Lecture Notes on Moral Hazard

Debasis Mishra

April 15, 2010

1 The Principal-Agent Problem

In the adverse selection and signaling models, we noticed how asymmetries in information exist between agents at the time of *contracting* (for example, the productivity of a worker is unknown to the firm before he is hired). In this section, we look at the problem where information asymmetry arise subsequent to the signing of the contract. This may happen even though there is no information asymmetry between the agents at the time of signing the contract. For example, though the firm may know the productivity of the worker, it may not be able to observe the effort put in by the worker after he is hired. Similarly, the worker will know more about the opportunities available to the firm than the owner of the firm itself.

The objective of this section is to anticipate such informational asymmetry and eliminate the problems by appropriate contracts. Such problems typically arise in situations where one agent hires another "agent" to take actions on his behalf. The hired agent is called the agent and the employer is called the "principal", and hence the name *the principal-agent problem*. The principal agent problem can be classified into two classes:

1. Hidden Action - Moral Hazard Problem. The moral hazard problem is illustrated by the fact that the owner of the firm is unable to observe the effort level of the worker.

2. Hidden Information - Monopolistic Screening Problem. The monopolistic screening problem is illustrated by the fact that a hired manager has more information about the opportunities of the firm than then owner of the firm himself.

We give some examples to illustrate the wide applicability of the principal-agent problem.

1. The owner and the manager. The owner of a firm hires a manager who gets to know more about the opportunities to the firm. Further, the effort level of the manager is unobservable to the owner of the firm.

2. THE INSURANCE COMPANY AND THE INSURED INDIVIDUAL. The insurance company cannot observe how much precaution is taken by the insured individual.

3. THE MANUFACTURERS AND THE DISTRIBUTORS. The distributor observes the market condition better than the manufacturer.

4. THE BANKS AND THE BORROWERS. The bank may not observe how the funds were used by the borrowers.

Here, we will take the firm as the principal and the manager as the agent.

2 Moral Hazard

A firm hires a manager for a project. The profit from the project is observable, and is denoted by π. The project's success depends on the action chosen by the manager. We denote the action of the manager as e, and the set of all possible actions as E. In this section, we treat e to be one dimensional, and hence, $E \subseteq \mathbb{R}$. However, one can easily adapt the analysis to deal with multidimensional action sets (various dimensions can be, for example, how hard the manager works, how much time he spends in consumer interaction etc.). We refer to any $e \in E$ as the *effort choice* or *effort level* of the manager.

If the manager's action is unobservable, then it should not be deducible from the profit of the project. Hence, we assume that although the profits from the project is influenced by effort level of the manager, it is not entirely dependent on it. In particular, we assume that the profit from the project to lie in $[L, H]$ and there is a conditional density function $f(\pi|e)$ such that $f(\pi|e) > 0$ for all $\pi \in [L, H]$ and for all $e \in E$.

For simplicity, we assume that $E = \{e_l, e_h\}$, where $e_l < e_h$. Here, e_h is a higher effort level of the manager and leads to higher profits than effort level e_l. Of course, the manager likes to put lower effort level than higher effort level.

More specifically, we assume that conditional distribution $f(\pi|e_h)$ first-order stochastically dominates the conditional distribution $f(\pi|e_l)$, i.e., the distribution functions $F(\pi|e_l)$ and $F(\pi|e_h)$ satisfy $F(\pi|e_h) \leq F(\pi|e_l)$ for all $\pi \in [L, H]$, with strict inequality holding on some open set of $[L, H]$. An implication of this is that the level of expected profits when the manager chooses e_h is higher than that from e_l. So,

$$\int \pi f(\pi|e_h) d\pi > \int \pi f(\pi|e_l) d\pi.$$

The manager is an expected utility maximizer with a utility function over w and w. We assume that

$$u(w, e) = v(w) - g(e),$$

where $v(w)$ is the value of wage w and $g(e)$ is the cost of effort level e. We assume that $u_w(w,e) > 0$ and $u_{ww}(w,e) \leq 0$ at all (w,e), where subscripts denote partial derivatives, and $u(w,e_h) < u(w,e_l)$. This implies that $v'(w) > 0$, $v''(w) \leq 0$, and $g(e_h) < g(e_l)$. Further, he has a *reservation utility* of $\bar{u}$, i.e., if he does not accept the contract then he gets this level of utility.

The owner of the firm is risk neutral. His payoff is the profit made from the project minus the wage paid to the manager. If the manager rejects the contract, then the owner receives zero payoff.

2.1 Observable Effort

We begin our analysis by looking at the optimal contract when the effort is observable. The owner offers a contract, and the manager may accept or reject it.

Definition 1 *A* **contract** *is a tuple* (e, w)*, where* $e \in \{e_h, e_l\}$ *is the effort level and* $w : [L, H] \to \mathbb{R}_+$ *is the wage schedule with respect to observed profits.*

We assume throughout that the owner will make an offer that the manager will find worthwhile to accept. Hence, the **optimal contract** is a contract which (a) maximizes the expected utility of the owner and (b) gives at least the reservation value to the manager.

$$\max_{e,w(\cdot)} \int_L^H [\pi - w(\pi)] f(\pi|e) d\pi$$
$$\text{s.t.} \qquad \int_L^H v(w(\pi)) f(\pi|e) d\pi - g(e) \geq \bar{u}.$$

The optimal contract problem is then easy to imagine in two stages. First, we fix an effort level e, and find out the optimal wage schedule. Then, we choose among various effort levels to maximize expected utility.

Suppose we fix the effort level at e. Then, the objective function now simplifies to

$$\max_{w(\cdot)} - \int_L^H w(\pi) f(\pi|e) d\pi.$$

So, the objective is to minimize the expected wage of the firm. Note that the constraint now is

$$\int_L^H v(w(\pi)) f(\pi|e) d\pi \geq g(e) + \bar{u}.$$

This constraint must bind at the optimum as the owner can reduce expected wage by lowering the wage schedule a little bit if the constraint is not binding. Let γ denote the Lagrange multiplier of the constraint. Hence, the problem can be rewritten as

$$\max_{w(\cdot),\gamma} - \int_L^H w(\pi)f(\pi|e)d\pi + \gamma[\int_L^H v(w(\pi))f(\pi|e)d\pi - g(e) - \bar{u}].$$

For every π, the optimal wage $w(\pi)$ must satisfy the first order condition

$$-f(\pi|e) + \gamma v'(w(\pi))f(\pi|e) = 0.$$

This gives us for all π,

$$\gamma = \frac{1}{v'(w(\pi))}.$$

We consider two cases. In the first case, if $v'(w)$ strictly decreasing in w, then the implication of this is that $w(\pi)$ has to be a constant. So, the manager must be offered a fixed wage. The intuition is that given that the manager is risk averse, the risk neutral owner must insure against variations in profit by giving the manager a fixed wage at a given effort level. So, given an effort level e, the optimum wage w_e^* must satisfy

$$v(w_e^*) = g(e) + \bar{u}.$$

Note that since $g(e_h) > g(e_l)$, the manager's wage is higher at higher effort level.

The other case is when the manager is risk neutral, say $v(w) = w$. Then, the first order condition is satisfied for any $w(\cdot)$ function. Any compensation function which ensures an expected wage of $g(e) + \bar{u}$ at effort level e will work. This also includes the fixed wage of $g(e) + \bar{u}$.

Now, consider the optimal choice of e. The owner must now maximize

$$\int_L^H \pi f(\pi|e)d\pi - v^{-1}(g(e) + \bar{u}).$$

Whether e_h or e_l will be optimal depends on the incremental increase in expected profit to firm from e_l to e_h and the disutility to the manager. This leads to the following theorem.

THEOREM **1** *In the principal-agent problem with observable managerial effort, an optimal contract specifies that the manager chooses the effort e^* that maximizes $[\int_L^H \pi f(\pi|e)d\pi - v^{-1}(g(e) + \bar{u})]$ and pays the manager a fixed wage $w^* = v^{-1}(g(e) + \bar{u})$. This is the unique optimal contract if $v''(w) < 0$ for all w.*

3 Unobservable Effort

When the effort is observable, the optimal contract specifies an effort level to the manager and insures him against risks associated with profit levels by providing a constant wage. When the effort level is not observable, these two events are often in conflict - to make the manager work hard involves relating his wage to profits, which is random. We first study the case when the manager is risk neutral.

3.0.1 A Risk Neutral Manager

Suppose the manager is risk neutral and $v(w) = w$ for all w. Hence, the optimal effort level e^* when effort level is observable solves

$$\max_{e \in \{e_l, e_h\}} \int_L^H \pi f(\pi|e) d\pi - (g(e) + \bar{u}). \tag{1}$$

The owner's expected utility is this expression evaluated at e^*, and the manager receives an expected utility equal to $\bar{u}$. We now show that when the effort is not observable same expected utility levels can be achieved as in the full observable case.

Note here that here a contract is only a wage schedule $w : [L, H] \to \mathbb{R}_+$ since the effort level is not observed any more.

Theorem **2** *In the principal-agent model with unobservable managerial effort and a risk neutral manager, an optimal contract generates the same effort choice and expected utilities for the manager and the owner as when the effort is observable.*

Proof: We show that there is a contract that the owner can offer to the manager that gives him everyone the same level of utilities that everyone receives under full information. This contract must therefore be optimal for the owner as the owner can never do better than than his utility with full observable effort level. To see this, note that this non-observable effort contract is always a feasible contract when the effort is not observable.

Suppose the owner offers a wage schedule of the form $w(\pi) = \pi - \alpha$ for all $\pi \in [L, H]$, where α is some constant. This contract can be thought of "selling the project to the manager", in the sense that the manager runs the firm, keeps all the profit except α, which he returns to the owner. So, α is like the sale price of the project. Let us see what the manager's optimal response to this contract. If the manager accepts this contract, he will choose an effort level which maximizes his expected utility, given by

$$\int_L^H w(\pi) f(\pi|e) d\pi - g(e) = \int_L^H \pi f(\pi|e) d\pi - \alpha - g(e).$$

Comparing with Equation 1, we see that e^* maximizes this expression. Thus, this contract induces the same level of effort as with full observable effort. The manager is willing to accept this contract if it gives him his reservation utility:

$$\int_L^H \pi f(\pi|e^*)d\pi - \alpha - g(e^*) \geq \bar{u}.$$

Let α^* be the level of α at which the above inequality binds (in an optimal contract, this constraint will bind). Hence,

$$\alpha^* = \int_L^H \pi f(\pi|e^*)d\pi - g(e^*) - \bar{u}.$$

Since α^* is the expected utility of the owner, and this expression is the same as the optimal value of expression in Equation 1, we get the desired result. ■

3.0.2 A Risk Averse Manager

When the manager is strictly risk averse, then the optimal contract gets complicated. Now, incentives for high effort can be provided by exposing the manager to some risks of profits. We now characterize the optimal contract. We do so in two steps. In the first step, we characterize the optimal wage scheme for a given effort level that the owner might want the manager to select. Next, we consider which effort level is optimal for the owner.

The optimal wage for implementing a given effort level e minimizes the owner's expected wage. There are two constraints : (1) *participation constraint* - the manager must get his reservation utility from this contract, and (2) *incentive compatibility constraint* - the given effort level must maximize the manager's expected utility over all possible effort levels.

$$\min_{w(\pi)} \int_L^H w(\pi) f(\pi|e) d\pi$$

$$\text{s.t.}$$

$$\int_L^H v(w(\pi)) f(\pi|e) d\pi - g(e) \geq \bar{u}$$

$$\int_L^H v(w(\pi)) f(\pi|e) d\pi - g(e) \geq \int_L^H v(w(\pi)) f(\pi|e') d\pi - g(e') \qquad \forall\ e' \in \{e_l, e_h\}.$$

A wage schedule w **implements** an effort level e if it solves the above optimization problem at e. First, we ask the question if there is a wage schedule to implement each of the effort levels.

Implementing e_l: If the owner wants to implement e_l, he can do so by giving a fixed wage: $w_{e_l} = v^{-1}(\bar{u} + g(e_l))$. Note that the wage here is independent of profit. Hence, it is optimal for the manager to select the effort level which is the lowest level. Secondly, the utility from this wage is exactly $\bar{u}$. Hence, participation constraint also holds. The total expected wage from this contrast is w_{e_l}, which is the same expected wage in the case when the effort is observable. Since the owner cannot do better than the full observable effort case (formally, the feasible set is larger with full observable effort since with non-observability, we also have the incentive compatibility constraints), this is indeed the optimal contract.

Implementing e_h: If the owner decides to implement e_h, then the wage schedule must solve

$$\min_{w(\pi)} \int_L^H w(\pi) f(\pi|e_h) d\pi$$

s.t.

$$\int_L^H v(w(\pi)) f(\pi|e_h) d\pi - g(e_h) \geq \bar{u}$$

$$\int_L^H v(w(\pi)) f(\pi|e_h) d\pi - g(e_h) \geq \int_L^H v(w(\pi)) f(\pi|e_l) d\pi - g(e_l).$$

Letting $\gamma \geq 0$ and $\mu \geq 0$ denote the Lagranger multipliers for both the constraints, and taking KKT first order conditions, we get that for all π, we must satisfy

$$-f(\pi|e_h) + \gamma v'(w(\pi)) f(\pi|e_h) + \mu v'(w(\pi))[f(\pi|e_h) - f(\pi|e_l)] = 0.$$

Equivalently,

$$\frac{1}{v'(w(\pi))} = \gamma + \mu\big[1 - \frac{f(\pi|e_l)}{f(\pi|e_h)}\big]. \tag{2}$$

Lemma 1 *In any solution to the above optimization problem, $\gamma > 0$ and $\mu > 0$ (i.e., both constraints must bind).*

Proof: If $\mu = 0$, then, as we did earlier, the optimal solution is a fixed wage schedule for every profit realization. But we know that this will lead the manager to choose e_l, and not e_h. Hence, incentive compatibility will be violated. Hence, $\mu > 0$.

Assume for contradiction $\gamma = 0$. Because $F(\pi|e_h)$ stochastically dominates $F(\pi|e_l)$, for some open set $X \subset \{L, H\}$, we have $\frac{f(\pi|e_l)}{f(\pi|e_h)} > 1$ at all $\pi \in X$. But if $\gamma = 0$, this implies that $v'(w(\pi)) \leq 0$ for all $\pi \in X$ (since $\mu \geq 0$). This is impossible since the manager is risk averse. Hence, $\gamma = 0$. ∎

Given Lemma 1, we can get useful insights into the implementation of e_h. First both the incentives and participation constraints must bind at optimality. Consider the fixed wage payment $\hat{w}$ such that $\frac{1}{v'(\hat{w})} = \gamma$. According to condition in Equation 2, we get two cases.

$$w(\pi) > \hat{w} \qquad \text{if } \frac{f(\pi|e_l}{f(\pi|e_h)} < 1 \tag{3}$$

$$w(\pi) < \hat{w} \qquad \text{if } \frac{f(\pi|e_l}{f(\pi|e_h)} > 1. \tag{4}$$

So, the optimal wage pays more than $\hat{w}$ for outcomes that are statistically relatively more likely to occur under e_h than under e_l and pays less than $\hat{w}$ for outcomes that statistically less likely to occur under e_l than under e_h. By structuring wages like this, the manager is provided incentives to produce higher effort level.

The main point is that the optimal wage may not be monotonic with profits. For the optimal wage to be monotonically increasing, the likelihood ratio $\frac{f(\pi|e_l)}{f(\pi|e_h)}$ must be decreasing in π: as π increases, the likelihood of getting profit level π if effort is e_h relative to the likelihood of getting profit level π if if effort is e_l must increase. To see this, take $\pi > \hat{\pi}$. If $w(\pi) \geq w(\hat{\pi})$, then, because of concave $v(\cdot)$, we must have $v'(w(\pi)) \leq v'(w(\hat{\pi}))$. Hence,

$$\gamma + \mu[1 - \frac{f(\pi|e_l)}{f(\pi|e_h)}] \geq \gamma + \mu[1 - \frac{f(\hat{\pi}|e_l)}{f(\hat{\pi}|e_h)}].$$

This implies that

$$\frac{f(\hat{\pi}|e_l)}{f(\hat{\pi}|e_h)} \geq \frac{f(\pi|e_l)}{f(\pi|e_h)}.$$

This property is called the **monotone likelihood ratio property**, and **is not implied** by first-order stochastic dominance.

The optimal contract is therefore not simple. Finally, the expected wage paid by the owner must be strictly greater than the fixed wage payment in the observable case ($w^*_{e_h} = v^{-1}(\bar{u} + g(e_h))$). Intuitively, the manager has to be insured against risk in profit levels and this insurance is higher for high effort levels. To see this, first

$$\int_L^H v(w(\pi))f(\pi|e_h)d\pi = E[v(w(\pi))|e_h] = \bar{u} + g(e_h) = v(w^*_{e_h}).$$

Using the fact that $v''(\cdot) < 0$ and Jensen's inequality (The discrete version of Jensen's inequality is easy to observe. Suppose there are two values of π - π_1 and π_2. Then, $E[v(w(\pi))|e_h] = v(w(\pi_1))f(\pi_1|e_h) + v(w(\pi_2))f(\pi_2|e_h) < v(w(\pi_1)f(\pi_1|e_h) + w(\pi_2)f(\pi_2|e_h))$, where the inequality is due to concavity.) [1], we get that

$$v(w^*_{e_h}) = E[v(w(\pi))|e_h]) < v(E[w(\pi)|e_h]).$$

[1] The general proof is also simple. It goes as follows. Let $E[w(\pi)|e_h] = z$. Since v is concave, we can draw a supporting hyperplane on $v(\cdot)$ at z such that it lies above the function $v(\cdot)$. Let this supporting hyperplane be $\lambda(w(\pi) - z) + v(z)$. Hence, we get that $E[\lambda(w(\pi) - z) + v(z)|e_h] > E[v(w(\pi))|e_h]$. But $E[\lambda(w(\pi) - z) + v(z)|e_h] = v(z)$. Hence, $v(E[w(\pi)|e_h]) > E[v(w(\pi))|e_h]$.

Since v is increasing, we get that $w^*_{e_h} < E[w(\pi)|e_h]$.

Optimal Contract: Table 1 summarizes our findings in this section. From the preceding analysis, we can say that the expected wage for implementing e_l remains the same but the expected wage to implement e_h goes up. Hence, unobservable level of efforts may lead to inefficient levels of effort. If e_l was optimal when effort was observable, it will still be optimal when it is unobservable. If e_h was optimal when effort was observable, it may be optimal to implement e_h using an incentive scheme that faces the manager with risk or the risk-bearing costs may be high enough such that the owner finds it optimal to implement e_l. In either case, the the welfare to the owner is lower in the case with non-observable effort level.

Cases	Expected Wage	Owner's Expected Payoff
Observable effort Risk neutral manager	$g(e^*) + \bar{u}$	$\max_{e\in\{e_l,e_h\}} \int_L^H \pi f(\pi\|e)d\pi - [g(e) + \bar{u}]$
Unobservable effort Risk neutral manager	$g(e^*) + \bar{u}$	$\max_{e\in\{e_l,e_h\}} \int_L^H \pi f(\pi\|e)d\pi - [g(e) + \bar{u}]$
Observable effort Risk averse manager	$v^{-1}(g(e^*) + \bar{u})$	$\max_{e\in\{e_l,e_h\}} \int_L^H \pi f(\pi\|e)d\pi - v^{-1}(g(e) + \bar{u})$
Unobservable effort (e_l) Risk averse manager	$v^{-1}(g(e_l) + \bar{u})$	$\int_L^H \pi f(\pi\|e_l)d\pi - v^{-1}(g(e_l) + \bar{u})$
Unobservable effort (e_h) Risk averse manager	$> v^{-1}(g(e_h) + \bar{u})$	$< \int_L^H \pi f(\pi\|e_h)d\pi - v^{-1}(g(e_h) + \bar{u})$

Table 1: Payoffs in various cases of the moral hazard problem