INTRODUCTION TO PRINCIPAL-AGENT THEORY UNDER MORAL HAZARD
Moral Hazard:

- Moral Hazard occurs in the Principal-Agent relationship when some *actions* of the agent are not perfectly observable.

- Contractual payment cannot depend on variables that are not observable by the Principal, the Agent, and by an outside (or third) party: a Judge.

- This poses the problem of performance measures. They may be more or less imprecise "signals" of the Agent’s true activity.
Moral Hazard and Risk

- *Important property.* In situations of *Moral Hazard*, the probabilities of outcomes, success or failure of projects, etc., depend on hidden actions of the agents.

- Uninformed parties (*i.e.*, Principals) must take into account this fact.

- To create effort incentives, contracts will typically involve an element of profit-sharing, and therefore impose risk upon the agent.

- Risky compensation hurts the risk-averse agent. This generates transactions costs.
Application of the Theory


- Delegation and Outsourcing; Public Procurement.

- Theory of Labor Contracts.

- Theory of Financial Intermediation (Banking), Insurance Markets,...

A Model of the Principal-Agent Relationship:


- We follow the presentation by David Kreps (1990).

- Agent chooses an action (or effort) $a$ in a set $A$.

- Principal observes an outcome $s$ in a set $S$.

- $A = \{a_1, a_2, ..., a_N\}$ ($A$ is a discrete finite set).

- $S = \{s_1, s_2, ..., s_M\}$ ($S$ is also finite).
Formal Representation of Moral Hazard and Risk:

- Define $\pi_{nm} = \text{Prob}[s_m \mid a_n]$

- $\sum_m \pi_{nm} = 1$ for all $n$.

- ASSUMPTION 1: We assume that $\pi_{nm} > 0$ for all $n$ and $m$. Any outcome is possible, under every action.

- The Agent is potentially risk averse. We assume,

$$U(w, a) = u(w) - a$$

where $w = \text{wage}$.

- ASSUMPTION 2: Risk aversion: $u : \mathbb{R} \to \mathbb{R}$ is strictly increasing, concave and continuously differentiable.
Definition of a Contract:

- A contract is a function $w : S \to \mathbb{R}$ or equivalently: $w = (w_1, w_2, ..., w_m)$, where $w_m$ is compensation if outcome $s_m$ is observed.

- Note: It is crucial here that the outcome $s_m$ is observable and verifiable. The Principal, the Agent, and possibly a Judge, can observe $s_m$ and check if $w_m = w(s_m)$ is paid to the Agent.
The Principal’s Objective and Information

- The Principal is risk neutral and observes the outcome (signal) $s_m$ only.

- His objective is expected profit, net of expected wages $E(w \mid a)$, that is,
  \[
  B(a) - E(w \mid a)
  \]

- $B(a_n) = \sum_m \pi_{nm}s_m = E(s \mid a_n)$

- Note $E(w \mid a_n) = \sum_m \pi_{nm}w_m$
The Principal-Agent Problem

• Choose \((a, w) \in A \times \mathbb{R}^M\) so as to maximize

\[ B(a) - E(w \mid a) \]

subject to the Agent’s participation constraint,

\[ E(u(w) \mid a) - a \geq U_o, \]

and subject to the incentive constraint,

\[ E(u(w) \mid a) - a = \max_{\alpha \in A} \{ E(u(w) \mid \alpha) - \alpha \}. \]

• Note: \(U_o\) is the Agent’s best outside option.
First-Best Effort:

- Compute the minimal certain reward for $a_n$, denoted $C^o(a_n)$
  \[ C^o(a_n) = u^{-1}[a_n + U_0] \]

- $C^o(a_n)$ is the cost of action $a_n$ if action $a$ is observable.

- The First-Best effort is a solution of the problem
  \[ \max_{a \in A} [B(a) - C^o(a)] \]

- Let $a^*$ be the first-best effort.
Transformation of the Principal-Agent Problem:

- Define, $x_m = u(w_m)$.

- Denote, $v = u^{-1}$.

- We have $w_m = v(x_m)$, and $v$ is convex.

- Then,

\begin{align*}
E(w_m \mid a_n) &= \sum_m \pi_{nm} v(x_m) \\
E(u(w_m) \mid a_n) &= \sum_m \pi_{nm} x_m.
\end{align*}
Cost of Incentives:

- The cost minimization problem: for a given action $a_n$,

$$\text{Minimize } \sum_{m} \pi_{nm} v(x_m)$$

subject to the participation constraint,

$$\sum_{m} \pi_{nm} x_m \geq U_0 + a_n \quad (IR)$$
and s.t. the *incentive constraints*,

\[
\sum_{m} \pi_{nm} x_m - a_n \geq \sum_{m} \pi_{\nu m} x_m - a_{\nu} \quad (IC_{\nu})
\]

for all \( \nu = 1 \ldots N \).

- Minimization is with respect to

\[
x = (x_1, x_2, \ldots, x_M).
\]
The solution of the above problem is a cost function $C : A \to \mathbb{R}_+$

$C(a_n) = \min_w \{ E(w \mid a_n) \mid (IR) \text{ and } (IC_\nu) \}$,

the smallest expected wage required to obtain an effort level $a_n$ from the Agent.

**Second-Best Effort** : a solution of

$$\max_{a \in A} [B(a) - C(a)]$$

We will show that the second-best solution may be different from $a^*$ since $C(a) > C^o(a)$. 
Efficient Risk-Sharing

• Assume that $a_n$ is observable, then, the optimal contract problem is just

$$\underset{x}{\text{Min}} \sum_{m} \pi_{nm} v(x_m)$$

s.t.,

$$\sum_{m} \pi_{nm} x_m \geq U_o + a_n$$
• Write the first-order conditions ($\lambda$ is a Lagrange multiplier):

$$v'(x_m) = \lambda \text{ for all } m$$

(Borch’s equation).

• This implies $x_1 = x_2 = \ldots = x_M = x^*$

(Full Insurance) or equivalently, $w_1 = w_2 = \ldots = w_M = w^*$.

•

$$x^* = U_o + a_n$$

$$w^* = v(U_o + a_n).$$
Proposition 1: If the agent is strictly risk averse (i.e., \( u'' < 0 \)), then, \( C^o(a_n) < C(a_n) \) for any action \( a_n \) that is more costly than some other action, i.e., such that \( a_n > \min_{A} a_{\nu} \).

Proof: Simple. If agent is risk averse, the unique efficient risk-sharing implies Full Insurance (as shown above). A constant wage implies that chosen effort is \( \min_{A} a_{\nu} \). Thus, \( IC_{\nu} \) is binding for some \( \nu \).
Proposition 2: The IR constraint is always binding at the optimum.

Proof: If not $\sum m \pi_{nm} x_m > U_o + a_n$ choose $\varepsilon > 0$ and $x'_m = x_m - \varepsilon$ for all $m$. If $\varepsilon$ small enough, IR and IC$_\nu$ are still satisfied. But expected wage $\sum m \pi_{nm} v(x'_m)$ is smaller: contradiction.

• Remark: $u(w)$ is not bounded below here.
Taking $IC_\nu$ constraints into account

- Lagrangian:
  \[ L(x, \lambda, \mu) = -\sum_m \pi_{nm}v(x_m) + \lambda \left( \sum_m \pi_{nm}x_m - a_n - U_0 \right) \]
  \[ + \sum_{\nu\neq n} \mu_\nu \left[ \sum_m (\pi_{nm} - \pi_{\nu m})x_m - a_n + a_\nu \right] \]

- First-Order Conditions: for all $m$,
  \[ v'(x_m) = \lambda + \sum_{\nu\neq n} \mu_\nu \left( 1 - \frac{\pi_{\nu m}}{\pi_{nm}} \right) \]
  (Mirrlees’ equation)

- Interpretation: variable wage; $w_m = \text{base wage} + \text{bonuses or penalties}$...
Special Case: Risk-Neutral Agent.

- Assume \( u(w) \equiv w \) (risk-neutrality)

**Proposition 3**: If the Agent is risk neutral, then, the first-best effort can be implemented,

\[
\max_{a \in A} (B(a) - C(a)) = \max_{a \in A} (B(a) - C^o(a))
\]

and

\[
w_m = s_m - B(a^*) + C^o(a^*)
\]
• **Proof**: Under the proposed contract, the Agent chooses $a_n$ so as to maximize,

$$E[U(w, a_n)] = \sum_m \pi_{nm} s_m - (B(a^*) - C^o(a^*)) - a_n$$

$$= B(a_n) - a_n - \text{constant},$$

and, $C^o(a_n) = a_n + U_o$,

$$E[U(w, a_n)] = B(a_n) - C^o(a_n) + U_o - \text{constant}.$$

This proves that the agent will choose $a_n = a^*$.

Q.E.D.
We have proved that

\[ C(a^*) = B(a^*) - B(a^*) + C^o(a^*) \]

\[ = C^o(a^*). \]

Since: \( B(a) - C(a) \leq B(a) - C^o(a) \)

and: \( B(a^*) - C^o(a^*) = \max_A (B(a) - C^o(a)) \).

We proved that:

\[ \max_A (B(a) - C(a)) = \max_A (B(a) - C^o(a)). \]

Q.E.D.
Further Properties of the Optimal Contract

- The optimal contract trades off insurance and incentives: If the Agent is better insured, the power of incentives is reduced.

- But the contract also exploits the information conveyed by the signal $s_m$ on effort $a_n$.

- Is it true that the optimal contract is a profit-sharing contract in the sense that:

  $w_1 < w_2 < \ldots < w_M$

  if

  $s_1 < s_2 < \ldots < s_M$?

- Assume now that signals are ranked: $s_k < s_{k+1}$.  

• Assume that effort levels are ranked: $a_1 < a_2 < \ldots < a_N$.

• Define $\Pi_{nm} = \text{Prob}(s \geq s_m \mid a_n)$

• **ASSUMPTION 3**: If $\nu > n$ then $\Pi_{nm} \leq \Pi_{\nu m}$ for all $m = 1 \ldots M$, with a strict inequality at least for some $m$.

• **First-Order Stochastic Dominance**:

Increasing effort increases the probability of getting a higher outcome. This implies that $B(a_n)$ is increasing in $n$. 
Example (Kreps)

• Three levels of profit:

\[ s_1 = 1, s_2 = 2, s_3 = 10,000. \]

• Two levels of effort:

\[ a_1 = 1 \text{ and } a_2 = 2. \]

• The probabilities \( \pi_{nm} \) are:

<table>
<thead>
<tr>
<th>s</th>
<th>( Pr(s \mid a = 1) )</th>
<th>s</th>
<th>( Pr(s \mid a = 2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10,000</td>
<td>0.2</td>
<td>10,000</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>0.3</td>
<td>2</td>
<td>0.1</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>0.4</td>
</tr>
</tbody>
</table>

• Check that \( \pi_{nm} \) satisfy the FOSD assumption.
• Suppose that \( a_2 \) is the optimal effort level.

• Write Mirrlees’ equations:

\[
\begin{align*}
  v'(x_1) &= \lambda + \mu \left(1 - \frac{0.5}{0.4}\right) = \lambda - \frac{\mu}{4} \\
  v'(x_2) &= \lambda + \mu \left(1 - \frac{0.3}{0.1}\right) = \lambda - 2\mu \\
  v'(x_3) &= \lambda + \mu \left(1 - \frac{0.2}{0.5}\right) = \lambda + \frac{3}{5}\mu 
\end{align*}
\]

• \( x_1 = u(w_1), x_2 = u(w_2) \), etc...

• Remark that \( w_m \) is non monotonic: \( w_2 < w_1 \) and \( w_3 > w_2 \)!
• We need two additional assumptions.

• ASSUMPTION 4 : (MLRP)

If \( a_n < a_\nu \) and \( s_m < s_\mu \), we have

\[
\frac{\pi_\nu \mu}{\pi_n \mu} \geq \frac{\pi_\nu m}{\pi_n m}
\]

• The Monotone-Likelihood Property. Can we now prove \( w_{k+1} > w_k \)? Not yet!
• Using Mirrlees’ equations, we get,

\[ v'(x_m) - v'(x_\mu) = \sum \mu_{\nu} \left[ \frac{\pi_{\nu\mu}}{\pi_{n\mu}} - \frac{\pi_{\nu m}}{\pi_{n m}} \right] \]

and \( \mu_{\nu} \geq 0 \) (multipliers are \( \geq 0 \)).

• We want \( w_\mu > w_m \) or \( x_\mu > x_m \) \( \Rightarrow v'(x_\mu) > v'(x_m) \).

The difference above should be negative.

• We obtain the desired results if \( \mu_{\nu} = 0 \) for all \( \nu > n \).
**Convexity of Distribution Function:**

- **ASSUMPTION 5 : (CDF)**

The cumulative distribution function of outcomes $s$ is convex with respect to effort:

$$\Pi_{nm} = \text{Prob}(s \geq s_m | a_n)$$

is a concave function of $a$.

- **Interpretation** : increases in effort have decreasing marginal impact on the probabilities of better outcomes.
Proposition 4: If $u$ is strictly concave, $(\pi_{nm})$ satisfy FOSD, MLRP and CDF, then the optimal wage-incentive scheme has wages that are nondecreasing functions of the level of firm profits, i.e.,

$$w_1 \leq w_2 \leq ... \leq w_M.$$ 

Proof: (see Kreps (1990))
Readings:


- The source for the theory presented here: