

Moral Hazard

In contractual arrangements in which the principal offers the contract, we distinguish between

hidden information - ‘adverse selection’

and

hidden action - ‘moral hazard’

The main trade-off in adverse selection is between efficient allocation and informational rent.

In moral hazard settings it is between risk sharing and incentives.

In this lecture, we discuss the basic moral hazard setting.

As the principal tries to infer from the output of the agent her effort choice, the principal engages in a statistical inference problem.

Basic Idea

Agent takes an action that affects the utility of the principal and the agent.

Principal observes an outcome x and possibly some signal s but not action a of agent

If there is no contract, the agent has no incentive to exert costly effort.

The principal influences the agent's action choice by offering transfer contingent on outcome.

Example 1 *Unobservable work effort , Property insurance (fire, theft), Financing of new projects etc.*

FB: agents action is observable (risk sharing)

SB: agents action is unobservable (risk sharing-incentive)

Simple Example

The agent can take action $a \in \{a_l, a_h\}$ at cost $c \in \{c_l, c_h\}$.

The outcomes $x \in \{x_l, x_h\}$ occur randomly, where the probabilities are governed by the action as follows.

$$p_l = \Pr(x_h | a_l) < p_h = \Pr(x_h | a_h)$$

The principal offers a wage, contingent on the outcome $w \in \{w_l, w_h\}$ to the agent and the utility of the agent is

$$u(w_i) - c_i$$

and of the principal it is

$$v(x_i - w_i)$$

where we assume that u and v are strictly increasing and weakly concave.

First Best

We consider initially the optimal allocation of risk between the agents in the presence of the risk and with observable actions.

If the actions are observable, then the principal can induce the agent to choose the preferred action a^* by

$$w(x_i, a) = -\infty$$

if $a \neq a^*$ for all x_i .

As the outcome is random and the agents have risk-averse preference, the optimal allocation will involve some risk sharing.

The optimal solution is characterized by

$$\max_{\{w_l, w_h\}} \{p_i v(x_h - w_h) + (1 - p_i) v(x_l - w_l)\}$$

subject to

$$p_i u(w_h) + (1 - p_i) u(w_l) - c_i \geq U, (\lambda)$$

which is the individual rationality constraint. Here we define

$$w(x_i, a^*) \triangleq w_i$$

This is a constrained optimization problem, and the first order conditions from the Lagrangian

$$L(w_l, w_h, \lambda) =$$

$$p_i v(x_h - w_h) + (1 - p_i) v(x_l - w_l) + \lambda (p_i u(w_h) + (1 - p_i) u(w_l) - c_i)$$

can be written as:

$$\frac{V'(x_i - w_i)}{U'(w_i)} = \lambda.$$

This relation is called Borch's rule.

Second Best

Consider now the case in which the action is unobservable and therefore

$$w(x_i, a) = w(x_i)$$

for all a . Suppose the principal wants to induce high effort, then the *incentive constraint* is:

$$p_h u(w_h) + (1 - p_h) u(w_l) - c_h \geq p_l u(w_h) + (1 - p_l) u(w_l) - c_l$$

or

$$(p_h - p_l) (u(w_h) - u(w_l)) \geq c_h - c_l \quad (1)$$

as $p_l \rightarrow p_h$ $w_h - w_l$ must increase, incentives become more high-powered.

The principal also has to respect a *participation constraint* (or

individual rationality constraint)

$$p_h u(w_h) + (1 - p_h) u(w_l) - c_h \geq U \quad (2)$$

We first show that both constraints are binding at optimum.

$$\max_{\{w_h, w_l\}} p_h (x_h - w_h) + (1 - p_h) (x_l - w_l)$$

subject to (1) and (2).

For the participation constraint, principal could lower both payments while satisfying the incentive constraint. To see that the incentive constraint holds as an equality suppose to the contrary. Then it is possible to subtract

$$\frac{(1 - p_h) \varepsilon}{u'(w_h)}$$

from w_h and add

$$\frac{p_h \varepsilon}{u'(w_l)}$$

to w_l .

The incentive constraint would still hold for ε sufficiently small.

This change in wages subtracts

$$(1 - p_h) \varepsilon$$

from $u(w_h)$ and adds

$$p_h \varepsilon$$

to $u(w_l)$, so that the expected utility remains constant.

But the expected wage bill is reduced by

$$\varepsilon p_h (1 - p_h) \left(\frac{1}{u'(w_h)} - \frac{1}{u'(w_l)} \right) > 0,$$

since $w_h > w_l$ and since u is concave.

Using the incentive constraint and the individual rationality

constraint, we get:

$$u(w_l) = U - \frac{c_h p_l - p_h c_l}{p_h - p_l}$$

and

$$u(w_h) = U - \frac{c_h p_l - p_h c_l}{p_h - p_l} + \frac{c_h - c_l}{p_h - p_l}$$

More Generally

Suppose

$$x_i \in \{x_1, \dots, x_I\}$$

and

$$a_j \in \{a_1, \dots, a_J\}$$

and the probability

$$p_{ij} = \Pr(x_i | a_j)$$

the utility is $u(w) - a$ for agent and $x - w$ for the principal. The

principal's problem is then to

$$\max_{\{w_i\}_{i=1}^I, j} \left\{ \sum_{i=1}^I (x_i - w_i) p_{ij} \right\}.$$

Given the wage contract the agent selects a_j if and only if

$$\sum_{i=1}^I u(w_i) p_{ij} - a_j \geq \sum_{i=1}^I u(w_i) p_{ik} - a_k \quad (\mu_k)$$

and

$$\sum_{i=1}^I u(w_i) p_{ij} - a_j \geq \underline{U} \quad (\lambda)$$

Fix a_j , then the Lagrangian is

$$\mathcal{L}(w, \lambda, \mu) = \left\{ \sum_{i=1}^I (x_i - w_i) p_{ij} \right\} + \sum_{k \neq j} \mu_k \left\{ \sum_{i=1}^I u(w_i) (p_{ij} - p_{ik}) - (a_j - a_k) \right\} + \lambda \left\{ \sum_{i=1}^I u(w_i) p_{ij} - a_j \right\}$$

Differentiating with respect to w_i yields

$$\frac{1}{u'(w_i)} = \lambda + \sum_{k \neq j} \mu_k \left(1 - \frac{p_{ik}}{p_{ij}} \right), \quad \forall i \quad (3)$$

With a risk-averse principal the condition (3) would simply be modified to:

$$\frac{v'(x_i - w_i)}{u'(w_i)} = \lambda + \sum_{k \neq j} \mu_k \left(1 - \frac{p_{ik}}{p_{ij}} \right), \quad \forall i \quad (4)$$

In the absence of an incentive problem, or $\mu_k = 0$, (4) reproduces the Borch rule of optimal risk sharing:

$$\frac{v'(x_i - w_i)}{u'(w_i)} = \lambda, \forall i.$$

which requires that the ratio of marginal utilities is equalized across all states i .

In the handout on moral hazard, I outline the conditions required to show that the optimal wage is increasing in output.

Linear contracts with normally distributed performance and exponential utility

This constitutes another “natural” simple case. Performance is assumed to satisfy $x = a + \varepsilon$, where ε is normally distributed with zero mean and variance σ^2 . The principal is assumed to be risk neutral, while the agent has a utility function :

$$U(w, a) = -e^{-r(w-c(a))}$$

where r is the (constant) degree of absolute risk aversion ($r = -U''/U'$), and $c(a) = \frac{1}{2}ca^2$.

We restrict attention to linear contracts:

$$w = \phi x + \beta.$$

A principal trying to maximize his expected payoff will solve :

$$\max_{a, \phi, \beta} E(x - w)$$

subject to :

$$E(-e^{-r(w-c(a))}) \geq U(\bar{w})$$

and

$$a \in \arg \max_a E(-e^{-r(w-c(a))})$$

where $U(\bar{w})$ is the default utility level of the agent, and \bar{w} is thus its certain monetary equivalent.

Even though we only cover the simplest one-shot single task problem in these notes, it should be pointed out that the main reason for adopting these functional assumptions is the ease with which the setup generalizes. The two papers by Milgrom and Holmström apply

this framework to the study of optimal contracts in a dynamic setting and in a model with multiple tasks. These articles are essential reading for anybody that wishes to study contract theory further.

Certainty Equivalent

The certainty equivalent w of random variable x is defined as follows

$$u(w) = \mathbb{E}[u(x)]$$

The certainty equivalent of a normally distributed random variable x under CARA preferences, i.e. the w which solves

$$-e^{-rw} = \mathbb{E}(-e^{-rx})$$

has a particularly simple form, namely^a

$$w = \mathbb{E}[x] - \frac{1}{2}r\sigma^2. \quad (5)$$

The difference between the mean of random variable and its certain equivalent is referred to as the risk premium:

$$\frac{1}{2}r\sigma^2 = \mathbb{E}[x] - w$$

^aTo see this recall that a Normal variable with mean μ and variance σ^2 has density $k(\sigma^2) \exp\left\{-\left(\frac{x-\mu}{\sigma}\right)^2\right\}$. Hence the expected utility of a random variable

$$\tilde{w} \text{ with normal distribution is simply } k(\sigma^2) \int \exp\{-\gamma x\} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx$$

$$= k(\sigma^2) \int \exp\left\{-\gamma x - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx$$

$$= k(\sigma^2) \int \exp\left\{-\left(\frac{x-\mu+\gamma\sigma^2}{\sigma}\right)^2 - \mu\gamma + \frac{1}{2}\gamma^2\sigma^2\right\} dx$$

$$= \exp\left\{-\gamma\left(\mu - \frac{1}{2}\gamma\sigma^2\right)\right\}.$$

Rewriting Incentive and Participation Constraints

Maximizing expected utility with respect to a is equivalent to maximizing the certainty equivalent wealth $\hat{w}(a)$ with respect to a , where $\hat{w}(a)$ is defined by

$$-e^{-r\hat{w}(a)} = \mathbb{E}(-e^{-r(w-c(a))})$$

Hence, the optimization problem of the agent is equivalent to:

$$a \in \arg \max \{ \hat{w}(a) \} =$$
$$\arg \max \left\{ \phi a + \beta - \frac{1}{2}ca^2 - \frac{r}{2}\phi^2\sigma^2 \right\}$$

which yields

$$a^* = \frac{\phi}{c}$$

Inserting a^* into the participation constraint

$$\phi \frac{\phi}{c} + \beta - \frac{1}{2}c \left(\frac{\phi}{c} \right)^2 - \frac{r}{2}\phi^2\sigma^2 = \bar{w}$$

yields an expression for β ,

$$\beta = \bar{w} + \frac{r}{2}\phi^2\sigma^2 - \frac{1}{2}\frac{\phi^2}{c}$$

This gives us the agent's effort for any performance incentive ϕ . The principal then solves :

$$\max_{\phi} \left\{ \frac{\phi}{c} - \left(\bar{w} + \frac{r}{2}\phi^2\sigma^2 + \frac{1}{2}\frac{\phi^2}{c} \right) \right\}$$

The first order conditions are

$$\frac{1}{c} - \left(r\phi\sigma^2 + \frac{\phi}{c} \right) = 0,$$

which yields :

$$\phi^* = \frac{1}{1 + rc\sigma^2}$$

Effort and the variable compensation component thus go down when c (cost of effort), r (degree of risk aversion), and σ^2 (randomness of performance) go up, which is intuitive. The constant part of the compensation will eventually decrease as well as r , c or σ^2 become large, as

$$\beta = \bar{w} + \left(\frac{1}{2} \frac{r\sigma^2 - \frac{1}{c}}{(1 + rc\sigma^2)^2} \right).$$

Additional Signals

Return to the general finite model.

Suppose besides the outcome, the principal can observe some other signal, say $y \in Y = \{y_1, \dots, y_l, \dots, y_L\}$, which could inform him about the performance of the agent. Denote then the probability of an outcome (x_k, y_l) given action choice a_i by p_{ki}^l and the corresponding wage by w_k^l . The form of the optimal contract can then be verified to be

$$\frac{1}{u'(w_k^l)} = \lambda + \sum_{j \neq i} \mu_j \left(1 - \frac{p_{kj}^l}{p_{ki}^l} \right).$$

Thus the contract should integrate the additional signal y if there exists some x_j, x_k and y_l such that for some a_i

$$\frac{p_{ki}^l}{p_{ji}^l} \neq \frac{p_{ki}}{p_{ji}},$$

but the inequality simply says that x is not a sufficient statistic for (x, y) with respect to a . Recall that x is sufficient for (x, y) with respect to a if we can write the conditional probability of (x, y) given a_j as follows

$$f(x_i, y_l | a_j) = h(x_i, y_l) g(x_i | a_j),$$

where $h(x_i, y_l)$ is an arbitrary function. Thus, Hart-Holmström (1987) write:

The additional signal s will necessarily enter an optimal contract if and only if it affects the posterior assessment of what the agent did; or perhaps more accurately if and only if s influences the likelihood ratio.

Multi Agent Moral Hazard

Deterministic Team Models

One of the best studied forms of collective moral hazard problems comes under the name of partnership problems or team problems. The basic model there is due to early work by Radner and Marschak, and the modern treatment emphasizing the role of incentives is due to Holmström (1982).

N players must take an unobservable effort a_i at private cost $c_i(a_i)$ in order to produce output x .

The output is observable and verifiable, but depends on the collective efforts of all the players according to the production function:

$$x = f(a_1, \dots, a_N).$$

We assume that f is increasing, concave and differentiable.

We will also assume that c_i are twice continuously differentiable, strictly increasing and convex for all i .

The partners (or the team members) must design a sharing rule, i.e. the output x must be divided into $(x_1(x), \dots, x_N(x))$ in such a way that budget is balanced.

$$x = \sum_{i=1}^N x_i. \quad (6)$$

The preferences of the players are then given by:

$$u_i(x_i, a_i) = x_i - c_i(a_i). \quad (7)$$

The first best level of effort is given by the vector $a^* = (a_1^*, \dots, a_N^*)$, where

$$c'_i(a_i^*) = f_i(a^*) \text{ for all } i. \quad (8)$$

The basic question is then if it is possible to design the sharing rule in such a way that all players i have the right incentives to supply effort at the first best level.

If the first best level of effort for each player is strictly positive, then this question can be answered in the negative.

To see the basic intuition, observe that at the first best level of effort as determined in 8, the marginal cost of effort is positive.

In order to provide player i with the proper incentives, it must be the case that x_i increases by $f_i(a^*)$ units for each marginal increase of a single unit in a_i .

But since

$$\sum_j x_j(x(a)) = x(a) \text{ for all } a,$$

we have by differentiation w.r.t. a_i and evaluating at a^* :

$$\sum_j x'_j(x) f_i(a^*) = f_i(a^*).$$

Hence if the incentives for i are right, it must be the case that $x'_j(x) = 0$ for all $j \neq i$.

But then it is not optimal for any $j \neq i$ to choose a_j^* . Hence it is impossible to provide the right incentives for the players.

This argument gives only the intuition rather than an actual proof since it assumes differentiability of the sharing schemes.

Differentiability may play a role in the theory since it is well known that for discontinuous production functions, first best may be achievable.

To see how this can make a difference, consider an example where the effort of player 1 is nonproductive, i.e. $f_1(a) = 0$ for all a , and the dependence of player j 's effort for $j > 1$ is symmetric.

Then the following scheme achieves the first best. Let $x^* = f(a^*)$.

$$x_1(x) = \begin{cases} 0 & \text{if } x \geq x^*, \\ x & \text{if } x < x^*. \end{cases},$$
$$x_j(x) = \begin{cases} \frac{x}{N-1} & \text{if } x \geq x^*, \\ 0 & \text{if } x < x^*. \end{cases} \quad \text{for } j > 1.$$

In this example then player 1 becomes effectively the principal for the model, and the incentive scheme is a milder form of the shoot the agents scheme where a collective punishment is imposed when a deviation is observed.

Since there is no need to provide 1 with any incentives, this scheme overcomes the inherent difficulties in team problems.

The balanced budget requirement in 6 can be circumvented easily,

and in a sense this scheme should be regarded as selling the partnership to an outsider.

For the complete argument in the case where $a_i^* > 0$, consider a point $x' < x^*$, where x^* , where x^* denotes again the first best level of production. Since $f_i(a^*) > 0$ for all i ,

$$x_i(x') \leq x_i(x^*) - c'_i(a_i^*) \frac{(x^* - x')}{f_i(a^*)} \text{ for all } i,$$

if x' is sufficiently close to x^* . (Can you supply the proof?)

But then the argument for the differentiable case can be applied to consider the changes from x' to x^* .

Observe that the key difference to the example above is that it would be impossible to reward one of the players in the case of a fall in the output since all of the players can induce this fall through their own actions.

Principal and Many Agents

In this section, we maintain much of the machinery from the previous section, but we switch attention back to the principal agent formulation for moral hazard models. Assume now that x depends on a in a stochastic manner.

In other words, the vector of efforts a parametrizes the distribution of the output:

$$x \sim F(x | a).$$

We assume that all of the distribution functions are sufficiently many times differentiable. Denote by z the vector of stochastic variables whose realizations are contractible and depend on the effort choice a according to:

$$z \sim G(z | a).$$

In many models, it will be convenient to assume that x is a subvector of z , but z may also contain other information that is correlated with the effort levels.

Since we no longer require budget balance, we switch back to the usual notation on wages $w_i(z)$, and we also reintroduce the possibility for risk aversion for the agents. The problem of the principal is then to solve the following maximization problem:

$$\max_{a,w} \int_x (E(x|a,z) - \sum_{i=1}^N w_i(z)) g(z|a) dz$$

subject to

$$\int_x u_i(w_i(z)) g(z|a) dz - c_i(a_i) \geq U_i \text{ for all } i,$$

$$a_i \in \arg \max_{a'_i \in a_i} \int_x u_i(w_i(z)) g(z|a_{-i}, a'_i) dz - c_i(a'_i) \text{ for all } i.$$

The constraints are again the IR and IC constraints for the agents, but it should be noted that each agent takes the effort choice of other agents as given.

Hence the assumption is that the principal gets to choose the equilibrium that the agents play, and the analysis is thus in accordance with our analysis in the adverse selection case.

It is important to observe that the efforts chosen by one player may make information about other players' choices more accurate. Thus a modification of the single person sufficiency notion is required.

Definition 2 *A statistic $T_i(z)$ is sufficient for z with respect to a_i if there are positive functions ϕ and ζ such that*

$$g(z|a) = \zeta_i(z, a_{-i}) \phi_i(T_i(z), a) \text{ for all } z \text{ and } a.$$

The following theorem is proved exactly as the corresponding result in the single agent case.

Theorem 3 *If $T_i(z)$ is sufficient for z with respect to a_i for all i , then there for each wage profile $\{w_i(z)\}_{i=1}^N$, there is another wage profile $\{\tilde{w}_i(z)\}_{i=1}^N$ that Pareto dominates $\{w_i(z)\}_{i=1}^N$.*

The intuition of this result runs as follows. From Blackwell's theorem, we know that a random variable is sufficient for another if and only if the latter is a garbling of the first. In other words, the second must be the first plus noise.

As long as the agents are risk averse adding this noise to the compensation package makes the agent strictly worse off. A converse to this theorem is also possible.

Definition 4 *A statistic $T_i(x)$ is globally insufficient for i if for all a, T_i*

$$\frac{g_{a_i}(z|a)}{g(z|a)} \neq \frac{g_{a_i}(z'|a)}{g(z'|a)} \text{ for almost all } z, z' \in \{z | T_i(z) = T_i\}.$$

Holmström (1982) proves the following theorem.

Theorem 5 *If $\{w_i(T_i(z))\}$ is a wage profile where $T_i(z)$ is globally insufficient for some i , then there exist wage schedules $\{\hat{w}_i(z)\}$ that yield a strict Pareto improvement on $\{w_i(T_i(z))\}$ and induce the same action.*

These two theorems are useful in analyzing optimal compensation schemes in environments with many agents.

For concreteness, we may think about many salesmen selling a new product in a number of different geographic areas.

Suppose that the effort of each salesman results in a level of sales given by

$$x_i(a_i, \theta_i),$$

where θ_i is a random variable that represents the business conditions in the area of salesman i .

Assume also that z contains (x_1, \dots, x_N) .

In this case, it is possible for the principal to base individual contracts on individual sales (since the efforts of others do not affect the sales). This will, however, be in general suboptimal.

The reason for this is that if the θ_i are correlated (as is presumably the case for this particular example), x_j is statistically dependent on x_i even if the effort choices are known (as is the case in equilibrium analysis).

Hence x_i is globally insufficient for i and the theorem above shows that the compensation scheme can be improved.

This observation is generally used as the justification for compensations based on relative performances within organizations.