CHAPTER 2: Moral Hazard with Multiple Agents*

This chapter analyzes optimal incentive schemes under moral hazard when a principal interacts with many agents. In a partnership/firm the output $Q = (q_1, \ldots, q_n)$ is jointly affected the efforts $a = (a_1, \ldots, a_n)$ of the n agents, where q_i is the individual output of agent i. Let $F(Q|a)$ be the joint conditional distribution of outputs. Throughout we will assume that the principal is risk neutral and agent i has separable preferences over salary w_i and effort a_i as follows

$$U_i(w, a) = u_i(w) - \psi_i(a), \text{ with } u_i''(.) \leq 0.$$

We will analyze the optimal incentive schemes under two different situations: when Q, the aggregate output is observed publicly, and when the individual outputs, q_i's are observed.

1 Moral Hazard in Teams

This section is based on Hölmstrom (1982). In a firm only the single aggregate and deterministic output is observed publicly, which is given by the production function $Q = Q(a)$. We assume that $Q_i(a) > 0$, $Q_{ii}(a) < 0$, $Q_{ij} \geq 0$ and the Hessian matrix $\nabla Q(a)$ is negative semi-definite. Since the output is deterministic, without loss of generality we can assume that the agents are risk neutral. A partnership is defined by the following scheme of sharing rules

$$w(Q) = [w_1(Q), \ldots, w_n(Q)].$$

To start with we impose the following budget balance condition

$$\sum_{i=1}^{n} w_i(Q) = Q, \text{ for all } Q. \tag{BB}$$

Here arises the free riding problem: someone else works hard, I gain. Let the first best efforts are given by a^* which are given by

$$a^* = \text{argmax}_{a \in [0,\infty)^n} \left\{ Q(a) - \sum_{i=1}^{n} \psi_i(a_i) \right\}. \tag{1}$$

The first order conditions are given by

$$Q_i(a^*) = \psi_i'(a_i^*), \text{ for all } i = 1, \ldots, n. \tag{FB}$$

*These notes are heavily borrowed from (Bolton and Dewatripont, 2005, Chapter 8).

Further, let a^N be the efforts chosen by the agents in a Nash equilibrium, which are given by

$$a_i^N = \operatorname{argmax}_{a_i \in [0,\infty)} \{w_i(Q(a)) - \psi_i(a_i)\} \text{ for all } i = 1, \dots, n. \tag{2}$$

The NE efforts are characterized by

$$w_i'(Q(a^N))\, Q_i(a^N) = \psi_i'(a_i^N), \text{ for all } i = 1, \dots, n. \tag{NE}$$

Now the question is whether $a^* = a^N$. Clearly, (FB) and (NE) are not compatible unless $w_i'(Q(a)) = 1$. This condition boils down to

$$w_i'(Q) = Q(a) + C_i, \text{ for all } i = 1, \dots, n. \tag{3}$$

For the above to be compatible with (BB) one thus needs to introduce a third party, called the budget breaker, who can write binding contracts with the n agents to receive transfers $t_i = -C_i$ from agent i. Hölmstrom (1982) claims that such a scheme exists and implements a^* in a Nash equilibrium. In order to show this we relax (BB) to a *no deficit* condition

$$\sum_{i=1}^{n} w_i(Q) \leq Q, \text{ for all } Q. \tag{ND}$$

From (3) and (ND) we have

$$\sum_{i=1}^{n} w_i(Q(a^*)) = nQ(a^*) - \sum_{i=1}^{n} t_i \leq Q(a^*). \tag{4}$$

Also

$$t_i = Q(a^*) - w_i(Q(a^*)) \leq Q(a^*) - \psi_i(a_i^*). \tag{5}$$

Thus for such t_i's to exist we need

$$\sum_{i=1}^{n} t_i \leq nQ(a^*) - \sum_{i=1}^{n} \psi_i(a_i^*) \leq \sum_{i=1}^{n} t_i + Q(a^*) - \sum_{i=1}^{n} \psi_i(a_i^*). \tag{6}$$

Notice that (4) and (5) imply that such a scheme is profitable for both the budget breaker and the agents. From the last equation, such transfers $t = (t_1, \dots, t_n)$ exist since $Q(a^*) - \sum_{i=1}^{n} \psi_i(a_i^*)$ is strictly positive, otherwise a^* would not be efficient. Notice that if the firm performs better then it hurts the budget breaker, since $w_{BB}(Q) = \sum_i t_i + Q - nQ$ implying that $w_{BB}'(Q) = -(n-1) < 0$ at $Q(a^*)$.

Are there other ways to support first best? Mirrlees contract does. Consider the following bonus scheme

$$w_i(Q) = \begin{cases} b_i & \text{if } Q = Q(a^*), \\ -k & \text{if } Q < Q(a^*). \end{cases}$$

Choose b_i's such that $b_i - \psi_i(a_i^*) > -k$ and $\sum_{i=1}^{n} b_i = Q(a^*)$. This is possible because of the fact that $Q(a^*) - \sum_{i=1}^{n} \psi_i(a_i^*) > 0$, otherwise a^* would not be efficient. Thus a^* is a Nash equilibrium. Mirrlees' contract can be interpreted as debt financing by the firm. Firm commits to repay debts of $D = Q(a^*) - \sum_{i=1}^{n} b_i$, and b_i to each i. If it cannot, creditors collect Q and each agent pays a penalty k.

Example 1 (Implementing the first best) Consider $n = 3$ with $a_i \in \{0, 1\}$ and $\psi_i(1) > \psi_i(0)$ for all $i =$

1, 2, 3. Let $Q^1 = Q(0, 1, 1)$, $Q^2 = Q(1, 0, 1)$ and $Q^3 = Q(1, 1, 0)$. Thus $a^* = (1, 1, 1)$. If $Q^1 \neq Q^2 \neq Q^3$, then the single shirker is identified and punished, and the other two are rewarded. Thus, first best efforts are implemented. If $Q^1 = Q^2 \neq Q^3$, then also it is possible to identify and reward the non-shirker, and hence first best can be implemented. ■

Example 2 (Approximate efficiency) Let $n = 2$, $A_1 = A_2 = [0, \infty)$, $Q = a_1 + a_2$ and $\psi_i(a_i) = a_i^2/2$. Then $a^* = (1, 1)$. Consider the following incentive scheme. For $Q \geq 1$, $w_1(Q) = (Q-1)^2/2$ and $w_2(Q) = Q - w_1(Q)$, and for $Q < 1$, $w_1(Q) = Q$ and $w_2(Q) = -k$. For sufficiently high k this scheme supports an NE where agent 2 chooses $a_2 = 1$ and agent 1 chooses $a_1 = 1$ with probability $1 - \varepsilon$ and $a_1 = 0$ with probability ε. To see that this is indeed an NE note the following. Given the action choice $a_2 = 1$, agent 1's best response is found by solving

$$\max_{a_1} \left\{ w_1(a_1 + 1) - \frac{a_1^2}{2} \right\} = \max_{a_1} \left\{ \frac{a_1^2}{2} - \frac{a_1^2}{2} \right\} = 0.$$

Hence, agent 1 is indifferent among any actions. Thus the above randomization is his best response. As for agent 2, $a_2 = 1$ guarantees $Q \geq 1$. His payoff is given by

$$(1-\varepsilon)\left(2 - \frac{1}{2}\right) + \varepsilon(1-0) - \frac{1}{2} = 1 - \frac{\varepsilon}{2}.$$

If agent 2 chooses $a_2 \in [0, 1)$, then $Q < 1$ with probability ε. Thus his payoff is

$$(1-\varepsilon)\left(1 + a_2 - \frac{a_2^2}{2}\right) - \varepsilon k - \frac{a_2^2}{2} \leq 1 + a_2 - a_2^2 - \varepsilon k,$$

which is maximized at $a_2 = \frac{1}{2}$. Thus at $a_2 = \frac{1}{2}$, agent 2's payoff is $(5/4) - \varepsilon k$. Therefore, $a_2 = 1$ is optimal if

$$k \geq \frac{1}{2} + \frac{1}{4\varepsilon}.$$

This proves the assertion that the first best can be implemented approximately. This example is based on Legros and Matthews (1993). ■

2 Relative Performance Evaluation

Consider the following model as in Hölmstrom (1982).

$$q_1 = a_1 + \varepsilon_1 + \alpha\varepsilon_2,$$
$$q_2 = a_2 + \varepsilon_2 + \alpha\varepsilon_1,$$

where ε_1 and ε_2 are iid Normal random variables with mean zero and variance σ^2. The agents have CARA utility functions.

$$u(w_i, a_i) = -e^{-\eta[w_i - \psi(a_i)]}, \text{ where } \psi(a_i) = \frac{ca_i^2}{2} \text{ for } i = 1, 2.$$

Linear incentive schemes:

$$w_1 = z_1 + v_1 q_1 + u_1 q_2,$$
$$w_2 = z_2 + v_2 q_2 + u_2 q_1,$$

The absence of relative performance evaluation implies $u_1 = u_2 = 0$. Given the symmetry of the principal's problem we need to solve only for an individual optimal scheme (a_i, w_i). This the principal will solve

$$\max_{\{a_i, z_i, v_i, u_i\}} E(q_i - w_i) \tag{M}$$

$$\text{subject to } E\left(-e^{-\eta[w_i - \psi(a_i)]}\right) \geq u(\bar{w}), \tag{IR$_i$}$$

$$a_i = \operatorname{argmax}_{\hat{a}_i} E\left(-e^{-\eta[w_i - \psi(\hat{a}_i)]}\right). \tag{IC$_i$}$$

The certainty equivalent wealth of agent i with respect to a, $\hat{w}_i(a)$, is defined by:

$$-e^{-\eta \hat{w}_i(a)} = E\left(-e^{-\eta[w_i - \psi(a_i)]}\right).$$

Notice that

$$Var[v_i(\varepsilon_i + \alpha\varepsilon_j) + u_i(\varepsilon_j + \alpha\varepsilon_i)] = \sigma^2\left[(v_i + \alpha u_i)^2 + (u_i + \alpha v_i)^2\right].$$

Then agent i's incentive constraint boils down to choosing a_i to maximize the certainty equivalent wealth:

$$\max_{\hat{a}_i}\left\{z_i + v_i\hat{a}_i + u_i a_j - \frac{c\hat{a}_i^2}{2} - \frac{\eta\sigma^2}{2}\left[(v_i + \alpha u_i)^2 + (u_i + \alpha v_i)^2\right]\right\}. \tag{IC$'_i$}$$

The above implies that

$$a_i = \frac{v_i}{c}.$$

Substituting for a_i in the principal's objective function and the agent's binding participation constraint one can reduce the principal's problem to

$$\max_{\{z_i, v_i, u_i\}}\left\{\frac{v_i}{c} - \left(z_i + \frac{v_i^2}{c} + \frac{u_i v_j}{c}\right)\right\}.$$

$$\text{subject to} \qquad z_i + \frac{v_i^2}{2c} + \frac{u_i v_j}{c} - \frac{\eta\sigma^2}{2}\left[(v_i + \alpha u_i)^2 + (u_i + \alpha v_i)^2\right] = \bar{w}.$$

Or, substituting for z_i from the participation constraint,

$$\max_{\{v_i, u_i\}}\left\{\frac{v_i}{c} - \frac{v_i^2}{2c} - \frac{\eta\sigma^2}{2}\left[(v_i + \alpha u_i)^2 + (u_i + \alpha v_i)^2\right]\right\}.$$

The principal's problem can be solved sequentially. (1) For a given v_i, determine u_i to minimize the variance. (2) The variable v_i is then set optimally to trade off risk sharing and incentives. Minimizing the variance with respect to u_i yields

$$u_i = -\left(\frac{2\alpha}{1+\alpha^2}\right)v_i.$$

The above formula implies that the optimal u_i is negative when the two agents' outputs are positively correlated, i.e., $\alpha > 0$. In other words, an agent is penalized for the better performance of the other agent. A better performance by agent j is likely to be due to a high realization of ε_j, which also positively affect the output of agent i. By setting u_i negative, the optimal incentive scheme reduces agent i's exposure to a common shock affecting both agents' output, and thus reduces the variance of agent i's compensation.

In the second step, substitute the above formula and solve for optimal v_i:

$$\max_{\{v_i\}} \left\{ \frac{v_i}{c} - \frac{v_i^2}{2c} - \frac{\eta\sigma^2\left(1-\alpha^2\right)^2 v_i^2}{2\left(1+\alpha^2\right)} \right\}.$$

The first order condition with respect to v_i yields

$$v_i = \frac{1+\alpha^2}{1+\alpha^2+\eta c\sigma^2\left(1-\alpha^2\right)^2}.$$

For $\alpha = 0$, the above formula reduces to the formula for the optimal share in the one-agent case. Notice that $v_i(\alpha)$ is convex, $v_i(-1) = v_i(1) = 1$ and the minimum is reached at $\alpha = 0$. A situation of perfect correlation is similar to first best, and hence the agent gets the full share. The reason for using relative performance evaluation is not to induce higher effort through greater competition, but to induce higher effort by lowering their risk exposure. Interested students should refer to Mookherjee (1984) for a more general model of relative performance evaluation contracts under moral hazard with many agents.

3 Tournaments

This subsection is based on Lazear and Rosen (1981), which consider a situation two risk neutral agents 1 and 2 produce individual outputs that are independently distributed. Obviously, following the previous analysis there is no reason why tournaments would be efficient incentive schemes. However, Lazear and Rosen (1981) show that the first-best outcome can be implemented using a tournament. Let

$$q_i = a_i + \varepsilon_i \;\text{ for }\; i = 1, 2, \;\text{ where }\; \varepsilon_i \sim F(0, \sigma^2).$$

The first-best efforts are given by

$$\psi'(a_1^*) = \psi'(a_2^*) = 1.$$

Consider the following incentive scheme:

$$w_i = z + q_i$$

At the first-best the expected utility of agent i is given by

$$z + E(q_i) - \psi(a^*) = z + a^* - \psi(a^*) = \bar{u}.$$

Now consider a symmetric "tournament": If $q_i > q_j$, then agent i is paid $z+W$ and agent j is paid only z. Under this scheme, the expected payoff for an agent i for efforts (a_i, a_j) is

$$z + pW - \psi(a_i),$$

where p is the probability that agent i is the winner, which is given by:

$$p = \text{Prob.}[q_i > q_j] = \text{Prob.}[a_i - a_j > \varepsilon_i - \varepsilon_j] = H(a_i - a_j),$$

where $H(\cdot)$ is the cumulative distribution of $(\varepsilon_i - \varepsilon_j)$ which has mean 0 and variance $2\sigma^2$. The best response for an agent under tournament is given by

$$W\frac{\partial p}{\partial a_i} = \psi'(a_i) \implies Wh(a_i - a_j) = \psi'(a_i).$$

In a symmetric NE, effort levels by both agents are identical. Thus in order to implement a^*, the prize must be

$$W = \frac{1}{h(0)},$$

The fixed wage z can be set to satisfy the following condition:

$$z + \{\text{Prob.}[0 = a_i^* - a_j^* > \varepsilon_i - \varepsilon_j]\}W - \psi(a^*) = z + \frac{H(0)}{h(0)} - \psi(a^*) = \bar{u}.$$

This is the same as the first-best with the piece rate scheme.

References

Bolton, Patrick and Mathias Dewatripont (2005), *Contract Theory*. The MIT Press.

Hölmstrom, Bengt (1982), "Moral Hazard in Teams." *Bell Journal of Economics*, 13, 324–340.

Lazear, Edward and Sherwin Rosen (1981), "Rank-Order Tournaments as Optimum Labor Contracts." *Journal of Political Economy*, 89, 841–864.

Legros, Patrick and Stephen Matthews (1993), "Efficient and Nearly Efficient Partnerships." *The Review of Economic Studies*, 60, 599–611.

Mookherjee, Dilip (1984), "Optimal Incentive Schemes with Many Agents." *The Review of Economic Studies*, 51, 433–446.