The Principal Agent Problem: Moral Hazard

Dilip Mookherjee

Boston University

Ec 703b Lectures 10-12 (text: MWG Ch 14 Section B)

1. Moral Hazard: Introduction

- Moral Hazard: a variant of the P-A problem involving *hidden action* rather than *hidden information*
- Contexts where some payoff-relevant actions of the agent are not observed by the Principal, or cannot be written into contracts (i.e., payments to the agent cannot be conditioned on such actions)
- Agent and Principal have conflicting preferences over these actions (*incentive problem*), generating a problem of *control*

Moral Hazard: Examples

- Provision of effort/care in various relationships:
  - Employer-employee
  - Landlord-tenant in sharecropping
  - Shareholder-manager
  - Lender-borrower
  - Insurance company-customer
  - Regulation/procurement customer versus suppliers/utility
  - Service customers versus providers (doctors, lawyers, consultants, teachers)
- Care enhances Principal's profit, but is costly for Agent
- Can also involve choice of project risk/direction, e.g. relationship between finance company and portfolio traders, between shareholders and managers, between firm owner and R&D investigators

Distinction between Hidden Action (HA) and Hidden Information (HI)

- Consider a general setting where HA and HI co-exist: output or profit generated for the Principal depends on agent productivity θ, effort e and luck ϵ:

$$\pi = f(\theta, e, \epsilon)$$

- Agent has personal cost $g(e)$ of effort with $g' > 0, g" > 0$
- Agent privately *knows* realization of θ, *selects* effort, while Nature determines ϵ
- P observes only realization of π, has prior over θ, ϵ
- A knows realization of θ, selects e, has prior over ϵ
- P chooses a contract determining payment to A as a function only of realization of π

Distinction between Hidden Action (HA) and Hidden Information (HI), contd.

- If luck plays no role, this reduces to a pure HI model: e.g., if $\pi = \theta + e$, we have $e = \pi - \theta$ and A's cost is $C(\pi, \theta) = g(\pi - \theta)$, with $C_{\pi\theta} < 0$, so e 'drops out' of the analysis
- If productivity plays no role but effort and luck do, this becomes a pure HA model — observing outcome π does not allow P to infer respective roles of effort and luck, so limits P's capacity to reward effort
- Many problems involve both HA and HI, but in the remaining lectures we will focus on pure HA

2. Pure HA Model

- Production Function: $\pi = f(e, \epsilon)$, which is equivalently represented by probability distribution $F(\pi; e)$ over P's profit conditional on A's effort
- A's effort disutility $g(e)$, increasing in e in some feasible set $\mathcal{E}$ of real numbers
- Discrete formulation: $\pi \in \{\pi_1, \pi_2, .., \pi_m\}$ with $\pi_i > \pi_{i-1}$ for all i; conditional density of π denoted $f_i(e)$
- Also wlog assume for any i: $f_i(e) > 0$ for some $e \in \mathcal{E}$
- If f is increasing in each argument, $F(.; e)$ FOSD $F(.; e')$ if $e > e'$
- Stronger property: Monotone Likelihood Ratio Property (MLRP): $\frac{f_i(e)}{f_{i-k}(e)}$ increasing in e, for all i, k

Pure HA Model, contd.

- P observes only realization of π_i, does not observe e or ϵ or any other signal of e
- Payment to A: $w_i \equiv w(\pi_i)$ depends only on observed outcome
- Limited Liability (LL) constraint: $w_i \geq \underline{w}$ (limit to punishments for poor performance, defined by law or custom)
- Sequence of Moves: (1) P offers contract (and recommended effort); (2) A accepts/rejects contract; (3) if accepts, A selects effort; (4) Nature selects ϵ

Payoffs

- P (risk neutral) payoff or net profit: $\sum_i f_i(e)[\pi_i - w_i]$
- A (risk-averse) exp payoff: $\sum_i f_i(e)v(w_i) - g(e)$, where $v' > 0, v" \leq 0$
- A's outside option payoff $\underline{U}$
- Assume $\underline{U} > v(\underline{w}) - \min_{e \in \mathcal{E}} g(e)$, which in most versions of the problem will ensure that the LL constraint will not bind (call this LLN from now on)

Math Programming Problem

Choose $(\{w_i\}_i, e \in \mathcal{E})$ to maximize net profit

$$\sum_i f_i(e)[\pi_i - w_i]$$

subject to:

$$\sum_i f_i(e)v(w_i) - g(e) \geq \underline{U} \qquad (PC)$$

$$\sum_i f_i(e)v(w_i) - g(e) \geq \sum_i f_i(e')v(w_i) - g(e') \quad \text{for all} \quad e' \neq e \qquad (IC)$$

$$w_i \geq \underline{w} \qquad (LL)$$

First-Best Benchmark

- Consider hypothetical situation where P could actually observe e and condition contracts based on it
- Then the contract could specify payments $w_i(e)$
- To ease IC as much as possible, would make sense to set for any $e' \neq e$: $w_i(e') = \underline{w}$
- This would ensure IC would not bind, given LLN assumption
- So IC can be dropped in P's problem

First-Best Problem, Risk Neutral Agent ($v(w) \equiv w$)

- Observe that PC must bind in the first-best problem (otherwise, there is i with $f_i(e) > 0$ and $w_i > \underline{w}$, and w_i can be lowered slightly)
- If A is risk-neutral: binding PC implies $\sum_i w_i f_i(e) = \underline{U} + g(e)$, so P's profit reduces to $\sum_i f_i(e)\pi_i - \underline{U} - g(e)$
- *First-best effort:* e^F maximizes $[B(e) - g(e)]$, where $B(e) \equiv \sum_i f_i(e)\pi_i$
- First-best payments: any $w_i \geq \underline{w}$ such that $\sum_i w_i f_i(e^F) = \underline{U} + g(e^F)$
- In particular can select $w_i = w^*$ where $w^* \equiv \underline{U} + g(e^F)$ (provided $f_i(e^F) > 0$ and $\underline{w}$ otherwise)

Second-Best Problem, Risk Neutral Agent

- Now return to the actual problem with moral hazard
- **Proposition (P14.B.2 in MWG)**: *If A is risk-neutral, P can achieve first best net profit, despite moral hazard*
- **Proof:** select $w_i = \pi_i - K$, for all i, where $K \equiv \sum_i f_i(e^F)\pi_i - g(e^F) - \underline{U}$ (*'Sell' the firm at price of* K)
- Check IC is satisfied: e chosen by A to maximize $\sum_i f_i(e)\pi_i - g(e) - K \equiv [B(e) - g(e)] - K$, so optimal for A to choose e^F
- Check PC: A attains exp. payoff $[B(e^F) - g(e^F)] - K = \underline{U}$
- P earns exp net profit of $K = B(e^F) - g(e^F) - \underline{U} = B(e^F) - w^*$, same as in first-best □

Intuition for Risk-Neutral case

- There is no incentive problem (conflict of objectives) between P and A if both are risk-neutral
- P can impose any degree of risk on A, and latter will be willing to accept the contract as long as mean payment is adequate to cover disutility of working at the first-best level
- So P can sell the firm to A at a fixed price, after which A perfectly internalizes the effect of own effort on the firm's profit, so selects the first-best effort ('franchise' contracts with high powered incentives)

Intuition for Risk-Neutral case, contd.

- In an insurance setting, A has no need for any insurance, so there is no moral hazard problem of exercising too little care to prevent an accident
- The moral hazard problem is interesting only if A is risk-averse (i.e., there is a need for P to insure A, but perfect insurance will create an incentive problem: A will exercise too little effort), assuming LLN holds

First-Best Problem, Risk Averse Agent

- Now suppose $v" < 0$; same argument as before (using LLN) implies that PC must bind
- Binding PC implies $\sum_i v(w_i)f_i(e) = \underline{U} + g(e)$
- A is now risk-averse, so cares about dispersion of wages $\{w_i\}_i$ rather than just the mean
- Optimal for P to shield A from risk, so set $w_i = w^*$ where $v(w^*) = \underline{U} + g(e)$ (provided $f_i(e) > 0$, and $\underline{w}$ otherwise) *(Pareto efficient risk-sharing a la Arrow-Borch)*
- Expected cost to P of implementing e is $C^F(e) \equiv v^{-1}(\underline{U} + g(e))$
- e^F maximizes $[B(e) - C^F(e)]$

Second-Best Problem, Risk Averse Agent, Two-Effort Case

- Suppose $v" < 0$, $e \in \mathcal{E} \equiv \{e_L, e_H\}$ with $e_L < e_H$ ('shirk' versus 'work')
- One situation where the FB profit can be attained by P despite moral hazard:
  - where $e^F = e_L$, i.e., $g(e_H) - g(e_L) > B(e_H) - B(e_L)$
  - P wants A to shirk
  - can insure A perfectly and A wants to shirk anyway (IC satisfied by FB contract)
- So focus henceforth on the case where $e^F = e_H$, where A is expected to work instead of shirk

Second-Best Problem, Risk Averse Agent, Two-Effort Case, $e^F = e_H$

- Can the first-best profit still be attained by P?
- First-best requires A to be perfectly insured, following high effort: $w_i = w^* = v^{-1}(\underline{U} + g(e_H))$ if $f_i(e_H) > 0$
- If $f_i(e_H) = 0$, impose maximal penalty: $w_i = \underline{w}$
- If (and only if) IC is satisfied, the first-best profit can still be attained

Second-Best Problem, Risk Averse Agent, Two-Effort Case, $e^F = e_H$, contd.

- When is IC satisfied by $w_i = w^*$ if $f_i(e^F) > 0$ and $\underline{w}$ otherwise, when:

$$v(w^*) - g(e_H) \geq p_F v(\underline{w}) + (1 - p_F) v(w^*) - g(e_L)$$

(where $p_F \equiv \sum_{i \in S} f_i(e_L)$, $S \equiv \{i | f_i(e_L) > 0 = f_i(e_H)\}$)

- Necessary that 'shifting support-strong punishment' condition (SSSP) holds:

$$p_F > 0 \qquad (SS)$$

$$v(\underline{w}) \leq \underline{U} + g(e_H) - \frac{g(e_H) - g(e_L)}{p_F} \qquad (SP)$$

- SSSP is also sufficient to ensure (IC)

Second-Best Problem, Risk Averse Agent, Two-Effort Case, $e^F = e_H$, contd.

- Hence: *first-best cannot be attained if and only if: (i) A is risk-averse, (ii)* $e^F = e_H$*, and (iii) SSSP is violated.*
- Suppose, from now on, that this is true ('there is an incentive problem')
- Consider the case where (SS) does not hold, i.e., $f_i(e) > 0$ for all i, for both $e \in \{e_L, e_H\}$
- Every outcome is possible, under either effort: call this condition (NSS)

Optimal Second-best Contract under NSS

- Proceed in two steps (a la Grossman-Hart (Ecta, 1983)):
    - Find optimal exp. cost minimizing contract that implements e_H, and corresponding exp cost $C(e_H)$
    - Choose second-best effort, which maximizes $B(e) - C(e)$, where $C(e_L) = C^F(e_L) = w^* \equiv v^{-1}(\underline{U} + g(e_L))$
- Problem is interesting only insofar as e_H is the second-best effort, so suffices to focus on the first-stage problem
- Grossman-Hart also assumed $v(\underline{w}) = -\infty$ so (LL) never binds

First-Stage Problem: Optimal Implementation of e_H

Choose $v_i \equiv v(w_i)$ to minimize (where $h \equiv v^{-1}(.)$ is strictly convex)

$$\sum_i f_i(e_H)h(v_i)$$

subject to:

$$\sum_i f_i(e_H)v_i - g(e_H) \geq \underline{U} \qquad (PC)$$

$$\sum_i f_i(e_H)v_i - g(e_H) \geq \sum_i f_i(e_L)v_i - g(e_L) \qquad (IC)$$

First-Stage Problem, contd.

- Can apply Kuhn Tucker theory (constraints are linear so satisfy Constraint Qualification conditions, objective function is strictly convex)
- Optimal payments (in 'utils'):

$$h'(v_i)(\equiv \frac{1}{v'(w_i)}) = \lambda + \mu[1 - \frac{f_i(e_L)}{f_i(e_H)}] \qquad (FOC)$$

where λ is KT multiplier for PC, and μ for IC

Key Feature of Optimal Contract

Lemma: $\lambda > 0, \mu > 0$.

Proof: If $\mu = 0$, A is perfectly insured (w_i does not vary with i), whence (IC) would be violated, a contradiction.

And if $\lambda = 0$, we can drop (PC) and obtain the same solution.

But (PC) must be binding: since LL does not bind in any state, we can reduce v_i slightly by some uniform amount for all i and lower the exp cost further without violating (IC), a contradiction. □

Properties of Optimal Contract

1. A must bear risk: *w_i varies with i.*

Because FOSD implies $\frac{f_i(e_L)}{f_i(e_H)}$ must vary with i.

'Controllability' Principle does not hold: Poor performance (low i) could happen even if A chose e_H, owing to factors beyond A's control, and yet A must be 'punished' if this happens

Properties of Optimal Contract, contd.

2. Monotonicity: *w_i is rising in i if MLRP holds, but not necessarily under FOSD.*

Proof: MLRP implies $\frac{f_i(e_L)}{f_{i-1}(e_L)} < \frac{f_i(e_H)}{f_{i-1}(e_H)}$, or $\frac{f_i(e_L)}{f_i(e_H)} < \frac{f_{i-1}(e_L)}{f_{i-1}(e_H)}$, and w_i is rising in i.

Consider following example satisfying FOSD but not MLRP, with three outcomes (low, medium, high performance) and probabilities $\frac{1}{2}, \frac{1}{4}, \frac{1}{4}$ under e_L, and $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ under e_H

Ratio $\frac{f_i(e_L)}{f_i(e_H)}$ for low performance is 2, medium performance is $\frac{1}{2}$, and high performance is 1, so A should be paid more for medium rather than high performance!

Properties of Optimal Contract, contd.

- Should payment to A depend on variables other than A's performance?
- Such as, performance of other agents (managers of other divisions/firms) — benchmarking/yardstick competition (Medicare reimbursement), relative performance (percentile scores, tournaments)
- Let $\delta_k, k = 1, 2, ...$ denote performance of some other agent B with production function $\delta = d(e_b, \epsilon_b)$, whose effort e_b is expected to be some e_b^*, and ϵ_b is the 'luck' of agent B
- A's performance π_i generated by A's effort $e = e_A$ and luck $\epsilon = \epsilon_A$
- When would it be valuable to base A's compensation on δ_k besides π_i?

The Informativeness Principle

- Holmstrom (BJE 1979, 1982) and Shavell (QJE 1979) posed and answered this question: it is valuable to base A's compensation on δ_k besides π_i **if and only if ϵ_A and ϵ_B are correlated**
- **Argument:** Extending the model to allow w to depend on both i and k (realizations of π and δ respectively), we obtain

$$h'(v_i)(\equiv \frac{1}{v'(w_{ik})}) = \lambda + \mu[1 - \frac{f_{ik}(e_{AL}, e_B^*)}{f_{ik}(e_{AH}, e_B^*)}] \qquad (FOC')$$

  where $f_{ik}(e_A, e_B)$ denotes the probability of the joint event $\{\pi = \pi_i, \delta = \delta_k\}$ conditional on e_A, e_B
- If ϵ_A, ϵ_B are independent, we have

$$\frac{f_{ik}(e_{AL}, e_B^*)}{f_{ik}(e_{AH}, e_B^*)} = \frac{f_i(e_{AL})}{f_i(e_{AH})} \frac{f_k(e_B^*)}{f_k(e_B^*)} = \frac{f_i(e_{AL})}{f_i(e_{AH})}$$

  is independent of agent B's performance k, so w_{ik} should not vary with k

The Informativeness Principle, contd.

- Conversely, if ϵ_A and ϵ_B are correlated, the likelihood ratio $\frac{f_{ik}(e_{AL}, e_B^*)}{f_{ik}(e_{AH}, e_B^*)}$ varies with k, so w_{ik} should vary with k
- *Intuition:* If A and B are working in a 'similar' production or marketing environment, performance of B provides a useful benchmark to evaluate A's performance
- B's performance δ_k provides some information about realization of ϵ_B, which in turn provides some information about realization about ϵ_A, and hence about the effort chosen by A
- More generally, A's compensation should be a function of a 'sufficient statistic' in P's inference of A's action: *every* informative signal should matter

3. Limited Liability and Moral Hazard with Risk Neutral Agents

- So far we have focused on moral hazard problems arising from conflict between insurance provision and effort incentives
- The analysis abstracted from problems created by **limited liability**, which imposes limits on punishments for poor performance
- Limits on liability arise owing to law or custom, which restrict scope of punishments
- E.g., bankruptcy laws restrict scope for lenders to extract assets of defaulting borrowers, labor laws restrict punishments that can be imposed on non-performing workers

Limited Liability and Moral Hazard with Risk Neutral Agents, contd.

- The Grossman-Hart model assumed that as payments to the agent approach their lower limit, utility of the agent approaches minus infinity, so LL constraint never binds
- If worst punishment is bounded, then the LL constraint could bite
- Then the limited scope of punishments can create an incentive problem, *even if the agent is risk-neutral*
- Illustrate this in the context of a simple model of credit, where a borrower/entrepreneur with limited wealth seeks a loan to finance an investment project

Limited Liability and Moral Hazard with Risk Neutral Agents, contd.

- Both borrower and lender are risk-neutral; borrower's effort affects the likelihood of project success
- If the project fails, limited liability restricts punishments and borrower defaults on the loan
- The model illustrates how poor entrepreneurs could encounter *credit rationing*

Credit Contract Model

- P is a lender, A an entrepreneur seeking funding for an investment project requiring an upfront investment of I
- A has assets of $W < I$, so needs to borrow $I - W$ to operate the project
- Project return is uncertain: could be successful and generate a given return of R, or fail and generate 0 return
- Probability of success $p(e)$ depends on effort $e \in \{e_l, e_h\}$ of A, where $e_l < e_h$ and $p_l = p(e_l) < p_h = p(e_h)$
- *Moral Hazard:* e is unobservable by P, high effort ('work') imposes higher personal cost on A than low effort ('shirk'): $g_l = g(e_l) = 0 < g_h = g(e_h)$

Credit Contract Model, contd.

- The project has a positive NPV if A works, but not if A shirks:

$$p_h R - I - g_h > 0 > p_l R - I$$

- Both P and A are *risk-neutral*, so there is no need for P to provide A any insurance
- *Limited Liability:* $w_s = R - r_s \geq 0, w_f = -r_f \geq 0$, where r_s, r_f is amount repaid by A in states s, f resp.
- Ex post payoffs:
  - *Success:* $\Pi_P = r_s - (I - W)$, $\Pi_A = w_s = R - r_s$
  - *Failure:* $\Pi_P = r_f - (1 - W)$, $\Pi_A = w_f = -r_f$
- Limited Liability imposes lower bound of 0 on borrower's net assets after failure, and forces lender to bear a loss in that state

Credit Contract: Payoffs and Constraints

- Contract parameters to be decided: (r_s, r_f, e)
- Ex ante payoffs: $E\Pi_P = p(e)r_s + (1 - p(e))r_f - (I - W)$, $E\Pi_A = p(e)[R - r_s](1 - p(e))r_f - g(e)$
- *Participation Constraints:*

$$p(e)r_s + (1 - p(e))r_f \geq (I - W) \qquad (PPC)$$

$$p(e)[R - r_s] + (1 - p(e))r_f - g(e) \geq W \qquad (APC)$$

- *Incentive Constraint:*

$$e \in \arg\max_{e'}\{p(e')[R - r_s] + (1 - p(e'))r_f - g(e')\} \qquad (IC)$$

- *Limited Liability:*

$$r_s \leq R, r_f \leq 0 \qquad (LL)$$

Results

Claim 1: *There is no feasible contract (satisfying PPC, APC, IC, LL) where the agent shirks ($e = e_l$).*

Proof: Add APC, PPC to obtain $p(e)R - I - g(e) \geq 0$, which is possible only if the agent works. □

Claim 2: *Given any feasible contract with $r_f < 0$ which induces A to work, there is another with $r'_f = 0$ which generates same ex ante payoffs to both and induces A to work.*

Proof: Select another contract with $r'_f = 0$ and $r'_s = r_s + \frac{1-p_h}{p_h} r_f < r_s$, which satisfies LL. Then observe that $r'_s - r'_f < r_s - r_f$ so (IC) is preserved. □

Results, contd.

- So it suffices to focus on contracts that induce A to work and $r_f = 0$. To simplify notation, we now use r to denote r_s.
- Only r remains to be determined, with constraints:

$$R - r \geq \frac{g_h}{p_h - p_l} \qquad (IC)$$

$$p_h[R - r] - g_h \geq W \qquad (APC)$$

$$p_h r \geq I - W \qquad (PPC)$$

- (IC) gives an upper bound while (PPC) generates a lower on r:

$$\frac{I - W}{p_h} \leq r \leq R - \frac{g_h}{p_h - p_l}$$

Necessary Condition for a Feasible Loan Contract

- So a feasible loan contract exists only if $R - \frac{g_h}{p_h - p_l} \geq \frac{I - W}{p_h}$, or

$$W \geq \underline{W} \equiv I - p_h[R - \frac{g_h}{p_h - p_l}]$$

- Borrower must meet a minimum wealth requirement
- Those with lower wealth cannot get a loan, irrespective of how the interest rate is set: one form of **credit rationing**

Market Equilibrium Outcomes for Qualifying Borrowers

- Consider now borrowers that meet the wealth requirement $W \geq \underline{W}$
- If lender has monopoly power, will set interest rate at the highest possible level, where either (APC) or (IC) binds:

$$r^m = R - \max\{\frac{g_h}{p_h - p_l}, \frac{g_h + W}{p_h}\} \qquad (UB)$$

- If the loan market is competitive, the interest rate at the lowest feasible level, where (PPC) binds:

$$r^c = \frac{I - W}{p_h}$$

Credit Rationing Among Qualifying Borrowers

- In either case, r^m is an upper bound to the interest rate
- If supply of loanable capital available to lenders is limited relative to demand (loan applications), the interest rate under competition would rise, but only upto the upper bound r^m
- If IC binds but APC does not bind at r^m, there will be **credit rationing** *a la* Stiglitz-Weiss (AER 1981):
  - some loan applicants would not be able to get loans while others with the same characteristics will succeed
  - those getting loans will be strictly better off than those not getting loans
  - despite this, the former will not be able to out-compete the latter group by offering to pay a higher interest rate