ECON 7050: TOPIC 5. THE PRINCIPAL-AGENT PROBLEM

In many economic situations, an individual may hire another individual to perform some tasks. This is often called a Principal-agent relationship. Typically, the agent has better information about the exact actions she would take and/or the environment she works in. For instance, a manager of a firm, who works for an owner, usually knows better than the owner about how hard she is working and what are the changes in the market conditions. In the first case, where a principal cannot observe the actions of an agent, the problem is referred to as hidden actions or moral hazard. In the second case, where the agent has superior information about the environment, the problem is usually referred to as hidden information. The study of such principal-agent problems has been an active area of research in microeconomics and industrial organization theory in recent years.

1 Hidden Actions (Moral Hazard)

The owner of a firm hires a manager for a one-time project. The profit of the firm depends on both the manager’s actions and some random factors. Think of the manager’s actions as effort, denoted by $e$. The firm’s profit can take values on $[\pi, \bar{\pi}]$, with conditional density function as $f(\pi | e)$ and $f(\pi | e) > 0$ for all $e$ and all $\pi \in [\pi, \bar{\pi}]$. Assume that $e \in \{e_H, e_L\}$, with $e_H > e_L \geq 0$. Also assume that $\pi$ increases in $e$ in the sense of first-order stochastic dominance; that is, the distribution functions satisfy $F(\pi | e_H) \leq F(\pi | e_L)$ for all $\pi \in [\pi, \bar{\pi}]$ with strict inequality on some open subset of $[\pi, \bar{\pi}]$. This implies that the level of expected profits when the manager chooses $e_H$ is higher than that from $e_L$.

$\int \pi f(\pi | e_H) d\pi > \int \pi f(\pi | e_L) d\pi$.

The manager’s utility function is assume to be $u(w, e) = v(w) - g(e)$, where $w$ is the wage she receives, with $v'(w) > 0$, $v''(w) < 0$, and $g(e_H) > g(e_L)$. Thus the manager is risk averse. (You should also read in the book about the case where the
manager is risk-neutral, in which case the problem of providing incentives is easy to solve.) The reservation utility of the manager is \( \pi \).

The owner is risk neutral and aims to maximize the expected value of \( \pi - w \). The owner is able to make a take-it-or-leave-it offer to the manager.

**The optimal contract when effort is observable.**

To begin with, consider the benchmark where \( e \) is observable. The owner then offer a contract \((w^*(\pi), e^*)\) that solves the following problem:

\[
\max_{e \in \{e_H, e_L\}, w(\pi)} \int [\pi - w(\pi)] f(\pi \mid e) d\pi
\]

subject to

\[
\int v(w(\pi)) f(\pi \mid e) d\pi - g(e) \geq \pi.
\]

It is convenient to solve this problem in two stages: first, at each given level of \( e \), what is the optimal \( w(\pi) \)? And second, what is the optimal \( e \) the owner wants from the manager?

For any given \( e \), the above problem is equivalent to

\[
\min_{w(\pi)} \int w(\pi) f(\pi \mid e) d\pi
\]

subject to

\[
\int v(w(\pi)) f(\pi \mid e) d\pi - g(e) \geq \pi.
\]

Note that the constraint must be binding in the optimal. The first-order condition is

\[- \int f(\pi \mid e) d\pi + \gamma \int v'(w(\pi)) f(\pi \mid e) d\pi = 0,
\]

or

\[
\int [-f(\pi \mid e) + \gamma v'(w(\pi)) f(\pi \mid e)] d\pi = 0.
\]

Since \( w(\pi) \) needs to be optimal for each \( \pi \), the first-order condition must hold for each \( \pi \), and thus the first-order condition becomes

\[-f(\pi \mid e) + \gamma v'(w(\pi)) f(\pi \mid e) = 0,
\]
or
\[
\frac{1}{v'(w(\pi))} = \gamma,
\]
where \(\gamma\) is the Lagrangian multiplier.

Since \(v''(w) < 0\), there can only be one \(w_e\) that solves the f.o.c., independent of \(\pi\), for any given \(e\). In other words, the optimal contract must be such that the manager be paid with a fixed wage. This result is quite intuitive: since \(e\) is observable, there exists no problem of providing incentives, and thus to implement any \(e\) it is best for the risk-neutral owner to insure the risk-averse manager.

Thus, given any \(e\), in the optimal contract the owner offers a fixed wage \(w^*_e\) such that
\[
v(w^*_e) - g(e) = \bar{u},
\]
or
\[
w^*_e = v^{-1}(\bar{u} + g(e))
\]
The optimal choice of \(e\) then solves
\[
\max_e \int \pi f(\pi | e) d\pi - v^{-1}(\bar{u} + g(e)).
\]

**The optimal contract when effort is not observable.**

When effort is unobservable, there arises a conflict between providing insurance and providing incentives to the manager. Since the manager is risk averse, optimal insurance requires that the manager be paid with a fixed wage. But then the manager will have little incentive to exert any effort beyond the minimum level. This can lead to inefficiencies in the optimal contract (which would not arise if effort is observable or if the manager is risk neutral).

We again study the optimal contract in two stages: how to optimally implement any specific level of \(e\) and what is the optimal \(e\) that the owner should implement.
To implement a specific level of \( e \), the owner solves

\[
\min_{w(\pi)} \int w(\pi)f(\pi \mid e)d\pi
\]

\[
s.t. \ (i) \ \int v(w(\pi))f(\pi \mid e)d\pi - g(e) \geq \pi
\]

\[
(ii) \ e \text{ solves } \max_{e'} \int v(w(\pi))f(\pi \mid e')d\pi - g(e').
\]

The difference of this program from the one under observable \( e \) is that, in addition to the individual rationality constraint, there is also an incentive compatibility condition given by (ii) now. The incentive compatibility condition ensures that the manager will indeed choose \( e \) as desired by the owner.

Suppose, first, that the owner wants to implement \( e = e_L \). Then it is easy to see that the optimal wage offer should be \( w^*_e = v^{-1}(\pi + g(e_L)) \), the same contract as if \( e \) is observable. The manager will accept the offer and indeed choose \( e_L \), so both (i) and (ii) above are satisfied. Moreover, \( w^*_e \) must be optimal for the owner since it is optimal in the program same as above except when (ii) is not required.

Next, suppose that the owner wants to implement \( e = e_H \). Constraint (ii) can then be written as

\[
(ii_H) \ \int v(w(\pi))f(\pi \mid e_H)d\pi - g(e_H) \geq \int v(w(\pi))f(\pi \mid e_L)d\pi - g(e_L).
\]

Let \( \gamma \geq 0 \) and \( \mu \geq 0 \) denote the multipliers on constraints (i) and (ii\(_H\)), respectively, \( w(\pi) \) must satisfy the following Kuhn-Tucker first order conditions at every \( \pi \in [\bar{\pi}, \overline{\pi}] \):

\[
-f(\pi \mid e_H) + \gamma v'(w(\pi))f(\pi \mid e_H) + \mu[f(\pi \mid e_H) - f(\pi \mid e_L)]v'(w(\pi)) = 0,
\]

or

\[
\frac{1}{v'(w(\pi))} = \gamma + \mu \left[1 - \frac{f(\pi \mid e_L)}{f(\pi \mid e_H)}\right].
\]
Lemma 1 In any solution to the owner’s problem of choosing the optimal \( w(\pi) \) with \( e = e_H \), both \( \gamma > 0 \) and \( \mu > 0 \).

Proof. Suppose that \( \gamma = 0 \). Because \( F(\pi \mid e_H) \) first order stochastically dominates \( F(\pi \mid e_L) \), there must exist an open subset of \([\pi, \pi]\) such that \( f(\pi \mid e_L) > f(\pi \mid e_H) \) for all \( \pi \) on this subset. But if \( \gamma = 0 \), the Kuhn-Tucker first-order condition then implies that \( v'(w(\pi)) \leq 0 \) at any such \( \pi \), which is impossible. Hence, \( \gamma > 0 \).

On the other hand, if \( \mu = 0 \), then the optimal \( w(\pi) \) would be a fixed wage payment for every \( \pi \) realization. But then the manager would choose \( e_L \) rather than \( e_H \), violating condition (ii). Hence, \( \mu > 0 \).

This tells us that both constraints (i) and (ii) bind when \( e = e_H \). This also gives us some additional insights on the properties of the optimal compensation for \( e = e_H \).

Consider the fixed wage payment \( \hat{w} \) such that \( \frac{1}{v'(w)} = \gamma \). The optimal compensation has the following feature:

\[
\begin{align*}
w(\pi) > \hat{w} & \text{ if } \frac{f(\pi \mid e_L)}{f(\pi \mid e_H)} < 1 \\
\text{and} \\
w(\pi) < \hat{w} & \text{ if } \frac{f(\pi \mid e_L)}{f(\pi \mid e_H)} > 1
\end{align*}
\]

Intuitively, the optimal compensation scheme pays more than \( \hat{w} \) for profits that are more likely to occur under \( e_H \) than under \( e_L \) in the sense that the likelihood ratio \( \frac{f(\pi \mid e_L)}{f(\pi \mid e_H)} \) is less than 1. Similarly, it offers less compensation for profits that are more likely to occur under \( e_L \). This provides the manager with an incentive to choose \( e_H \).

If the likelihood ratio decreases monotonically in \( \pi \), then the optimal \( w(\pi) \) monotonically increases in \( \pi \). When the likelihood ratio is not monotonic, however, \( w(\pi) \) will not be monotonic either. In this case, a higher profit realization may call for a lower wage payment in the optimal compensation scheme. An example is shown below.
The optimal compensation is generally a quite complicated function of \( \pi \). This does not square well with the empirical evidence that most compensation scheme is rather simple and often a linear function of \( \pi \). However, Holmstrom and Milgrom (1987) has considered a more general model and shown that linear compensation contracts can actually be optimal.

One should also notice that the expected value of the manager’s wage payment for \( e = e_H \) must exceed the fixed wage payment when \( e \) is observable, \( w^*_{e_H} = v^{-1}(\bar{\pi} + g(e_H)) \). This is because the manager now bears some risk, thus she must be paid a higher expected wage to compensate the risk.

Now that we have found how the owner can optimally implement different levels of \( e \), the remaining question is what level of \( e \) the owner should implement. This is easy. Let the optimal compensation implementing \( e = e_H \) be \( w^*(\pi) \), then implementing \( e_H \) is optimal for the owner if and only if

\[
\int [\pi - w^*(\pi)] f(\pi \mid e_H) d\pi \geq \int [\pi - v^{-1}(\bar{\pi} + g(e_L))] f(\pi \mid e_L) d\pi.
\]
2. Hidden Information

We next consider optimal contract design by a principal when the agent has private information about his “types”.

Again suppose that the owner wants to hire a manager to run a one-time project. The manager’s effort level, $e$, is now fully observable, but different managers may have different costs to exert the same effort. That is, the manager has private information about his costs of effort, or abilities.

Assume:

Gross profit from the project is $\pi(e)$, with $\pi'(e) > 0$, and $\pi''(e) < 0$ for all $e$.

The manager’s utility is

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

where $g(0, \theta) = 0$ for all $\theta$,

$$g_e(e, \theta) \begin{cases} > 0 & \text{for } e > 0 \\ = 0 & \text{for } e = 0 \end{cases}$$

$$g_{ee}(e, \theta) > 0 \text{ for all } e$$

$$g_{\theta}(e, \theta) < 0 \text{ for all } \theta$$

$$g_{e\theta}(e, \theta) \begin{cases} < 0 & \text{for } e > 0 \\ = 0 & \text{for } e = 0 \end{cases}.$$

Thus, the manager’s utility decreases in effort at a decreasing rate, and the cost and the marginal cost of effort decreases in $\theta$. Assume $\theta \in \{\theta_L, \theta_H\}$, with $\theta_L < \theta_H$, and $\text{prob}(\theta = \theta_H) = \lambda \in (0, 1)$. $\theta$ is realized after the contract is signed.

Also assume $v'(\cdot) > 0$ and $v''(\cdot) < 0$. 

The State of $\theta$ is observable.—

Consider a benchmark where $\theta$ is observable. A contract in this case consists of two wage-effort pairs: $(w_H, e_H)$ for $\theta_H$ and $(w_L, e_L)$ for $\theta_L$.

The owner’s problem is

$$\max_{w_L, e_L \geq 0, w_H, e_H \geq 0} \lambda [\pi(e_H) - w_H] + (1 - \lambda) [\pi(e_L) - w_L]$$

$$\text{s.t. } \lambda v(w_H - g(e_H, \theta_H)) + (1 - \lambda) v(w_L - g(e_L, \theta_L)) \geq \bar{u}$$

In any solution $[(w_H^*, e_H^*), (w_L^*, e_L^*)]$ to the problem, the participation constraint must be binding. Let $\gamma \geq 0$ be the multiplier for the constraint. The solution satisfies

$$-\lambda + \gamma \lambda v'(w_H^* - g(e_H^*, \theta_H)) = 0.$$  \hspace{1cm}(1)

$$-(1 - \lambda) + \gamma (1 - \lambda) v'(w_L^* - g(e_L^*, \theta_L)) = 0.$$  \hspace{1cm}(2)

$$\lambda \pi'(e_H^*) - \gamma \lambda v'(w_H^* - g(e_H^*, \theta_H)) g_{e} (e_H^*, \theta_H) \leq 0, \text{ with } "=\text{" if } e_H^* > 0.$$  \hspace{1cm}(3)

$$(1 - \lambda) \pi'(e_L^*) - \gamma (1 - \lambda) v'(w_L^* - g(e_L^*, \theta_L)) g_{e} (e_L^*, \theta_L) \leq 0, \text{ with } "=\text{" if } e_L^* > 0.$$  \hspace{1cm}(4)

Combining (1) and (2),

$$v'(w_H^* - g(e_H^*, \theta_H)) = v'(w_L^* - g(e_L^*, \theta_L))$$

so the manager’s marginal utility from income is equalized across states. Further,

$$w_H^* - g(e_H^*, \theta_H) = w_L^* - g(e_L^*, \theta_L),$$

which implies

$$v(w_H^* - g(e_H^*, \theta_H)) = v(w_L^* - g(e_L^*, \theta_L)),$$

thus the manager’s utility is the same in both states. Given that the participation constraint is binding, this utility must equal to $\bar{u}$. 

8
Now consider the optimal effort levels in the two states. Since $g_e(0, \theta) = 0$ and $\pi'(0) > 0$, conditions (3) and (4) must hold in equality and $e_i^* > 0$ for $i = H, L$. Combining (1) with (3) and (2) with (4), we have

$$\pi'(e_i^*) = g_e(e_i^*, \theta_i) \text{ for } i = H, L.$$ 

Thus the optimal effort is such that marginal cost equals marginal benefit from the effort.

Thus the optimal contract under observable states is for the owner to fully insure the manager and require the manager to choose the effort level which equates marginal costs with marginal benefits.

Since $g_e e (e, \theta) < 0$ for $e > 0$, or the marginal cost of effort is lower for the higher type, we must have $e_H^* > e_L^*$.

**The State of $\theta$ is observed only by the manager.**

When $\theta$ is not observable to the owner, the owner has to balance the need to provide optimal insurance and the need to elicit the proper level of effort from the manager. If the manager is fully insured, or receiving the same utility in both states, then since

$$v(w_L^* - g(e_L^*, \theta_H)) > v(w_L^* - g(e_L^*, \theta_L)) = v(w_H^* - g(e_H^*, \theta_H)),$$

the $\theta_H$ type would lie to be the $\theta_L$ type and select the effort level designed for the low type. Thus in general the optimal contract cannot fully insure the manager.

But what would be the optimal contract? To answer this question one needs to consider the set of contracts that are possible. In general, this set is very big. Fortunately, we can limit ourselves to a much smaller set, due to an important result known as

The Revelation Principle: Denote the set of possible states by $\Theta$. In searching for an optimal contract, the owner can without loss restrict himself to contracts of the following form:
1. After the state $\theta$ is realized, the manager is required to announce which state has occurred.

2. The contract specifies an outcome $[w(\hat{\theta}), e(\hat{\theta})]$ for each possible announcement $\hat{\theta} \in \Theta$.

3. In every state $\theta \in \Theta$, the manager finds it optimal to report the state truthfully.

We next simplify our analysis by assuming that the manager has infinite risk aversion: for the manager to be willing to accept the contract, he has to have a utility at least $\bar{u}$ in each state. The owner’s problem, called (PB), then becomes:

$$\max_{w_L, e_L \geq 0, w_H, e_H \geq 0} \lambda [\pi(e_H) - w_H] + (1 - \lambda) [\pi(e_L) - w_L]$$

s.t. (i) $w_H - g(e_H, \theta_H) \geq v^{-1}(\bar{u})$

(ii) $w_L - g(e_L, \theta_L) \geq v^{-1}(\bar{u})$

(iii) $w_H - g(e_H, \theta_H) \geq w_L - g(e_L, \theta_H)$

(iv) $w_L - g(e_L, \theta_L) \geq w_H - g(e_H, \theta_H)$

Conditions (i) and (ii) are the individual rationality constraints, and (iii) and (iv) are the incentive compatibility constraints.

**Lemma 2** We can ignore constraint (i) in solving (PB).

**Proof.** Whenever (ii) and (iii) are satisfied, we have

$$w_H - g(e_H, \theta_H) \geq w_L - g(e_L, \theta_H) \geq w_L - g(e_L, \theta_L) \geq v^{-1}(\bar{u}),$$

and thus (i) is satisfied. ■

Thus, in general (i) is not binding. The high type will receive positive rents in the optimal contract (as long as $e_L > 0$).
Lemma 3 In any optimal contract, \( w_L - g(e_L, \theta_L) = v^{-1}(\bar{u}) \).

Proof. If not, \( w_L - g(e_L, \theta_L) > v^{-1}(\bar{u}) \). Consider a small reduction of \( \varepsilon \) for both \( w_L \) and \( w_H \). (ii), (iii), (iv) will all still be satisfied, but the owner’s expected payoff is increased, contradicting the optimality of the original contract.

Thus the low type receives the reservation utility at the optimal contract. However, the high type will generally receive utility above the reservation level.

Lemma 4 In any optimal contract: \( e_L \leq e^*_L \) and \( e_H = e^*_H \), where \( e^*_L \) and \( e^*_H \) are the effort levels that would have been optimal if \( \theta \) is observable.

We can construct a graphical proof. The basic idea is that if \( (w_L, e_L) \) is such that \( e_L > e^*_L \), a change to \( (\hat{w}_L, e^*_L) \) will still satisfy the constraints but will increase the owner’s profit in the \( \theta_L \) state without changing its profit in the \( \theta_H \) state. On the other hand, given \( e_L \leq e^*_L \), \( (\hat{w}_H, e^*_H) \) will satisfy all the constraints and maximize the owner’s profit in the \( \theta_H \) state. This is a typical result called ”no distortion in the top”.
Lemma 5  In any optimal contract, $e_L < e^*_L$.

This is because if $e_L = e^*_L$, Lemma 3 implies $w_L = w^*_L$. But at $(w^*_L, e^*_L)$, a small decrease in $e_L$ and a corresponding decrease in $w_L$ has second-order effect on the owner’s profit in the $\theta_L$ state, but it relaxes the incentive constraint at state $\theta_H$ that has a first-order effect on the owner’s profit at the $\theta_H$ state. This is the distortion in effort choice when $\theta$ is not observable.

We can summarize the above discussion in a proposition.

How low should $e_L$ be? It needs to balance the trade off between decreasing profits in the $\theta_L$ state and increasing profits in the $\theta_H$ state. If we solve the problem by Kuhn-tucker conditions, we can obtain:

$$[\pi'(e_L) - g_e(e_L, \theta_L)] + \frac{\lambda}{1-\lambda} [g_e(e_L, \theta_H) - g_e(e_L, \theta_L)] = 0$$

(5)

In addition, we know in an optimal contract $e_H = e^*_H$, or

$$\pi'(e_H) = g_e(e_H, \theta_H)$$

(6)
These two equations will determine the optimal $e_L$ and $e_H$. We can then determine the optimal $w_L$ and $w_H$ by

$$w_L - g(e_L, \theta_L) = v^{-1}(\bar{u}),$$

and

$$w_H - g(e_H, \theta_H) = w_L - g(e_L, \theta_H).$$

The principal-agent problem is part of a more general problem of mechanism design: a principal designs a mechanism that induces agents to take certain actions that are most favorable to the principal. Some of the areas this approach has found many applications include: the design of optimal selling mechanisms by a monopolist or the government; design of regulation regimes; corporate governance; provision of public goods; design of optimal trade instruments; optimal compensation and labor contracts.