Part III

Hidden Action

Chapter 5

The Principal-Agent Model

5.1 Setup

- 2 players: The principal, owner of the firm; and the agent, manager/worker in the firm
- The principal hires the agent to perform a task
- The agent chooses an action, $a \in A$
- Each $a \in A$ yields a distribution over payoffs (revenues) for the firm, q, according to the function $q = f(a, \varepsilon)$ where ε is some random variable with known distribution (Note that the principal "owns" the rights to the revenue q.)
- $a \in A$ is *not observable* to the principal; q is *observable and verifiable* (so q can be a basis for an enforceable contract)

Figure Here (time line)

Examples:

	Principal	Agent	Action
1	Firm owner	Manager	choice of (risky) project
2	Employer/Manager	employee/worker	effort in job
3	Regulator	Regulated firm	cost reduction research
4	Insurer	Insuree	care effort

Question: What makes the problem interesting?

A conflict of interest over $a \in A$ between the principal and the agent.

5.1.1 Utilities

We assume that the principal is risk neutral, and the agent is risk averse. This is the standard formulation of the P-A problem.

- **Agent's utility:** The agent has a vNM utility function defined on his action, a, and the income he receives, I,

$$U(a, I) = v(I) - g(a),$$

  where $v' > 0$, and $v'' < 0$ guarantee risk aversion. (As for $g(\cdot)$, we assume that effort is costly, $g' > 0$, and we usually assume that the marginal cost of effort increases with effort, $g'' > 0$, which seems "reasonable.") Note that this utility function is additively separable. This simplification is very helpful for the analysis and its implication is that the agent's preference over income lotteries is independent of his choice of action Finally, assume that the agent's reservation utility is given by $\overline{u}$ (which is determined by some alternative option.)

- **Principal's utility:** The principal's utility is just revenue less costs, and if we assume that the only cost is compensating the agent then this is trivially given by $q - I$.

Question: Why is it reasonable to assume that the principal is risk neutral while the agent is risk averse?

1. If the agent is risk neutral, and the agent has no limits on wealth, the problem will be trivial (as we will later see why).

2. If both are risk averse the analysis is more complicated, but we get the same general issues and results, so it is unnecessarily more complex.

3. An appealing rationale is that the principal is wealthy and has many investments, so this firm is only a small fraction of his portfolio and risk is idiosyncratic. (Caveat: usually shareholders don't design incentive schemes!)

4. **Another modelling possibility:** Agent is risk neutral but has limited wealth - say, in the form of limited liability - so that he cannot suffer large losses. This gives a "kink" at an income level of zero, which gives the necessary concavity of the agent's utility function which yields the same kind of results.

Further Assumptions

We begin our analysis by following the work of Grossman and Hart (1983) (G-H hereafter), and later go back to the earlier formulations of the principal-agent problem.

- $\tilde{q}$ (the random variable) is discrete: $q \in \{q_1, q_2, ..., q_n\}$, and w.l.o.g., $q_1 < q_2 < \cdots < q_n$.
- $A \subset \Re^k$ is compact and nonempty.
- given $a \in A$, the mapping $\pi : A \to S$ maps actions into distributions over outcomes, where $S = \{x \in \Re^n : x_i \geq 0 \ \forall i \text{ and } \sum_{i=1}^{n} x_i = 1\}$ is the n-dimensional simplex. $\pi_i(a)$ denotes the probability that q_i will be realized given that $a \in A$ was chosen by the agent. As usual, we assume that the distribution functions are common knowledge.
- $v(\cdot)$ is continuous, $v' > 0$, $v'' < 0$, and $v(\cdot)$ is defined over $(\underline{I}, \infty) \in \Re$ where $\lim_{I \to \underline{I}} v(I) = -\infty$ (This guarantees that we need not worry about corner solutions.) For example: $v(\cdot) = \ln(\cdot)$, in which case $\underline{I} = 0$. (This is Assumption A1 in G-H.)
- $g(\cdot)$ is only assumed to be continuous.

5.2 First Best Benchmark: Verifiable Actions

Assume that $a \in A$ is *observable and verifiable* so that the principal can basically "choose" the best $a \in A$, and contract on it while promising the agent some compensation. Assume that the principal has all the bargaining power. (This is a standard simplification, which can be relaxed without affecting the qualitative results.) She then solves:

$$F.B. \begin{cases} \max\limits_{\substack{a \in A \\ I \in \Re}} & \sum\limits_{i=1}^{n} \pi_i(a) q_i - I \\ \text{s.t.} & v(I) - g(a) \geq \overline{u} \quad \text{(IR)} \end{cases}$$

where $I \in \Re$ is the payoff to the agent. Note that we considered I to be a fixed payoff, which brings us to the following question: Could a random payment $\tilde{I}$ be optimal? The answer is clearly no, which follows from the agent's risk aversion. We can replace any $\tilde{I}$ with its *Certainty Equivalent*, which will be less costly to principal since she is risk neutral.

Claim: (IR) binds at the solution

This claim is clearly trivial, for if (IR) would not bind we can reduce I, and lower the principal'ss costs.

We can define

$$C_{FB}(a) \equiv v^{-1}(\overline{u} + g(a))$$

to be the cost to the principal of compensating the agent for choosing $a \in A$. This is the *First-Best* (FB) cost of implementing an action a, since risk is optimally shared between the risk neutral principal and the risk averse agent. Using this formalization, the principal solves:

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

The *First-Best Solution* yields a FB action, $a^*_{FB} \in A$ (which may not be unique), and this action is implemented by the FB contract: (a^*_{FB}, I^*_{FB}) where $I^*_{FB} = C_{FB}(a^*_{FB})$.

5.3 Second Best: non-observable action

Once the action is not observable, we can no longer offer contracts of the form (a^*_{FB}, I^*_{FB}). One can then ask, if it is not enough just to offer I^*_{FB} and anticipate the agent to perform a^*_{FB} ? The answer is no, since if I^*_{FB} is offered, then the agent will choose $a \in A$ to minimize $g(a)$, and it may not be that "likely" that a^*_{FB} will achieve the agent's goal.

Question: What can the principal do to implement an action $a \in A$?

The principal can resort to offering an *Incentive Scheme* which rewards the agent according to the level of revenues, since these are assumed to be observable and verifiable. That is, the agent's compensation will be a function, $I(q)$.

By the finiteness of $\tilde{q}$, restrict attention to $I \in \{I_1, ..., I_n\}$. The principal will choose some a^*_{SB}, together with an incentive scheme that implements this action. (That is, an incentive scheme that causes the agent to choose a^*_{SB} as his "best response.") Clearly, at the optimum it must be true that a^*_{SB} is implemented at the *lowest possible cost* to the principal. This implies that we can decompose the principal's problem to a two stage problem as follows:

1. First, look at the lowest cost to implement any $a \in A$ (i.e., for each $a \in A$, find $(I^*_1(a), ..., I^*_n(a))$ which is the lowest cost incentive scheme needed to implement $a \in A$)

2. Given $\{(I^*_1(a), ..., I^*_n(a))\}_{a \in A}$ choose a^*_{SB} to maximize profits.

The Second Stage

If the principal has decided to implement a^*, then it must be implemented at the lowest cost. That is, $(I_1, ..., I_n)$ must solve

$$\begin{cases} \min\limits_{I_1,...,I_n} & \sum\limits_{i=1}^{I} \pi_i(a^*) I_i & \\ \text{s.t.} & \sum\limits_{i=1}^{n} \pi_i(a^*) v(I_i) - g(a^*) \geq \overline{u} & \text{(IR)} \\ & \sum\limits_{i=1}^{n} \pi_i(a^*) v(I_i) - g(a^*) \geq \sum\limits_{i} \pi_i(a) v(I_i) - g(a) \;\; \forall a \in A & \text{(IC)} \end{cases}$$

Note that we have a program with a linear objective function, and concave constraints. For mathematical convenience, we can transform the program into one with a convex objective function and with linear constraints. To do this, we work with "utils" instead of income: Let $h(\cdot) \equiv v^{-1}(\cdot)$ be the inverse utility function with respect to the agent's income, and consider the principal's choice of $(v_1, ..., v_n)$, where $I_i = h(v_i)$. (The existence of such an inverse function $h(\cdot)$ is guaranteed by G-H assumption A2.) We know that

$h(\cdot)$ is convex since, $v' > 0$ and $v'' < 0$ imply that $h' > 0$ and $h'' > 0$, and the assumption that $\lim_{I \to \underline{I}} v(I) = -\infty$ implies that $v_i \in (-\infty, \overline{v})$ where $\overline{v}$ can be ∞. The program can therefore be written as,

$$\begin{cases} \min_{v_1,\ldots,v_n} & \sum_{i=1}^{I} \pi_i(a^*)h(v_i) & \\ \text{s.t.} & \sum_{i=1}^{n} \pi_i(a^*)v_i - g(a^*) \geq \overline{u} & \text{(IR)} \\ & \sum_{i=1}^{n} \pi_i(a^*)v_i - g(a^*) \geq \sum_i \pi_i(a)v_i - g(a) \ \ \forall a \in A & \text{(IC)} \end{cases} \tag{5.1}$$

which is a "well behaved" program. (If A is finite, we can use Kuhn-Tucker.)

Question: When will we have a solution?

If the constrained set is empty, we clearly won't have one, so this is not an interesting case. Assuming that the constrained set is non-empty, then if we can show that it is closed and bounded, then a solution exists (Weirstrass). The question is, therefore, when will it be true?

Assumption 4.1: (A3 in G-H) $\pi_i(a) > 0$ for all $i \in \{1, ..., n\}$ and for all $a \in A$.

Proposition 4.1: Under the assumptions stated above, a solution is guaranteed.

Sketch of Proof: The idea is that we can bound the set of v's, thus creating a compact constrained set. Assume in negation that we cannot: $\exists$ an unbounded sequence $\{v_1^k, ..., v_n^k\}_{k=1}^{\infty}$ such that some components go to $-\infty$. This implies that I_i^k will also be unbounded for some i where $I_i^k = h(v_i^k)$. Since $\pi_i(a) > 0$ for all i, then if some $v_i^k \to -\infty$ without some $v_j^k \to +\infty$ then the agent's utility will be going to $-\infty$. So, if $v_i \in (-\infty, \overline{v})$, where $\overline{v} < \infty$ we are done since we cannot have some $v_j^k \to +\infty$. Assume, therefore, that $v_i \in (-\infty, \infty)$ so that we could have some components going to $-\infty$ and others going to $+\infty$. Now risk aversion will come into play: the variance of the incentive scheme I^k goes to infinity, and to compensate the agent for this risk, the mean must go to infinity as well. Thus, the principal's expected payment to agent goes to $+\infty$, which is worse than not implementing the action at all. Therefore, we can put an upper bound on $(-\infty, \infty)$, which is calculated so that if payoffs reach the bound then the principal prefers no action. But then we're back in $(-\infty, \overline{v})$. $\square$

Since we have guaranteed a solution, we can now state some facts about the solution itself. Let $C_{SB}(a^*)$ be the value function of the program, i.e., the second best cost of implementing a^*. This function is well defined, and has the following features:

1. $C_{SB}(a^*)$ can be $+\infty$ for some $a^* \in A$ (if the constrained set is empty for this a^*)
2. $C_{SB}(a^*)$ is *lower semi-continuous,* which implies that it attains a minimum in the constrained set. (A function $f(\cdot)$ is Lower Semi-Continuous at x if $\liminf_{k\to\infty} f(x_k) \geq f(x)$.)
3. If $v'' < 0$ then $C_{SB}(a^*)$ has a unique minimizer.

We also have the following straightforward result:

Lemma 4.1: At the optimum IR binds.

Proof: Assume not. Then, we can reduce all v_i's by $\varepsilon > 0$ such that (IR) is still satisfied. Notice that this does not affect the incentive constraint, and thus still implements a^* at a lower cost to the principal - a contradiction. *Q.E.D.*

Notes:

1. If $U(a, I) = v(I)g(a)$ (multiplicative separability) then (IR) still binds at a solution.. (If not, scale v_i's down by some proportion $\alpha < 1$ and the same logic goes through.)
2. If $U(a, I) = g(a) + v(I)k(a)$ then (IR) may not bind and we may have the agents expected utility exceeding $\overline{u}$. (We can't use any of the above arguments.) In this case we get some "efficiency wage."

The First Stage

After finding $C_{SB}(a)$ the principal solves:

$$\max_{a \in A} B(a) - C_{SB}(a) \tag{5.2}$$

where $B(a) \equiv \sum_{i=1}^{n} \pi_i(a) q_i$ is continuous and $-C_{SB}(a)$ is upper-semi-continuous (because $C_{SB}(\cdot)$ is lower semi-continuous). Thus, since A is compact, we have a "well behaved" program. Let a^*_{SB} solve (5.2) above, we call a^*_{SB} the *second best optimal solution* where $\{I^*_i\}_{i=1}^n$ is given by the solution to the first program, (5.1) above.

Question: When is the SB solution also the FB solution?

Each one of the following is a sufficient condition (from parts of Proposition 3 in G-H):

1. $v'' = 0$ and the agent has unlimited wealth. In this case the principal and agent share the same risk-attitude, and the principal can "sell" the project (or firm) to the agent. Since the agent would then maximize

   $$\max_{a \in A} \sum_{i=1}^{n} \pi_i(a) q_i - v^{-1}(g(a) + \overline{u}) \ ,$$

   which is the expected profits less the cost of effort (and less the outside option), then the principal can ask for a price equal to the value of the agent's maximization program. This results in the principal getting the same profits as in a FB situation, and a^*_{FB} solves the agent's problem after he purchases the firm.

2. If a^*_{FB} is also the solution to $\min_{a \in A} g(a)$. In this case there is no "conflict of interest" between the principal's objectives and the agent's cost-minimizing action.

3. If A is finite, and there exists some a^*_{FB} such that for some i, $\pi_i(a^*_{FB}) = 0$ and $\pi_i(a) > 0$ for all $a \neq a^*_{FB}$. In this case it happens to be that if a^*_{FB} is chosen, then there is some outcome q_i that cannot occur, and if the agent chooses any other $a \in A$, then q_i can happen with some positive

probability. Thus, by letting $I_i = -\infty$ and $I_j = v^{-i}(g(a^*_{FB}) + \overline{u})$ for all $j \neq i$, we implement a^*_{FB} at the FB cost of $v^{-i}(g(a^*_{FB}) + \overline{u})$. This is called the case of "shifting support", since the support of the probability distribution changes with the chosen action.

In cases (1)-(3) above there is no trade-off between optimal risk-sharing and giving the agent incentives to choose a^*_{FB}. (There are 3 more cases in proposition 3 of G-H). But in general, it turns out that we will have such a trade-off, as the following result easily demonstrates, (part of proposition 3 in G-H)

Proposition 4.2: If $\pi_i(a) \geq 0$ for all i and for all $a \in A$, if $v''(\cdot) < 0$, and if $g(a^*_{FB}) > \min_{a\in A} g(a)$ for all a^*_{FB} , then the principal's profits in the SB solution are lower than in FB solution.

Proof: $C_{SB}(a^*_{FB}) > C_{FB}(a^*_{FB})$ because we need incentives ($I_i \neq I_j$ for some i, j) for the agent to choose $a^*_{FB} \neq \arg\min_{a\in A} g(a)$. Therefore, we cannot have optimal risk sharing, which implies that the solution to (5.2) is worse than the FB solution. $Q.E.D.$

5.3.1 The Form of the Incentive Scheme

Assume for simplicity that A is finite and contains m elements, $|A| = m$, so that the Lagrangian of program (5.1) (the cost-minimization program) is:

$$\begin{aligned}\mathcal{L} &= \sum_{i=1}^{n} \pi_i(a^*)h(v_i) \\ &\quad - \sum_{a_j\neq a^*} \mu_j \left[\sum_{i=1}^{n} \pi_i(a^*)v_i - g(a^*) - \sum_i \pi_i(a_j)v_i + g(a_j)\right] \\ &\quad -\lambda \left[\sum_{i=1}^{n} \pi_i(a^*)v_i - g(a^*) - \overline{u}\right]\end{aligned}$$

(Note that there are $m-1$ (IC) constraints and one (IR) constraint.) Assume that the program is convex so that the FOCs are necessary & sufficient for a solution, then we have,

$$\pi_i(a^*)h'(v_i) - \sum_{a_j\neq a^*} \mu_j[\pi_i(a^*) - \pi_i(a_j)] - \lambda\pi_i(a^*) = 0 \;\; \forall\, i = 1, ..., n$$

where $\mu_j \geq 0 \; \forall j \neq a^*$ and $\lambda \geq 0$ (where $\mu_j > 0$ implies that IC_j binds.) Since Assumption 4.1 guarantees that $\pi_i(a^*) > 0$, we can divide the FOC by $\pi_i(a^*)$ to obtain,

$$h'(v_i) = \lambda + \sum_{a_j \neq a^*} \mu_j - \sum_{a_j \neq a^*} \mu_j \frac{\pi_i(a_j)}{\pi_i(a^*)} \qquad \forall i = 1, ..., n$$

The following result is proposition 6 in G-H:

Proposition 4.3: If $a^*_{SB} \notin \arg\min_{a \in A} g(a)$ and if $v'' < 0$, then $\mu_j > 0$ for some j with $g(a_j) \leq g(a^*_{SB})$.

Proof: Assume not, i.e., the agent strictly prefers a^*_{SB} to all a_j which satisfy $g(a_j) \leq g(a^*_{SB})$. Define the set

$$A' = A \backslash \{a_j : a_j \neq a^*_{SB} \text{ and } g(a_j) \leq g(a^*_{SB})\},$$

and solve the program for $a \in A'$. Since none of the elements in $A \backslash A'$ were chosen when we solved for $a \in A$, we must still get a^*_{SB} as the solution when $a \in A'$. But observe that now $a^*_{SB} \in \arg\min_{a \in A'} g(a)$, which implies that $I_1 = I_2 = \cdots = I_n$ at the solution (full insurance). But this is the case for the FB solution, which contradicts proposition 4.2 above. $Q.E.D.$

This proposition implies that the agent will be indifferent between choosing a^* and choosing some a_j such that $g(a_j) \leq g(a^*_{SB})$. That is, will be indifferent between working "optimally" and working "less", if higher costs of effort are associated with higher levels of effort. Thus, the main point of Proposition is that we must have some "downward" binding incentive constraints.

1. The proof depends on the finiteness of A. In the infinite case this result holds only locally.
2. This result does not rule out some "upward" binding IC's which will be somewhat "problematic" as we will see shortly.

Recall that $h' > 0$, and $h'' > 0$. Thus, we have,

Corollary 4.1: If $\mu_j > 0$ for only one $a_j \neq a^*_{SB}$ then $I_i > I_k$ if and only if $\frac{\pi_i(a^*_{SB})}{\pi_i(a_j)} > \frac{\pi_k(a^*_{SB})}{\pi_k(a_j)}$. That is, I is monotonic in the likelihood ratio.

This follows from the fact that $h'(\cdot)$ is increasing ($h'' > 0$) and from the FOC:

$$h'(v_i) = \lambda + \mu_j - \mu_j \frac{\pi_i(a_j)}{\pi_i(a^*_{SB})}\,. \tag{5.3}$$

If $\frac{\pi_i(a^*_{SB})}{\pi_i(a_j)} > \frac{\pi_k(a^*_{SB})}{\pi_k(a_j)}$, then clearly $\frac{\pi_i(a_j)}{\pi_i(a^*_{SB})} < \frac{\pi_k(a_j)}{\pi_k(a^*_{SB})}$, which implies that for i, the last term in (5.3) becomes less negative compared to k. This implies that $h'(v_i) > h'(v_k)$, which implies that $v_i > v_k$ (since $h'' > 0$.)

We can relate this result to an appealing property of the probability distributions induced by the different actions as follows:

Definition 4.1: Assume that $\pi_i(a) > 0$ for all $i \in \{1, ..., n\}$ and for all $a \in A$. (This was assumption 4.1 above.) The *monotone likelihood ratio condition (MLRC)* is satisfied if $\forall\, a, a' \in A$ such that $g(a') \leq g(a)$, we have $\frac{\pi_i(a)}{\pi_i(a')}$ is nondecreasing in i.

Corollary 4.2: Assume MLRC. Then, $I_{i+1} \geq I_i\ \forall\, i = 1, ..., n-1$ if either,

1. $\mu_j > 0$ for only one $a_j \neq a^*_{SB}$ ($\Rightarrow g(a_j) < g(a^*_{SB})$ from Proposition 4.3)
2. $A = \{a_L, a_H\}$, $g(a_L) < g(a_H)$ and $a^*_{SB} = a_H$

Case (1) follows immediately from Corollary 4.1. Case (2) does as well but it is worth mentioning since this is the "simple" 2-action case. We focus on this kind of "monotonicity" since it seems realistic in the sense that higher output leads to higher payments. We are therefore interested in exploring under what assumptions this kind of result prevails.

1. The solution to the principal-agent problem seems to have a flavor of a statistical-inference problem (the MLRC result). Note, however, that this is *not* a statistical inference problem, but rather an *equilibrium model* for which we found a subgame-perfect equilibrium. In equilibrium the principal has *correct beliefs* as to what the agent chooses and does not need to infer it from the outcome.

2. MLRC $\Rightarrow$FOSD but FOSD $\not\Rightarrow$ MLRC: Recall that the distribution $\pi(a^*)$ *First Order Stochastically Dominates* the distribution $\pi(a)$ if

$$\sum_{i=1}^{k} \pi_i(a) \geq \sum_{i=1}^{k} \pi_i(a^*) \quad \forall\, k = 1, ..., n$$

i.e., lower output is more likely under a than it is under a^*.(Don't go through the following in class.)

Claim: MLRC$\Rightarrow$FOSD. That is, $\frac{\pi_i(a^*)}{\pi_i(a)}$ increases (weakly) in i implies that

$$\sum_{i=1}^{k} \pi_i(a) \geq \sum_{i=1}^{k} \pi_i(a^*) \;\; \forall\, k = 1, ..., n$$

Proof: (I) we can't have $\frac{\pi_i(a^*)}{\pi_i(a)} > 1\; \forall\, i$. To see this, assume in negation that $\frac{\pi_i(a^*)}{\pi_i(a)} > 1\; \forall\, i$. $\Rightarrow$

$$1 = \sum \pi_i(a^*) = \sum \frac{\pi_i(a^*)}{\pi_i(a)} \cdot \pi_i(a) > \sum \pi_i(a) = 1 \text{ a contradiction.}$$

(II) Let $k^* = \max\{i = 1, ..., n \mid \frac{\pi_i(a^*)}{\pi_i(a)} \leq 1\}$. Define

$$\varphi_k = \begin{cases} 0 \text{ for } k = 0 \\ \sum\limits_{i=1}^{k} \pi_i(a) - \sum\limits_{i=1}^{k} \pi_i(a^*) \text{ for } k = 1, ..., n \end{cases}$$

note that $\varphi_0 = \varphi_n = 0$, for all $k \leq k^*$ φ_k is increasing in k, and for all $k \geq k^*$ φ_k is decreasing in k. $\Rightarrow \sum\limits_{i=1}^{k} \pi_i(a) - \sum\limits_{i=1}^{k} \pi_i(a^*)$ $\geq 0 \forall k = 1, ..., n$. *Q.E.D.*

Claim: FOSD $\not\Rightarrow$MLRC.

Example: 3 outcomes: q_1, q_2, q_3, two actions: a, a^*, with probabilities

$$\begin{array}{ll} \pi_1(a) = 0.4 & \pi_1(a^*) = 0.2 \\ \pi_2(a) = 0.4 & \pi_2(a^*) = 0.6 \\ \pi_3(a) = 0.2 & \pi_3(a^*) = 0.2 \end{array}$$

Figure Here

$\pi(a^*)$ FOSD's $\pi(a)$ by taking some probability from q_1 to q_2 without changing the probability of q_3. However, MLRC is violated:

$$\frac{\pi_1(a^*)}{\pi_1(a)} = \frac{1}{2} < \frac{\pi_2(a^*)}{\pi_2(a)} = \frac{3}{2} > \frac{\pi_3(a^*)}{\pi_3(a)} = 1$$

(Again, this follows the intuition of an inference problem - trying to verify that a^* was chosen is by identifying which outcome has a higher likelihood ratio.)

3. If $\mu_j > 0$ for more than one j then MLRC is *not enough* for monotonicity. In this case the IC's can bind in different directions and this causes "trouble" in the analysis. G-H use the *Spanning Condition* as a sufficient condition for monotonicity (see G-H proposition 7.)

4. *Robustness of Monotonicity:* Without MLRC or the Spanning condition, G-H are still are able to show that some monotonicity exists:

   Proposition 5 (G-H): In the SB solution without MLRC and without the Spanning condition, when the SB solution is worse than the FB, then:

   (i) $\exists\ i,\ 1 \leq i < n$ such that $I_{i+1} > I_i$,

   (ii) $\exists\ j, 1 \leq j < n$ s.t. $q_j - I_j < q_{j+1} - I_{j+1}$

   Idea: (i) I is monotonic somewhere (that is, we can't have "perverse" incentive schemes.) (ii) "$I'(\cdot) < 1$" somewhere: There is some increase in output that induces an increase in the principal's share (again, can't have "perverse" profit sharing).

5. *Enriching the agent's action space restricts the set of incentive schemes:*

   (I) First, allowing for *free disposal*: implies that the slope of the incentive scheme must be non-negative. That is, if the agent can "destroy" output q, then we must have $I'(\cdot) \geq 1$. This implies monotonicity.

   (II) Second, allowing the agent to borrow q with no restrictions (from a third party, say a bank) implies that $I'(\cdot) \leq 1$ (or else, the agent will borrow, present a higher output to the principal, get more than his loan and repay the loan at a profit, assuming negligible interest rates for very short loans.) Therefore, the principal's share must increase in i.

So, in remark 3 we saw that $I'(\cdot) \geq 0$, and $I'(\cdot) \leq 1$ must hold somewhere, and with a realistic enrichment of the agent's action space we get these conditions holding everywhere.

6. *Random incentive schemes don't help.* This is straightforward: If $\tilde{I}_i$ is the random income that the agent faces after q_i is realized, define $\tilde{v}_i = v(\tilde{I}_i)$ and $\overline{v}_i = E\tilde{v}_i$ so that $\overline{I}_i = h(\overline{v}_i)$ is the *certainty equivalent* of $\tilde{I}_i$. Now have the principal offer $\{\overline{I}_1, ..., \overline{I}_n\}$ instead of $\{\tilde{I}_1, ..., \tilde{I}_n\}$. This contract has no effect on (IC) or (IR), and since $\overline{I}_i = h(\overline{v}_i) < Eh(\tilde{v}_i) = E\tilde{I}_i$, then the principal implements the same action at a lower cost.

5.4 The Continuous Model

Let the utilities be specified as before, but now assume that $a \in [\underline{a}, \overline{a}]$ is a continuous, one-dimensional effort choice, and assume that q is continuous with density $f(q|a)$. (That is, the density function is conditional on the choice of a.) The principal's problem is now:

$$\begin{cases} \max\limits_{a, I(\cdot)} & \displaystyle\int_q [q - I(q)] f(q|a) dq \\ \text{s.t.} & \displaystyle\int_q v(I(q)) f(q,a) dq - g(a) \geq \overline{u} \qquad \text{(IR)} \\ & a \in \arg\max \left\{ \displaystyle\int_q v(I(q)) f(q,a') dq - g(a'),\ a' \in A \right\} \qquad \text{(IC)} \end{cases}$$

This is the general (and *correct*) way of writing the problem. However, this is not a form that we can do much with. As we will see, there is a simple way of reducing this problem to a "manageable" program, but this will require some extra assumptions for the solution to be correct. We begin by analyzing the first-best benchmark.

5.4.1 First Best Benchmark: Verifiable Actions

As before, in this case we only have (IR), which will bind at a solution, so the Lagrangian is,

$$\mathcal{L} = \int_q [q - I(q) + \lambda v(I(q))] f(q|a) dq - \lambda g(a) - \lambda \overline{u}$$

Assuming interior solution, the (point-wise) FOC with respect to $I(\cdot)$ yields,

$$\frac{1}{v'(I(q))} = \lambda \ \forall q. \tag{5.4}$$

Denote by $f_a \equiv \frac{\partial f(q|a)}{\partial a}$, then the FOC with respect to a yields,

$$\int_q [q - I(q) + \lambda v(I(q))] f_a(q|a) dq = \lambda g'(a) \tag{5.5}$$

and the (IR) constraint will bind (the usual argument.)

1. The first FOC with respect to. $I(\cdot)$ is known as the Borch rule where optimal risk sharing occurs. If the principal were not risk neutral but would rather have some risk averse utility function $u(\cdot)$, then the numerator of the LHS of the FOC (5.4) above would be $u'(q - I(q)]$.
2. Notice that the Borch rule is satisfied for all q and not on average since this is what risk sharing is all about.

5.4.2 Second Best: non-observable action

The First Order Approach:

In the hidden information framework we saw that under some conditions we can replace global incentive compatibility with local incentive compatibility. The question is, can we restrict attention to local incentive compatibility in the moral hazard framework?

This is known as the *first order approach,* an approach that was popular in the 70's until Mirrlees (1975) showed that it is flawed, unless we impose additional restrictions on $f(\cdot|\cdot)$. The first order approach simplifies the problem by replacing (IC) above with the agent's FOC of his optimization problem, that is, of choosing his optimal action $a \in A$. The agents FOC condition is,

$$\int_q v(I(q))f_a(q|a)dq - g'(a) = 0 \qquad ((\text{IC}^F))$$

We proceed to solve the principal's problem subject to (IC^F) and (IR), and later we will check to see if the agents SOC is satisfied, that is, if

$$\int_q v(I(q))f_{aa}(q|a)dq - g''(a) \leq 0\,.$$

Also, we will have to check for global IC, a condition for which the FOC and SOC of the agent are neither necessary nor sufficient.

The Lagrangian of the principal's problem is,

$$\begin{aligned} \mathcal{L} &= \int_q [q - I(q)]f(q|a)dq + \lambda\left[\int v(I(q))f(q|a)dq - g(a) - \overline{u}\right] \\ &\quad +\mu\left[\int v(I(q))f_a(q|a)dq - g'(a)\right] \end{aligned}$$

and maximizing with respect to $I(\cdot)$ point-wise yields the FOC,

$$\frac{1}{v'(I(q))} = \lambda + \mu\frac{f_a(q|a)}{f(q|a)} \text{ a.e.} \qquad (5.6)$$

Notice that this looks very similar to the FOC for the case of A being finite, with only one (downward) binding IC. (This is also like the 2-action case with $a^*_{SB} = a_H$.) In the previous formulation we had $h'(\cdot) = \frac{1}{v'(\cdot)}$, and $\frac{f_a}{f}$ is "similar" to the continuous version of the likelihood ratio for small changes in a, which is $\frac{f(q|a+\delta)}{f(q|a)}$. This can be seen as follows: Notice that $\frac{f(q|a+\delta)}{f(q|a)} - 1$ is a monotonic transformation of $\frac{f(q|a+\delta)}{f(q|a)}$, and dividing this by δ is yet another monotonic transformation, which then yields,

$$\lim_{\delta\to 0}\frac{f(q|a+\delta) - f(q|a)}{\delta f(q|a)} = \frac{f_a(q|a)}{f(q|a)}\,.$$

Definition 4.2: $f(q|a)$ satisfies MLRC if $\frac{f_a(q|a)}{f(q|a)}$ increases in q.

Proposition 4.4: Assume that the first order approach is valid, that MLRC is satisfied, and that $\mu > 0$. Then $I'(q) \geq 0$.

Proof: Follows directly from the FOC (**??**) that we derived above and from $v'' < 0$: As q increases, $\mu \frac{f_a}{f}$ increases, $\Rightarrow \frac{1}{v'}$ increases, $\Rightarrow v'$ decreases, $\Rightarrow I(q)$ increases. $Q.E.D.$

Therefore, in addition to assuming MLRC, and that the first order approach is valid, we need to ask ourselves when is $\mu > 0$ guaranteed? This is answer is that if the first order approach is valid, then we must have $\mu > 0$. This is demonstrated in the following proposition (Holmstrom (1979) proposition 1):

Proposition 4.5: If the first order approach is valid, then at the optimum, $\mu > 0$.

Proof: (i) Assume in negation that $\mu < 0$. If the first order approach is valid, then from the FOC (**??**) above, we get the solution $I^*(q)$ which is decreasing in $\frac{f_a}{f}$. Define $\widehat{I} = I(q)$ for those q which satisfy $\frac{f_a(q|a)}{f(q|a)} = 0$. That is, since I is a function of q, and given a each q determines $\frac{f_a(q|a)}{f(q|a)}$, then we can think of I as a function of $\frac{f_a(q|a)}{f(q|a)}$, s shown in the following figure:

Figure Here

Now consider the first term of the agent's FOC (IC^F) above:

$$\begin{aligned} & \int_q v(I(q)) f_a(q|a) dq \\ = & \int_{\{q:\, f_a \geq 0\}} v(I(q)) f_a(q|a) dq + \int_{\{q:\, f_a < 0\}} v(I(q)) f_a(q|a) dq \\ < & \int_{\{q|fa \geq 0\}} v(\hat{I}) f_a(q|a) dq + \int_{\{q:\, f_a < 0\}} v(\hat{I}) f_a(q|a) dq \\ = & \, v(\hat{I}) \int_q f_a(q|a) dq \\ = & \, 0 \end{aligned}$$

(The last equality follows from $\int_q f(q|a)dq = 1 \ \forall a$.) But this contradicts

$$\int_q v(I(q))f_a(q|a)dq = g'(a) > 0.$$

(ii) Assume in negation that $\mu = 0$. From the FOC (**??**) above we get,

$$\frac{1}{v'(I(q))} = \lambda \ \forall q \, ,$$

which implies that the agent is not exposed to risk. This in turn implies that the agent chooses his action to minimize his cost $g(a)$, which is generally not the solution to the principal's program. $Q.E.D.$

Caveat: It turns out that we could have $\mu = 0$ and the FB is almost achieved. The following example is due to Mirrlees (1974):

Example 4.1: Assume that output is distributed according to $q = a + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2)$. That is, for two distinct actions $a_1 < a_2$, we get the distributions of q to "shift" as shown in the following figure:

Figure Here

Assume also that $a \in [\underline{a}, \overline{a}]$, and that the agents utility is given by

$$U(I, a) = \ell n I - a \, .$$

This looks like a "well behaved" problem but it turns out that the principal can achieve profits that are arbitrarily close to the FB profits: Let $a^*_{FB} \in (\underline{a}, \overline{a})$, and calculated $\overline{I}$ such that

$$\ell n \overline{I} - a^*_{FB} = \overline{u}.$$

Set $\underline{q}$ to be "very" negative, and let the incentive scheme be

$$I(q) = \begin{cases} \overline{I} > \underline{q} \\ \delta \leq \underline{q} \end{cases}$$

where δ is "very" small. Mirrlees showed that

$$\lim_{\underline{q} \to -\infty} F(\underline{q}) = \lim_{\underline{q} \to -\infty} \Pr\{q < \underline{q}\} = 0,$$

but that

$$\lim_{\underline{q}\to-\infty} F_a(\underline{q}) = \lim_{\underline{q}\to-\infty} \frac{d}{da}\left(\Pr\{q < \overline{q}\}\right) \neq 0,$$

which implies that the agent will not "slack," his expected utility is close to $\overline{u}$, and he is almost not exposed to risk (To get (IR) satisfied, we need to add ε to $\overline{I}$, which measures the departure from FB costs to the principal.) This example is a continuous approximation of the "shifting support" discrete case.

1. Again, this looks like a statistical inference problem but it is not. (Notice that $\frac{f_a}{a}$ is the derivative of the log-likelihood function, $\ell n f(q,a)$, with respect to a. This turns out to be the gradient for the Maximum Likelihood Estimator of a given q, which is the best way to infer a from the "data" q, if this *were* a statistical inference problem.)
2. Holmstrom (1979, section 6) shows that this approach is valid when the agent has private information (i.e., a type) as in the hidden-information models. The wage schedule will then be $I(q,\theta)$, where $\theta \in \Theta$ is the agent's type. Now the multiplier will be $\mu(\hat{\theta})$ and it may be negative at a solution.
3. Intuitively we may think that $a^*_{SB} < a^*_{FB}$. This is not a generic property.

Validity of the First Order Approach

Mirrlees (1974) observed that the solution to the "Relaxed program," i.e., using the first order approach, *may not be* the true SB solution to the correct unrelaxed program. The problem is more serious than the standard one, in which case the FOC's (of the principal's program, not of the agent's program) being necessary but not sufficient. It turns out that the first order approach suffers from a much deeper problem; it may be the case that when the first order approach is used, then the principal's FOC is not only insufficient, but it *may not be necessary.*

This can be seen using the following graphical interpretation of incentive schemes.

2 figures here

Given $I(\cdot)$, the agent maximizes

$$EU(a|I) = \int_q v(I(q))f(q|a)dq - g(a)\,.$$

Then, using the first order approach, we maximize the principal's profits subject to the constraint,

$$\frac{dEU(a|I(\cdot))}{da} = 0\,,$$

that is, given an incentive scheme $I(\cdot)$, the agent's FOC determines his choice of a. But notice that $EU(a|I(\cdot))$ need *not be* concave in a, which implies that there may be more than one local maximizer to the agent's problem (we ignore the problem of no solutions.) To see the problem, imagine that we can "order" the space of incentive schemes $I(\cdot)$ using the ordering "$<$"as follows: Consider a scheme $\tilde{I}$ such that for all schemes $I < \tilde{I}$, there is only one local maximizer for $Eu(a|I)$. Also consider a scheme $\hat{I}$ such that for all schemes $I > \hat{I}$ there is only one local maximizer. Assume that $\tilde{I}$ and $\hat{I}$ have one inflection point, (a point which is not a local maximizer or minimizer, but has the FOC of the agent satisfied,) so that all schemes I such that $\tilde{I} < I < \hat{I}$ have two local maximizers, and one local minimizer. In particular, consider some scheme $\overline{I}(\cdot)$, $\hat{I}(\cdot) < \overline{I}(\cdot) < I$, such that the agent is indifferent between the two local maximizers $\overline{a}_0$ and $\overline{a}_1$, but the principal prefers $\overline{a}_1$. Then, it may be optimal for principal to move to $\langle I^*(\cdot), a_1^*\rangle$ at which both the principal's *and* agent's FOC's are satisfied. But under $I^*(\cdot)$, the action a_1^* is a local maximizer, while a_0^* is the *global maximizer*, in which case $\langle I^*(\cdot), a_1^*\rangle$ is not implementable. When we solve the *true program* we maximize the principal's profits subject to the agents *true* choices, so given the graphical description, $\left\langle \overline{I}(\cdot), \overline{a}_1\right\rangle$ is the solution to the true program, and at this point the principal's FOC in the reduced first order approach program is not even satisfied.

There are two ways of dealing with this problem. The first, is to ignore it and check if the solution to the first order approach is a true solution to the SB problem, and if it is not, then the true problem needs to be solved. The second, is to find cases for which the first order approach is valid. We will explore two such cases:

Case 1: LDFC: Let $a \in A = [0, 1]$. We say that the *Linear Distribution Function Condition* (LDFC) is satisfied if there exist two density functions, $f_H(q)$, and $f_L(q)$, such that

$$f(q|a) = af_H(q) + (1-a)f_L(q)$$

(this is also called the *Spanning Condition* - see G-H 1983.) In this case

$$EU(a|I(\cdot)) = \int v(I(q))f(q|a)dq - g(a)$$
$$= a\int v(I(q))f_H(q)dq + (1-a)\int v(I(q))f_L(q)dq - g(a)$$
$$= ak_1 + (1-a)k_2 - g(a)$$

where k_1 and k_2 are constants (given the incentive scheme $I(\cdot)$,)and since $g' > 0$ and $g'' > 0$, this is a well behaved concave function which guarantees that there is no problem, and LDFC is a sufficient condition for the first order approach to work.

Case 1: MLRC + CDFC: We say that the cumulative distribution function $F(q|a)$ satisfies the Convexity of the Distribution Function Condition (CDFC) if it is convex in a; that is, for all $\lambda \in [0,1]$,

$$F(q|\lambda a + (1-\lambda)a') \leq \lambda F(q|a) + (1-\lambda)F(q|a') .$$

Recall that MLRC implies that $\frac{f_a(q|a)}{f(q|a)}$ increases in q. Take the agent's expected utility and integrate it by parts:

$$EU(a|I(\cdot)) = \int_{\underline{q}}^{\overline{q}} v(I(q))f(q|a)dq$$
$$= [v(I(q)) \cdot F(q|a)|_{\underline{q}}^{\overline{q}} - \int_{\underline{q}}^{\overline{q}} v'(I(q))I'(q)F(q|a)dq$$
$$= v(I(\overline{q})) - \int_{\underline{q}}^{\overline{q}} v'(I(q))I'(q)F(q|a)dq - g(a)$$

where the last equality follows from $F(\overline{q}|a) = 1$ and $F(\underline{q}|a) = 0$ (note that $\underline{q}$ or $\overline{q}$ need not be bounded.) Now consider the second derivative of $EU(a|I(\cdot))$ with respect to a:

$$[EU(a|I(\cdot))]'' = -\int_{\underline{q}}^{\overline{q}} v'(I(q))F_{aa}(q|a)dq - g''(a) < 0$$

which follows from $v' > 0$, $I'(q) > 0$ (which is implied by MLRC,) $F_{aa} > 0$ (which is implied by CDFC,) and finally $g'' > 0$. Thus, MLRC and CDFC together are sufficient for the first order approach to be valid.

1. Recall that in our proof that MLRC $\Rightarrow I'(\cdot) > 0$ we used the fact that $\mu > 0$. But this fact is true only if the first order approach is valid. Thus, the proof we have given (for MLRC and CDFC together to be sufficient for the first order approach to be valid) contains a "circular" mistake. See Rogerson (1985) for a complete and correct proof.
2. Jewitt (1988) provides alternative sufficient conditions for validity of the first order approach which avoids CDFC (CDFC turns out to be very restrictive), and puts conditions on $v(\cdot)$, namely that $-\frac{v''(\cdot)}{[v'(\cdot)]^3}$ is non-decreasing, that is, risk aversion does not decrease too quickly. CARA is an example that works with various "commonly used" distributions.

5.5 The Value of Information: The Sufficient-Statistic Result

Assume now that there are more signals above and beyond q. For example, let y denote some other parameter of the project (say, some intermediate measure of success) and assume that q and y are jointly distributed given the agent's action a.

Question: When should y be part of the contract in addition to q?

This question was answered independently by Holmstrom (1979) and Shavel (1979). Rewrite the Lagrangian (assuming that the first order approach is valid) by taking (q, y) to be the two-dimensional signal, and $I(q, y)$

is the incentive scheme:

$$\begin{aligned}\mathcal{L} &= \int_y \int_q [q - I(q,y)] f(q,y|a) dq dy \\ &\quad + \lambda \left[\int_y \int_q v(I(q,y)) f(q,y|a) dq dy - g(a) - \overline{u} \right] \\ &\quad + \mu \left[\int_y \int_q v(I(q,y) f_a(q,y|a) dq dy - g'(a) \right]\end{aligned}$$

and the FOC with respect to $(I\cdot,\cdot)$ is:

$$\frac{1}{v'(I(q,y))} = \lambda + \mu \frac{f_a(q,y|a)}{f(q,y|a)}$$

Now, to answer the question, we can modify it and ask, "When can we ignore y?" The answer is clearly that we can ignore y and have $I(q)$ if and only if the first order condition is independent of y, which gives us the same FOC as in our analysis without the y signal. This will be satisfied if $\frac{f_a(q,y|a)}{f(q,y|a)}$ is independent of y.

Definition 4.3: q is a *Sufficient Statistic for* (q, y) *with respect to* $a \in A$ if and only if the conditional density of (q, y) is multiplicative separable in y and a:

$$f(q,y|a) = g(q,y) \cdot h(q|a) \, .$$

We say that y is informative about a if q is not a sufficient statistic as defined above.

This is a statistical property which says that when we want to make an inference about the random variable. $\tilde{a}$, if q is a sufficient statistic as above then we can ignore y. (Again, remember that here a is *known* in equilibrium, but we have the same flavor of a statistical inference problem.) We thus have the following proposition:

Proposition 4.6: Assume that the first order approach is valid. y should be included in the incentive scheme if and only if y is informative about a.

Consider the case where q is a sufficient statistic for (q, y) as defined above. We can interpret this result as follows. Given a choice $a \in A$, we can think of q being a r.v. whose distribution is dependent on a, and then, once q is realized (but maybe not yet revealed,) then y is a r.v. whose distribution is dependent on q. This can be depicted using the following "causality" diagram,

$$a \overset{\widetilde{\varepsilon}}{\to} q \overset{\widetilde{\eta}}{\to} y \, ,$$

that is, given a, some random shock $\widetilde{\varepsilon}$ determines q, and given the realized value of q, some random shock $\widetilde{\eta}$, which is independent of a, determines y. Thus, y is a noisier signal of a compared to q, or we say that y is a *garbling* of q.

Corollary 4.3: If q is a sufficient statistic then, if the principal is restricted to contract on one signal then $I(q)$ is better than $I(y)$ for the principal.

Corollary 4.4: Random compensation schemes are not optimal (for the separable utility case.)

The second corollary follows since a random incentive scheme $\widetilde{I}(x)$ is a payment based on a r.v. y which is independent of a, making q a sufficient statistic. (If the agent's utility isn't separable in q and a, then the agent's risk attitude depends on a, and randomizations of the incentive scheme may be beneficial for the principal.)

5.6 Incentives in Teams: Group Production

We now explore the situation in which several agents together produce some output. The following model is based on section 2 in Holmstrom (1982a):

- Consider a group of n agents, each choosing an action $a_i \in A_i \subset \Re$, for $i \in \{1, 2, ..., n\}$.
- Output is given by $x(a_1, a_2, ..., a_n, \varepsilon) \in \Re$, where ε is some random noise. For now we will perform the analysis with no noise, that is, set $\varepsilon = 0$. We assume that $\frac{\partial x}{\partial a_i} > 0 \; \forall i$, and $\forall a_i$, that is, output is increasing in each agent's action (effort.) Finally, assume that $x(\cdot)$ is concave so that we can restrict attention to interior solutions.

- Agent's utilities are given by $u_i(m_i, a_i) = m_i - g_i(a_i)$ where m_i denotes monetary income, and $g_i(a_i)$ is the private cost of effort, with $g_i' > 0$, and $g_i'' > 0$. (Note that agents are risk neutral in money.) Let $\overline{u}_i$ be outside option (which determines each agent's IR).

5.6.1 First Best: The Planner's Problem

Assume that there are no incentive problems, and that a planner can force agents to choose a particular action (so we are also ignoring IR constraints.) The first best solution maximizes total surplus, and solves,

$$\max_{a_1,...,a_n} x(a_1, ..., a_n) - \sum_{i=1}^{n} g_i(a_i)$$

which yields the FOCs,

$$\frac{\partial x(a)}{\partial a_i} = g_i'(a_i) \;\; \forall\, i = 1, ..., n\,. \tag{5.7}$$

That is, the marginal benefit from agent i's action equals the marginal private cost of agent i.

5.6.2 Second Best: The "Partnership" Problem

Consider the partnership problem where the agents jointly own the output. Assume that the actions are not contractible and the agents must resort to an incentive scheme $\{s_i(x)\}_{i=1}^n$ under the restriction of a balanced budget ("split-the-pie") rule:

$$\sum_{i=1}^{n} s_i(x) = x \;\; \forall\, x\,, \tag{5.8}$$

and we also impose a "limited liability" restriction,

$$s_i(x) \geq 0 \;\; \forall\, x\,. \tag{5.9}$$

(We will also assume that the $s_i(x)$ functions are differentiable in x. This is not necessary and we will comment on this later.)

We solve the partnership problem using Nash Equilibrium (NE), and it is easy to see that given an incentive scheme $\{s_i(x)\}_{i=1}^n$,any NE must satisfy,

$$s_i'(x) \cdot \frac{\partial x}{\partial a_i} = g_i'(a_i)$$

Question: Can the Partnership achieve FB.?

For the partnership to achieve FB efficiency, we must find an incentive scheme for which the NE coincides with the FB solution. That is, from the FOC of the FB problem, (5.7) above, we must have $s_i'(x) \equiv 1 \; \forall \, x$, and $\forall \, i$, which implies that

$$\sum_{i=1}^{n} s_i'(x) = n.$$

But from (5.8) we have

$$\sum_{i=1}^{n} s_i'(x) = 1 \; \forall \, x \, ,$$

which implies that a budget balanced partnership cannot achieve FB efficiency. The intuition is standard, and is related to the "Free riding" problem common to such problems of externalities.

Question: How can we solve this? (in the deterministic case)

One solution is to violate the budget balanced constraint, (5.8) above, by letting $\sum s_i(x) < x$ for some levels of output x.

Example 4.1 Consider the following incentive scheme:

$$s_i(x) = \begin{cases} s_i^* \text{ if } x = x_{FB}^* \\ 0 \text{ if } x \neq x_{FB}^* \end{cases}$$

where s_i^* is arbitrarily chosen to satisfy: $\sum_{i=1}^{n} s_i^* = x_{FB}^*$, and $s_i^* > g_i(a_i^*) \; \forall \, i$. (This can be done if we assume that $\overline{u}_i = 0 \; \forall i$.) It is easy to see that this scheme will yield the FB as a NE for the case where $g_i(0) = 0 \; \forall \, i$, $x(0, ..., 0) = 0$, and $\sum_{i=1}^{n} g_i'(0) < \sum_{i=1}^{n} \frac{\partial x(0,...,0)}{\partial a_i}$. (these are just Inada conditions that guarantee the solution.) □

1. One problem with the scheme above is that there are multiple NE. For example, $a_i = 0 \; \forall \, i$ is also a NE given the scheme above.
2. A second problem is that this scheme is not credible (or, more precisely, not renegotiation proof.) If one agent "slacks" and $x < x^*$ is realized, then all agents have an incentive to renegotiate. That is, in the scheme above, $\sum_{i=1}^{n} s_i < 0$ is not *ex-post efficient* off the equilibrium path. Thus, with renegotiation we cannot achieve FB efficiency for partnerships.

3. As mentioned above, we do not need $s_i(x)$ to be differentiable in x. See the appendix in Holmstrom 1982a. (That is, the inability to achieve FB is more general.)

5.6.3 A solution: Budget-Breaker Principal

We now introduce a new, $(n+1)^{th}$ agent to the partnership who will play the role of the "Budget-Breaker." This is a theoretical foundation for the famous paper by Alchian and Demsetz (1972). The idea in Alchian and Demsetz is that if we introduce a monitor to the partnership problem, then this monitor (or "principal") will make sure that the agents do not free ride. The question then is, who monitors the monitor? Alchian and Demsetz argue that if the monitor has residual claims, then there will be no need to monitor him. One should note, however, that there is *no monitoring* here in Holmstrom's model. This is just another case of an "unproductive" principal who helps to solve the partnership problem.

Example 4.2: Consider a modification of Example 4.1 where we add a principal (budget breaker), denoted by $n+1$, and modify the incentive scheme as follows:

$$s_i(x) = \begin{cases} s_i^* \text{ if } x \geq x_{FB}^* \\ 0 \text{ if } x < x_{FB}^* \end{cases} \quad \text{for } i = 1, ..., n;$$

$$s_{n+1}(x) = \begin{cases} x - x_{FB}^* & \text{if } x \geq x^* \\ x & \text{if } x < x^* \end{cases}$$

In equilibrium,

$$\sum_{i=1}^{n+1} s_i(x) = x_{FB}^*,$$

and $s_{n+1}(x_{FB}^*) = 0$. □

Add Uncertainty: $\varepsilon \neq 0$

If agents are risk neutral, then it is easy to extend the previous analysis to see that a partnership cannot achieve FB efficiency in a NE. (As before, we will get under-provision of effort in any NE.) It turns out that adding a Budget-Breaker will help achieve FB efficiency as the following analysis shows. Denoted the principal by $n+1$, and consider the incentive scheme,

$$\begin{aligned} s_i(x) &= x - k \;\; \forall\, i = 1, ..., n \\ s_{n+1}(x) &= nk - (n-1)x \end{aligned}$$

where k is chosen so that

$$(n-1)\int_x x f(x, a^*_{FB})dx = nk\,,$$

which guarantees that at the FB solution, the $(n+1)^{th}$ agent breaks even in expectation. Under this scheme, each agent $i = 1, ..., n$ solves,

$$\max_{a_i} E[x(a_1, ..., a_n)] - k - g(a_i)\,,$$

which yields the FOC,

$$\frac{\partial Ex(a_1, ..., a_n)}{\partial a_i} = g_i'(a_i)\;\;,$$

which is precisely the FB solution.

The intuition is simple: This is exactly like a *Groves Mechanism.* Each agent captures the full extent of the externality since the principal "sells" the *entire* firm to each agent for the price k.

1. If the principal is risk-averse we cannot achieve FB efficiency as demonstrated above in the case with uncertainty. This follows since the principal will be exposed to some risk, and since he breaks even in the sense of expected utility then there is a reduction in total expected social surplus.

2. A problem in this setup is *collusion*: Agent i can go to the principal and say: "If I choose a_i^* you get zero, so we can split x after I choose $a_i = 0$." This is very different from the problem of renegotiation proofness in the partnership problem. For collusion to be a problem we must have "secret" side-contracts between the principal and some agent. These issues are dealt with in the collusion literature.

3. Legros-Matthew (1993) extended Holmstrom's results, and they established necessary and sufficient conditions for a Partnership (with no budget-breaker) to achieve FB. Some interesting cases in their analysis are:

   Case 1: A_i finite and $x(\cdot)$ being a generic function. In this case if only one agent deviates then it will be clear who it was, so we can achieve FB without then problem of renegotiation. For example, consider the incentive scheme (ignoring IR constraints,)

$$s_i(x) = \begin{cases} \frac{x}{n} & \text{if } x = x^* \\ \frac{1}{n-1}(F + x) & \text{if } x \neq x^* \text{ and } j \neq i \text{ deviated} \\ -F & \text{if } x \neq x^* \text{ and } i \text{ deviated} \end{cases}$$

   (In fact, it is enough to know *who didn't deviate* to do something similar.)

   Case 2: $A_i = [\underline{a_i}, \overline{a_i}] \subset \Re$ is compact, and $a^* \in (\underline{a_i}, \overline{a_i})$ (i.e., an interior FB solution). In this case we can have one agent, say i, randomize between choosing a_i^* with high probability and some other action with low probability, and all other agents will a_j^* for sure. We can now achieve actions that are arbitrarily close to the FB solution with appropriate schemes. As $\Pr\{a_i = a_i^*\} \to 1$, we approach the FB solution. Note, however, that there are two criticisms to this case: First, as $\Pr\{a_i = a_i^*\} \to 1$, we need fines for "bad" outcomes that approach $-\infty$ to support the FB choice of actions, and second, do we really think that agents randomize?

5.6.4 Relative Performance Evaluations

This is the second part of Holmstrom (1982a). The model is setup as follows:

- Risk neutral principal
- n risk averse agents, each with utility over income/action as before, $u_i(m_i, a_i) = u_i(m_i) - g_i(a_i)$, with $u' > 0$, $u'' < 0$, $g' > 0$, and $g'' > 0$.
- $y = (y_1(a), \ y_2(a), ..., y_m(a))$ is a vector of random variables (e.g., outputs, or other signals) which is dependent on the vector of actions, $a = (a_1, a_2, ..., a_n)$.

- $E(x|y,a)$ denotes the principal's expected profit given the actions and the signals. (Thus, we can think of y as signals which are interdependent through the a_i's and some noise, and x maybe part of y.)
- $G(y,a)$ denotes the distribution of y as a function of a with $g(y,a)$ being the density.

The principal's problem is therefore,

$$\begin{cases} \max\limits_{a,s_1(y),...,s_n(y)} & \int\limits_y \left[E(x|y,a) - \sum\limits_{i=1}^{n} s_i(y) \right] dG(y,a) \\ \text{s.t.} & \int\limits_y u_i(s_i(y))dG(y,a) - g_i(a_i) \geq \overline{u}_i \ \ \forall\ i \qquad \text{(IR)} \\ & a_i \in \arg\max\limits_{a_i' \in A_i} \int\limits_y u_i(s_i(y))dG(y,(a_i',a_{-i})) - g_i(a_i') \ \ \forall\ i \qquad \text{(IC)} \end{cases}$$

(Note that if x is part of y then $E(x|y,a) \equiv x$.)

Definition 4.4: A function $T_i(y)$ is a sufficient statistic for y with respect to a_i if there exist functions $h_i(\cdot) \geq 0$ and $p_i(\cdot) \geq 0$ such that,

$$g(y,a) = h_i(y,a_{-i})p_i(T_i(y),a) \ \ \forall\,(y,a) \in \text{support}(g(\cdot,\cdot))$$

The vector $T(y) = (T_1(y),...,T_n(y))$ is sufficient for y with respect to a if each $T_i(y)$ is sufficient for y with respect to a_i.

This is just an extended version of the sufficient statistic definition we saw for the case of one agent. For example, if each $T_i(y)$ is sufficient for y with respect to a_i, we can intuitively think of this situation as one where each a_i generates a random variable $T_i(y)$, and y is just a garbling of the vector of random variables, $T(y) = (T_1(y),...,T_n(y))$, or as the figure describes the process,

$$\left.\begin{matrix} a_1 \overset{noise}{\rightarrow} T_1(y) \\ \vdots \qquad \vdots \\ a_n \overset{noise}{\rightarrow} T_n(y) \end{matrix}\right\} \overset{noise}{\rightarrow} y$$

Proposition 4.7: (Holmstrom, Theorem 5) Assume $T(y) = (T_1(y), ..., T_n(y))$ is sufficient for y with respect to a. Then, given any collection of incentive schemes $\{s_i(y)\}_{i=1}^n$, there exists a set of schemes $\{\widetilde{s}_i(T_i(y))\}_{i=1}^n$, that weakly Pareto dominates $\{s_i(y)\}_{i=1}^n$

Proof: Let $\{s_i(y)\}_{i=1}^n$ implement the Nash equilibrium $(a_1, ..., a_n)$, and consider changing i's scheme from $s_i(y)$ to $\tilde{s}_i(T_i)$ as defined by:

$$u_i(\tilde{s}_i(T_i)) \equiv \int\limits_{\{y:T_i(y)=T_i\}} u_i(s_i(y))\frac{1}{p_i(T_i, a)} \cdot g(y, a)dy$$
$$= \int\limits_{\{y:T_i(y)=T_i\}} u_i(s_i(y))h_i(y, a_{-i})dy$$

By definition, (IC_i) and (IR_i) are not changed since agent i's expected utility given his choice a_i is unchanged. Also,

$$\tilde{s}_i(T_i) \leq \int\limits_{\{y:T_i(y)=T_i} s_i(y) \cdot h_i(y, a_{-i})dy\,,$$

because $u'' < 0$, and T_i is constant whereas $s_i(y)$ is random *given* T_i. Integrating over T_i :

$$\int\limits_y \tilde{s}_i(T_i(y))g(y, a)dy \leq \int\limits_y s_i(y)g(y, a)dy\,.$$

This can be done for each $i = 1, ..., n$ while setting the actions of $j \neq i$ as given (to preserve the NE solution), and by offering $\{\tilde{s}_i(T_i)\}_{i=1}^n$ the principal will implement $(a_1, ..., a_n)$ at a (weakly) lower cost. $Q.E.D.$

The intuition is the same as for single agent in Holmstrom (1979): The collection $(T_1, ..., T_n)$ gives better information than y. Thus we can think of y as a garbling (or even a mean-preserving spread) of the vector $T(y)$.

Yet again, this looks like a statistical inference problem, but it is not; it is an equilibrium problem. It turns out that the "mechanics" of optimal incentives look like the mechanics of optimal inference. This is like "reverse engineering"; we use to the statistical inference properties that a would cause on $T(y)$, and use incentive based on these properties to make sure that the agents will choose the "correct" a.

Application: Yardstick Competition

We return to the simple case of $x = y$, so that the signal is profits, but consider the restricted case in which:

$$x(a,\theta) = \sum_{i=1}^{n} x_i(a_i,\theta_i).$$

That is, total profits equal the sum of individual profits generated by each agent individually, and each agent's profits are a function of his effort and some individual noise θ_i, where θ_i, θ_j may be correlated.

Proposition 4.8: If θ_i and θ_j are independent for all $i \neq j$, and x_i is increasing in θ_i, then $\{s_i(x_i)\}_{i=1}^{n}$ is the optimal form of the incentive schemes.

Proof: Let $f_i(x_i, a_i)$ be the density of x_i given a_i. Then define:

$$g(x,a) = \prod_{j=i}^{n} f_j(x_1, a_1) \ ,$$
$$p_i(x_i, a) = f_i(x_i, a_i) \, ,$$
$$h_i(x, a_{-i}) = \prod_{j \neq i} f_j(x_j, a_j) \, ,$$

and apply Proposition 4.7. *Q.E.D.*

Now consider a different scenario: Let $x_i = a_i + \varepsilon_i + \eta$, where

$$\varepsilon_i \sim N(0, \frac{1}{\tau_i}) \ \ \forall \, i = 1, ..., n$$

is an idiosyncratic noise that is independent across the agents, and

$$\eta \sim N(0, \frac{1}{\tau_0})$$

is some common shock. Therefore, we don't have independence as in the previous paragraph. (Note, we use the notion of *precision*, which is expressed by the τ_i's. These are the inverse of variance, $\tau_i \equiv \frac{1}{\sigma_i^2}$; the more precision a signal has, the less is its variance.)

Proposition 4.9: Let

$$\alpha_i = \frac{\tau_i}{\sum_{j=1}^{n} \tau_j}\ , \quad \forall\, i = 1, ..., n\ ,$$

$$\overline{x} = \sum_{i=1}^{n} \alpha_i x_i$$

then, in this scenario, $s_i(x_i, \overline{x})$ is the optimal form of the incentive scheme.

Proof: We prove this using the sufficient statistic result. Since $x_i = a_i + \eta + \varepsilon_i$ then $\varepsilon_i = x_i - a_i - \eta$, and we can write:

$$F(\hat{x}_1, ..., \hat{x}_n, a) =$$

$$k \int_{-\infty}^{\infty} \left[\underbrace{\int_{-\infty}^{\hat{x}_1 - a_1 - \eta} e^{-\frac{1}{2}\tau_1 \varepsilon_1^2} d\varepsilon_1}_{I_1} \cdot \underbrace{\int_{-\infty}^{\hat{x}_2 - a_2 - \eta} e^{-\frac{1}{2}\tau_2 \varepsilon_2^2} d\varepsilon_2}_{I_2} ... \cdots \underbrace{\int_{-\infty}^{\hat{x}_n - a_n - \eta} e^{1\frac{1}{2}\tau_n \varepsilon_n^2} d\varepsilon_n}_{I_n} \right] e^{-\frac{1}{2}\tau_0 \eta^2} d\eta$$

each of the inner integrals, I_i, can be written as:

$$I_i = \int_{-\infty}^{\hat{x}_i} e^{-\frac{1}{2}\tau_i (x_i - a_i - \eta)^2} dx_i$$

To obtain $f(\hat{x}_1, ..., \hat{x}_n, a)$ we need to partially differentiate $F(\cdot)$ with respect to $\hat{x}_i$, for $i = 1, ..., n$ sequentially, which yields:

$$f(x, a) = \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left[\sum_{j=1}^{n} \tau_j (x_j - a_j - \eta)^2 + \tau_0 \eta^2\right]} d\eta$$

Let $\overline{\tau}_{-i} = \sum\limits_{j\neq i} \tau_j$ and $\overline{z}_{-i} = \sum\limits_{j\neq i} \frac{\tau_j}{\overline{\tau}_{-i}}(x_j - a_j)$ and note that:

$$\sum_{j=1}^{n} \tau_j(x_j - a_j - \eta)^2$$

$$= \sum_{j\neq i} \tau_j[\overbrace{(x_j - a_j - \overline{z}_{-i})}^{A_j} + \overbrace{(\overline{z}_{-i} - \eta)}^{B}]^2 + \tau_i(x_i - a_i - \eta)^2$$

$$= \sum_{j\neq i} \tau_j(x_j - a_j - \overline{z}_{-i})^2 + \sum_{j\neq i} \tau_j(\overline{z}_{-i} - \eta)^2 + \tau_i(x_i - a_i - \eta)^2 - \overbrace{\sum_{j\neq i} \tau_j 2A_j B}^{=\mathbf{0}}$$

where the last term is equal to zero because B is independent of j so, $\sum\limits_{j\neq i} \tau_j 2A_j B = 2B \sum\limits_{j\neq i} \tau_j A_j$, and it is easy to check that $\sum\limits_{j\neq i} \tau_j A_j = 0$. So we have,

$$f(x,a) = \int_{-\infty}^{\infty} e^{\frac{1}{2}[\sum\limits_{j\neq i}(x_j - a_j - \overline{z}_{-i})^2} \cdot e^{-\frac{1}{2}[\sum\limits_{j\neq i}(\overline{z}_{-i} - \eta)^2 + \tau_i(x_i - a_i - \eta)^2 + \tau_0\eta^2]} d\eta$$

$$= \underbrace{e^{\frac{1}{2}[\sum\limits_{j\neq i}(x_j - a_j - \overline{z}_{-i})^2]}}_{h(x,a_{-i})} \cdot \underbrace{\int_{-\infty}^{\infty} e^{-\frac{1}{2}[\sum\limits_{j\neq i}(\overline{z}_{-i} - \eta)^2 + \tau_i(x_i - a_i - \eta)^2 + \tau_0\eta^2]} d\eta}_{\hat{p}_i(\overline{z}_{-i}, x_i, a_i, \eta)}$$

Letting $\overline{\tau} \equiv \sum\limits_{j=1}^{n} \tau_j$, we can write,

$$\overline{z}_{-i} = \frac{1}{\overline{\tau}_{-i}} \sum_{j\neq i} \tau_j x_j - \sum_{j\neq i} \frac{\tau_j a_j}{\overline{\tau}_{-j}}$$

$$= \frac{\overline{\tau}}{\overline{\tau}_{-i} \cdot \overline{\tau}} \left[\left(\sum_{j=1}^{n} \tau_j x_j \right) - \tau_i x_i \right] - \sum_{j\neq i} \frac{\tau_j a_j}{\overline{\tau}_{-i}}$$

$$= \frac{\overline{\tau}\,\overline{x} - \tau_i x_i}{\overline{\tau}_{-i}} - \sum_{j\neq i} \frac{\tau_j a_j}{\overline{\tau}_{-i}}$$

So we can rewrite $\hat{p}_i(\overline{z}_{-i}, x_i, a_i, \eta)$ as $p_i(\overline{x}, x_i, a, \eta)$ and we can now apply the sufficient statistic result from Proposition 4.7. $Q.E.D.$

(Note: having the ε_i's with mean is not crucial, we can have $\varepsilon_i = \mu_i + \varepsilon_i'$, where $\varepsilon_i' \sim N(0, \frac{1}{\tau_i})$ which gives us any mean we wish.)

$>$From the two scenarios we saw, we can conclude that competition (via rank-order tournaments, e.g., $s(x_i, \overline{x})$ as the incentive scheme) is not useful per-se but only as a way of getting more information. With independent shocks relative performance schedules only add noise and reduce the principal's profits. Thus, if we believe that Mutual fund managers face some common shock as well as an idiosyncratic shock, then we should see "yardstick competition." In a firm, however, if two divisions have unrelated technology (no common shock), then the division managers should not have relative evaluation schemes.

1. There is a literature on tournaments, e.g., Lazear-Rosen (1981), which showed that rank-order tournaments can increase effort [also, Green-Stokey (1983), Nalebuff-Stiglitz (1983)]. However, from Holmstrom we see that only under very restrictive conditions will rank-order tournaments be *optimal* (for this, ordinal rankings need to be a sufficient statistic.)

2. In all the multi-agent literature it is assumed that the principal chooses her "preferred" Nash Equilibrium if there are several. Mookherjee RES 1984 demonstrated that there may be a NE that is better for the agents. Consider the following example: $i \in \{1, 2\}$, $a_i \in \{a_L, a_m, a_H\}$, $g_i(a_L) < g_i(a_m) < g_i(a_H)$, and $x_i = a_i + \eta$. (This is an "extreme" case of a common shock with no individual shock.) Assume that (a_m, a_m) is FB optimal choice for the principal. The principal can implement the FB with a rank-order tournament:

$$s_i = \begin{cases} u_i^{-1}(g_i(a_m) + \overline{u}) & \text{if } x_i \geq x_j \\ u_i^{-1}(g_i(a_L) + \overline{u} - \delta) & \text{if } x_i < x_j \end{cases}$$

It is easy to see that (a_m, a_m) is a NE implemented at the FB cost (agent's IR is binding). Notice, however, that (a_L, a_L) is also a NE, and the agents strictly prefer this because then,

$$u_i = g_i(a_m) + \overline{u} - g_i(a_L) > \overline{u}\,.$$

(We can overcome this problem if we can find two payments, $s_{\text{win}} > s_{\text{tie}}$, such that

$$g_i(a_m) - g_i(a_L) \leq u_i^{-1}(s_{\text{win}}) - u_i^{-1}(s_{\text{tie}}) \leq g_i(a_H) - g(a_m)$$

and we can use the scheme $s_{\text{win}} > s_{\text{tie}} > s_{\text{lose}}$.)

3. Ma (1988) suggested a way of getting around the multiple NE problem: use an *indirect mechanism* and employ subgame-perfect-implementation (can also use Nash implementation with integer-games à la Maskin):

$$\begin{aligned}\text{Stage 1}: &\ \text{players choose } (a_1, a_2) \text{ which is observable to them} \\ &\ \text{but not to the principal} \\ \text{Stage 2}: &\ \text{player 1 can "protest" (p) or "not protest" (np)} \\ \text{Stage 3}: &\ x_i\text{'s realized, } s_i\text{'s payed to agents.}\end{aligned}$$

where,

$$s_i(x_1, x_2, \text{np}) = \begin{cases} u_i^{-1}(g_i(a_m) + \overline{u} + \gamma) & \text{if } x_i > x_j \\ u_i^{-1}(g_i(a_m) + \overline{u}) & \text{if } x_i = x_j \\ u_i^{-1}(g_i(a_L) + \overline{u} - \delta) & \text{if } x_i < x_j \end{cases}$$

$$s_1(x_1, x_2, \text{p}) = s_1(x_1, x_2, \text{np}) + \overbrace{\alpha[x_2 - E(x_2|a_2 = a_m)]}^{>0 \text{ iff } a_2 = a_H}(\alpha \text{ small})$$

$$s_2(x_1, x_2, \text{p}) = u_2^{-1}(g_2(a_L) + \overline{u} - \beta),\ (\beta \text{ large})$$

So, player 1's best response (BR) to $a_2 \in \{a_L, a_m\}$ is "NP", and $a_1 = a_m$ would be the BR ex-ante. Look at Normal Form, and it is easy to see that $((a_m, \text{"}NP\text{"}), a_m)$ is the unique NE (also SPE).

figure here

Note: It is important that (a_1, a_2) are observable by (at least) agent 1. If not, we can't use the standard implementation approach and can't get FB (see Ma).

5.7 Dynamic Models

5.7.1 Long-Term Agency Relationship

In the Adverse-Selection (hidden information) models we saw that long-term commitment contracts will do better than short-term contracts (or Long-Term renegotiable contracts).

Stylized facts: Sequences of short-term contracts are common, e.g., piece-rates for laborers and sales commissions for sales representatives. That is, many schemes pay for per period performance.

Question: When will Short-Term contracts be as good as Long-Term?

This question is addressed by Fudenberg-Holmstrom-Milgrom (1990) (FHM hereafter.) We will outline a simplified version of FHM. Consider a 2-period relationship. In each period $t \in \{1, 2\}$ we have the sequence of events as described by the following time line:

Figure here

The *technology* is given by the following distributions: In period t, the agent exerts effort (action) a_t, and the output, x_t, is distributed according to $x_1 \sim F_1(x_1|a_1)$, and $x_2 \sim F_2(x_2|a_1, a_2, x_1, \sigma_1)$, where σ_1 is a signal observed by the agent. That is, second period technology can depend on all previous variables.

1. No discounting (no interest on money)
2. Agent has all the bargaining power. (Note that this will cause our program to look different with respect to individual rationality, but the essence of the moral hazard problem is unchanged, and we still need an incentive constraint for the agent. Now, however, the individual rationality constraint will be for the principal.)

Given the dynamic nature of the problem, let $a = (a_1, a_2(a_1, \sigma_1, x_1))$ be an *action plan* for the agent that has his second period action dependent on all first period observables. (These ae observables to him, not necessarily to the principal.) Let $c = (c_1(\sigma_1, x_1), c_2(a_1, \sigma_1, \sigma_2, x_1, x_2))$ be a (contingent) *consumption plan* for the agent.

Assumption A1: x_t and s_t are observable and verifiable (contractible)

(This is also A1 in FHM.) Assumption A1 implies that a payment plan (incentive scheme) will take the form $s = (s_1(x_1), s_2(x_1, x_2))$.

We allow the agent's utility function to take on the most general form,

$$U(a_1, a_2, c_1, c_2, \sigma_1, \sigma_2),$$

and the principal's utility is given by (undiscounted) profits,

$$\pi = x_1 - s_1 + x_2 - s_2\,.$$

A *Long-Term-Contract (LTC)* is a triplet, $\Delta = (a, c, s)$, where (a, c) are the agent's "suggested" plans, and s is the payment plan. The agent's expected utility, and the principal's expected profits *given* a LTC Δ are:

$$\begin{aligned} U(\Delta) &= E[U(a, c, \sigma)|a] \\ \pi(\Delta) &= E[x_1 - s_1(x_1) + x_2 - s_2(x_1, x_2)|a] \end{aligned}$$

Definition 4.5: We say that a LTC Δ is:

1. *Incentive Compatible* (IC) if $(a, c) \in \arg\max_{\hat{a},\hat{c}} E[U(\hat{a}, \hat{c}, \sigma)|\hat{a}]$
2. *Efficient* if it is (IC) and if there is no $\tilde{\Delta}$ such that $\pi(\tilde{\Delta}) \geq \pi(\Delta)$ and $U(\tilde{\Delta}) \geq U(\Delta)$ with at least one strict inequality.
3. *Sequentially Incentive Compatible* (SIC) if given any history of the first period, the continuation of Δ is IC in the second period.
4. *Sequentially Efficient* if it is (SIC), and, given any history of the first period, the continuation of Δ is efficient in the second period.

(*Note:* (3) and (4) in the definition above need to be formally defined with continuation utilities and profits. This is done in FHM, but since the idea is quite clear we will skip the formalities.)

We will now set up a series of assumptions that will guarantee that ST contracts (to be defined) will do as well as LTC's: (The numbering of the assumptions are as in FHM).

Assumption A3: The agent and the principal *have equal access to banks* between periods at the competitive market rate ($\delta = 1$).

Implication: the agent has a budget constraint:

$$c_1 + c_2 = s_1(x_1) + s_2(x_1, x_2).$$

Assumption A4: At the beginning of each period t, there is *common knowledge of technology.*

Implication: $F(x_2|a_1, a_2, x_1, \sigma_1) = F(x_2|a_2, x_1)$, or, (x_1, a_2) is a sufficient statistic for $(x_1, a_2, a_1, \sigma_1)$ with respect to x_2. That is, information provided by x_1 is sufficient to determine how a_2 affects x_2. A simple example of this assumption is the common time-separable case: $F_2(x_2|a_2)$, which implies that the periods are independent in technology.

Assumption A5: At the beginning of each period there is *common knowledge of the agent's preferences* over the continuation plans of any (a, c, s).

Implications:

1. $U(a, c, \sigma)$ cannot depend on σ (that is, there is no hidden information, or "types", in period 2.)
2. The continuation plan at $t = 2$ cannot depend on a_1.
3. The continuation plan at $t = 2$ cannot depend on c_1, if c_1 is not observable to the principal.

For simplicity we will assume that the agents utility function is given by (this satisfies A5):

$$U(\cdot, \cdot) = v_1(c_1) - g_1(a_1) + v_2(c_2) - g_2(a_2)$$

with the standard signs of the derivatives, $v_t^{'} > 0$, $v_t^{''} < 0$, $g_t^{'} > 0$, and $g_t^{''} > 0$. Thus, from now on we can ignore σ as if it did not exist (to satisfy A4 and A5.)

The importance of A4 and A5 is that they guarantee no "adverse selection" at the negotiation stage in period $t = 2$. (That is, at the re-contracting stage of any LTC.) We will later see examples of violations of these two assumptions, and the problems that are caused by these violations.

Given any continuation of a SIC contract, let $UPS(a_t, x_t)$ denote the *utility possibility set* given the history (a_t, x_t). (For a_0, x_0 this is not history dependent since there is no history before $t = 1$.) Let $\pi = UPF(u|a_1, x_1)$ denote the *frontier* of the UPS.

Assumption A6: For every (a_t, x_t), the function $UPF(u|a_t, x_t)$ is strictly decreasing in u.

Implication: The full set of incentives can be provided by efficient contracts (If the frontier were not strictly decreasing, we cannot keep the agent's utility fixed at u' and move to an efficient point.)

2Figures here

We now turn to the analysis of LTC's and STC's.

Definition 4.6: We say that a LTC Δ is *optimal* if it solves,

$$\begin{cases} \max\limits_{s,a,c} & E_{x_1,x_2}[v_1(c_1(x_1)) - g_1(a_1) + v_2(c_2(x_1,x_2)) - g_2(a_2)|a] \\ \text{s.t.} & (a,c) \in \arg\max\limits_{a_1,a_2,c_1,c_2} E_{x_1,x_2}[v_1(c_1)) - g_1(a_1) + v_2(c_2) - g_2(a_2)|a] \quad (\text{IC}_A) \\ & s_1(x_1) + s_2(x_1,x_2) = c_1(x_1) + c_2(x_1,x_2) \quad (\text{BC}_A) \\ & E_{x_1 x_2}[x_1 - s_1(x_1) + x_2 - s_2(x_1,x_2)|a] = 0 \quad (\text{IR}_P) \end{cases}$$

Notes:

1. (IR_P) must bind because of A6 (downward sloping UPF)
2. A6 and (IR_P) binding imply that if the solution is optimal then it must be efficient.
3. We restrict attention to choices as functions only of the x_t's, since we ignore the σ's to satisfy A4 and A5.

Short-Term Contracts

A sequence of *short-term contracts* (STC) will specify one contract, $\Delta_1 = (a_1, c_1(x_1), s_1(x_1))$, at $t = 1$, and given a *realization* of x_1, the parties will specify a second contract, $\Delta_2 = (a_2, c_2(x_2), s_2(x_2))$, at $t = 2$.

Fact: Since Δ_2 depends on the realization of x_1, if the parties are rational, and A4-A5 are satisfied, then parties can *foresee the contract* Δ_2 for every realization of x_1.

This fact is straightforward, and it implies that for convenience, we can think of Δ as a complete contingent plan, $\Delta = (\Delta_1, \Delta_2)$, where,

$$\begin{aligned} \Delta_1 &= (a_1, c_1(x_1), s_1(x_1)) \,, \\ \Delta_2 &= (a_2(x_1), c_2(x_1, x_2), s_2(x_1, x_2)) \,. \end{aligned}$$

Question: What is the difference between such a complete contingent plan and a LTC?

The answer is that in a LTC, the agent *can commit* to Δ_2 ex-ante, whereas with STC's this is impossible. However, players can foresee Δ_2 that will arise in a sequence of STC's.

To solve for the *optimal sequence* of STC's we will work backward (i.e., use dynamic programming.) At $t = 2$, for every (x_1, s_1, c_1) the agent finds the optimal Δ_2^* by solving:

$$\begin{cases} \max\limits_{\{a_2(x_1), c_2(x_1,\cdot) s_2(x_1,\cdot)\}} & E_{x_2}[v_2(c_2(x_1, x_2)) - g_2(a_2(x_1))|a_2(x_1)] \\ \text{s.t.} & a_2(x_1), c_2(x_1, \cdot) \in \arg\max\limits_{a_2, c_2} E_{x_2}[v_2(c_2) - g_2(a_2))|a_2] \quad (\text{IC}_A^2) \\ & s_1 + s_2(x_1, \cdot) = c_1 + c_2(x_1, \cdot) \quad (\text{BC}_A) \\ & E_{x_2}[x_2 - s_2(x_1, x_2)|a_2(x_1)] = 0 \quad (\text{IR}_P^2) \end{cases}$$

Since the agent has perfect foresight at date $t = 1$, then anticipating $c_2(x_1, \cdot), a_2(x_1, \cdot)$ correctly he finds the optimal Δ_1^* by solving:

$$\begin{cases} \max\limits_{\Delta=(s,c,a)} & E_{x_1,x_2}[v_1(c_1(x_1)) - g_1(a_1) + v_2(c_2(x_1, x_2)) - g_2(a_2(x_1))|a] \\ \text{s.t.} & (a_1, c_1(x_1)) \in \arg\max\limits_{a_1, c_1} E_{x_1 x_2}[v_1(c_1) - g_1(a_1) + v_2(c_2(x_1, x_2)) - g_2(a_2(x_1))|a] \quad (\text{IC}_A^1) \\ & E_{x_1}[x_1 - s_1(x_1)|a_1] = 0 \quad (\text{IR}_p^1) \end{cases}$$

Notes:

1. The agent's budget constraint, (BC_A), is only relevant at $t = 2$.

2. (IR_P^1) depends only on $x_1, s_1(x_1)$, since perfect foresight implies that $E[\pi_2] = 0$. (This need not hold for LTC's since the principal must break even over the whole relationship, and may have positive expected profits in one period for some histories, and negative expected profits in the other cases.)

3. Expectations about Δ_2^* are correct.

Proposition 4.10: (Theorem 2 in FHM) Under (A1) [verifiability], (A4) and (A5) [common knowledge of technology and preferences] and (A6) [decreasing UPF] any *efficient* LTC can be replaced by a *sequentially efficient* LTC which provides the same initial expected utility and profit levels.

Proof: Suppose Δ is an efficient LTC that is not sequentially efficient. $\Rightarrow \exists\ \hat{x}_1$ such that the continuation $a_2(\hat{x}_1), c_2(\hat{x}_1, x_2), s_2(\hat{x}_1, x_2)$ is not efficient. $\Rightarrow \exists\ \Delta_2'$ (a different continuation contract) which Pareto dominates Δ_2 after $\hat{x}_1$. Now, (A6) $\Rightarrow \exists\ \Delta_2'$ that gives the same expected continuation utility to the agent with a higher expected profit to the principal.

Figure Here

Construct a new LTC $\widetilde{\Delta}$ s.t. $\widetilde{\Delta}_1 = \Delta_1$, and,

$$\tilde{\Delta}_2(x_1) = \begin{cases} \Delta_2'(x_1) & \text{if } x_1 = \hat{x}_1 \\ \Delta_2(x_1) & \text{if } x_1 \neq x_1 \end{cases}$$

i.e., the same as Δ, but with continuation Δ' after $\hat{x}_1$. $\Rightarrow$ the agent's continuation utilities are unchanged, $\Rightarrow$(IC_A) is preserved, and this is common knowledge from (A4) and (A5). The principal strictly prefers the new continuation contract so (IR_P) is preserved since the principal is ex-ante weakly better off. (If she were strictly better off then Δ could not have been efficient.) So, $\tilde{\Delta}$ is sequentially efficient and gives the same U_0, π_0. *Q.E.D.*

Intuition: Just like complete Arrow-Debreu contingent markets.

Note: The reason the principal is only weakly better off (i.e., indifferent) is that the efficient LTC can be not sequetially efficient only for zero-probability histories. (e.g., continuous output space with a finite number of such histories, or histories that are off the equilibrium path.)

Question: What is the difference between an optimal sequentially efficient LTC and a series of optimal STC's?

In the series of optimal STC's the principal's rents are zero in *every period,* regardless of the history, whereas along the "play" of an optimal sequentially efficient LTC the principal may have different expected continuation profits for different histories at times $t \neq 1$.

Definition 4.7: A sequentially efficient LTC which gives the principal zero expected profits conditional on any history is called *sequentially optimal.*

Observation: If Δ is a sequentially optimal LTC, then its per-period continuation contracts constitute a sequence of optimal STC's.

Proposition 4.11: (Theorem 3 in FHM) Under the assumptions of Proposition 4.10, plus assumption A3 (equal access to banking,) then any optimal LTC can be replaced by a sequence of optimal STC's.

Proof: From Proposition 4.10 we know that there exists a sequentially efficient LTC that replaces the optimal LTC, so we are left to construct a sequentially efficient LTC with zero-expected profits for the principal at time 2, for any history of the first period. Given a sequentially efficient contract Δ, define:

$$\pi_2(x_1) = E_{x_2}[x_2 - s_2(x_1, x_2)|a_2(x_2)]$$

and assume that for some histories $\pi_2(x_1) \neq 0$ (if $\pi_2(x_1) = 0$ for all x_1, then we are done.) Using A3, we can define a new contract $\widehat{\Delta}$, such that

$$\begin{aligned} \hat{s}_1(x_1) &= s_1(x_1) - \pi_2(x_1), \\ \hat{s}_2(x_1, x_2) &= s_2(x_1, x_2) + \pi_2(x_1)\ \forall\,(x_1, x_2)\,, \end{aligned}$$

and leave $c_t(\cdot)$'s and $a_2(\cdot)$ unchanged. By construction we have the following four conditions:

1. (BC_A) is satisfied (both for the LTC $\widehat{\Delta}$ at $t = 1$, and for the resulting STC's at $t = 2$.)
2. For all x_1 :

$$\begin{aligned} E_{x_2}[x_2 - \hat{s}_2(x_1, x_2)|a_2(x_1)] &= E_{x_2}[x_2 - s_1(x_1, x_2) - \pi_2(x_1)|a_2(x_1)] \\ &= \pi_2(x_1) - \pi_2(x_1) \\ &= 0. \end{aligned}$$

3. The agent's incentives are unchanged since the $c_t(\cdot)$'s are unchanged. (And, they need not be changed since (BC_A) is satisfied.)

4. ex ante we have,

$$\begin{aligned} E_{x_1}[x_1 - \hat{s}_1(x_1)|a_1] &= E_{x_1}[x_1 - s_1(x_1) + \pi_2(x_1)|a_1] \\ = E_{x_1 x_2}[x_1 - x_1(x_1) + x_2 - &s_2(x_1, x_2)|a_1, a_2(x_1)] \\ &= 0 \end{aligned}$$

   where the last equality follows from Δ being an optimal sequentially efficient LTC.

But notice that (1)-(4) above imply that $\widehat{\Delta}$ is sequentially optimal. *Q.E.D.*

Violating the Assumptions:

Case 1: Consumption not observable. This violates A5, that is, there is now no common knowledge of the agent's preferences at $t = 2$. This will cause a violation of Proposition 4.11, and an example is given in FHM, example 2.

Case 2: Agent cannot access bank. This is the case in Rogerson (1985). Restrict $s_t \equiv c_t$, so that the agent cannot borrow or save. In this case the agent would optimally like to "smooth" his consumption across time, so the principal is performing two tasks: First, she is giving the agent incentives in each period, and second, she is acting as a "bank" to smooth the agent's consumption. Rogerson looks at a stationary model and shows that under these assumptions optimal LTC's have $s_2(x_1, x_2)$ and not $s_2(x_2)$, that is, memory "matters". The intuition goes as follows: If for a larger x_1 the principal wants to give larger compensation, then both $s_1(x_1)$ and $s_2(x_1, \cdot)$ should rise. We need LTC's to *commit* to this incentives scheme, because with STC's the principal cannot commit to increase $s_2(\cdot)$ when x_1 is larger. Rogerson also looks at consumption paths given different conditions on the agent's utility function.

Case 3: No common knowledge of tecnology. That is,

$$x_2 \sim F_2(x_2|x_1, a_1, a_2)$$

in which case the agent's action in the first period affects the second period's technology. Consider the simple case where $x_1 = a_2 \equiv 0$, and $a_1 \in \{a_L, a_H\}$ with $g(a_H) > g(a_L)$.

Figure Here

Suppose that $a_1 = a_H$ is the optimal second-best choice the principal wants to implement, and at the time of negotiation 2 (or *renegotiation,*) the principal does not observe a_1. The optimal LTC is one in which $s(x_2)$ has the agent exposed to some risk, so that he has incentives to choose a_H. Note, however, that the optimal sequence of STC's is as follows: At $t = 1$, s_1 is a constant (there is nothing to condition on). At $t = 2$, s_2 must be a constant as well, which follows from the fact that the effort was *already taken,* and efficiency dictates that the principal should bear all the risk (*ex-post efficient* renegotiation.) This implies that if renegotiation is possible *after* effort has been exerted, but *before* outcome has been realized, then the *only ex-post credible* (RNP) contract has $s(x)$ being a constant, and no incentives are provided. This is exactly what Fudenberg and Tirole (1990) analyze. The optimal LTC of the standard second best scenario is not sequentially efficient, or, is not RNP. This implies that with renegotiation we cannot have the agent choosing a_H with probability 1. The solution is as follows: The agent chooses a_H with probability $p_H < 1$, and a_L with probability $1 - p_H$. At the renegotiation stage the principal plays the role of a monopolistic insurer (à la Stiglitz), and a menu of contracts is offered so that it is RNP ex-post (at the stage of negotiation 2 in the figure above). This is demonstrated by the following figure:

Figure Here

Other Results on Renegotiation in Agency

Hermalin and Katz (1991) look at following case (and more...):

Figure Here

The optimal LTC without renegotiation is the standard second-best $s(x)$. Any sequence of STC's has the same problem as Fudenberg and Tirole (1990). Hermalin and Katz consider a combination of ex-ante LTC in which renegotiation occurs *on* the equilibrium path. In the case where the agent's action is observable and verifiable (in contrast to some signal, y, of the action a) then FB is achieved with LTC and renegotiation. The procedure goes as follows:

1. At $t = 0$ the principal offers the agent a risky incentive scheme, $s(x)$, that implements a^*_{FB}. That is, the solution to the first stage in the Grossman-Hart decomposed process, which is *not* the SB standard contract but rather the lowest cost contract to implement a^*_{FB}.

2. Given *any* choice $a \in A$ that the agent actually chose, at $t = 2$ the principal offers the agent a constant $\hat{s}$ that is the certainty-equivalent of $s(x)$. That is, she offers $\hat{s} = E[v(s(x)|a]$, which implies that the agent's continuation utility is unchanged, so choosing a^*_{FB} after $t = 0$ is still optimal, and there is no ex-post risk. Thus, the FB is achieved.

Hermalin and Katz (1991) also looks at a signal $y \neq a$ being observable and not verifiable. If y is a sufficient statistic for x with respect to a, then we can implement $s(y)$ in the same fashion and improve upon $s(x)$.

Question: If renegotiation outside the contract helps, does this mean that the RNP principal is not applicable here?

The answer is no. We can think of renegotiation as a bargaining game where the principal makes take-it-or-leave-it offers. Then, this process can be written into the contract as follows: The principal offers $s(x)$ and the agent accepts/rejects, then the principal offers $\hat{s}$ and agent accepts/rejects. Thus, we put renegotiation into the contract. (Note that we can alternatively have *message-game* that replicates the process of renegotiation: after $s(x)$ has been offered, and a has been chosen, both the principal and agent announce $\hat{a}$, and if their announcements coincide then $\hat{s}$ is awarded. If the announcements are different, both parties are penalized.)

Other papers: **Ma** (1991)

Figure Here

can get $a_1 = a_H$ with probability 1, but this may not be optimal...

Segal-Tadelis (1996):

Figure Here

σ_1 is a signal that is always observed by agent, while the principal may choose to observe it at no cost, or commit not to observe it. In the optimal contract the principal may choose not to observe the signal even if x_2 is not a sufficient statistic. (Intuition: create endogenous asymmetric information at the renegotiation stage.)

Matthews (1995): Looks at renegotiation where the agent has all the bargaining power and the action is unobservable. Using a forward induction argument (note that agent makes the renegotiation offers, and we are thus in a signalling game) the unique RNP contract is one where the principal "sells" the firm to the agent.

5.7.2 Agency with lots of repetition

One might expect that as a relationship is repeated more often, then there is room to achieve efficiency. Indeed, Radner (1981) and Rubinstein & Yaari (....) show that if an agency relationship is repeated for a long time then we can arbitrarily approach the FB as $T \to \infty$. The idea goes as follows: Consider the agent's utility as:

$$U = \sum_{t=1}^{T} \delta^t [v(s_t(x^t)) - g(a_t)]$$

where $x^t = (x_1, x_2, ..., x_t)$, and assume that the model is stationary with i.i.d. shocks (this is the formulation in Radner(1981)). Then, as $T \to \infty$ and $\delta \to 1$ the principal can use a "counting scheme," i.e., count the outcomes and pay "well" if the distribution over x is commensurate with a^*_{FB}, and punish the agent otherwise. (The intuition is indeed simple, but the proof is quite complicated.)

1. From Rogerson (1985) we know that if the agent has access to a bank (lending and borrowing) then the FB is attainable *without* a principal; the agent can insure himself across time as follows: choose a^*_{FB} in each period, consume x^* in each period, lend (or borrow if negative) $\ell_t = x_t - x^*$ in each period, and

$$\lim_{T\to\infty} \frac{\ell_1 + \ell_2 + ... + \ell_T}{T} = 0 \text{ a.s.}$$

(We need to assure no bankruptcy; Yaari (....) deals with this issue.)

2. We need to ask ourselves if this kind of repetition, and the underlying conditions constitute a leading case? How interesting is this result (i.e., FB achieved with or without a principal.) The answer seems to be that these conditions are rather hard to justify which puts these results in some doubt.

5.7.3 Complex Environments and Simple Contracts

What have we learned so far from agency theory with moral hazard? We can roughly summarize it in two points: First, agency trades off risk with incentives. This is a "nice" result since it allows us to understand better what these relationships entail in terms of these trade-offs.. Second, we learned that the form of an optimal SB contract can be *anything.* For example, even a simple real-world observation like monotonicity is far from general. This lesson, that the optimal SB contract may look very strange, is less attractive as a result given its real world implications. Furthermore, in a very simple example which seems to be "well behaved", Mirrlees has shown that there is no SB solution (recall from Example 4.1 above.)

If we try and perform a *reality check,* it is quite clear that strange contracts (like that in Mirrlees's example) are not observed, whereas simple linear contracts seem to be common, where the linearity is applied to some aggregate measure of output. For example,

1. Piece rate per week (not per hour);

2. Sales commission per week/month/quarter;

3. Stocks or options with clauses like "can't sell before date t", which is similar to holding some percentage of the firm's value at date t.

Holmstrom and Milgrom (1987) (H-M hereafter) make the point that "strange" and complicated contracts are due to the simplistic nature of the models. In reality, things are much more complex, and the agent has room for manipulation. Thus, we would like to have incentive schemes that are robust to these complexities. As H-M show, it turns out that linear incentive schemes with respect to aggregates will be optimal in a complex environment that has some realistic flavors to it, and are robust to small changes in the environment.

Simplified version of Holmstrom and Milgrom (1987)

Consider a principal-agent relationship that lasts for T periods:

Figure Here

We make the following assumptions:

1. Let $x_t \sim f_t(x_t|a_t)$, which implies common knowledge of technology.

2. At time t, the agent observes past outcomes $(x_1, ..., x_{t-1})$, which implies that he can "adjust" effort a_t, so that the choice of effort is a history dependent strategy:

$$a_1, a_2(x_1), ..., a_t(x_1, x_2, ..., x_{t-1}), a_T(x_1, ..., x_{T-1})$$

3. Assume zero interest rate (no discounting). This implies that the agent cares only about total payment, $s(x_1, x_2, ..., x_T)$, since he can consume at the end of the relationship.

4. Utilities are given as follows: For the principal, profits (utility) are:

$$\pi = \sum_{t=1}^{T} x_t - s(x_1, ..., x_T)\,,$$

and for the agent, utility is given by:

$$U = -e^{-r[w+s-\sum_{t=1}^{T} c(a_t)]}\,,$$

where $c(a_t)$ denotes the cost of effort level a_t,and w denotes the agent's initial wealth. (Note that this is a Constant Absolute Risk Aversion (CARA) utility function, which implies that both preferences and and the optimal compensation are independent of w. Therefore, w.l.o.g. assume that $w = 0$.)

5. Agent has all the bargaining power. This is not how H-M '87 proceed (they have principal with all the bargaining power) but here we will use the F-H-M '90 results to simplify the analysis.

6. For simplicity, let $x_t \in \{0, 1\}$ for all t (e.g., "sale" or "no sale").

The following proposition is analogous to Theorem 5 in H-M.

Proposition 4.12: The optimal LTC takes on the form:

$$s(x_1, ..., x_T) = \alpha \sum_{t=1}^{T} x_t + \beta.$$

i.e., compensation is linear in aggregate output.

Proof: The assumptions of F-H-M (1990) are satisfied: (A1) x's are verifiable; (A3) Same access to bank with 0 interest rate; (A4) $x_t \sim f_t(x_t|a_t)$ implies common knowledge of technology; (A5) in each period t the agent's utility is:

$$-e^{-r[-\sum_{\tau=1}^{t-1} c(a_\tau)]} \cdot e^{-r[s-\sum_{\tau=t}^{T} c(a_\tau)]},$$

where the first term is a constant that is unknown to the principal, and the second term is known to the principal. Thus, the agent's preferences are common knowledge; (A6) we clearly have a downward sloping UPF. Now, using Proposition 4.11 (Theorem 3 in F-H-M), an optimal LTC can be implemented with a sequence of STC's:

$$\{s_1(x_1), s_2(x_1, x_2), ..., s_T(x_1, ..., x_T)\},$$

and we can just define $s \equiv \sum_{t=1}^{T} s_t(\cdot)$. To find the optimal sequence of STC's we solve backward: At $t = T$:

$$U = -\underbrace{e^{-r\left[\sum_{t=1}^{T-1}(s_t(\cdot)-c_t(a_t))\right]}}_{\text{constant}} \cdot e^{-r[s_T(\cdot)-c_T(a_T)]}$$

Thus, what happened in periods $t = 1, ..., T-1$ is irrelevant, and the agent solves the one-shot program:

$$\begin{cases} \max\limits_{s_T(\cdot), a_T} & E_{x_T}[-e^{-r[s_T(\cdot)-c(a_T)]}|a_T] \\ \text{s.t.} & a_T \in \arg\max\limits_{a} E_{xT}[-e^{-r(s_T(\cdot)-c(a)]}|a] \\ & E[x_T - s_T(\cdot)|a_T] = 0 \end{cases}$$

and the solution is: $\langle a^*, s^*(x_T = 0), s^*(x_T = 1)\rangle$. Now move back to period $T-1$, in which the agent's utility is given by

$$U = -e^{-r[\sum_{t=1}^{T-2}(s_t - c(a_t))]} \cdot e^{-r[s_{T-1}-c(a_{T-1})]} \cdot e^{-r[s_T - c(a_T)]} .$$

When we solve the program the first term is constant, and being forward-looking the agent maximizes:

$$\max_{s_{T-1},a_{T-1}} E_{x_{T-1},x_T}[-e^{-r[s_{T-1}-c(a_{T-1})]} \cdot e^{-r[s^*(x_T)-c(a^*)]} | a_{T-1}, a_T = a^*] .$$

But since period T does not depend on the history (due to the sequence of STC's), the objective function can be rewritten as:

$$\max_{s_{T-1},a_{T-1}} E_{x_{T-1}}[-e^{-r[s_{T-1}-c(a_{T-1})]} | a_{T-1}] \cdot \underbrace{E_{x_T}[e^{-r[s^*(x_T)-c(a^*)]} | a^*]}_{\text{constant}} ,$$

and again the agent solves a one-shot program, which is the same as for T. Thus, $a_{T-1} = a^*$, $s_{T-1} = s^*(x_{T-1})$. This procedure repeats itself for all t and we get:

$$\begin{aligned} s(x_1, ..., x_T) &= \sum_{t=1}^{T} s_t(x_t) \\ &= \#\{t : x_t = 1\} \cdot s^*(1) + \#\{t : x_t = 0\} \cdot s^*(0) \\ &= (\sum_{t=1}^{T} x_t) \cdot s^*(1) + (T - \sum_{t=1}^{T} x_t) s^*(0) \\ &= \underbrace{[s^*(1) - s^*(0)]}_{\alpha} \sum_{t=1}^{T} x_t + \underbrace{T s^*(0)}_{\beta} \end{aligned}$$

Q.E.D.

Intuition of Linearity: Due to the CARA utility function of the agent, and the separability, each period is like the "one shot" single period problem and we have an optimal "slope" of the one-shot incentive scheme, $s^*(1) - s^*(0)$, which is constant and will give the agent incentives to choose a^* which is time/history independent.

1. It turns out that the principal need not observe $(x_1, ..., x_T)$, but it is enough for her to observe $\sum_{t=1}^{T} x_t$ (and of course, this needs to be verifiable). In Holmstrom-Milgrom this is the assumption. (That is, the principal only observes aggregates, and these are verifiable.) Two common examples are first, a **Salesman,** whose compensation only depends on total sales (not when they occurred over the measurement period), and second, a **Laborer**, whose piece-rate depends on the number of items produced per day/week, etc. (not on when they were produced during the time period.)

2. With a non-linear scheme the following will occur. Since the agent observes the past performance (and the principal does not) then he can "calibrate" his next effort level:

   Figure Here

   If the past performance up to a certain time is relatively low, then the agent will work harder to compensate for that past. Similarly, if output is relatively high, then the agent will work less. However, this is not optimal since a^* in each period is optimal. (This is true assuming a concave $s(\cdot)$ as in the figure above. If it were convex then the reverse will happen.)

Question: What happens of we *fix time,* and let $T \to \infty$?

The idea goes as follows: Take a *fixed time horizon* and increase the number of periods while making the length of each period shorter, thus keeping total time fixed.

Fact: From the Central Limit Theorem, the average of many independent random variables will be normally distributed.

Notice, that by doing the above exercise, we are practically taking the average of more and more i.i.d. random variables. (This intuitively follows from the fact that each carries less "weight" since the time period for each is shorter.)

Question: If for the situation described above, $\lim_{T\to\infty} \sum_{t=1}^{T} x_t$ is normally distributed, are we in the "Mirrlees example" case? i.e., can we approximate the FB?

The answer is "yes" if the agent *cannot* observe $x_1, ..., x_T$ but only $\sum x_t$. If, however, the agent observes all the x_t's the theorem we proved implies a linear scheme and the answer is "no." Thus, we can conclude that letting the agent do "more" (namely, observe outcome and calibrate effort) gives us a simple scheme, and a well defined SB solution.

The Brownian-Motion Approximation

Consider our simple model, $x_t \in \{0, 1\}$, and $a_t \in A$ affects the probability distribution of x_t. Then, fix the time length of the relationship to the interval [0,1], and let $T \to \infty$, with the length of each period, $\frac{1}{T}$, going to zero. In the limit we will get a *Brownian Motion* (one dimensional) $\{x(t),\, t \in [0, 1]\}$, with

$$dx(t) = \mu(t)dt + \sigma dB(t)\,,$$

where $\mu(t)$ is the drift rate, σ is the standard deviation, and $B(t)$ is the *standard Brownian motion* (zero drift, unit variance) and $x(t)$ is the "sum" of all the changes up until time t.

Fact: If $\mu(t) = \mu$ for all t, then $x(1) \sim N(\mu, \sigma^2)$.

That is, if the drift is constant and equal to μ, then the value of x when $t = 1$ will be normally distributed with mean μ and variance σ^2. So, we can think of this as the continuous approximation of our earlier exercise, and if agent *does not* observe $x(t)$, for all $t < 1$ then we are "stuck" in a Mirrlees case of no SB solution. But, if $x(t)$ is observed by the agent then the Holmstrom-Milgrom results hold, and we get a nice linear scheme in $x(1)$, since this normal distribution is the result of a dynamic stochastic process. (The Mirrlees problem arises only in the static contract.)

Analysis of the Brownian-Motion Approximation

We assume that the Brownian Motion model is described as follows:

- The agent controls μ (the drift) by choice of $a \in A$ where "a" stands for the constant choice over the time length [0,1] at total cost $c(a)$. This implies that $x(1) \sim N(a, \sigma^2)$. From now on we will consider $x \equiv x(1)$.

- The principal offers the contracts (In this case the agent has CARA utility, so that having the principal get all the surplus does not change the optimal SB choice a^*; only the division of surplus is changed.)

First Best: The agent will be fully insured, and the principal wants to maximize the mean of x subject to the agent's (IR). Assuming that $\overline{u} = 0$ for the agent, the objective function is,

$$\max_a a - c(a)$$

and the FOC is, $c'(a) = 1$.

Second Best: We know from our previous analysis that the optimal scheme is a linear scheme: $s(x) = \alpha x + \beta$, where $x \sim N(a, \sigma^2)$. The agent maximizes:

$$\max_a \; Ex[e^{-r(\alpha x+\beta-c(a))}|a]$$

We can simplify this by having the agent maximize his *certainty equivalent* instead of maximizing expected utility, that is, maximize

$$CE = \underbrace{\alpha a + \beta}_{\text{mean}} - \underbrace{\frac{r}{2}\alpha^2\sigma^2}_{\text{risk permium}} - c(a)$$

(Note: CARA $\Rightarrow$ the risk premium is independent of the agent's income.)

Assume that $c'(\cdot) > 0$, and $c''(\cdot) > 0$, so that the first-order approach is valid, and since the agent maximizes,

$$\max_a \; \alpha a + \beta - \frac{r}{2}\alpha^2\sigma^2 - c(a)\,,$$

and the FOC is

$$c'(a) = \alpha\,.$$

The principal will set the agent's (IR) to bind, i.e., $CE = 0$:

$$\alpha a + \beta - \frac{r}{2}\alpha^2\sigma^2 - c(a) = 0$$

or,

$$\beta = c(a) + \frac{r}{2}\alpha^2\sigma^2 - \alpha a\,.$$

We can now substitute for β into the incentive scheme,

$$\begin{aligned} s(x) &= \alpha x + \beta \\ &= \alpha x + c(a) + \frac{r}{2}\alpha^2\sigma^2 - \alpha a, \end{aligned}$$

and the principal maximizes her expected profits, $E[x - s(x)|a]$, subject to the agent's (IC). Since $E[x] = a$, the principal's problem can be written as

$$\begin{cases} \max\limits_{a,\alpha} \; a - c(a) - \frac{r}{2}\alpha^2\sigma^2 \\ \text{s.t.} \quad c'(a) = \alpha \end{cases}$$

By substituting $c'(a)$ for α in the objective function, the principals necessary (but not sufficient) FOC with respect to a becomes,

$$1 = c'(a) + rc'(a) \cdot c''(a)\sigma^2$$

Let's assume that the SOC is satisfied, and we have a unique maximizer, in which case we are done.

Example 4.2: A simple case is when $c(a) = \frac{k}{2}a^2$, and the principal's FOC is

$$ka + rk^2a\sigma^2 = 1$$

which yields,

$$a = \frac{1}{k + rk^2\sigma^2} \quad ; \quad \alpha = \frac{1}{1 + rk\sigma^2}$$

and we get a nice closed form solution with "realistic" results as follows:

1. $c'(a) < 1 \Rightarrow$ less effort in SB relative to FB.
2. $0 < \alpha < 1 \Rightarrow$ a sharing rule that "makes sense."
3. Appealing comparative statics: $\alpha \downarrow$ and $a \downarrow$ if either: (i) $r \uparrow$ (more risk aversion) or, (ii) $\sigma^2 \uparrow$ (more exogenous variance) That is, more risk implies less effort and a "flatter" (more insured) scheme.

Remark: A nice feature of the model is that if agent owns the firm he will choose $a = a^*_{FB}$ but he will be exposed to risk (follows from CARA).

1. Complicated environment $\Rightarrow$ simple optimal contracts
2. Nice tractable model, generalizes easily to $x = (x_1, ..., x_n)$ and $a = (a_1, ..., a_n)$ vectors.

5.8 Nonverifiable Performance

5.8.1 Relational Contracts (Levin)

5.8.2 Market Pressure (Holmstrom)

5.9 Multitask Model

Holmstrom-Milgrom (1991) analyze a model in which the agent has *multiple tasks*, for example, he produces some output using a machine, and at the same time needs to care for the machine's long-term quality. Using their model, H-M '91address the following questions:

1. Why are many incentive schemes "low powered?" (i.e., "flat" wages that do not depend on some measure of output.)

2. Why are certain verifiable signals left out of the contract? (assuming that the ones left in are not sufficient statistics.)

3. Should tasks be performed "in house" or rather purchased through the market?

Using the multitask model, H-M '91 show that incentive schemes not only create risk and incentives, but also allocate the agent's efforts among the various tasks he performs.

5.9.1 The Basic Model

- A risk averse agent chooses an effort vector $t = (t_1, ..., t_n) \geq 0$ at cost $c(t) \geq 0$, where $c(y)$ is strictly convex.

- Expected gross benefits to principal is $B(t)$. (The principal is risk neutral.)

- The agent's effort t also generates a vector of information signals: $x \in \Re^k$ given by,

  $$x = \mu(t) + \varepsilon \, ,$$

  where, $\mu : \Re^n_+ \to \Re^k$ is concave, and the noise is multi-normal, $\varepsilon \sim N(0, \Sigma)$, $0 \in \Re^k$ is the vector of zeros, and Σ is a $k \times k$ covariance matrix.

- Given wage w and action t, the agent has exponential CARA utility given by,
$$u(w,t) = e^{-r[w-c(t)]}$$

- Following Holmstrom-Milgrom (1987) we assume that this is a final stage of a Brownian Motion model, so that the optimal scheme is linear, and given by,
$$w(x) = \alpha^T x + \beta \, ,$$
$$\alpha^T x = \sum_{n=1}^{k} \alpha_n x_n \, ,$$
and the agent's expected utility is equal to the certainty equivalent,
$$CE = \alpha^T \mu(t) + \beta - \frac{r^2}{2}\alpha^T \Sigma \alpha - c(t)$$

(where $\alpha^T \Sigma \alpha$ is the variance of $\alpha^T \varepsilon$.)

First Best: The principal ignores (IC) and only need to compensate the agent for his effort, so the principal'ss program is,
$$\max_t B(t) - c(t) \, .$$

Second Best: The principal maximizes,
$$\begin{cases} \max\limits_t & B(t) - \alpha^T \mu(t) - \beta \\ \text{s.t.} & t \in \arg\max \alpha^T \mu(t) - c(t) \qquad \text{(IC)} \\ & \alpha^T \mu(t) + \beta - \frac{r^2}{2}\alpha^T \Sigma \alpha - c(t) \geq 0 \quad \text{(IR)} \end{cases}$$

As before (in H-M '87) (IR) binding gives β as a function of (α, t), so we can substitute this into the objective function, and get,
$$\begin{cases} \max\limits_t B(t) - c(t) - \frac{r^2}{2}\alpha^T \Sigma \alpha \\ \text{s.t.} \quad t \in \arg\max \alpha^T \mu(t) - c(t). \quad \text{(IC)} \end{cases}$$

(Note: $B(t)$ need not be part of x. For example, $B(t)$ can be a private benefit of the principal, or due to inaccurate accounting, we can have x being an inaccurate signal of true output.)

We can introduce the following simplification: $\mu(t) = t$. This implies that $x \in \Re^n$ (one interpretation is that there is one signal per task.) This is the case of *full dimensionality*. (Note that this is not really a special case: If $\mu(t) \in \Re^m$, $m > n$, then we can "reduce" the dimensionality by some combination of signals, and if $m < n$ then we can add signals with variance of infinity.)

>From this simplification:

$$CE = \alpha^T t + \beta - \frac{r}{2}\alpha^T \Sigma \alpha - c(t)$$

and the agent's FOC's are (we assume that we get an interior solution with $t >> 0$):

$$\alpha_i = c_i(t) \; \forall \, i = 1, ..., n$$

Following the first-order approach, we can substitute these FOC's into the principal's objective function where we use these FOCs as (IC) constraints. First, note that from the agent's FOCs we have (in vector notation):

$$\alpha(t) = \nabla c(t)$$

which implies that $\nabla \alpha(t) = [c_{ij}]$ which is the $n \times n$ matrix of the second derivatives of $c(t)$. Using the Inverse Function Theorem we get $\nabla t(\alpha) = [c_{ij}]^{-1}$ which we use later to perform comparative statics on $t(\cdot)$.

The principal maximizes,

$$\max_t \; B(t) - c(t) - \frac{r}{2}\alpha(t)^T \Sigma \alpha(t)$$

and since $\alpha(t) = \nabla c(t)$ from the agent's FOC, we can write $\alpha_i(t) = c_i(t)$ for each $i = 1, ...n$ (where $c_i(t) = \frac{\partial c}{\partial t_i}$,) and we get the principal's FOCs with respect to t,

$$B_i(t) = \alpha_i(t) + r \sum_{k=1}^{n} \sum_{j=1}^{n} \alpha_j(t) \delta_{jk} C_{ki}(t) \; \forall \, i = 1, ..., n$$

or in vector form:

$$\nabla B(t) = [I + r[c_{ij}]\Sigma]\alpha \, ,$$

where I is the identity matrix, and thus we have,

$$\alpha = [I + r[c_{ij}]\Sigma]^{-1} \nabla B(t) \, . \tag{5.10}$$

Assuming that $\nabla c(t)^T \Sigma \nabla c(t)$ is a convex function of t will give sufficiency of the FOCs.

Benchmark: Stochastic and Technological Independence

To simplify we assume stochastic and technological independence which is given by,

- Σ is diagonal $\Rightarrow \sigma_{ij} = 0$ if $i \neq j$, (errors are stochastically independent.)
- $[c_{ij}]$ is diagonal $\Rightarrow c_{ij} = 0$ if $i \neq j$, (technologies of the different tasks are independent.)

under these assumptions the solution to (5.10) yields,

$$\alpha_i = \frac{B_i}{1 + rc_{ii}\sigma_i^2} \;\forall\, i = 1, ..., n \tag{5.11}$$

Observe that this solution implies:

1. "commissions" for the different tasks are independent (not surprising given the independence assumptions.)
2. α_i decreases in risk (higher r or higher σ^2)
3. α_i decreases in c_{ii} (if c_{ii} is larger, then the agent is less responsive to incentives for task i.)

A Special Case

Consider the case where $n = 2$, and only action t_1 can be measured:

$$x_i = t_i + \varepsilon_i \,,$$

where $var(\varepsilon_2) = \infty$, and $0 < var(\varepsilon_1) < \infty$, that is, $0 < \sigma_1^2 < \infty$, $\sigma_2^2 = \infty$, and $\sigma_{12} = \sigma_{21} = 0$.

>From the FOC (5.11) above we have that $\alpha_2 = 0$, and (assuming an interior solution $t_1, t_2 > 0$,)

$$\alpha_1 = \frac{B_1 - \frac{B_2 C_{12}}{C_{22}}}{1 + r\sigma_1^2(C_{11} - \frac{C_{12}^2}{C_{22}})} \tag{5.12}$$

We can now ask what happens if t_1, t_2 are complements ($c_{12} < 0$) or substitutes ($c_{12} > 0$)? (i.e., via the agents cost function.) To answer this we can look at (5.12) above, and start with $c_{12} = 0$. As we change to $c_{12} > 0$ we see that α_1 decreases, and as we change to $c_{12} < 0$, α_1 increases.

Caveat: This is a *local argument* since C_{12} is a function of (t_1, t_2), which are in turn functions of α_1 through the incentives.

We can explain the intuition for the two directions above as follows:

1. Making C_{12} positive is making the tasks *substitutes in costs* for the agent. If we want both t_1 and t_2 to be performed, and we can only give incentives to t_1 through α_1, then increasing α_1 "kills" incentives for t_2 and increases incentives for t_1. This may be undesirable. (In fact, it is undesirable around $C_{12} = 0$.)

2. With C_{12} negative, the reverse happens: an increase in α increases $t_1$and reduces c_2 so that t_2 increases. Thus, α_1 gives incentives to both tasks. This result is actually global and does not depend on the local analysis performed above.

5.9.2 Application: Effort Allocation

Consider the case in which the agent's cost of efforts is a function of the *sum* of all efforts. This is a special case in which we assume that the efforts are extreme substitutes: The agent is indifferent between which tasks he performs, as long as his *total* effort is unchanged. We simplify by assuming that there are only two tasks, and further restrict attention t the following special case:

1. $c(t_1, t_2) = c(t_1 + t_2)$

2. There exists $\bar{t} > 0$ such that $c'(\bar{t}) = 0,\ c''(t) > 0\ \forall\, t.$

Figure Here

The idea behind assumption (2) above is that people will work $\bar{t}$ without incentives, not caring how $\bar{t}$ is allocated, as long as $t_1 + t_2 = \bar{t}$. However, providing incentives will affect the choice of effort. Note that this is a somewhat unorthodox assumption, however it is not extremely unrealistic. One way to think about this is that an agent will perform some minimal amount of performance either due to the prospect of getting fired, or due to the alternative of boredom..

Missing Incentives

Now assume that we cannot measure the effort for the first task. For example, the agent can be a contractor that is remodelling the principal's house, and t_1 can be courtesy, or attention to detail. On the other hand, we assume that t_2 is (imperfectly) measurable. Using the contractor example, t_2 can be time to completion, or how close the original plan was followed.

To formalize this assume that $\mu(t_1, t_2) = \mu(t_2) \in \Re$, a one dimensional effect of both tasks, and following the previous notation, let the measureable (verifiable) signal be given by,

$$x = \mu(t_2) + \varepsilon \ ,$$

and the linear compensation scheme be

$$s(x) = \alpha_2 x + \beta$$

Assumption: $B(t_1, t_2) > 0$ and increasing in both components, and $B(0, t_2) = 0 \; \forall \, t_2$ (t_1 is "essential" for the project to have value.)

Proposition 4.13: In the above set-up, $\alpha_2 = 0$ is optimal (even if the agent is risk neutral.)

Proof: With $\alpha_2 = 0$ ($\alpha_1 = 0$ since t_1 is not measurable) then principal maximizes: $B(t_1, \bar{t} - t_1)$ and due to his indifference, the agent will accommodate any solution. In this case there is no risk, and total surplus is

$$S^* = B(t_1^*, \bar{t} - t_1^*) - c(\bar{t}).$$

If $\alpha_2 > 0$, then the agent's choices are $t_1 = 0, \; t_2 = \hat{t}_2 \neq \bar{t}$, and social surplus is

$$\overbrace{B(0, \hat{t}_2)}^{0} - \overbrace{c(\hat{t})}^{>c(\bar{t})} - \overbrace{\frac{r}{2}\alpha_2^2\sigma^2}^{\geq 0} < S^*.$$

If $\alpha_2 < 0$, then $t_2 = 0$, and $t_1 < \bar{t}$ (since $c'(t_1) < 0 = c'(\bar{t})$), and total surplus is

$$\overbrace{B(t_1, 0)}^{<B(\bar{t},0)} - \overbrace{c(t_1)}^{>c(\bar{t})} - \overbrace{\frac{r}{2}\alpha^2\sigma^2}^{\geq 0} < S^*$$

Q.E.D.

1. It is important that $B(\cdot,\cdot)$ is a "private benefit,"or else principal can "sell" the project to the agent. The house contracting example is a very nice one, as is the example of a teacher having incentives to teach children both skills of succeeding in tests (measurable), and of creative thinking (not measurable) where the parents (or local government) have a private benefit from the children's' education.
2. This result is not "robust": it relies both on $B(0, t_2) = 0$ and $C(t_1, t_2) = C(t_1 + t_2)$. But, the intuition is very appealing..

Low powered incentives in firms

Williamson (1985) observed that inside firms incentives are "low-powered" (e.g., wages to workers) compared to "high-powered" incentives offered to independent contractors. Also, employees work with the principal's assets while contractors work with their own assets. Using a variant of the previous set-up, this can be explained by multi-tasking.

Assume that,

$$B(t_1, t_2) = B(t_1) + v(t_2)\,,$$

with $B' > 0$, $v' > 0$, $B'' < 0$, $v'' < 0$ and $B(0) = v(0) = 0$. We can interpret $B(t_1)$ to be the current expected profit from activity t_1, e.g., effort in production, while $v(t_2)$ is the future value of the "assets" from activity t_2, e.g. preventive maintenance, etc.

Now let t_1 be measurable with the signal

$$x = \mu(t_1) + \varepsilon_x.$$

Let the change in the asset's value be $v(t_2) + \varepsilon_v$, and it is important to assume that the actual value accrues to the owner of the asset (for example, there can be some private benefit, imperfect markets, etc.) We finally assume that ε_x and ε_v are independent shocks.

As before, the incentive scheme will be linear, and given by

$$s(x) = \alpha x + \beta\,.$$

We now consider two alternatives for the relationship: Either the principal and agent enter a *contracting relationship*, in which case the agent owns asset, or they enter an *employment relationship*, in which case the principal owns the asset.

Define,

$$\begin{aligned}\pi^1 &= \max_{t_1} B(t_1) - C(t_1)\,,\\ \pi^2 &= \max_{t_2} v(t_2) - C(t_2)\,,\\ \pi^{12} &= \max_{t_1} B(t_1) + v(\bar{t} - t_1) - C(\bar{t})\,.\end{aligned}$$

That is, π^1 is maximal total surplus when only t_1 can vary, π^2 is maximal total surplus when only t_2 can vary, and π^{12} is maximal total surplus when we can choose t_1 subject to $t_1 + t_2 = \bar{t}$. We have the following proposition:

1. If $\pi^{12} > \max\{\pi^1, \pi^2\}$ then the optimal employment contract has $\alpha = 0$.
2. If contracting is optimal then $\alpha > 0$.
3. There exist $(r, \sigma_v^2, \sigma_\varepsilon^2)$ for which employment is optimal and other parameters for which contracting is optimal.
4. If employment is optimal for some $(r, \sigma_v^2, \sigma_\varepsilon^2)$, it is also optimal for larger values. The reverse for contracting.

1. If $v(t_2)$ accrues to the principal then $\alpha > 0 \Rightarrow t_2 = 0,\ c'(t_1) = \alpha$, and total surplus is

   $$B(t_1) - c(t_1) - \frac{r}{2}\alpha^2\sigma_x^2 < \pi^1 \le \pi^{12}$$

   whereas π^{12} can be achieved with $\alpha = 0$.
2. Idea: with no incentives, agent will not care about t_1 but rather only about t_2. With $\alpha > 0$, agent still chooses the same t_2 because he gets $v(t_2)$. $\Rightarrow \alpha > 0$ is optimal.
3. This is just to say that the π's can be ordered in any way depending on the parameters.
4. Intuition: if employment is optimal then more risk aversion implies that there is a stronger case for no risk in contract, so employment must still be optimal.

In chapter 5 we will discuss some theories of ownership, and this is an interesting model tat has implication to these issues.

5.9.3 Limits on outside activities

The following observation has been made by xxx: Some employees (usually high level) have more freedom to engage in "personal business" than others (e.g., private telephone conversations, undefined lunch breaks, etc.) This is another observation to which the multitask model adds some insight. To analyze this issue consider the following modifications to the model:

- **Tasks:** there are $k+1$ tasks, $(t, t_1, ..., t_k) \in \Re_+^{k+1}$, where only the first task, t, benefits the principal:

$$B(t, t_1, ..., t_k) = p \cdot t$$

  (e.g., some market price for an output.)

- **Agent's costs:**

$$c(t, t_1, ..., t_k) = c(t + \sum_{i=1}^{k} t_i) - \sum_{i=1}^{k} v_i(t_i)$$

where $v_i(t_i)$ is the agent's private benefit from task i, with $v' > 0$, and $v'' < 0$. (e.g., having access to non pecuniary tasks like outside phone lines, long breaks, taking care of errands during work hours, etc.)

- Every personal task can be either *completely excluded* or *completely allowed* in the contract, but no personal task can be restricted to a level (i.e., "all-or-nothing".)
- $x = \mu(t) + \varepsilon$ is the signal, $s(x) = \alpha x + \beta$ is the incentive scheme.
- The principal can choose a contract that includes $A \subset \{1, ..., k\}$ of "allowable" personal tasks, and α, β for the incentive scheme.

To solve for the optimal contract we consider a *two stage solution process* which is similar in spirit to the two stage program of Grossman and Hart (1983): For every α, find $A(\alpha)$ that is optimal, then given $A(\alpha)$ choose α optimally.

Assume that an interior solution exists, so that given (α, A) the agent's FOC's yield:

$$\alpha = c'(t + \sum_{i=1}^{k} t_i),$$
$$\alpha = v'(t_i) \, \forall \, i \in A$$

Given α, use the following "cost-benefit" argument to determine the tasks that should not be excluded:

$$A = \{i : v_i(t_i(\alpha)) > p \cdot t_i(\alpha)\}$$

Figure Here

To understand the idea, look at figure above. For each task i, there exists some $\hat{t}_i$ such that for $t_i < \hat{t}_i$ the private benefit to the agent is larger than $p \cdot t_i$ and for $t_i > \hat{t}$ the reverse holds. From a social-surplus point of view, if $t_i(\alpha) < \hat{t}_i$, it is better to have the agent exert $t_i(\alpha)$ into his private benefit $v_i(t_i)$ rather than in the principal's private benefit $p \cdot t_i(\alpha)$. Thus, in figure, task 1 should be allowed and task 2 should not.

We can also see that an increase in α will (weakly) cause an increase in the set A. We also get:

Proposition 4.14: Assume $t(\alpha)$ is optimal then:

1. $\alpha = \frac{p}{1+r\sigma^2 dt/d\alpha}$,
2. If measurement is easier (σ^2 decreases), or if the agent is less risk averse (r decreases), then α and $A(\alpha)$ will be larger.
3. tasks that are excluded in the FB contract are also excluded in the SB contract, but for high $r\sigma^2$ some tasks that are included in a FB contract will be excluded in a SB contract.

(For a proof see the paper.)

The part of the proposition that is most interesting is part (2): The set A gets smaller (and incentives weaker) when we have measurement problems over t. A nice application is that without measurement problems, an outside sales force (independent contractors with no exclusions) is optimal, whereas with measurement problems an inside sales force is optimal (e.g., can't sell competitor's products, etc.)

5.9.4 Allocating tasks between two agents:

This is the last section of H-M '91. The results are:

1. It is never optimal to assign two agents to the same task.
2. The principal wants "information homogeneity": the hard-to-measure tasks go to one agent and the easy-to-measure tasks go to the other.

The intuition goes as follows: (1) we don't have "team" problem if agents are assigned to different tasks; (2) We avoid the multitask problems mentioned earlier; The agent with hard-to-measure tasks gets low incentives ("insider") and the other gets high incentives ("outsider").