1}^{n} y_i = 1\}$ and assume that there exists a twice continuously differentiable function $\pi : \mathbb{R} \times A \rightarrow S$ (relaxation of the differentiability restriction is also discussed in Section 4). Let $\theta \in \mathbb{R}$ be a signal about the agent's action which the principal observes in addition to output level $q \in \mathbb{R}$. The probabilities of outcomes $q_1, ..., q_n$ are therefore $\pi_1(a, \theta), ..., \pi_n(a, \theta)$. The principal is able to make a payment to the agent I which can be conditioned on the pair (q, θ).

Let the agent's von Neumann-Morgenstern utility function be given by the following:

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

Assumption A1. *V is a continuous, strictly increasing, real-valued, concave function on a subset 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 $K(a)$ is strictly positive.*

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 the following assumption is important in ensuring that an optimal second best action and incentive scheme exists.

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.

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

I now outline the two steps involved in solving the problem, and show how to determine whether the optimal action depends on the signal θ.

2.2 Solution: Step One

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.

Formally this is as follows:

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

$$\text{subject to}$$

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

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

Following Grossman and Hart (1983), I proceed in two stages. First I 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. I then consider, given this, what is the optimal a^* which she wishes to implement. To that end write the problem as:

$$\min_{I_1,...,I_n} \left\{ \sum_{i=1}^{n} \pi_i(a^*,\theta) I_i \right\} \tag{2}$$

$$\text{subject to}$$

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

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

$$I_i \in \mathcal{I}, \forall i$$

Now define $\upsilon_1 = V(I_1), ..., \upsilon_n = V(I_n)$ and $h \equiv V^{-1}$. These will be used as the control variables. The problem can now be stated as:

$$\min_{\upsilon_1,\ldots,\upsilon_n}\left\{\sum_{i=1}^{n}\pi_i(a^*,\theta)h(\upsilon_i)\right\} \tag{3}$$

subject to

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

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

$$\upsilon_i\in\mathcal{U},\forall i$$

The constraints in (3) are linear in the υ_js and since V is concave h is convex. Consequently the problem in (3) is simply to minimize a convex function subject to a set of linear constraints. Grossman and Hart (1983) show that existence follows from Weierstrass's theorem.

Definition 1. *A vector $(\upsilon_1,...,\upsilon_n)$ which satisfies the constraints in (3) is said to Implement action a^*.*

The following definition states the cost of implementing a given action.

Definition 2. *Let:*

$$C(a^*,\theta)=\inf\left\{\sum_{i=1}^{n}\pi_i(a^*,\theta)h(\upsilon_i)|\boldsymbol{\upsilon}=(\upsilon_1,...,\upsilon_n)\text{ implements }a^*\right\}$$

which implements a^* if the constraint set in (3) is non-empty. If the constraint set is empty then let $C(a^*)=\infty$.

2.3 Solution: Step Two

The second step is to choose the action which maximizes the difference between the expected benefits, and cost of implementation. The principal solves:

$$\max_{a\in A}\left\{\sum_{i=1}^{n}\pi_i(a,\theta)q_i-C(a,\theta)\right\}. \tag{4}$$

2.4 Comparative Statics

Denote the (possibly set valued) solution to the problem in (4) as a^{**}. We will be interested in conditions under which this set is increasing in the Strong Set Order, which is defined below.

Definition 3. *A function $f : \mathbb{R}^2 \to \mathbb{R}$ has "Increasing Differences" if for all $x'' > x$, $f(x'', \theta) - f(x', \theta)$ is nondecreasing in θ.*

Definition 4. *Let $X, Y \subset \mathbb{R}^n$. Then X is higher than Y in the Strong Set Order ($X \geq_S Y$) iff for any $\mathbf{x} \in X$ and $\mathbf{y} \in Y$, $\max\{\mathbf{x}, \mathbf{y}\} \in X$ and $\min\{\mathbf{x}, \mathbf{y}\} \in Y$.*

Holden (2005) shows that, provided there are interior maximizers for all values of θ, a^{**} is increasing in the strong set order ("SSO") in θ if and only if:

$$\sum_{i=1}^{n} q_i \frac{\partial^2 \pi_i(a, \theta)}{\partial a \partial \theta} - \frac{\partial^2 C(a, \theta)}{\partial a \partial \theta} > 0 \tag{5}$$

and decreasing in the SSO if and only if:

$$\sum_{i=1}^{n} q_i \frac{\partial^2 \pi_i(a, \theta)}{\partial a \partial \theta} - \frac{\partial^2 C(a, \theta)}{\partial a \partial \theta} < 0. \tag{6}$$

To see this denote the solution to problem (4), above, as a^{**}. Then from the Monotonicity Theorem of Milgrom and Shannon (1994), and the results of Edlin and Shannon (1998), it follows that a necessary and sufficient condition for a^{**} to be strictly increasing in the SSO as a function of θ is, provided that C is differentiable and there exist interior maximizers in A for all θ, that:

$$\sum_{i=1}^{n} \pi_i(a, \theta) q_i - C(a, \theta)$$

has strictly increasing differences in (a, θ). This is precisely:

$$\sum_{i=1}^{n} q_i \frac{\partial^2 \pi_i(a, \theta)}{\partial a \partial \theta} - \frac{\partial^2 C(a, \theta)}{\partial a \partial \theta} > 0.$$

(6) follows by identical arguments for strictly decreasing differences.

3 The Result

The main result of this paper is that the optimal action depends on the signal θ if and only if θ 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, θ) with respect to a.

Definition 5 (DeGroot, 1970, 2004)**.** *Let the posterior distribution of A be $\xi(\cdot|x)$ and let $x_1 = (q_1, \theta_1)$ and $x_2 = (q_2, \theta_2)$. We say that q is a Sufficient Statistic for the family of pdfs $\{f(q, \theta|a), a \in A\}$ if $\xi(q, \theta|x_1) = \xi(q, \theta|x_2)$ 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 θ through q. Since the posterior distribution is assumed to be computed according to Bayes Rule, this definition places no restriction on events which occur on a set of measure zero. I now show that this is equivalent to a restriction on the cross partial of $\pi_i(a, \theta)$.

Lemma 1. *q is a Sufficient Statistic for (q, θ) with respect to a if and only if:*

$$\frac{\partial^2 \pi_i(a, \theta)}{\partial a \partial \theta} = 0, \forall i \ .$$

Proof. q is a sufficient statistic for (q, θ) with respect to a if and only

$$\frac{f(q, \theta|a)}{g(q|a)} \text{ is independent of } a \text{ for any } (q, \theta) \in \mathbb{R}^2.$$

(see Casella and Berger, 2001). Now note that $g(q|a) = \int_{\mathbb{R}} g(q|a, u) z(u) du = \int_{\mathbb{R}} \pi_i(a, u) z(u) du$. Since $f(q, \theta|a) = h(q|\theta, a) z(\theta)$ we have:

$$\begin{aligned} \frac{f(q, \theta|a)}{g(q|a)} &= \frac{\pi_i(a, \theta) z(\theta)}{g(q|a)} \\ &= \frac{\pi_i(a, \theta) z(\theta)}{\int_{\mathbb{R}} \pi_i(a, u) z(u) du}. \end{aligned}$$

Therefore if q is a sufficient statistic for (q, θ) with respect to a then:

$$z(\theta) \frac{\int_{\mathbb{R}} \pi_i(a, u) z(u) du \frac{\partial \pi_i(a, \theta)}{\partial a} - \int_{\mathbb{R}} \pi_i(a, u) \frac{\partial \pi_i(a, u)}{\partial a} z(u) du}{\left(\int_{\mathbb{R}} \pi_i(a, u) z(u) du\right)^2} = 0.$$

This requires:

$$\int_{\mathbb{R}} \pi_i(a,u)z(u)du\frac{\partial \pi_i(a,\theta)}{\partial a} - \pi_i(a,u)\int_{\mathbb{R}} \frac{\partial \pi_i(a,u)}{\partial a} z(u)du = 0.$$

Differentiating the above w.r.t. θ yields:

$$\int_{\mathbb{R}} \pi_i(a,u)z(u)du\frac{\partial^2 \pi_i(a,\theta)}{\partial a \partial \theta} - \frac{\partial \pi_i(a,u)}{\partial \theta}\int_{\mathbb{R}} \frac{\partial \pi_i(a,u)}{\partial a} z(u)du = 0.$$

Note, however, that since $\int_{\mathbb{R}} \pi_i(a,u)z(u)du = 1$ we have $\int_{\mathbb{R}} \frac{\partial \pi_i(a,u)}{\partial a} z(u)du = 0$. Subsituting these two expression into the above yields:

$$\frac{\partial^2 \pi_i(a,\theta)}{\partial a \partial \theta} = 0.$$

This completes the proof. ■

Lemma 2. *If q is a Sufficient Statistic for (q,θ) with respect to a it follows that:*

$$\frac{\partial^2 C(a,\theta)}{\partial a \partial \theta} = 0.$$

Proof. Taking the cross partial of $C(a,\theta)$, from Definition 2, w.r.t. a,θ and appying the Envelope Theorem yields:

$$\sum_{i=1}^{n} \frac{\partial^2 \pi_i(a,\theta)}{\partial a \partial \theta} I_i^*$$

By Lemma 1 we then have $\partial^2 C(a,\theta)/\left(\partial a \partial \theta\right) = 0$. ■

Definition 6. *If q is a sufficient statistic for (q,θ) with respect to a then we say that the signal is Uninformative, and Informative otherwise.*

Holmström's result can then be stated as follows.

Theorem 1 (Holmström, 1979)**.** *Suppose that the first-order approach is valid, $A \subseteq \mathbb{R}$ and $U = u(I) - c(a)$. Then the optimal action depends on a signal, θ, if and only if the signal is informative.*

Theorem 2. *Assume A1-A3 and that $\pi_i(a,\theta)$ is twice continuously differentiable in (a,θ). Then, except on a set of measure zero, the optimal action, a^{**}, depends on θ if and only if q is not a sufficient statistic for a with respect to (q,θ).*

Proof. The proof is by contraposition. Hence we will show that a^{**} does not depend on $\theta \Leftrightarrow q$ is a sufficient statistic for a with respect to (q, θ).
$\Rightarrow$ Direction: From (5) and (6) it follows that a^{**} does not depend on θ implies the following:

$$\sum_{i=1}^{n} q_i \frac{\partial^2 \pi_i(a,\theta)}{\partial a \partial \theta} - \frac{\partial^2 C(a,\theta)}{\partial a \partial \theta} = 0.$$

This can only be satisfied on a set of positive measure if:

$$\frac{\partial^2 \pi_i(a,\theta)}{\partial a \partial \theta} = 0 \wedge \frac{\partial^2 C(a,\theta)}{\partial a \partial \theta}.$$

By Lemmas 1 and 2 this implies that q is a sufficient statistic for a with respect to (q, θ).

$\Longleftarrow$ Direction: by Lemma 1 we have:

$$\frac{\partial^2 \pi_i(a,\theta)}{\partial a \partial \theta} = 0, \forall i$$

and by Lemma 2 we have:

$$\frac{\partial^2 C(a,\theta)}{\partial a \partial \theta} = 0.$$

Together these imply that a^{**} does not depend on θ since:

$$\sum_{i=1}^{n} q_i \frac{\partial^2 \pi_i(a,\theta)}{\partial a \partial \theta} - \frac{\partial^2 C(a,\theta)}{\partial a \partial \theta} = 0.$$

■

4 Concluding Remarks

I have assumed throughout that $A \subseteq \mathbb{R}$. This can be relaxed so that A is a compact set which is dense in $\mathbb{R}^n$. Holden (2005) shows that for a^{**} to be increasing in θ requires that:

$$\sum_{i=1}^{n} \pi_i(a,\theta) q_i - C(a,\theta)$$

have strictly increasing differences in (a, θ), provided that there exist interior maximizers. When $A \subseteq \mathbb{R}$ this reduces to the convenient form used above.

The first-order approach has proved to be a convenient technique for analyzing principal-agent problems, but also a controversial one. It carries with it strong assumptions. Using the approach of Grossman and Hart (1983) and Holden (2005), I have shown that one of the most important results of agency theory, Holmström's Sufficient Statistic Theorem, does not depend on the assumptions of the first-order approach. The result holds in a substantially more generally setting.

References

[1] Bertrand, Marianne and Sendhil Mullainathan. (2001). "Are CEOs Rewarded for Luck? The Ones Without Principals Are," *Quarterly Journal of Economics*, 116(3), pp.901-932.

[2] DeGroot, Morris. (1970, 2004). *Optimal Statistical Decisions.* Whiley. Hoboken, NJ.

[3] Casella, George and Roger L. Berger. (2001). *Statistical Inference.* 2nd ed. Duxbury. New York, NY.

[4] Edlin, Aaron S. and Chris Shannon. (1998). "Strict Monotonicity in Comparative Statics," *Journal of Economic Theory* 81. pp.201-219.

[5] Grossman, Sanford J. and Oliver D. Hart (1983). "An Analysis of The Principal-Agent Problem," *Econometrica,* 51, pp.7-45.

[6] Hart, Oliver D. and Bengt Holmström (1987). "The Economics of Contracts", in *Advances in Economic Theory: Proceedings of the Fifth World Congress of the Econometric Society.*

[7] Holden, Richard T. (2005). "A Monotonicity Based Approach for Analyzing Principal-Agent Problems." mimeo. Harvard University Department of Economics.

[8] Holmström, Bengt (1979). "Moral Hazard and Observability," *Bell Journal of Economics,* 10, pp.74-91.

[9] Jewitt, Ian (1988). "Justifying the First-Order Approach to Principal-Agent Problems," *Econometrica,* 56, pp.1177-1190.

[10] Keeney, Ralph L. (1973). "Risk Independence and Multiattributed Utility Functions," *Econometrica,* 41. pp.27-34.

[11] Milgrom, Paul and Chris Shannon (1994). "Monotone Comparative Statics," *Econometrica,* 62, pp.157-180.

[12] Mirrlees, James A. (1975). "The Theory of Moral Hazard and Unobservable Behavior: Part I," Nuffield College, Oxford, mimeographed.

[13] Pollak, Robert A. (1973). "The Risk Independence Axiom," *Econometrica,* 41, pp.35-39.