Contract Theory
## Contents

1 Hidden Information: Screening .......................................................... 6

1.1 A Simple Model of Price Discrimination ........................................... 8

1.1.1 Full Information Benchmark: First-Best Outcome or Perfect Price ... 8

1.1.2 Asymmetric Information: Linear Pricing and Two-Part Tariffs ....... 9

1.1.3 Asymmetric Information: Second-Best Outcome or Optimal Nonlinear Pricing ......................................................... 12

1.2 Modeling Issues in Contracting Problem .......................................... 16

1.2.1 Ex-Ante Contracting ................................................................. 16

1.2.2 Countervailing Incentives ......................................................... 18

1.3 Other Applications .......................................................................... 21

1.3.1 Regulation ............................................................................ 21

1.3.2 Collateral as a Screening Device in Loan Markets .................... 23

1.3.3 Credit Rationing .................................................................... 27
### 2 Hidden Action: Moral Hazard

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 A Simple $2 \times 2$ Model</td>
<td>36</td>
</tr>
<tr>
<td>2.1.1 First-Best: Complete Information Optimal Contract</td>
<td>37</td>
</tr>
<tr>
<td>2.1.2 Risk-Neutral Agent and the First-Best Contract</td>
<td>38</td>
</tr>
<tr>
<td>2.1.3 Limited Liability and Second-Best Contract</td>
<td>39</td>
</tr>
<tr>
<td>2.1.4 Risk-Averse Agent and Second-Best Contract</td>
<td>40</td>
</tr>
<tr>
<td>2.2 Extensions</td>
<td>42</td>
</tr>
<tr>
<td>2.2.1 More than 2 Outcomes</td>
<td>42</td>
</tr>
<tr>
<td>2.2.2 Comparing Information Structures</td>
<td>43</td>
</tr>
<tr>
<td>2.2.3 Moral Hazard and Renegotiation</td>
<td>45</td>
</tr>
<tr>
<td>2.3 Applications</td>
<td>47</td>
</tr>
<tr>
<td>2.3.1 Moral Hazard in Insurance Markets</td>
<td>47</td>
</tr>
<tr>
<td>2.3.2 Moral Hazard in Teams</td>
<td>50</td>
</tr>
</tbody>
</table>

### 3 Incomplete Contracts

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 The Holdup Problem</td>
<td>56</td>
</tr>
<tr>
<td>3.2 Organizational Solution</td>
<td>57</td>
</tr>
<tr>
<td>3.3 Contractual Solution</td>
<td>59</td>
</tr>
<tr>
<td>3.4 Cooperative Investment and Failure of Contracts</td>
<td>62</td>
</tr>
</tbody>
</table>
4 Auction Theory

4.1 Second-Price (or English) Auction ........................................ 67
4.2 First-Price Auction .......................................................... 68
4.3 Revenue Equivalence and Optimal Auction ............................. 71
4.4 Extensions ................................................................. 75
  4.4.1 Risk Averse Bidders ..................................................... 76
  4.4.2 Budget Constrained Bidders ......................................... 78
  4.4.3 Optimal Auction with Correlated Values .......................... 79
4.5 Appendix: Envelope Theorem .......................................... 81

5 Matching Theory ............................................................. 83

5.1 Stable Matchings .......................................................... 85
  5.1.1 Gale-Shapley Algorithm ............................................... 87
  5.1.2 Incentive Problem ..................................................... 90
  5.1.3 Extension to Many-to-One Matching ............................... 92
5.2 Efficient Matching ....................................................... 95
  5.2.1 Top Trading Cycles Algorithm ..................................... 96
Introduction

• What is contract? - ‘A specification of actions that named parties are supposed to take at various times, generally as a function of the conditions that hold’ (Shavell, 2004).
  – An example is insurance contract under which (risk-averse) insureds pay premiums and are covered against risk by an insurer.
  – Some party might have incentive to behave opportunistically at the expense of others.
  – In an ideal world, people can write a complete contingent contract that induces all the parties to take the ‘right’ actions in every possible state of world, which leads to a Pareto efficient outcome.

• Contract theory studies what will or should be the form of contracts in less than ideal worlds, where there exist
  – Hidden action (or moral hazard): when the involved party’s behavior cannot be perfectly monitored by others.
  – Hidden information (or adverse selection): when the involved party has private information which is not known to others.
  – Contractual incompleteness: when contracts do not deal with all relevant contingencies.

• There are many applications of contract theory, among which the followings are important:
– Labor contracts

– Regulation

– Price discrimination

– Optimal taxation.

– Financial contracts

– Auctions

• In most applications, one party, called ‘principal’, offers a contract to the other party, called ‘agent’.

– One principal + one agent

– One principal + multiple agents, e.g. auction

– Multiple principals + one agent (common agency), e.g. lobby groups influencing a government agency

– Multiple principals + multiple agents
Chapter 1

Hidden Information: Screening

• General setup

  – An agent, informed party, who is privately informed about his type.

  – A principal, uninformed party, who designs a contract in order to screen different types of agent and maximize her payoff.

  – This is a problem of hidden information, often referred to as screening problem.

• A list of of economic issues where this setup is relevant.

  – Price discrimination: A monopolist selling to a buyer whose demand is unknown. (often called ‘second-degree price discrimination’)

  – Credit rationing in financial markets: A lender investing in a project whose profitability is unknown.

  – Optimal income taxation: A government designing the tax scheme for the people whose income generating ability is unknown.
– Implicit labor contract: An employer offering the wage contract to an employee whose productivity is unknown.

In these examples, contracts correspond to the pricing scheme, investment decision, tax scheme, and wage contract.

• To describe the contractual situations in general, we adopt the following time line:

  – At date 0, the agent learns his type.
  – At date 1, the principal offers a contract to the agent.
  – At date 2, the agent accepts or rejects the contract. If rejects, then the principal and agent get their outside utilities, which are normalized to zero. If accepts, then they go to date 3.
  – At date 3, the transaction (or allocation) occurs, following the contract

We then ask what is the optimal contract in the principal’s perspective, or the contract which maximizes the principal’s payoff. Usually, however, there are so many contracts one can conceive of that it is not feasible to try all possible contract one by one.

• Revelation principle

  – To determine the optimal contract among all possible ones, it suffices to consider the contracts which specify one allocation for each type of agent. (e.g. the insurance example)
  – Need to make sure that each type has an incentive to select only the allocation that is destined to him.
1.1 A Simple Model of Price Discrimination

Consider a transaction between a buyer (agent) and a seller (principal), where the seller does not know perfectly how much the buyer is willing to pay.

- The buyer has utility function is given by $u(q, T, \theta) = \theta v(q) - T$
  
  - $q =$ the number of units purchased and $T =$ the total amount paid to the seller.
  
  - The buyer’s characteristic is represented by $\theta$, which is only known to the buyer: $\theta = \theta_L$ with probability $\beta \in (0, 1)$ and $\theta = \theta_H > \theta_L$ with $(1 - \beta)$. Define $\Delta \theta := \theta_H - \theta_L$.
  
  - We impose technical assumptions as follows: $v(0) = 0$, $v'(q) > 0$, and $v''(q) < 0$ for all $q$.
  
  - Outside utility of the buyer is fixed at some level $\bar{u} = 0$.

- The seller’s preference is given by $\pi = T - cq$, where $c$ is the seller’s production cost per unit.

1.1.1 Full Information Benchmark: First-Best Outcome or Perfect Price Discrimination

Suppose that the seller is perfectly informed about the buyer’s type. The seller can treat each type of buyer separately and offer a type-dependent contract: $(q_i, T_i)$ for type $\theta_i$ ($i = H, L$).

- The seller solves

$$\max_{q_i, T_i} T_i - cq_i \quad \text{subject to} \quad \theta_i v(q_i) - T_i \geq 0.$$
− The constraint is called *participation constraint* (or individual rationality, *(IR)*, constraint): If this constraint is not satisfied, then the agent would not participate in the contractual relationship.

− At the optimal solution, *(IR)* constraint must be satisfied as an equality or *binding*. (Why?)

− A solution to this problem is \((\hat{q}_i, \hat{T}_i)\) given by

\[
\hat{T}_i = \theta_i v(\hat{q}_i) \quad \text{from binding *(IR)* constraint}
\]

and

\[
\theta_i v'(\hat{q}_i) = c. \quad \text{from first order condition}
\]

− This quantity is efficient since it maximizes the total surplus:

\[
\hat{q}_i = \arg \max_q \theta_i v(q) - cq \quad \text{for } i = H, L.
\]

− With this solution, the seller takes all the surplus equal to

\[
\beta(\theta_L v(\hat{q}_L) - c\hat{q}_L) + (1 - \beta)(\theta_H v(\hat{q}_H) - c\hat{q}_H),
\]

while the buyer gets no surplus. It is why this solution is called *first-best* or *perfect price discrimination*.

1.1.2 Asymmetric Information: Linear Pricing and Two-Part Tariffs

Suppose from now on that the seller cannot observe the type of the buyer, facing the adverse selection problem. The first-best contract above is no longer feasible. (Why?) The contract
set is potentially large since the seller can offer any combination of quantity-payment pair 
\((q, T(q))\). In other words, the seller is free to choose any function \(T(q)\). Let us first consider 
two simple functions among others.

**Linear Pricing:** \(T(q) = Pq\)

This is the simplest contract in which the buyer pays a uniform price \(P\) for each unit he 
buys.

- Given this contract, the buyer of type \(\theta_i\) chooses \(q\) to maximize

\[
\theta_i v(q) - Pq, \quad \text{where } i = L, H.
\]

- From the first-order condition, we obtain \(\theta_i v'(q) = P\).

- We can derive the demand function: \(q_i = D_i(P), i = L, H\). Note that \(D_H(P) > D_L(P)\).

- We can also calculate the buyer’s net surplus as \(S_i(P) := \theta_i v(D_i(P)) - PD_i(P), i = L, H\). Note that \(S_H(P) > S_L(P)\).

- Let us define

\[
D(P) := \beta D_L(P) + (1 - \beta)D_H(P) \quad \text{and} \quad S(P) := \beta S_L(P) + (1 - \beta)S_H(P).
\]

- With linear pricing, the seller solves

\[
\max_P (P - c)D(P).
\]

- From the first-order condition, we obtain the optimal price \(P^m\)

\[
P^m = c - \frac{D(P^m)}{D'(P^m)}.
\]
With this solution, the buyer obtains positive rents (why?) and consumes inefficiently low quantities since
\[ \theta_i v'(q_i^m) = P^m > c = \theta_i v'(\hat{q}_i). \]

**Two-Part Tariff:** \( T(q) = F + Pq \)

With two-part tariff, the seller charges a fixed fee \( F \) up-front, and a price \( P \) for each unit purchased. For any given price \( P \), the maximum fee the seller can charge up-front is \( F = S_L(P) \) if he wants to serve both types. (\( : S_L(P) < S_H(P) \))

- The seller chooses \( P \) to maximize
  \[
  \beta [S_L(P) + (P - c)D_L(P)] + (1 - \beta) [S_L(P) + (P - c)D_H(P)] \\
  = S_L(P) + (P - c)D(P).
  \]

- From the first order condition, we obtain the optimal unit price \( P^t \) solving
  \[
  P^t = c - \frac{D(P^t) + S'_L(P^t)}{D'(P^t)}.
  \]

- Since \( S'_L(P^t) = -D_L(P^t) \) by the envelope theorem, we have \( D(P^t) + S'_L(P^t) > 0; \) in addition, \( D'(P^t) < 0 \), so \( P^t > c \). Again, the quantities are inefficiently low, or \( q^t_i < \hat{q}_i \) for \( i = L, H \).

- But, the inefficiency reduces compared to the linear pricing, or \( q^t_i > q^m_i \), since \( P^t < P^m \).

To see it, we observe
\[
\frac{d}{dP} \left[ (P - c)D(P) + S_L(P) \right] \bigg|_{P=P^m} = S'_L(P^m) = -D_L(P^m) < 0
\]
so that the seller is better off lowering the price from \( P^m \). (Intuition?)
• Let $B_i := (q_i^t, S_L(P_t) + P_t q_i^t)$ denote the bundle for type $\theta_i$ for $i = L, H$.

  – The $\theta_H$-type strictly prefers the bundle $B_H$ to $B_L$.

  – Perhaps, the seller could raise the payment of $\theta_H$-type beyond $S_L(P_t) + P_t q_i^t$ to offer some other bundle $B'_H$ so that the $\theta_H$-type is indifferent between $B_L$ and $B'_H$.

### 1.1.3 Asymmetric Information: Second-Best Outcome or Optimal Nonlinear Pricing

Here, we look for the best pricing scheme among all possible ones. In general, the pricing scheme can be described as $(q, T(q))$, where the function $T(q)$ specifies how much the buyer has to pay for each quantity $q$. We do not restrict the function $T(\cdot)$ to be linear or affine as before.

• The seller’s problem is to solve

\[
\max_{T(q)} \beta(T(q_L) - cq_L) + (1 - \beta)(T(q_H) - cq_H)
\]

subject to

\[
\theta_i v(q_i) - T(q_i) \geq \theta_i v(q) - T(q) \text{ for all } q \text{ and } i = L, H \quad (IC)
\]

and

\[
\theta_i v(q_i) - T(q_i) \geq 0 \text{ for } i = L, H. \quad (IR)
\]

  – The first two constraints are called *incentive compatibility* constraint, which guarantees that each type selects the bundle that is designed for him.
- The next two constraints are called *individual rationality* or participation constraint, which guarantees that each type is willing to participate in the seller’s contract.
- We solve this adverse selection problem step-by-step in the below.

- **Step 1: Apply the revelation principle.**

- We can use the revelation principle to restrict our attention to a couple of bundles, $(q_L, T(q_L))$ for type $\theta_L$ and $(q_H, T(q_H))$ for type $\theta_H$. Then, defining $T_i := T(q_i)$ for $i = L, H$, the above problem can be rewritten as

\[
\max_{(q_L, T_L), (q_H, T_H)} \beta(T_L - cq_L) + (1 - \beta)(T_H - cq_H) \tag{1.1}
\]

subject to

\[
\theta_H v(q_H) - T_H \geq \theta_H v(q_L) - T_L \quad (IC_H)
\]
\[
\theta_L v(q_L) - T_L \geq \theta_L v(q_H) - T_H \quad (IC_L)
\]
\[
\theta_H v(q_H) - T_H \geq 0 \quad (IR_H)
\]
\[
\theta_L v(q_L) - T_L \geq 0. \quad (IR_L)
\]

- Note that the incentive compatibility constraint has been greatly simplified.

- **Step 2: **(IR_L) must be binding at the optimal solution.

- If not, then the seller can raise $T_L$ and $T_H$ by small $\epsilon > 0$, which does not violate any constraint but raises the seller’s utility, a contradiction.

- Thus, we have

\[
T_L = \theta_L v(q_L). \tag{1.2}
\]
• Step 3: $(IC_H)$ must be binding at the optimal solution.

  - If not, then the seller can raise $T_H$ by small $\epsilon > 0$, which does not violate any constraint but raises the seller’s utility, a contradiction.

  - Thus, we have

    $T_H = \theta_H v(q_H) - (\theta_H v(q_L) - T_L) = \theta_H v(q_H) - \Delta \theta v(q_L)$. \hspace{1cm} (1.3)

  where $\Delta \theta$ represents the information rent.

• Step 4: $(IR_H)$ is automatically satisfied, provided that $(IC_H)$ and $(IR_L)$ are binding.

  - Due to the binding $(IC_H)$ and $(IR_L)$,

    $\theta_H v(q_H) - T_H = \theta_H v(q_L) - T_L \geq \theta_L v(q_L) - T_L = 0,$

    so $(IR_H)$ is satisfied.

  - We can thus ignore the $(IR_H)$ constraint.

• Step 5: Given that $(IC_H)$ is binding, $(IC_L)$ is satisfied if and only if $q_H \geq q_L$.

  - Given that $(IC_H)$ is binding, $(IC_L)$ constraint can be written as

    $\theta_L (v(q_L) - v(q_H)) \geq T_L - T_H = \theta_H (v(q_L) - v(q_H))$,

    which will be satisfied if and only if $q_H \geq q_L$.

  - Thus, we can replace $(IC_L)$ constraint by the constraint $q_H \geq q_L$.

• Step 6: Eliminate $T_L$ and $T_H$ using (1.2) and (1.3) and solve the problem without any constraint.

  - For the moment, let us ignore the constraint $q_H \geq q_L$, which will be verified later.
Substituting (1.2) and (1.3), the seller’s problem is turned into

$$\max_{q_L,q_H} \beta(\theta_L v(q_L) - cq_L) + (1 - \beta)(\theta_H v(q_H) - cq_H - \Delta \theta v(q_L)).$$

From the first-order condition, we obtain the optimal quantities, $q_H^*$ and $q_L^*$, solving

$$\theta_H v'(q_H^*) = c$$
$$\theta_L v'(q_L^*) = \frac{c}{1 - \frac{(1-\beta)\Delta \theta}{\beta \theta_L}} > c,$$

which implies $q_H^* = \hat{q}_H > \hat{q}_L > q_L^*$, as desired.

If $\beta \theta_L < (1 - \beta)\Delta \theta$, then the RHS of (1.4) becomes negative so we would have a corner solution, $q_L^* = 0$ and $q_H^* = \hat{q}_H$. (Try to interpret this)

The second-best outcome exhibits the following properties.

- No distortion at the top: $q_H^* = \hat{q}_H$
- Downward distortion below the top: $q_L^* < \hat{q}_L$
- Why? → To reduce the information rent. (Refer to the equation (1.3))
1.2 Modeling Issues in Contracting Problem

1.2.1 Ex-Ante Contracting

There are situations in which the agent can learn his type only after he signs a contract: For instance, an employee who is hired to work on some project, may not know whether his expertise is suited to the project until he starts working on it. This kind of situation can be modeled by modifying the timeline in page 7 and assuming that the agent privately learns his type between date 2 and 3. So, it is at the ex-ante stage that the contracting occurs. In the original model, by contrast, the contracting has occurred at the interim stage in which the agent is already informed of his type. We analyze the problem of designing optimal ex-ante contract in our basic model of price discrimination, though the results that follow are much more general, applying to any adverse selection model. To put forward the conclusion, it is possible for the principal to achieve the first-best outcome, despite the asymmetric information arising after the contracting stage.

- The principal’s problem is now to solve

\[
\max_{(q_L, T_L), (q_H, T_H)} \beta(T_L - cq_L) + (1 - \beta)(T_H - cq_H)
\]

subject to

\[
\beta(\theta_L v(q_L) - T_L) + (1 - \beta)(\theta_H v(q_H) - T_H) \geq 0 \quad (IR)
\]
\[
\theta_H v(q_H) - T_H \geq \theta_H v(q_L) - T_L \quad (IC_H)
\]
\[
\theta_L v(q_L) - T_L \geq \theta_L v(q_H) - T_H \quad (IC_L)
\]
− Note that there is only one \((IR)\) constraint that is for the agent to participate in the contract without knowing his type.

− Note also that \((IC_L)\) and \((IC_H)\) constraints are intact since the information the agent learns in the post-contracting stage remains to be private.

− As it turns out, the first-best outcome is achievable for the principal as follows:

  (i) \((IR)\) constraint must be binding (Why?).

  (ii) Substitute the binding \((IR)\) constraint into the object function to get

  \[
  \beta \theta_L v(q_L) + (1 - \beta) \theta_H v(q_H) - \beta cq_L - (1 - \beta) cq_H,
  \]

  which is maximized at \(q_H = \hat{q}_H\) and \(q_L = \hat{q}_L\), the first-best quantities.

  (iii) Choose \(T_L\) and \(T_H\) to make \((IR)\) binding and, at the same time, satisfy \((IC_H)\) and \((IC_L)\). (How?)

There are some lessons to be learned from the above results: First, it is not the lack of information itself but the asymmetry of information that causes the inefficiency in the previous section. Also, what generates a rent for the agent is the asymmetric information at the contracting stage. In the above setup where both parties symmetrically informed at the contracting stage, the principal only needs to satisfy the \textit{ex-ante} participation constraint, which enables the principal to push down the low type’s payoff below outside utility while guaranteeing the high type a payoff above outside utility. So, the agent without knowing about his type breaks even on average and thus is willing to participate. This \textit{ex ante} incentive, however, would not be enough for inducing the agent’s participation if the agent can opt out of the contractual relationship at any stage. In such case, both types should
be assured of their reservation payoff at least, which will revert the principal’s problem to that in the previous section.

1.2.2 Countervailing Incentives

So far we have assume that the agent’s outside utility is type-independent, being uniformly equal to zero. Often, this is not very realistic. In our price discrimination model, for instance, the high-type consumer to pay may be able to find an outside opportunity that is more lucrative than the low-type consumer does, thereby enjoying a higher outside utility than the latter does. This situation can easily be modeled by modifying the \((IR_H)\) as follows: for some \(\bar{u} > 0\),

\[
\theta_H v(q_H) - T_H \geq \bar{u} \quad (IR_H')
\]

We maintain all other constraints. This small change in the contracting problem leads to some dramatic change in the form of optimal contract.

- If \(\bar{u} \in [\Delta \theta v(\hat{q}_L), \Delta \theta v(\hat{q}_H)]\), then the seller as principal can achieve the first-best outcome.

  - To see it, note that the first-best outcome requires (i) the first-best quantities or \(q_i = \hat{q}_i, i = H, L\), and (ii) both \((IR_L)\) and \((IR_H')\) be binding so that \(T_L = \theta_L v(\hat{q}_L)\) and \(T_H = \theta_H v(\hat{q}_H) - \bar{u}\).

  - We only need to verify that the quantities and transfers given above satisfy both \((IC_L)\) and \((IC_H)\) conditions, which can be written as

\[
\theta_L(v(q_L) - v(q_H)) \begin{array}{c} \geq \end{array} T_L - T_H \begin{array}{c} \geq \end{array} \theta_H(v(q_L) - v(q_H)).
\]
By plugging the above numbers into this equation and rearranging, we obtain

\[ \Delta \theta v(\hat{q}_H) \geq \bar{u} \geq \Delta \theta v(\hat{q}_L), \]

which holds since \( \bar{u} \in [\Delta \theta v(\hat{q}_L), \Delta \theta v(\hat{q}_H)] \) as assumed.

What happens if \( \bar{u} > \Delta \theta v(\hat{q}_H) \)? In this case, the first-best contract given above violates (\( IC_L \)) only, which implies (\( IC_L \)) must be binding at the optimum. Also, analogously to Step 5 in page 14, \( q_H \geq q_L \) is necessary and sufficient for (\( IC_H \)) to be satisfied, given that (\( IC_L \)) is binding. As before, we will ignore the constraint \( q_H \geq q_L \) for the moment, which can be verified later. Another observation is that (\( IR'_H \)) also must be binding.\(^2\) We now consider two cases depending on whether (\( IR_L \)) is binding or not.

- Let us first focus on the case in which (\( IC_L \)) and (\( IR'_H \)) are binding while (\( IR_L \)) is not.
  - From two binding constraints, one can derive
    \[ T_H = \theta_H v(q_H) - \bar{u}, \quad T_L = \Delta \theta v(q_H) + \theta_L v(q_L) - \bar{u}. \]
  - By substitution, the principal’s problem can be turned into
    \[ \max_{q_L, q_H} \beta (\Delta \theta v(q_H) + \theta_L v(q_L) - \bar{u} - cq_L) + (1 - \beta)(\theta_H v(q_H) - \bar{u} - cq_H). \]
  - The first-order condition yields the optimal quantities satisfying
    \[ \theta_H v'(q_H^{**}) = \frac{c}{1 + \frac{\beta \Delta \theta}{(1 - \beta) \theta_H}} < c \]
    \[ \theta_L v'(q_L^{**}) = c, \]

\(^1\)The same analysis and result as in the part 1.1.3 follow in case \( \bar{u} \leq \Delta \theta v(q_L^*) \), including the case \( \bar{u} = 0 \). If \( \bar{u} \in (\Delta \theta v(q_L^*), \Delta \theta v(q_L)) \), the analysis will be slightly different but yield the same qualitative result as in 1.1.3.\(^2\) An argument to show this is a bit tedious and will be given in the class.
which implies that $q_{L}^{*} = \hat{q}_{L}$ and $q_{H}^{**} > \hat{q}_{H}$. In contrast to the standard case, the quantity here is distorted upward at the top.

- It remains to check that $(IR_{L})$ is satisfied. To do so, plug the above quantities and transfers into the non-binding $(IR_{L})$ to obtain

$$\bar{u} > \Delta \theta v(q_{H}^{**}).$$

(1.5)

So the above inequality is necessary (and sufficient) for having an optimal contract in which only $(IC_{L})$ and $(IR'_{H})$ are binding.

- If (1.5) is violated, that is $\bar{u} \in (\Delta \theta v(\hat{q}_{H}), \Delta \theta v(q_{H}^{**}))$, then, $(IR_{L})$ also must be binding.

- Plug $(IR'_{H})$ and $(IR_{L})$ (as equality) into $(IC_{L})$ (as equality) to obtain the optimal quantity for high type satisfying

$$\Delta \theta v(q_{H}^{***}) = \bar{u},$$

which then determines $T_{H}$ through the binding $(IR'_{H})$ constraint.

- Then plug $(IR_{L})$ into the objective function, whose first-order condition results in the optimal quantity for low type being equal to $\hat{q}_{L}$.

- Note that since $\Delta \theta v(\hat{q}_{H}) < \bar{u} = \Delta \theta v(q_{H}^{**})$, we must have $\hat{q}_{H} < q_{H}^{***}$, an upward distortion at the top again.

As seen above, if the high type entertains an attractive outside option, the participation constraint *countervails* the incentive constraint so $(IC_{H})$ does not play a role in the design of optimal contract. In other words, it is not much of a concern for the principal to save the information rent needed to make the high type tell the truth. This leads to no downward
distortion at the bottom but rather an upward distortion at the top. This upward distortion is the best way to satisfy the participation constraint for the high type while preventing the low type from mimicking the high type.

1.3 Other Applications

1.3.1 Regulation

The public regulators are often subject to an informational disadvantage with respect to the regulated utility or natural monopoly. Consider a regulator concerned with protecting consumer welfare and attempting to force a natural monopoly to charge the competitive price. The difficulty is that the regulator does not have full knowledge of the firm’s intrinsic cost structure.

- Consider a natural monopoly with an exogenous cost parameter $\theta \in \{\theta_L, \theta_H\}$ with $\Delta \theta = \theta_H - \theta_L > 0$, which is only observable to the monopolist.
  - The cost parameter is $\theta_L$ with probability $\beta$ and $\theta_H$ with $1 - \beta$.
  - The firm’s cost of producing good is observable (and contractible) and given by $c = \theta - e$, where $e > 0$ stands for the cost-reducing effort. Expending effort $e$ has cost $\psi(e) = e^2/2$.
  - To avoid the distortionary tax, the regulator tries to minimize its payment $P = c + s$ to the firm, where $s$ is a “subsidy” in excess of accounting cost $c$.

- As a benchmark, assume that there is no information asymmetry. For each type $i$, the
regulator solves

$$\min_{(e_i, s_i)} s_i + c_i = s_i + \theta_i - e_i$$

subject to

$$s_i - e_i^2 / 2 \geq 0.$$  \hfill (IR_i)

- The constraint \((IR_i)\) is binding at the solution, which is then given by \(\hat{e}_i = 1\) and \(\hat{s}_i = 1/2\) for both \(i = H\) and \(L\).

- Under the asymmetric information, this contract is vulnerable to type \(L\)’s pretending to be type \(H\): To generate cost \(c_H = \theta_H - 1\), type \(L\) only need to exert effort \(e = 1 - \Delta \theta\).

- Under the asymmetric information, letting \(\Delta \theta := \theta_H - \theta_L\), the regulator solves

$$\min_{(e_L, s_L)} \beta (s_L - e_L) + (1 - \beta) (s_H - e_H)$$

subject to

$$s_L - e_L^2 / 2 \geq 0$$ \hfill (IR_L)

$$s_H - e_H^2 / 2 \geq 0$$ \hfill (IR_H)

$$s_L - e_L^2 / 2 \geq s_H - (e_H - \Delta \theta)^2 / 2$$ \hfill (IC_L)

$$s_H - e_H^2 / 2 \geq s_L - (e_L + \Delta \theta)^2 / 2.$$ \hfill (IC_H)

- The binding constraints are \((IC_L)\) and \((IR_H)\). (Why?) Thus,

$$s_L - e_L^2 / 2 = s_H - (e_H - \Delta \theta)^2 / 2$$

$$s_H - e_H^2 / 2 = 0.$$
Substituting these equations, the optimization problem becomes

$$\min_{e_L, e_H} \beta \left( e_L^2/2 - e_L + e_H^2/2 - (e_H - \Delta \theta)^2/2 \right) + (1 - \beta) \left( e_H^2/2 - e_H \right),$$

whose solution is given by

$$e_L^* = 1 \quad \text{and} \quad e_H^* = 1 - \frac{\beta}{1 - \beta} \Delta \theta.$$

As before, the provision of effort is efficient at the top ($\theta_L$) while distorted downward at the bottom ($\theta_H$).

The downward distortion is severer as the cost differential is larger or the type is more likely to be efficient.

### 1.3.2 Collateral as a Screening Device in Loan Markets

Consider a loan market in which the bank (lender) offers loan contracts to the liquidity-constrained investors (borrowers). Investors are heterogeneous in terms of the probability their investment projects succeed. Though liquidity-constrained, borrowers own some asset that can be (only) used as collateral for a loan contract. We study how this collateral can help the lender screen different types of borrowers, using the simple model as follows:

- Each investor has a project with a random return $\tilde{y}$ that can either fail ($\tilde{y} = 0$) or succeed ($\tilde{y} = y > 0$).

- There are two types of investors who only differ in their failure probability: $\theta_H$ and $\theta_L < \theta_H$.

- The proportion (or probability) of type $k$ is given as $\beta_k, k = H, L$. 
Letting $U_k$ denote the reservation utility for type $k$, we assume that

$$\frac{U_L}{1 - \theta_L} > \frac{U_H}{1 - \theta_H}. \quad (1.6)$$

- There is a monopolistic lender who offers a menu of loan contracts $\{(C_k, R_k)\}_{k=L,H}$, where $R_k$ and $C_k$ denote the repayment (in case of success) and collateral for type $k$.

- If the project fails, then the lender can take the specified amount of collateral from the borrower and liquidate it to obtain $\delta C_k$ with $\delta < 1$. (Thus $(1 - \delta)C_k$ corresponds to a liquidation cost.)

- Thus, the borrower’s payoff from type $k$ is $R_k$ if the project succeeds and $\delta C_k$ if fails. So the expected payoff is given as

$$\sum_{k=H,L} \beta_k[(1 - \theta_k)R_k + \theta_k\delta C_k] \quad (1.7)$$

- The payoff of type $k$ borrower is $(y - R_k)$ if succeeds $-C_k$ if fails. So the expected payoff is given as

$$(1 - \theta_k)(y - R_k) - \theta_k C_k.$$ 

- Note that this model has some fundamental difference from the ones we have seen so far in the sense that the agent’s private information, $\theta$, directly enters into the principal’s utility function. In short, the agent’s information has an externality on the principal’s payoff. A model with this kind of feature is generally referred to as ‘common value model’.

- Consider as a benchmark that the lender is able to observe the borrower’s types.
– At an optimal contract, the individual rationality constraints given by

\[(1 - \theta_L)(y - R_L) - \theta_L C_L \geq U_L \quad (IR_L)\]

and \[(1 - \theta_H)(y - R_H) - \theta_H C_H \geq U_H \quad (IR_H)\]

must be binding.

– Under the constraints \((IR_L)\) and \((IR_H)\), \((1.7)\) is maximized by setting \(C_k = 0\) and \(R_k = y - \frac{U_k}{1-\theta_k}\) for each type \(k\): No collateral is required at the optimal contract since the liquidation of collateral is costly.

– If the lender faces the asymmetric information problem, then the above contract would not work since \(R_L = y - \frac{U_L}{1-\theta_L} < y - \frac{U_H}{1-\theta_H} = R_H\) from \((1.6)\) and thus both types will prefer announcing \(\theta_L\).

• Consider now the problem of finding the optimal menu of contracts for the lender who is \textit{uninformed} about the borrower types.

– We need to maximize \((1.7)\) under the \((IR_L)\), \((IR_H)\) and the \((IC)\) constraints given as

\[(1 - \theta_H)(y - R_H) - \theta_H C_H \geq (1 - \theta_H)(y - R_L) - \theta_H C_L \quad (IC_H)\]

\[(1 - \theta_L)(y - R_L) - \theta_L C_L \geq (1 - \theta_L)(y - R_H) - \theta_L C_H. \quad (IC_L)\]

– As can be seen above, what matters is to get rid of the incentive of type \(\theta_H\) to mimic \(\theta_L\), which means that \((IC_H)\) must be binding at the optimal contract.

– \((IC_H)\) and \((IC_L)\) constraints can be rewritten together as

\[
\frac{1 - \theta_L}{\theta_L}(R_L - R_H) \leq C_H - C_L \leq \frac{1 - \theta_H}{\theta_H}(R_L - R_H), \quad (1.8)
\]

which implies \(C_H - C_L \leq 0\) since \(\frac{1 - \theta_H}{\theta_H} < \frac{1 - \theta_L}{\theta_L}\).
Given that \((IC_H)\) is binding and \(C_L \geq C_H\), (1.8) implies that \((IC_L)\) is automatically satisfied and thus can be ignored.

- **At the optimal contract** \(\{(C_k, R_k)\}_{k=L,H}\), we must have \(C_H = 0\): If \(C_H > 0\), then the principal can offer an alternative contract \(\{(C_L', R_L'), (C_H', R_H')\}\), where \(C_H' = C_H - \epsilon\) and \(R_H' = R_H + \frac{\theta_H}{1-\theta_H} \epsilon\) for small \(\epsilon > 0\). With this contract, \((IC_H)\), \((IR_L)\), and \((IR_H)\) all remain the same as with the original contract and thus are satisfied. Since \(C_H' < C_H \leq C_L\) and \((IC_H)\) is binding, \((IC_L)\) is also satisfied by the previous argument. However, the principal’s profit increases by

\[
(1 - \theta_H) \frac{\theta_H}{1-\theta_H} \epsilon - \delta \theta_H \epsilon = \theta_H(1 - \delta) \epsilon > 0,
\]

contradicting the optimality of the original contract.

- **At the optimal contract**, \((IR_L)\) must be binding: Suppose it’s not binding or

\[
(1 - \theta_L)(y - R_L) - \theta_L C_L > U_L. \tag{1.9}
\]

We can then consider an alternative contract, \(\{(C_L', R_L'), (C_H, R_H)\}\), where \(C_L' = C_L - \epsilon\) and \(R_L' = R_L + \frac{\theta_H}{1-\theta_H} \epsilon\) for small \(\epsilon > 0\). With this contract, \((IC_H)\) and \((IR_H)\) remain the same as with the original contract and thus are satisfied. Also, \((IR_L)\) is satisfied since \(\epsilon\) is small. Since \(C_L' = C_L - \epsilon > C_H = 0\) and \((IC_H)\) is binding, \((IC_L)\) is satisfied for the same reason as before. However, the principal’s payoff increases by

\[
(1 - \theta_L) \frac{\theta_H}{1-\theta_H} \epsilon - \delta \theta_L \epsilon = \theta_L \left( \frac{\theta_H(1 - \theta_L)}{\theta_L(1 - \theta_H)} - \delta \right) \epsilon > 0,
\]

contradicting the optimality of the original contract.

- Thus, the lender’s problem is simplified to

\[
\max_{R_H, R_L, C_L} \beta_H(1 - \theta_H)R_H + \beta_L[(1 - \theta_L)R_L + \delta \theta_L C_L]
\]
subject to

\[(1 - \theta_H)(y - R_H) = (1 - \theta_H)(y - R_L) - \theta_H C_L \quad (1.10)\]

\[(1 - \theta_L)(y - R_L) - \theta_L C_L = U_L \quad (1.11)\]

\[(1 - \theta_H)(y - R_H) \geq U_H. \quad (1.12)\]

– One can check that if \(R_H\) is reduced by \(\epsilon > 0\), then \(R_L\) and \(C_L\) must be changed by \(\frac{\theta_L(1 - \theta_H)}{\theta_H - \theta_L}\epsilon\) and \(-\frac{(1 - \theta_L)(1 - \theta_H)}{\theta_H - \theta_L}\epsilon\), respectively. Thus, the corresponding change of the lender’s payoff is \(\epsilon\) times

\[(1 - \theta_H) \left[ -\beta_H + \frac{\beta_L(1 - \delta)(1 - \theta_L)\theta_L}{\theta_H - \theta_L} \right]. \quad (1.13)\]

– In the optimal contract, we have two cases: (i) if (1.13) is positive, then \((C_L, R_L) = (C_H, R_H) = (0, y - \frac{U_L}{1 - \theta_L})\) (and (1.12) is not binding); (ii) if (1.13) is negative, (1.12) is binding and \(R_H, R_L,\) and \(C_L\) are the solution of (1.10) to (1.12).

– As \(\delta\) is small or \(\beta_L\) big or \(\theta_H\) close to \(\theta_L\), no collateral for type \(\theta_L\) is optimal.

– We may conclude that

(a) High (low) risk type pays a high (low) interest rate but puts down no (some) collateral as long as the optimal contract is separating;

(b) In case collateral is costly or types are likely to be low or two types are close, no collateral is required and the optimal contract is pooling.

1.3.3 Credit Rationing

Adverse selection is typical of financial markets: A lender knows less than a borrower about the quality of a project he invests in. Because of this informational asymmetry, good
quality projects are denied credit so there arises inefficiency in the allocation of investment funds to project. This type of inefficiency is generally referred to as “credit rationing.”

- Consider a unit mass of borrowers who have no wealth but each own a project:
  
  - Each project requires an initial investment of $I = 1$.
  
  - There are two types of projects, safe and risky. A project $i = s, r$ yields a random return, denoted by $X_i$: $X_i = R_i > 0$ (success) with probability $p_i \in [0, 1]$ and $X_i = 0$ (failure) with $1 - p_i$.
  
  - Assume that $p_i R_i = m > 1$ while $p_s > p_i$ and $R_s < R_r$. That is, two types of projects have the same expected return, but type-$s$ is safer than type-$r$.
  
  - The proportion of “safe” borrowers is $\beta$ and that of “risky” borrowers is $1 - \beta$. The bank has a total amount $\alpha$ of funds: $\alpha > \max\{\beta, 1 - \beta\}$

What type of lending contract should the bank offer a borrower? Under symmetric information, the bank would be indifferent about financing either type of borrower: In exchange for the initial investment $\hat{I} = 1$, it requires a repayment of $\hat{D}_i = R_i$ in case of success from type-$i$ borrowers. What if there is asymmetric information problem between lender and borrowers?

- Let us first consider contracts where the bank specifies a fixed repayment, $D$.
  
  - If the bank decides to accommodate type-$r$ only, then setting $D = R_r$ is optimal, which gives the bank a profit of

  $$(1 - \beta)(m - 1). \quad (1.14)$$
– If the bank decides to accommodate both types, setting $D = R_s$ is optimal. Assuming that each applicant has an equal chance of being financed, the bank’s profit is equal to
\[
\alpha \left[ \beta (m - 1) + (1 - \beta)(p_r - R_s - 1) \right].
\] (1.15)

– Ceteris paribus, (1.15) will be higher than (1.14) when $1 - \beta$ is small enough, or when $p_r$ is close enough to $p_s$, in which case some “risky” borrowers cannot get credit even though they would be ready to accept a higher $D$.

Can’t bank do better by offering more sophisticated contracts?

• Consider a contract $(x_i, D_i)$ that offers financing with probability $x_i$ and repayment $D_i$. The bank’s problem is
\[
\max_{(x_s, D_s), (x_r, D_r)} \beta x_s(p_s D_s - 1) + (1 - \beta)x_r(p_r D_r - 1)
\]
subject to
\[
0 \leq x_i \leq 1 \quad \text{for all } i = s, r \quad (IR_i)
\]
\[
D_i \leq R_i \quad \text{for all } i = s, r \quad (IR_i)
\]
\[
x_i p_i (R_i - D_i) \geq x_j p_j (R_j - D_j) \quad \text{for all } i = s, r \quad (IC_i)
\]
\[
\beta x_s + (1 - \beta)x_r \leq \alpha.
\]

– The same line of reasoning as in the previous model shows (how?) that $(IR_s)$ and $(IC_r)$ constraints are binding, which implies that
\[
D_s = R_s \quad \text{and} \quad D_r = R_r - \frac{x_s}{x_r} (R_r - R_s).
\]

– As before, we can ignore $(IC_s)$ and $(IR_r)$ constraints.
• So, the proceeding problem becomes

\[
\max_{x_s,x_r} \beta x_s(p_s R_s - 1) + (1 - \beta) [x_r(p_r R_r - 1) - x_s p_r (R_r - R_s)]
\]

subject to

\[
\beta x_s + (1 - \beta) x_r \leq \alpha.
\]

- The solution of this problem is given by: \(x_r^* = 1\) and

\[
-x_s^* = \begin{cases} 
0 & \text{if } \beta(p_s R_s - 1) - (1 - \beta)p_r (R_r - R_s) < 0 \\
\frac{\alpha - (1 - \beta)}{\beta} & \text{otherwise.}
\end{cases}
\]

- The risky types are not rationed any more.

- The safe types are not fully funded but are indifferent about being funded.

- The informational asymmetry does not necessarily give rise to credit rationing.

- We restricted attention to a fixed repayment schedule, so called debt contract. Why not use a return-contingent repayment schedule, which is to set repayments equal to realized returns and is much more efficient than debt contracts?

### 1.3.4 Optimal Income Taxation

In this part, we study the trade-off between allocative efficiency and redistributive taxation in the presence of adverse selection. A lump-sum taxation does entail no distortion in general, but may not be feasible since its size is limited by the lowest income level. To be able to raise higher tax revenues, the tax need be based on an individual’s income-generating-ability, which is not observable, however so adverse selection arises.
Consider an economy where each individual from a unit mass of population is endowed with production function \( q = \theta e \), where \( q \), \( e \), and \( \theta \) denote income, effort, and ability, respectively.

- A proportion \( \beta \) of individuals has low productivity \( \theta_L \) while a proportion \((1 - \beta)\) has high productivity \( \theta_H > \theta_L \).

- All individuals have the same utility function \( u(q - t - \psi(e)) \), where \( t \) is the net tax paid to (or received from) the government and \( \psi(e) \) is an increasing and convex cost function.

- The government’s budget constraint is given by

\[
0 \leq \beta t_L + (1 - \beta)t_H, \tag{1.16}
\]

where \( t_i \) is the net tax from a type-\( i \) individual.

- Under symmetric information, a utilitarian government maximizing the sum of individual utilities, solves

\[
\max_{(e_L,t_L)} \beta u(\theta_L e_L - t_L - \psi(e_L)) + (1 - \beta)u(\theta_H e_H - t_H - \psi(e_H)),
\]

subject to (1.16).

- At the optimum, (1.16) is binding and the first order conditions yield

\[
u_L' := \theta_L u' (\theta_L e_L - t_L - \psi(e_L)) = \theta_H u' (\theta_H e_H - t_H - \psi(e_H)) =: u_H' \tag{1.17}
\]

\[\psi'(\hat{e}_L) = \theta_L\]

\[\psi'(\hat{e}_H) = \theta_H.\]
• Under asymmetric information, the government has extra conditions to satisfy:

\[ \begin{align*}
\theta_L e_L - t_L - \psi(e_L) & \geq \theta_H e_L - t_H - \psi \left( \frac{\theta_H e_H}{\theta_L} \right) , \\
\theta_H e_H - t_H - \psi(e_H) & \geq \theta_L e_L - t_L - \psi \left( \frac{\theta_L e_L}{\theta_H} \right) .
\end{align*} \tag{IC_L} \tag{IC_H} \]

- Note that the allocation satisfying (1.17) violates (IC_H) since

\[ \theta_H \hat{e}_H - t_H - \psi(\hat{e}_H) = \theta_L \hat{e}_L - t_L - \psi(\hat{e}_L) > \theta_L \hat{e}_L - t_L - \psi \left( \frac{\theta_L \hat{e}_L}{\theta_H} \right) . \]

Thus, in a second-best optimum, (IC_H) must be binding together with (1.16).

- Substituting these two binding constraints and applying the first order condition with respect to \( e_H \) and \( e_L \) yields

\[ \begin{align*}
\psi'(e_H^*) & = \theta_H \\
\psi'(e_L^*) & = \theta_L - (1 - \beta) \gamma \left[ \psi'(e_L^*) - \frac{\theta_L}{\theta_H} \psi' \left( \frac{\theta_L e_L^*}{\theta_H} \right) \right],
\end{align*} \]

where \( \gamma := (u'_L - u'_H)/[\beta u'_L + (1 - \beta) u'_H] \). Thus, \( e_H^* = \hat{e}_H \) but \( e_L^* < \hat{e}_L \), since \( Q > 0 \) (why?).

- We can interpret this result in terms of the income tax code: The marginal tax rate is equal to 0 at output \( q_H = \theta_H e_H^* \) while it is equal to \( \frac{Q}{\theta_L} \) at output \( q_L = \theta_L e_L^* \).

1.4 Appendix: Lagrangian Method

Let \( f \) and \( h_i, i = 1, \ldots, m \) be concave \( C^1 \) functions defined on the open and convex set \( U \subset \mathbb{R}^n \). Consider the following maximization problem:

Maximize \( f(x) \) subject to \( x \in D = \{ z \in U \mid h_i(z) \geq 0, i = 1, \ldots, m \} \).
Set up the Lagrangian for this problem, which is defined on $U \times \mathbb{R}_+^m$, as follow:

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x).$$

We call a pair $(x^*, \lambda^*) \in U \times \mathbb{R}_+^m$ saddle point of $L$ if it satisfies the following conditions:

$$Df(x^*) + \lambda^*_i Dh_i(x^*) = 0 \quad (1.18)$$

$$h_i(x^*) \geq 0 \text{ and } \lambda^*_i h_i(x^*) = 0, \forall i = 1, \cdots, m. \quad (1.19)$$

Here, (1.18) implies that given $\lambda^*$, $x^*$ maximizes $L$ while (1.19) implies that given $x^*$, $\lambda^*$ minimizes $L$, which is why we call $(x^*, \lambda^*)$ saddle point. Condition (1.19) is called complementary slackness condition.

**Theorem 1.1** (Kuhn-Tucker). $x^* \in U$ solves the above problem if and only if there is $\lambda^* \in \mathbb{R}_+^m$ such that $(x^*, \lambda^*)$ is a saddle point of $L$.

**Remark 1.** If functions $f$ and $h_i$ are quasi-concave, then $x^*$ maximizes $f$ over $\mathcal{D}$ provided at least one of the following conditions holds: (a) $Df(x^*) \neq 0$; (b) $f$ is concave.

**Remark 2.** If some of the constraints, say constraints $k$ to $m$, are not binding that is $h_i(x^*) \neq 0$ for $i = k, \cdots, m$, then $x^*$ must be a solution of the following relaxed problem also:

Maximize $f(x)$ subject to $x \in \mathcal{D}' = \{z \in U \mid h_i(z) \geq 0, i = 1, \cdots, k-1\}$. 

33
Chapter 2

Hidden Action: Moral Hazard

• Basic setup:
  
  – Agent takes an action that is not observable or verifiable and thus cannot be contracted upon.

  – The agent’s action is costly to himself but benefits the principal.

  – The agent’s action generates a random outcome, which is both observable and verifiable.

  – Principal designs a contract based on the observable outcome.

We refer this kind of setup as moral hazard problem.

• We are interested in

  – How to design a contract which gives agent an incentive to participate (participation constraint) and choose the ‘right’ action (incentive constraint).
– When and how the unobservability of agent’s action distorts the contract away from the first-best.

• The usual time line is as follows:

– At date 0, principal offers a contract.

– At date 1, agent accepts or rejects the contract. If rejects, then both principal and agent get their outside utilities, zero. If accepts, then they proceed.

– At date 2, agent chooses action.

– At date 3, outcome is realized.

– At date 4, the contract is executed.

• A list of examples includes

– Employer and employee: Induce employee to take a profit-enhancing action ← Base employee’s compensation on employer’s profits

– Plaintiff and attorney: Induce attorney to exert effort to increase plaintiff’s chance of prevailing at trial ← Make attorney’s fee contingent on damages awarded plaintiff.

– Landlord and tenant: Induce tenant to make investment that preserve property’s value to the landlord ← Make tenant post deposit to be forfeited if value declines too much.

\(^1\)These examples draw on Hermalin’s note.
2.1 A Simple $2 \times 2$ Model

- Suppose that there is an employer (principal) and an employee (agent)

  - Agent could be idle ($e = 0$ or low effort) or work hard ($e = 1$ or high effort), which is not observable to the principal.

  - The level of production, which is observable and verifiable, is stochastic, taking two values $\bar{q}$ and $\underline{q}$ with $\bar{q} > q$: Letting $\pi_e$ denote the probability that $\bar{q}$ realizes when agent exerts $e$, we assume $\pi_0 < \pi_1$.

  - The agent’s utility when exerting $e$ and paid $t$: $u(t) - ce$ with $u(\cdot)$ increasing and concave, and $u(0) = 0$. ($h := u^{-1}$, inverse function of $u(\cdot)$)

  - The principal’s utility when $q$ realizes and $t$ is paid to agent: $S(q) - t$.

- The principal can only offer a contract based on the level of $q$: $t(q)$ with $\bar{t} := t(\bar{q})$ and $\underline{t} := t(\underline{q})$. We also denote $\bar{S} := S(\bar{q})$ and $\underline{S} := S(\underline{q})$.

  - If agent exerts $e$, then principal receives the expected utility

    $$V_e = \pi_e(\bar{S} - \bar{t}) + (1 - \pi_e)(\underline{S} - \underline{t}).$$

  - To induce a high effort, the (moral hazard) incentive constraint is needed:

    $$\pi_1 u(\bar{t}) + (1 - \pi_1)u(\underline{t}) - c \geq \pi_0 u(\bar{t}) + (1 - \pi_0)u(\underline{t}).$$

  - To ensure that agent participates, the participation constraint is needed:

    $$\pi_1 u(\bar{t}) + (1 - \pi_1)u(\underline{t}) - c \geq 0.$$
2.1.1 First-Best: Complete Information Optimal Contract

In this benchmark case, we assume that the effort level can contracted upon since it is observable and verifiable.

- If principal wants to induce effort level $e$, his problem becomes

$$
\max_{\bar{t}, \bar{t}} \pi_e(S - \bar{t}) + (1 - \pi_e)(\bar{S} - \bar{t})
$$

s.t. \hspace{0.5cm} \pi_e u(\bar{t}) + (1 - \pi_e)u(\bar{t}) - ce \geq 0 \quad (2.1)

- Letting $\lambda$ denote the Lagrangean multiplier for the participation constraint (2.1), the first-order conditions w.r.t. $\bar{t}$ and $\bar{t}$ yield

$$
-\pi_e + \lambda \pi_e u'(\bar{t}^f) = 0
$$

$$
-(1 - \pi_e) + \lambda (1 - \pi_e) u'(\bar{t}^f) = 0
$$

From this, we immediately derive $\lambda = \frac{1}{u'(\bar{t}^f)} = \frac{1}{u'(\bar{t}^f)}$, or $\bar{t}^f = \bar{t}^f = \bar{t}^f$ for some $t^f \geq 0$.

- Since (2.1) must be binding at the optimum, $t^f = h(ce)$.

- So, the risk-neutral principal offers a full insurance to the risk-averse agent and then extracts the full surplus.

- Principal prefers $e = 1$ if

$$
\pi_1 S + (1 - \pi_1)S - h(c) \geq \pi_0 S + (1 - \pi_0)S
$$

or

$$
\Delta \pi \Delta S \geq h(c) \quad (2.2)
$$

and otherwise prefers $e = 0$. 

37
From now on, we assume (2.2) holds so principal would prefer \( e = 1 \) without moral hazard problem.

### 2.1.2 Risk-Neutral Agent and the First-Best Contract

Suppose that there is moral hazard problem and suppose also that agent is risk-neutral, i.e. \( u(t) = t \). In this case, the principal can achieve the first-best outcome.

- The principal’s optimal contract to induce \( e = 1 \) must solve the following problem:

\[
\begin{align*}
\max_{t, \pi} & \quad \pi_1(S - t) + (1 - \pi_1)(S - \bar{t}) \\
\text{s.t.} & \quad \pi_1 \bar{t} + (1 - \pi_1) \bar{t} - c \geq \pi_0 \bar{t} + (1 - \pi_0) \bar{t} \\
& \quad \pi_1 \bar{t} + (1 - \pi_1) \bar{t} - c \geq 0
\end{align*}
\]

(2.3) (2.4)

- One contract \((\bar{t}, \bar{\pi})\) that achieves the first-best can be found by making both (2.4) and (2.3) binding, which yields

\[
\bar{t}^* = -\frac{\pi_0}{\Delta \pi} c < 0 \quad \text{and} \quad \bar{\pi}^* = \frac{1 - \pi_0}{\Delta \pi} c > 0.
\]

So, agent is rewarded if output is high while punished if output is low. The agent’s expected gain of exerting a high effort is \( \Delta \pi (\bar{t}^* - \bar{t}^*) = c \).

- The principal obtains the first-best surplus, \( \pi_1 \bar{S} + (1 - \pi_1) \bar{S} - c \).

- In fact, there are many other contracts that achieve the first-best. Among them, the “selling the firm” contract is often observed in the real world.

- The idea is to let the agent buy out the principal’s firm at a fixed price \( T^* \) by setting

\[
\bar{t}^* = \bar{S} - T^* \quad \text{and} \quad \bar{t}^* = \bar{S} - T^* \quad \text{with} \quad T^* = \pi_1 \bar{S} + (1 - \pi_1) \bar{S} - c
\]

Easy to see that this achieves the first-best surplus for the principal.
Note that the agent’s incentive constraint is satisfied since

\[ \Delta \pi (\bar{t}' - \bar{t}''') = \Delta \pi \Delta S > h(c) = c, \]

which holds due to (2.2).

### 2.1.3 Limited Liability and Second-Best Contract

Let us assume that the agent has no wealth and is protected by limited liability constraint that the transfer received by the agent should not be lower than zero:

\begin{align*}
\bar{t} &\geq 0 \quad (2.5) \\
\bar{t} &\geq 0. \quad (2.6)
\end{align*}

- Then, the principal’s problem is written as

\[
\max_{\bar{t}} \pi_1(S - \bar{t}) + (1 - \pi_1)(S - \bar{t})
\]

subject to (2.3) to (2.6).

- One can argue that (2.3) and (2.6) must be binding, from which we obtain \( t^* = 0 \) and \( \bar{t}^* = \frac{c}{\Delta \pi} \). Then, it can be checked that other constraints are automatically satisfied.

- The agent’s rent is positive:

\[
\pi_1 \bar{t}^* + (1 - \pi_1)\bar{t}^* - c = \frac{\pi_0}{\Delta \pi} c > 0.
\]

- So, the principal’s payoff is equal to

\[
V_1^* := \pi_1(S - \bar{t}^*) + (1 - \pi_1)(S - \bar{t}^*) = \pi_1 S + (1 - \pi_1)S - \frac{\pi_1}{\Delta \pi} c,
\]

which is lower than her first-best payoff by as much as the agent’s rent.
The principal would then like to induce $e = 1$ if $V_1^* \geq V_0$ or

$$\frac{\Delta \pi \Delta S}{\Delta \pi} \geq \frac{\pi_1 c}{\Delta \pi} = c + \frac{\pi_0 c}{\Delta \pi}$$

and otherwise induce $e = 0$.

### 2.1.4 Risk-Averse Agent and Second-Best Contract

Let us now turn to the second source of inefficiency in a moral hazard context: Agent’s risk aversion.

- The principal’s problem is written as

\[
(P) \quad \max_{\tilde{t}} \pi_1(S - \tilde{t}) + (1 - \pi_1)(S - \tilde{t})
\]

subject to

\[
\pi_1 u(\tilde{t}) + (1 - \pi_1)u(t) - c \geq \pi_0 u(\tilde{t}) + (1 - \pi_0)u(t) \tag{2.7}
\]

\[
\pi_1 u(\tilde{t}) + (1 - \pi_1)u(t) - c \geq 0. \tag{2.8}
\]

- It is not clear that $(P)$ is a ‘concave program’ for which the Kuhn-Tucker condition is necessary and sufficient.

- Making the change of variables $\bar{u} := u(\tilde{t})$ and $\bar{u} := u(t)$ or equivalently $\tilde{t} = h(\bar{u})$ and $t = h(\bar{u})$, the problem $(P)$ turns into

\[
(P') \quad \max_{\bar{u}} \pi_1(S - h(\bar{u})) + (1 - \pi_1)(S - h(\bar{u}))
\]
subject to

\[ \pi_1 \bar{u} + (1 - \pi_1)u - c \geq \pi_0 \bar{u} + (1 - \pi_0)u \tag{2.9} \]
\[ \pi_1 \bar{u} + (1 - \pi_1)u - c \geq 0, \tag{2.10} \]

which is a concave program.

- Set up the Lagrangian function for \((P')\) as

\[ \mathcal{L} = \pi_1(\bar{S} - h(\bar{u})) + (1 - \pi_1)(S - h(u)) + \lambda(\Delta \pi (\bar{u} - u) - c) + \mu(\pi_1 \bar{u} + (1 - \pi_1)u - c). \]

- Differentiating \(\mathcal{L}\) with \(\bar{u}\) and \(u\) yields

\[ -\pi_1 h'(\bar{u}^*) + \lambda \Delta \pi + \mu \pi_1 = 0 \]
\[ -(1 - \pi_1)h'(u^*) - \lambda \Delta \pi + \mu (1 - \pi_1) = 0. \]

Rearranging, we obtain

\[ h'(\bar{u}^*) = \mu + \frac{\lambda \Delta \pi}{\pi_1} \tag{2.11} \]
\[ h'(u^*) = \mu - \frac{\lambda \Delta \pi}{(1 - \pi_1)}. \tag{2.12} \]

- One can argue that both \(\lambda\) and \(\mu\) are positive so that (2.9) and (2.10) are both binding:

(i) If \(\lambda = 0\), then (2.11) and (2.12) imply \(\bar{u}^* = u^*\), which violates (2.9), a contradiction.
(ii) If \(\mu = 0\), then (2.12) implies \(h'(u^*) < 0\), which is not possible since \(h(\cdot) = u^{-1}(\cdot)\) is an increasing function.

- From (2.9) and (2.10) as equality, we obtain

\[ \bar{u}^* = \frac{1 - \pi_0}{\Delta \pi} c \]
\[ u^* = \frac{-\pi_0}{\Delta \pi} c. \]
Using this, one can show that the second-best cost of inducing a high effort is higher than the first-best cost:

\[ C^* := \pi_1 \bar{t}^* + (1 - \pi_1) t^* = \pi_1 h(\bar{u}^*) + (1 - \pi_1) h(u^*) \]

\[ > h(\pi_1 \bar{u}^* + (1 - \pi_1) u^*) = h(c), \]

where the inequality holds due to the convexity of \( h(\cdot) \).

So, the principal would like to induce \( e = 1 \) if

\[ \Delta \pi \Delta S \geq C^* \]

and otherwise induce \( e = 0 \).

### 2.2 Extensions

#### 2.2.1 More than 2 Outcomes

Suppose that there are \( n \) possible output levels, \( \{q_1, \ldots, q_n\} \) with \( q_1 < q_2 < \cdots < q_n \). Each \( q_i \) is realized with probability \( \pi_{ie} \) given the effort level, \( e_i \), with \( \pi_e = (\pi_{1e}, \cdots, \pi_{ne}) \).

Now, the optimal contract that induces \( e = 1 \) must solve

\[
\max_{t_1, \ldots, t_n} \sum_{i=1}^{n} \pi_{i1} (S_i - t_i),
\]

subject to

\[
\sum_{i=1}^{n} \pi_{i1} u(t_i) - c \geq \sum_{i=1}^{n} \pi_{i0} u(t_i) \tag{2.13}
\]

\[
\sum_{i=1}^{n} \pi_{i1} u(t_i) - c \geq 0. \tag{2.14}
\]
Let $\lambda$ and $\mu$ denote the Lagrangian multipliers for (2.13) and (2.14), respectively. Then, the first-order condition with respect to $t_i$ is given as

$$\frac{\pi_{i1}}{u'(t_i^*)} = \mu \pi_{i1} + \lambda (\pi_{i1} - \pi_{i0}).$$

(2.15)

Analogously to the two-outcome case, both $\lambda$ and $\mu$ can be shown to be positive: An argument to show $\mu > 0$ is very similar and thus is omitted. To show $\mu > 0$, sum up equations (2.15) across $i = 1, \ldots, n$ and obtain

$$0 < \sum_{i=1}^{n} \frac{\pi_{i1}}{u'(t_i^*)} = \mu \sum_{i=1}^{n} \pi_{i1} + \lambda \sum_{i=1}^{n} (\pi_{i1} - \pi_{i0}) = \mu.$$

One may ask under what condition $t_i^*$ is increasing with $i$ or wage is increasing with output level. The condition for this is that the likelihood ratio is monotone, or $\frac{\pi_{i1}}{\pi_{i0}}$ is increasing with $i$. To see it, rewrite (2.15) as

$$\frac{1}{u'(t_i^*)} = \mu + \lambda \left(1 - \frac{\pi_{i0}}{\pi_{i1}}\right).$$

### 2.2.2 Comparing Information Structures

The performance of optimal contract in moral hazard setup is in part affected by how informative a contractible variable is about an agent’s hidden action. In the extreme case the contractible variable perfectly reveals the agent’s action, for instance, the principal will be able to achieve the first-best outcome. In this part, we will investigate how the principal’s payoff is affected by the informativeness of contractible variables in less than extreme cases.

- Consider two information structures, $\pi$ and $\bar{\pi}$, satisfying

$$\bar{\pi}_{ej} = \sum_{i=1}^{n} \pi_{ei} p_{ij} \text{ for all } j = 1, \ldots, n$$

for some $n \times n$ transition matrix $P = (p_{ij})$ with $\sum_{j=1}^{n} p_{ij} = 1$ for each $i = 1, \ldots, n$.  

43
We say that the information structure \( \pi \) is *Blackwell-sufficient* for the information structure \( \bar{\pi} \). In other words, the information structure \( \bar{\pi} \) is a *garbling* of the original information structure \( \pi \).

Let us define \( C^*(\pi) \) and \( C^*(\bar{\pi}) \) as the second-best costs of inducing \( e = 1 \) under the information structures \( \pi \) and \( \bar{\pi} \), that is

\[
C^*(\pi) = \sum_{i=1}^{n} \pi_{1i} t_i^* \quad \text{and} \quad C^*(\bar{\pi}) = \sum_{i=1}^{n} \bar{\pi}_{1i} \bar{t}_i^*
\]

- The information structure \( \pi \) is *more efficient* than \( \bar{\pi} \) if \( C^*(\pi) \leq C^*(\bar{\pi}) \), which is indeed true if \( \pi \) is Blackwell-sufficient for \( \bar{\pi} \), as shown in the following proof.

**Proof.** Consider a different wage schedule \((t'_1, \ldots, t'_n)\) defined by

\[
u(t'_i) = \sum_{j=1}^{n} p_{ij} u(\bar{t}_j^*).
\]

This wage schedule satisfies the incentive compatibility and individual rationality conditions under the original information structure \( \pi \):

\[
c = \sum_{j=1}^{n} (\bar{\pi}_{1j} - \bar{\pi}_{0j}) u(\bar{t}_j^*) = \sum_{j=1}^{n} \left( \sum_{i=1}^{n} (\pi_{1i} - \pi_{0i}) \right) p_{ij} u(\bar{t}_j^*) = \sum_{i=1}^{n} (\pi_{1i} - \pi_{0i}) \left( \sum_{j=1}^{n} p_{ij} u(\bar{t}_j^*) \right)
\]

and

\[
c = \sum_{j=1}^{n} \bar{\pi}_{1j} u(\bar{t}_j^*) = \sum_{j=1}^{n} \left( \sum_{i=1}^{n} \pi_{1i} p_{ij} \right) u(\bar{t}_j^*) = \sum_{i=1}^{n} \left( \sum_{j=1}^{n} p_{ij} u(\bar{t}_j^*) \right).
\]

Thus, we must have

\[
\sum_{i=1}^{n} \pi_{i} t'_{1i} \geq C^*(\pi).
\]
We can also obtain by the Jensen’s inequality

\[
\sum_{i=1}^{n} \pi_{1i} t_i' = \sum_{i=1}^{n} \pi_{1i} h \left( \sum_{j=1}^{n} p_{ij} u(\tilde{t}_j^*) \right) \leq \sum_{i=1}^{n} \pi_{1i} \left( \sum_{j=1}^{n} p_{ij} h(u(\tilde{t}_j^*)) \right) = \sum_{i=1}^{n} \pi_{1i} \left( \sum_{j=1}^{n} p_{ij} \tilde{t}_j^* \right) = \sum_{j=1}^{n} \tilde{\pi}_{1j} \tilde{t}_j^*.
\]

Thus, we conclude that \( C^*(\pi) \leq \sum_{i=1}^{n} \pi_{1i} t_i' \leq \sum_{j=1}^{n} \tilde{\pi}_{1j} \tilde{t}_j^* = C^*(\tilde{\pi}) \).

### 2.2.3 Moral Hazard and Renegotiation

Let us alternatively assume that the agent’s effort is not verifiable but can be observed by the principal. Also, after observing the effort choice by agent but before the output is realized, the principal can propose to renegotiate the initial contract. We investigate whether this change in the contractual environment affects the implementability of the first-best and then how.

It is more convenient for our analysis to generalize the setup as follows. The effort \( e \), which can be any nonnegative value, costs \( g(e) \). The principal’s profit, denoted \( S \), is drawn from \([0, \bar{S}]\) according to distribution \( p(S|e) \). The contract is now a function, \( t : [0, \bar{S}] \to \mathbb{R} \).

- Without renegotiation, the second-best contract \( t^*(\cdot) \) for inducing a target effort level \( e \) will satisfy \((IC)\) and \((IR)\) as follows:

\[
e = \arg \max_{e' \geq 0} \int_{0}^{\bar{S}} u(t^*(S)) p(S|e') dS - g(e') \quad (IC)
\]

and

\[
\int_{0}^{\bar{S}} u(t^*(S)) p(S|e) dS - g(e) = 0. \quad (IR)
\]
Suppose now that the principal is allowed to renegotiate after observing the agent’s effort but before the profit is realized.

- After \( e' \) has been chosen, the principal offers \( \tilde{t}(\cdot|e') \) in renegotiation that solves the following program:

\[
\max_{\tilde{t}(\cdot|e')} \int_0^S (S - t(S|e'))p(S|e') \\
\text{subject to} \int_0^S u(t(S|e'))p(S|e')dS \geq \int_0^S u(t^*(S))p(S|e')dS. \tag{2.16}
\]

- Clearly, \( \tilde{t}(\cdot|e') \) must be a full insurance contract that binds (2.16), that is

\[
\tilde{t}(S|e') = \tilde{t}(e') := h \left( \int_0^S u(t^*(S))p(S|e')dS \right) \text{ for all } S \in [0, S].
\]

- Note that \( \tilde{t}(e) \) corresponds to the minimum cost of inducing \( e \):

\[
\tilde{t}(e) = h \left( \int_0^S u(t^*(S))p(S|e)dS \right) = g(e).
\]

- It is important to ensure that the renegotiation does not change the agent’s incentive to choose \( e \) rather than \( e' \neq e \). For this, it suffices to see that

\[
\int_0^S u(\tilde{t}(S|e'))p(S|e')dS - g(e') = u(\tilde{t}(e')) - g(e') = \int_0^S u(t^*(S))p(S|e')dS - g(e').
\]

- We conclude that the principal can achieve the first-best outcome.

- What happens if the renegotiation occurs without principal observing the agent’s effort?
2.3 Applications

2.3.1 Moral Hazard in Insurance Markets

Moral hazard is pervasive in insurance markets. Let us consider a risk-averse agent with utility function \( u(\cdot) \) and initial wealth \( w \). Effort \( e \in \{0, 1\} \) is a level of safety care while \( ce \) is the cost of choosing \( e \). Given the effort level \( e \), an accident occurs with probability \( 1 - \pi_e \) with the damage worth \( d \) being incurred and otherwise no damage incurred. We assume that without insurance, the agent would choose \( e = 1 \) (or due care), that is

\[
\Delta \pi (u(w) - u(w - d)) \geq c,
\]

where \( \Delta \pi \equiv \pi_1 - \pi_0 > 0 \). Thus, the agent’s outside utility, denoted \( \hat{u} \), is given as

\[
\hat{u} = \pi_1 u(w) + (1 - \pi_1)u(w - d) - c.
\]

Let us assume that there is a competitive insurance market. In other words, there are many insurance companies offering insurance contracts to the agent. An insurance contract can be described by \((\bar{u}, u)\), where \( \bar{u} \) (\( u \), resp.) is the agent’s utility from his final wealth if no accident (accident, resp.) occurs.\(^2\) Because of the competition among insurance companies, the contract should maximize the agent’s expected utility subject to the standard incentive compatibility constraint

\[
\bar{u} - u \geq \frac{c}{\Delta \pi}, \tag{2.17}
\]

and the constraint that the profit of the insurance company be non-negative

\[
\pi_1(w - h(\bar{u})) + (1 - \pi_1)(w - d - h(u)) \geq 0. \tag{2.18}
\]

\(^2\)Note that since the total wealth is divided between agent and insurance company, the latter’s profit is equal to \( w - h(\bar{u}) \) if accident occurs and equal to \( w - d - h(u) \) if not.
Without moral hazard, the incentive compatibility constraint can be ignored so the equilibrium contract offered by insurance company must solve the following problem:

\[
\max_{(\pi, \underline{u})} \pi_1 \underline{u} + (1 - \pi_1) \underline{u} - c
\]
subject to (2.18).

- As in Part 2.1.1, the solution of the above problem yields \( \underline{u} = \pi \), which is denoted as \( u^f \). Letting \( U^f = u^f - c \) denote the agent’s utility at the first-best, we have

\[
h(U^f + c) = w - d(1 - \pi_1),
\]

since (2.18) must be binding.

- The market would not break down if the agent is (weakly) better off with insurance than without, or \( U^f \geq \hat{u} \), which is equivalent to

\[
u(w - d(1 - \pi_1)) - c \geq \pi_1 u(w) + (1 - \pi_1)u(w - d) - c. \tag{2.19}
\]

This always holds since \( u(\cdot) \) is a concave function.

Under moral hazard, assuming that \( e = 1 \) is chosen in the equilibrium, the equilibrium contract must solve the following problem:

\[
(P) \quad \max_{(\pi, \underline{u})} \pi_1 \underline{u} + (1 - \pi_1) \underline{u} - c
\]
subject to (2.17) and (2.18).

\[^3\text{Note that we need to check that the agent is willing to participate, and also that the insurance firms cannot do better by inducing the agent to choose } e = 0. \text{ We will later see whether and when they hold.}\]
Let $\lambda$ and $\mu$ denote the Langrangian multipliers for (2.17) and (2.18). Then, the first-order condition is given by

$$\pi_1 + \lambda = \mu \pi_1 h'(\bar{u}^*)$$

and

$$1 - \pi_1 - \lambda = \mu (1 - \pi_1) h'(u^*) .$$

By the same argument as in Part 2.1.4, one can show that both $\lambda$ and $\mu$ are positive so that (2.17) and (2.18) are binding.

Letting $U^* = \pi_1 \bar{u}^* + (1 - \pi_1) u^* - c$ denote the agent’s expected utility at the second-best, the binding constraint (2.17) yields

$$u^* = U^* - \frac{\pi_0}{\Delta \pi} c$$

and

$$\bar{u}^* = U^* + \frac{1 - \pi_0}{\Delta \pi} c .$$

Plugging (2.20) and (2.21) into the binding constraint (2.18) yields

$$\pi_1 h \left( U^* + \frac{1 - \pi_0}{\Delta \pi} \right) + (1 - \pi_1) h \left( U^* - \frac{\pi_0}{\Delta \pi} c \right) = w - d (1 - \pi_1) .$$

We need to check that (i) the agent has an incentive to participate or $U^* \geq \hat{u}$ and (ii) an insurance company can do better by offering a contract that induces the agent to choose $e = 0$.

(1) To see (i), it is enough to note that the pair of utility $(u(w), u(w - d))$ satisfies both (2.17) and (2.18) and also yields the expected utility $\hat{u}$, which means that $U^*$, the expected utility obtained from the problem $(P)$ above, cannot be smaller than $\hat{u}$, as desired.
(2) For (ii), an insurance company should not be able to make the agent better off by offering the optimal contract inducing \( e = 0 \), which is obtained from solving

\[
\max_{(\pi, u)} \pi_0 \bar{u} + (1 - \pi_0)u - c
\]

subject to

\[
\pi_0(w - h(u)) + (1 - \pi_0)(w - d - h(\bar{u})) \geq 0.
\]

The solution of this problem is \( \bar{u} = u = u(w - d(1 - \pi_0)) \) (Why?). Thus, (ii) will holds if

\[
U^* \geq u(w - d(1 - \pi_0)) - c.
\]

Otherwise, \( e = 0 \) will be chosen in equilibrium and the equilibrium contract will be \( \pi = u = u(w - d(1 - \pi_0)) \).

### 2.3.2 Moral Hazard in Teams

Consider \( n \) agents working in a partnership to produce aggregate output

\[
Q = Q(a_1, a_2, \ldots, a_n),
\]

where \( a_i \in [0, \infty) \) is the agent \( i \)'s action. Assume that the function \( Q(\cdot) \) satisfies

\[
\frac{\partial Q}{\partial a_i} > 0, \quad \frac{\partial^2 Q}{\partial a_i^2} < 0, \quad \text{and} \quad \frac{\partial^2 Q}{\partial a_i \partial a_j} \geq 0.
\]

Suppose that each agent is risk-neutral: With wage \( w_i \) and effort \( a_i \), the agent \( i \)'s utility is \( w_i - g_i(a_i) \) with \( g_i(\cdot) \) strictly increasing and being convex. Since each agent’s effort or individual output is not observable, a contract in the partnership can only depend on the aggregate output: \( w(Q) = (w_1(Q), w_2(Q), \ldots, w_n(Q)) \). We require each \( w_i(\cdot) \) to
be differentiable (almost everywhere). All agents in the partnership share the aggregate output in the following sense:

$$\sum_{i=1}^{n} w_i(Q) = Q \text{ for each } Q.$$  \hfill (2.23)

A key observation in this model is the externality among agents: If agents are rewarded for raising output, then one agent working hard to raise aggregate output will benefit others also. This provides a possible source for an inefficient outcome.

- The first-best profile of actions $\hat{a} = (\hat{a}_1, \cdots, \hat{a}_n)$ solves the following problem

$$\max_{(a_1, \cdots, a_n)} \sum_{i=1}^{n} (w_i(Q) - g_i(a_i)) = Q(a_1, \cdots, a_n) - \sum_{i=1}^{n} g_i(a_i).$$

- The corresponding first-order condition is given by

$$\frac{\partial Q(\hat{a})}{\partial a_i} = g'_i(\hat{a}_i) \text{ for each } i.$$  \hfill (2.24)

When actions are not contractible, the first-best is no longer achievable.

- Given other agents’ actions $a_{-i} = (a_1, \cdots, a_{i-1}, a_{i+1}, \cdots, a_n)$, agent $i$ will choose an effort $a_i$ that satisfies the first-order condition:

$$\frac{dw_i(Q(a_i, a_{-i}))}{dQ} \frac{\partial Q(a_i, a_{-i})}{\partial a_i} = g'_i(a_i) \text{ for each } i.$$

- If the first-best action profile $\hat{a}$ were a Nash equilibrium, then (2.24) would imply

$$\frac{dw_i(Q(\hat{a}_i, \hat{a}_{-i}))}{dQ} = 1 \text{ for each } i.$$  \hfill (2.25)

- However, (2.23) tells us that for all $a$,

$$\sum_{j=1}^{n} \frac{dw_j(Q(a))}{dQ} \frac{\partial Q(a)}{\partial a_i} = \frac{\partial Q(a)}{\partial a_i},$$

which contradicts with (2.25).
Here, the failure to achieve the first-best is partly due to the balanced-budget constraint (2.23). Indeed, if there is a third party who can act as a budget breaker, then it is possible to design a contract which achieves the first-best outcome.

- Let a third party pay \( w_i(Q) = Q \) to each of agents, who in return hand over the entire output \( Q \) and make an up-front payment \( z_i \) to the third party.

  - Given this contract, each agent \( i \) optimally chooses \( a_i = \hat{a}_i \), the first-best level.

  - Two participation constraints need to be satisfied: the third party not losing money requires

    \[
    Q(\hat{a}) + \sum_{i=1}^{n} z_i \geq nQ(\hat{a}),
    \]  

    while each agent not losing money requires

    \[
    z_i \leq Q(\hat{a}) - g_i(\hat{a}_i).  \tag{2.27}
    \]

  - There exist \( z_i \)'s that satisfy both (2.26) and (2.27) if

    \[
    nQ(\hat{a}) - \sum_{i=1}^{n} g_i(\hat{a}_i) \geq (n - 1)Q(\hat{a}) \quad \text{or} \quad Q(\hat{a}) - \sum_{i=1}^{n} g_i(\hat{a}_i) \geq 0,
    \]

    which has to be satisfied at the first-best.

There is some problem with the above contract: If the team performance is better than the first-best, then the third party will be forced to overpay the agents and lose money. Then, the agents might be able to exploit this opportunity by deciding collusively on an excessively high level of actions. There is an alternative contract, called Mirrlees contract, that avoids such problem and achieves the first-best.
• Each agent is rewarded with bonus $b_i$ if output level $Q(\hat{a})$ is realized and punished with a penalty $k$ to be paid to the budget breaker if any other output is realized.

– Under this contract, it is a Nash equilibrium for each agent to choose $\hat{a}_i$: given that other agents choose $\hat{a}_{-i}$, each agent $i$ will choose $\hat{a}_i$ if $b_i$ and $k$ satisfy

$$b_i - g_i(\hat{a}_i) \geq -k.$$ 

– However, this game might have other equilibria in which either only a few agents work or all shirk.
Chapter 3

Incomplete Contracts

So far, the parties could write a complete contract that identifies every contingency which will arise at a later stage.

• However, the parties’ ability to write such contract might well be limited for the following reasons

  – *Unforeseen contingencies*: “Parties cannot define ex ante the contingencies that may occur later on.”

  – *Cost of writing contracts*: “Even if one could foresee all contingencies, they might be so numerous that it would be too costly to describe them in a contract.”

  – *Cost of enforcing contracts*: “Courts must understand the terms of the contract and verify the contracted upon contingencies and actions in order to enforce the contract.”

• For the above reasons, the contracts could be *incomplete* in the sense that they leave out what to do in some contingencies, which has the following implications:
– The parties have to rely on ex post bargaining or renegotiation in those contingencies

– The parties’ opportunism at the ex post bargaining stage might deter each other from taking an efficient action at an earlier stage (e.g. GM-Fisher Body)

– The organizational arrangement such as allocation of ownership rights, authority, or financial control becomes important
3.1 The Holdup Problem

- Consider the following setup with one seller (S) and one buyer (B) where two parties are not integrated:

  - At date 0, the parties sign a contract

  - At date 1 (ex ante), each party decides on how much to invest, \( i \in [0, 1] \) at cost \( \phi(i) \) by S, and \( j \in [0, 1] \) at cost \( \psi(j) \) by B. (Assume that \( \phi(\cdot) \) and \( \psi(\cdot) \) are increasing and convex functions.)

  - At date 2 (ex post), the state of nature \( \theta = (v, c) \), where \( v \) is B’s value and \( c \) S’s cost, is realized as follows: \( v \in \{v_L, v_H\} \) with \( v_L < v_H \) and \( \text{Prob}(v_H) = j \), and \( c \in \{c_L, c_H\} \) with \( c_L < c_H \) and \( \text{Prob}(c_L) = i \).

  - At date 3, two parties trade a quantity \( q \in [0, 1] \) at price \( P \).

  - The resulting payoff levels are \( vq - P - \psi(j) \) for B and \( P - cq - \phi(i) \) for S.

  - Assuming that \( c_H > v_H > c_L > v_L \), the ex-post efficient level of trade is \( q = 1 \) if \( \theta = (v_H, c_L) \) and \( q = 0 \) otherwise. Given this, the ex-ante efficient investment solves

    \[
    \max_{i, j} i j (v_H - c_L) - \psi(j) - \phi(i),
    \]

    which yields the optimal investment levels \( i^* \) and \( j^* \) as follows:

    \[
    i^*(v_H - c_L) = \psi'(j^*) \quad \text{and} \quad j^*(v_H - c_L) = \phi'(i^*). \tag{3.1}
    \]

- What happens if \( i, j, \) and \( \theta \) are not contractible, that is the contracts are incomplete? Take an extreme case where there is no contract signed at date 0.
Two parties bargain over $q$ and $P$ once $\theta$ is realized. According to the Nash bargaining solution, two parties evenly split the (extra) surplus so the payoffs become

$$ij\left[\frac{1}{2}(v_H - c_L)\right] - \psi(j)$$

for $B$ and

$$ij\left[\frac{1}{2}(v_H - c_L)\right] - \phi(i)$$

for $S$, which lead to the first-order condition

$$\frac{1}{2}j(v_H - c_L) = \psi'(\hat{j}) \quad \text{and} \quad \frac{1}{2}\hat{j}(v_H - c_L) = \phi'(\hat{i}) \quad (3.2)$$

Comparing (3.1) and (3.2), one can see that $\hat{j} < j^\ast$ and $\hat{i} < i^\ast$, underinvestment.

### 3.2 Organizational Solution

One solution to the underinvestment problem due to the incomplete contract is organizational arrangement such as the integration of firms by (re)defining ownership rights. To illustrate the point, we modify the above model and assume that only buyer makes an investment while seller’s cost is fixed at $c$ with $v_H > c > v_L$. Also, $B$ and $S$ can find another, less efficient and competitive, trading partner: seller $S'$ with cost $c' \in (c, v_H)$ and buyer $B'$ with $v' \in (c, v_H)$. Defining $\chi = v_H - c$, $\chi_S = v' - c$ and $\chi_B = v_H - c'$, assume that

$$\chi > \chi_S + \chi_B \quad (3.3)$$

so $B$ and $S$ can generate a greater surplus by trading with each other than with outside partners. Also, (3.3) implies that the marginal effect of $B$’s investment is greater within the relationship or relationship-specific. We consider two other regimes in addition to the
no-integration regime: $B$-integration and $S$-integration. Under $i$-integration, the party $i$ has the ownership of both $B$ and $S$’s assets. The ownership title gives the owner the right to all revenues generated by the asset.

- Under no integration, $B$ and $S$ each hold their own assets so that if $B$ and $S$ fail to trade ex post, then they obtain the outside payoffs $\chi_B$ and $\chi_S$, respectively.
  
  Thus, $B$’s ex ante payoff is
  \[
j[jB + \frac{1}{2}(\chi - \chi_B - \chi_S)] - \psi(j),\]

  which leads to the first-order condition
  \[
  \chi_B + \frac{1}{2}(\chi - \chi_B - \chi_S) = \frac{1}{2}(\chi + \chi_B - \chi_S) = \psi(j^*_N).
  \]

- Under the $B$-integration, if $B$ and $S$ fail to trade ex post, $B$ obtains the outside payoff $\chi_B$ while $S$ obtains zero since he has no asset.
  
  Thus, $B$’s ex ante payoff is given by
  \[
j[jB + \frac{1}{2}(\chi - \chi_B)] - \psi(j),\]

  which leads to the first-order condition
  \[
  \chi_B + \frac{1}{2}(\chi - \chi_B) = \frac{1}{2}(\chi + \chi_B) = \psi(j^*_B).
  \]

- Under the $S$-integration, if $B$ and $S$ fail to trade ex post, then $S$ obtains the outside payoff $\chi_S$ while $B$ obtains zero.
  
  Thus, $B$’s ex ante payoff is given by
  \[
j[\frac{1}{2}(\chi - \chi_S)] - \psi(j),\]
which leads to the first-order condition

\[ \frac{1}{2}(\chi - \chi_s) = \psi'(j^*_S). \]  

(3.6)

Which regime is better in terms of generating a higher investment from \( B \)? Comparing (3.4), (3.5), and (3.6), it is straightforward to see that \( j^*_S < j^*_N < j^*_B \). The lessen here is that

1. The ownership protects the investing party against the ex post opportunism of other parties.

2. The equilibrium allocation of ownership rights will be determined by the relative value of each party’s ex ante specific investments.

The above model has many limitations. Among others, the ranking of investment may depend on the particular bargaining solution assumed. More importantly, the organization intervention may not be necessary for resolving the holdup problem, in particular when a simple contract suffices to do so.

### 3.3 Contractual Solution

In this section, we show that there exists a simple contract that can induce the ex-ante investment and resolve the holdup problem. Let us keep the model from the previous section and assume for simplicity that there is no outside partner. We consider two contracts: specific performance contract (SPC) and option contract (OC).

- The SPC specifies the following action to be taken unless renegotiation occurs: ‘\( B \) and
S trade a quantity \( \hat{q} \in (0, 1) \) at price \( \hat{P} \). We ask if there is some \( \hat{q} \) that induces \( B \) to choose \( j^* \). There are two cases to be considered.

- If \( \theta = (v_H, c) \), then the renegotiation will result in \( q = 1 \) since it is more efficient than \( \hat{q} < 1 \). In this case, the \( B \)'s payoff will be

\[
v_H \hat{q} - \hat{P} + \frac{1}{2}(1 - \hat{q})(v_H - c).
\]

- If \( \theta = (v_L, c) \), then the renegotiation will result in \( q = 0 \), which is more efficient than \( \hat{q} > 0 \). In this case, the \( B \)'s payoff will be

\[
v_L \hat{q} - \hat{P} + \frac{1}{2}\hat{q}(c - v_L).
\]

Thus, \( B \)'s ex ante payoff from investing \( j \) is

\[
j[v_H \hat{q} - \hat{P} + \frac{1}{2}(1 - \hat{q})(v_H - c)] + (1 - j)[v_L \hat{q} - \hat{P} + \frac{1}{2}\hat{q}(c - v_L)] - \psi(j).
\]

which leads to the first order condition

\[
\frac{1}{2}v_H (1 + \hat{q}) - \frac{1}{2}\hat{q}v_L - \frac{1}{2}c = \psi'(j). \tag{3.7}
\]

- If some \( \hat{q} \) can be set to make the LHS of (3.7) equal to \( v_H - c \), then \( j^* \), efficient level, will solve (3.7). Note that with \( \hat{q} = 0 \),

\[
\text{LHS of (3.7)} = \frac{1}{2}(v_H - c) < v_H - c
\]

while with \( \hat{q} = 1 \),

\[
\text{LHS of (3.7)} = v_H - \frac{1}{2}v_L - \frac{1}{2}c > v_H - c
\]

since \( v_L < c \). Thus, by continuity, there exists \( \hat{q} \in (0, 1) \) such that the LHS of (3.7) is equal to \( (v_H - c) \), as desired.
To consider the option contract, let us now be more specific about the bargaining protocol. After $\theta$ has been realized, the parties can simultaneously send to each other letters containing trading offers, which the receiving party can later present as evidence to the court. We make the simplifying assumption that there are only two trading possibilities, trade ($q = 1$) or no trade ($q = 0$).

- Option contracts are such that ‘$S$ receives $P_0$ in case of no trade, and has the option to trade at $P_1 = P_0 + K$, where $P_0$ and $K$ are set to satisfy $K > c$ and $v_H > P_1 = P_0 + K$’.

  - Note first that according to the contract, $S$ would always want to trade (if there is no renegotiation) since ‘trade’ will yield $P_1 - c$ to $S$ while ‘no trade’ $P_0 < P_1 - c$.

  - Consider first the case $\theta = (v_L, c)$. In this case, it is socially efficient to trade, which will indeed take place since (i) $S$ prefers ‘trade’ over ‘no trade’ and (ii) $B$ cannot make any renegotiation offer that induces $S$ to choose ‘no trade’ (that is makes $S$ better with ‘no trade’ than with ‘trade at $P_1$’) and at the same time makes himself better off.

  - Even in the case $\theta = (v_L, c)$, $S$ would like to trade for the above reason. Then, $B$ will send a letter offering $P_0' = P_1 - c(\epsilon)$ for ‘no trade’ so that his payoff can be $-P_0' = -P_1 + c$ instead $v_L - P_1 (< -P_1 + c)$. Also, $B$ can hide whatever offers has been sent by $S$.

  - Thus, $B$’s ex ante payoff is

$$j[v_H - P_1] + (1 - j)[-P_0'] - \psi(j) = jv_H + (1 - j)c - P_1 - \psi(j),$$

which leads to the first-order condition

$$v_H - c = \psi'(j),$$

resulting in $j = j^*$.
3.4 Cooperative Investment and Failure of Contracts

It has been assumed so far that investments are selfish in that one party’s investment has no direct impact on the other’s payoff. Often, however, investments are cooperative. For example, the Fisher Body’s decision to build a plant near GM may benefit both parties by reducing shipping costs and so on. This externality turns out to have an important consequence on the efficiency properties of contracts.

• For a simple model of cooperative investment, assume that only $S$ invests $i$, which only benefits $B$ so that $\text{Prob}(v_H) = i$ while $S$’s cost is fixed at $c \in (v_L, v_H)$.

  - The first-best level of investment is defined as before to satisfy
    \[ v_H - c = \phi'(i^*) \].
  
  - Without any contract, $S$’s ex ante payoff would be
    \[ i \left[ \frac{1}{2} (v_H - c) \right] - \phi'(i) \]
    so $S$ chooses $\hat{i}$ satisfying
    \[ \frac{1}{2} (v_H - c) - \phi(\hat{i}) = 0. \] \hspace{1cm} (3.8)

• Now consider the specific-performance contract $(\hat{P}, \hat{q})$. Once $\theta$ is realized, the renegotiation proceeds as follows:

  - If $\theta = (v_H, c)$ realizes, then $S$’s renegotiated payoff is
    \[ \hat{P} - c\hat{q} + \frac{1}{2} (1 - \hat{q}) (v_H - c). \]
If $\theta = (v_L, c)$ realizes, then $S$’s renegotiated payoff is

$$\hat{P} - c\hat{q} + \frac{1}{2}\hat{q}(c - v_L).$$

Thus, $S$’s ex ante payoff is given as

$$i[\hat{P} - c\hat{q} + \frac{1}{2}(1 - \hat{q})(v_H - c)] + (1 - i)[\hat{P} - c\hat{q} + \frac{1}{2}\hat{q}(c - v_L)] - \phi(i),$$

whose differentiation with $i$ yields

$$\frac{1}{2}[(1 - \hat{q})v_H + \hat{q}v_L - c] - \phi'(i) \leq \frac{1}{2}(v_H - c) - \phi'(i).$$

Comparing (3.8) and (3.9) shows that $S$’s investment is no higher with contracts than without.
Chapter 4

Auction Theory

• Why do we study auctions?

  – Practical significance; commonly used in many areas
  – Good welfare and revenue features
  – Serves as a “behavioral” foundation of markets
  – Empirical testability

• Different Environments

  – Identity of the auctioneer: Seller auction, buyer auction, or double auction
  – Number of objects: Single Unit or Multiunit
  – Information: Bidders have signals which convey information about the value of auctioned object

  (1) Private vs interdependent values (c.f., Mineral rights auction)
(2) Independent vs correlated signals

(3) Single-dimensional vs multidimensional signals

- Different Auction Rules
  - Single Unit: First-price, Second-price (Vickrey), Open-oral ascending, Dutch, All-pay, and many others
  - Multi Unit:
    1. Discriminatory price auction (pay your bid)
    2. Uniform price auction (pay highest losing bid)

- Our study focuses on the single-unit IPV model, whose basic setup is given as follows:
  - 1 unit on sale
  - $n \geq 2$ risk-neutral bidders. Let $N = \{1, \ldots, n\}$ denote the set of all bidders.
  - Bidder $i$’s value: $\theta_i \in \Theta_i := [\overline{\theta}_i, \underline{\theta}_i]$ drawn according to distribution $F_i(\cdot)$ with density $f_i(\cdot)$
  - Each bidder $i$ knows his value $\theta_i$ while others only know its distribution $F_i(\cdot)$.
  - Bidder $i$’s utility: $\theta_i x_i - t_i$, where $x_i$ is the probability of bidder $i$ winning the object and $t_i$ the bidder $i$’s payment. Each bidder’s reservation utility is zero.
  - Seller’s utility: $\sum_{i=1}^{n} t_i$

- Two criteria to evaluate an auction outcome
  - Efficiency: Whether the auctioned objects go to the bidders who value them most highly.
– Revenue: How much revenue the seller raise in expectation.
4.1 Second-Price (or English) Auction

The auction rule is as follows: The bidder who submits the highest bidder wins the object and pays the second highest bid. If there is a tie, then the winner is randomly determined among the highest bidders.

• It is a (weakly) dominant strategy for each bidder $i$ to submit a bid equal to his value $\theta_i$.

Proof. Consider bidder 1, say, and suppose that $p_1 := \max_{j \neq i} b_j$ is the highest competing bid. By bidding $b_1$, bidder 1 will win if $b_1 > p_1$ and not win if $b_1 < p_1$. It is straightforward that bidding $\theta_1$ is always (weakly) better than bidding some $b_1 < \theta_1$: If $p_1 \leq b_1 < \theta_1$ or $p_1 < b_1 < \theta_1$, then bidder 1 is indifferent between $\theta_1$ and $b_1$. If $b_1 \leq p_1 \leq \theta_1$, then $\theta_1$ is strictly better than $b_1$. A similar argument shows that bidding $\theta_1$ is always better than bidding some $b_1 > \theta_1$. Thus, it is weakly dominant to bid $\theta_1$.

− By the same logic, it is weakly dominant for a bidder in English auction to drop out when the price reaches his value.

− As a result, the object is allocated to the highest value bidder, who pays the second highest value. So, the equilibrium allocation is efficient.

− The seller’s revenue is

$$R_{SPA} := \mathbb{E}[\theta^{(2)}],$$

where $\theta^{(2)}$ is the second order statistic.

− Assuming that bidder are symmetric, that is $F_i(\cdot) = F(\cdot)$ for some $F : [\underline{\theta} ; \overline{\theta}] \to [0, 1]$, each bidder $i$ of type $\theta_i$ pays in expectation

$$\int_{\underline{\theta}}^{\theta_i} s(n - 1) f(s) F^{n-2}(s) ds, \quad (4.1)$$

67
which will prove useful for comparing the revenues between SPA and FPA.

- It doesn’t matter whether bidders know the others’ values or not since they play the weakly dominant strategy.

- There are other equilibria which are less reasonable: Bidder 1, say, always bids $\infty$ while others bid 0 → These equilibria might be utilized by the collusive bidders.

4.2 First-Price Auction

The rule is the same as in the second-price auction except that the winner pays his own bid. It is very difficult to fully characterize the equilibrium bidding strategy if bidders are asymmetric, that is $F_i(\cdot) \neq F_j(\cdot)$ for $i \neq j$. So, we assume that bidders are symmetric. We focus on (Bayesian) Nash equilibrium in which bidders use the symmetric and increasing bidding strategy, $\beta : [\theta, \bar{\theta}] \rightarrow \mathbb{R}_+$.\(^1\)

- In equilibrium, the optimal bid of bidder $i$ with value $\theta_i$ is

$$
\beta(\theta_i) = \arg \max_{b \in \mathbb{R}_+} (\theta_i - b) F_n^{-1}(\beta^{-1}(b)).
$$

This problem can be translated into

$$
\theta_i = \arg \max_{\theta \in [\theta, \bar{\theta}]} (\theta_i - \beta(\hat{\theta})) F_n^{-1}(\hat{\theta}), \quad (4.2)
$$

which must be maximized at $\hat{\theta} = \theta_i$.

- To differentiate (4.2) with $\hat{\theta}$ and set it equal to zero at $\hat{\theta} = \theta_i$ yields

$$
- \beta'(\theta_i) F_n^{-1}(\theta_i) + (\theta_i - \beta(\theta_i))(F_n^{-1}(\theta_i))' = 0,
$$

\(^1\)It can be shown that the equilibrium bidding strategy we will discuss is indeed unique.
which can be rearranged to yield

\[(\beta(\theta_i)F^{n-1}(\theta_i))' = \theta_i(F^{n-1}(\theta_i))',\]

and then

\[\beta(\theta_i)F^{n-1}(\theta_i) = \int_{\theta}^{\theta_i} s(F^{n-1}(s))'ds + K.\]

Clearly, \(K = 0\). Rearranging once more, we obtain

\[\beta(\theta_i) = \frac{\int_{\theta}^{\theta_i} s(F^{n-1}(s))'ds}{F^{n-1}(\theta_i)} = \frac{\theta_i F^{n-1}(\theta_i) - \int_{\theta}^{\theta_i} F^{n-1}(s)ds}{F^{n-1}(\theta_i)} = \theta_i - \frac{\int_{\theta}^{\theta_i} F^{n-1}(s)ds}{F^{n-1}(\theta_i)}, \tag{4.4}\]

where the second equality follows from the integration by parts.

- The equilibrium bidding function in (4.4) has resulted from only considering the first-order condition. To check that \(\beta(\theta)\) is indeed a global optimum (or second-order condition) for each type \(\theta_i \in [\theta, \bar{\theta}]\), substitute (4.4) into (4.2) to obtain

\[\theta_i \hat{\theta} F^{n-1}(\hat{\theta}) + \int_{\theta}^{\hat{\theta}} F^{n-1}(s)ds.\]

Differentiate this with \(\hat{\theta}\) to verify

\[(\theta_i - \hat{\theta})(F^{n-1}(\hat{\theta}))' \geq 0 \quad \text{if} \quad \hat{\theta} \leq \theta_i,\]

implying that \(\hat{\theta} = \theta_i\) achieves the global maximum.

- Applying the formula in (4.4) to the case of uniform distribution \(F(\theta_i) = \theta_i\) on \([0, 1]\) yields

\[\beta(\theta_i) = \theta_i - \frac{\int_{0}^{\theta_i} s^{n-1}ds}{\theta_i^{n-1}} = \frac{n - 1}{n} \theta_i.\]

- Note that the equilibrium allocation is efficient as in the second-price auction.
– The seller’s revenue is equal to

\[ R_{FPA} := \int_{\theta}^{\bar{\theta}} \beta(\theta)(F^n(\theta))'d\theta. \]

In case of the uniform distribution,

\[ R_{FPA} = \int_{\theta}^{\bar{\theta}} \frac{n-1}{n} \theta(\theta^n)'d\theta = \frac{n-1}{n+1}. \]

– The seller’s revenue is the same across SPA and FPA since in FPA, the amount each bidder \( i \) of type \( \theta_i \) pays in expectation is given as, using the second expression in (4.4),

\[ F^{n-1}(\theta_i)\beta(\theta_i) = \int_{\theta}^{\theta_i} s(F^{n-1}(s))'ds, \]

which is equal to (4.1).
4.3 Revenue Equivalence and Optimal Auction

Here, we ask two questions: (1) Can we generalize the observation that the first-price and second-price auctions generate the same revenue? (2) Does there exist the auction mechanism that maximizes the seller’s revenue among all possible auction mechanisms? It turns out that the answer to both questions are positive. To do so, we allow for bidders to be asymmetric.

• Let us consider an auction in its most general format: Given a type realization \( \theta = (\theta_1, \cdots, \theta_n) \), \( x_i(\theta) \) is the probability bidder \( i \) wins the object and \( t_i(\theta) \) is his payment.

  - For instance, at the equilibrium of the first-price auction with the uniform distribution, \((x_i(\theta), t_i(\theta)) = \begin{cases} (1, \frac{n-1}{n}\theta_i) & \text{if } \theta_i > \max_{j \neq i} \theta_j \\ (0, 0) & \text{otherwise.} \end{cases} \)

  - Call \( X_i(\theta_i) := \mathbb{E}_{\theta_{-i}}[x_i(\theta_i, \theta_{-i})] \) and \( T_i(\theta_i) := \mathbb{E}_{\theta_{-i}}[t_i(\theta_i, \theta_{-i})] \) interim winning probability and interim payment, respectively.

  - If bidder \( i \) of type \( \theta_i \) pretends to be of type \( \theta_i' \) (or uses the equilibrium strategy of type \( \theta_i' \)), then his expected payoff is given by \( u_i(\theta_i', \theta_i) := \theta_i X_i(\theta_i') - T_i(\theta_i') \). Thus, his equilibrium payoff is given by \( U_i(\theta_i) := u_i(\theta_i, \theta_i) \).

• There are two constraints for the above auction mechanism to work well:

\[
U_i(\theta_i) \geq u_i(\theta_i', \theta_i) \quad \text{for all } \theta_i \in [\underline{\theta}_i, \overline{\theta}_i] \text{ and } \theta_i' \in [\underline{\theta}_i, \overline{\theta}_i] \\
U_i(\theta_i) \geq 0 \quad \text{for all } \theta_i \in [\underline{\theta}_i, \overline{\theta}_i].
\]

  - The following result helps us express the above constraints in a simple form, which will be instrumental to the ensuing analysis.
Theorem 4.1. The constraint (IC) holds if and only if for all $i$,

$$X_i(\cdot) \text{ is non-decreasing} \quad (M)$$

$$U_i(\theta_i) = U_i(\bar{\theta}_i) + \int_{\bar{\theta}_i}^{\theta_i} X_i(s) ds \text{ for all } \theta_i. \quad (Env)$$

Proof. ($\rightarrow$) To first show that (IC) implies (M), observe that

$$\theta_i X_i(\theta_i) - T_i(\theta_i) = U_i(\theta_i) \geq u_i(\theta_i, \theta_i) = \theta_i X_i(\theta_i) - T_i(\theta_i')$$

and

$$\theta_i' X_i(\theta_i') - T_i(\theta_i') = U_i(\theta_i') \geq u_i(\theta_i, \theta_i') = \theta_i' X_i(\theta_i) - T_i(\theta_i),$$

which yields

$$\theta_i (X_i(\theta_i') - X_i(\theta_i)) \leq T_i(\theta_i') - T_i(\theta_i) \leq \theta_i'(X_i(\theta_i') - X_i(\theta_i)),$$

or

$$(\theta_i' - \theta_i)(X_i(\theta_i') - X_i(\theta_i)) \geq 0,$$

implying (M). To next show that (IC) implies (Env), observe that

$$\theta_i = \arg \max_{\theta_i' \in [\theta_i, \bar{\theta}_i]} u_i(\theta_i', \theta_i) \quad \text{and} \quad U_i(\theta_i) = \max_{\theta_i' \in [\theta_i, \bar{\theta}_i]} u_i(\theta_i', \theta_i). \quad (4.5)$$

Thus, applying the envelope theorem to (4.5), we obtain

$$\frac{dU_i(\theta_i)}{d\theta_i} = \frac{\partial u_i(\theta_i', \theta_i)}{\partial \theta_i} \bigg|_{\theta_i' = \theta_i} = X_i(\theta_i),$$

which implies (Env).

($\leftarrow$) Do this for yourself.  

- The seller’s revenue is given by

$$\max_{(x_i(\cdot), t_i(\cdot))_{i=1}^n} \mathbb{E} \left[ \sum_{i=1}^{n} t_i(\theta) \right] = \sum_{i=1}^{n} \mathbb{E} [T_i(\theta_i)]$$
Using \((Env)\), we can make the following substitution

\[
T_i(\theta_i) = \theta_i X_i(\theta_i) - \int_{\theta_i}^{\theta_i} X_i(s)ds - U_i(\theta_i).
\]  

(4.6)

Then,

\[
\sum_{i=1}^{n} \mathbb{E}[T_i(\theta_i)] = \sum_{i=1}^{n} \int_{\theta_i}^{\theta_i} \left( \theta_i X_i(\theta_i) - \int_{\theta_i}^{\theta_i} X_i(s)ds - U_i(\theta_i) \right) f_i(\theta_i) d\theta_i,
\]

which results in the following theorem:

**Theorem 4.2 (Revenue Equivalence).** If two auction mechanisms generate the same allocation and the same utility to the lowest type of each bidder, then they yield the same expected revenue to the seller.

Let us now search for the optimal auction which maximizes the seller’s revenue.

- The seller’s problem is to solve

\[
\max_{\{x_i(\cdot), t_i(\cdot)\}_{i=1}^{n}} \mathbb{E} \left[ \sum_{i=1}^{n} t_i(\theta) \right] = \sum_{i=1}^{n} \mathbb{E}[T_i(\theta_i)]
\]

subject to (IC) and (IR), or equivalently (M), \((Env)\), and (IR).

- From (4.6), we have

\[
\mathbb{E}[T_i(\theta_i)] = \int_{\theta_i}^{\theta_i} \left( \theta_i X_i(\theta_i) - \int_{\theta_i}^{\theta_i} X_i(s)ds - U_i(\theta_i) \right) f_i(\theta_i) d\theta_i
\]

\[
= \int_{\theta_i}^{\theta_i} \left( \theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} \right) X_i(\theta_i) f_i(\theta_i) d\theta_i - U_i(\theta_i)
\]

\[
= \mathbb{E} \left[ \left( \theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} \right) x_i(\theta) \right] - U_i(\theta_i)
\]

- The seller’s revenue can be expressed as

\[
\mathbb{E} \left[ \sum_{i=1}^{n} t_i(\theta) \right] = \mathbb{E} \left[ \sum_{i=1}^{n} J_i(\theta_i) x_i(\theta) \right] - \sum_{i=1}^{n} U_i(\theta_i),
\]

73
where
\[ J_i(\theta_i) = \theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)}, \]
called “virtual valuation.” We assume that \( J_i(\cdot) \) is non-decreasing, as is satisfied by most distribution functions.

- Let us temporarily ignore \((M)\). Then, the optimal auction mechanism requires that (i) the \((IR)\) constraint for the lowest type be binding or \( U_i(\theta_i) = 0 \) and (ii) the allocation rule be given as

\[
x_i^*(\theta) = \begin{cases} 
1 & \text{if } J_i(\theta_i) > \max\{\max_{j \neq i} J_j(\theta_j), 0\}, \\
0 & \text{otherwise}.
\end{cases}
\]

Note that this allocation rule satisfies \((M)\) since \( J_i(\cdot) \) is nondecreasing.

- Let us assume that bidders are symmetric or \( F_i(\cdot) = F(\cdot) \), which implies that \( J_i(\cdot) = J(\cdot) \). Then,

\[
x_i^*(\theta) = \begin{cases} 
1 & \text{if } \theta_i > \max_{j \neq i} \theta_j \text{ and } \theta_i > \hat{\theta}, \\
0 & \text{otherwise},
\end{cases}
\]

where \( \hat{\theta} := J^{-1}(0) \).

- In case of uniform distribution, \( J(\theta) = 2\theta - 1 \) and thus \( \hat{\theta} = 1/2 \). In the optimal auction, the object is sold to the highest value bidder only if his value is greater than \( 1/2 \). The optimal revenue is \( \frac{n-1}{n+1} + \frac{1}{n+1} \left( \frac{1}{2} \right)^n \).

- How can we practically implement the optimal auction mechanism above? \( \rightarrow \) Use the first-price or the second-price auction with a reserve price set at \( \hat{\theta} \). For instance, in the first-price auction with a reserve price \( \hat{\theta} \), the equilibrium bidding strategy is given as

\[
\beta(\theta_i) = \begin{cases} 
\theta_i - \frac{\int_{\hat{\theta}}^{\theta_i} F^{n-1}(s)ds}{F^{n-1}(\theta_i)} & \text{if } \theta_i \geq \hat{\theta} \\
0 & \text{otherwise}.
\end{cases}
\]
What will the optimal auction look like if bidders are asymmetric? Consider the following example with two bidders: \( F_1(\theta) = \theta \) and \( F_2(\theta) = \theta^2 \), so bidder 2 is stronger (why?). Note that

\[
\frac{1 - F_1(\theta)}{f_1(\theta)} = 1 - \theta \quad \leq \quad \frac{1 - \theta^2}{2\theta} = \frac{1 - F_2(\theta)}{f_2(\theta)}.
\]

Thus, the optimal auction favors the weaker bidder, or handicaps the stronger bidder.

The idea behind the optimal auction is not much different from that of the monopoly pricing. Consider a monopolist who faces \( n \) markets and assume that the market \( i \) is populated by consumers whose value distribution follows \( F_i(\cdot) \). So, the demand function and the total revenue in the market \( i \) are \( q = 1 - F_i(p) \) and \( qF_i^{-1}(1 - q) \), respectively. The marginal revenue is thus

\[
\frac{d}{dq}(qF_i^{-1}(1 - q)) = F_i^{-1}(1 - q) - q\frac{1}{f_i(F_i^{-1}(1 - q))} = p - \frac{1 - F_i(p)}{f_i(p)}.
\]

Thus, the virtual valuation of a bidder can be interpreted as a marginal revenue. The allocation rule of the optimal auction is to sell the good to the buyer associated with the highest marginal revenue as long as it is nonnegative.

### 4.4 Extensions

In this section, we relax three important assumptions we have adopted so far in order to make the model more realistic: (1) risk-neutrality of bidders, (2) bidders with unlimited budget, and (3) independent value distributions. We will see that relaxing the above assumptions leads us to quite different conclusion than before. It turns out that if (1) or (2) is relaxed, then two auctions are no longer equivalent with the seller favoring the
first-price auction. If (3) is relaxed, then the seller may be able to extract the entire surplus from bidders.

### 4.4.1 Risk Averse Bidders

So far, we have assumed that bidders are risk neutral in the sense that each bidders’ utilities are linear in net surplus (= value minus payment). We ask here what happens to the seller’s revenue if bidders are risk averse so that each bidder’s utility is represented by a strictly concave function $u : \mathbb{R}_+ \to \mathbb{R}$ satisfying $u(0) = 0$, $u' > 0$, and $u'' < 0$. Then, if bidder with value $\theta$ wins and pays $b$, his utility is given as $u(\theta - b)$. We maintain the assumption that bidders’ values are identically and independently distributed in the interval $[0, 1]$.

**Proposition 4.1.** With risk averse bidders, the expected revenue is higher in the first-price auction than in the second-price auction.

**Proof.** In the second-price auction, it is still a dominant strategy for each bidder to bid his or her value.

Let us examine the equilibrium strategy for the first-price auction. Letting $\gamma : [0, 1] \to \mathbb{R}_+$ denote the (symmetric) equilibrium bidding strategy, it must be that $\gamma(0) = 0$ (why?). Also, each bidder $i$ of type $\theta_i > 0$ must solve

$$\max_{\theta'_i} G(\theta'_i) u(\theta_i - \gamma(\theta'_i)),$$

where $G(\theta'_i) = F_{n-1}(\theta'_i)$. The first-order condition for this problem is

$$g(\theta_i) u(\theta_i - \gamma(\theta'_i)) - G(\theta'_i) \gamma'(\theta_i) u'(\theta_i - \gamma(\theta'_i)) = 0.$$
Setting $\theta'_i = \theta_i$ and rearranging yield
\[
\gamma'(\theta_i) = \frac{u(\theta_i - \gamma(\theta_i))}{u'(\theta_i - \gamma(\theta_i))} g(\theta_i) G(\theta_i).
\] (4.7)

If bidders were risk-neutral or $u(x) = x$, then (4.7) would become
\[
\beta'(\theta_i) = (\theta_i - \beta(\theta_i)) \frac{g(\theta_i)}{G(\theta_i)},
\]
which is the same expression as (4.3) after a rearrangement. Since $u$ being strictly concave with $u(0) = 0$ implies $u(x)/x > u'(x)$ or $u(x)/u'(x) > x$, we obtain the following: Whenever $\gamma(\theta_i) = \beta(\theta_i)$,
\[
\gamma'(\theta_i) = \frac{u(\theta_i - \gamma(\theta_i))}{u'(\theta_i - \gamma(\theta_i))} g(\theta_i) G(\theta_i) > (\theta_i - \gamma(\theta_i)) \frac{g(\theta_i)}{G(\theta_i)} = (\theta_i - \beta(\theta_i)) \frac{g(\theta_i)}{G(\theta_i)} = \beta'(\theta_i).
\]

Now, we make use of the following lemma:

**Lemma 4.1.** Suppose $f : \mathbb{R} \to \mathbb{R}$ is a continuously differentiable such that $f(x) \geq 0$. Suppose that for all $x \geq \bar{x}$, $f'(x) > 0$ whenever $f(x) = 0$. Then, $f(x) > 0$ for all $x > \bar{x}$.

Defining $f(\theta_i) := \gamma(\theta_i) - \beta(\theta_i)$, the function $f(\cdot)$ satisfies the condition in Lemma 4.1 so $f(\theta_i) = \gamma(\theta_i) - \beta(\theta_i) > 0$ for all $\theta_i \in (0, 1]$. To conclude, in the first-price auction, the risk aversion causes an increase in equilibrium bids while it does not in the second-price auction.

Why does risk aversion lead to higher bid in the first-price auction? Raising one’s bid slightly in the first-price auction is analogous to buying partial insurance: it reduces the probability of a zero payoff and increases the chance of winning at a lower profit margin. Risk averse bidders value this insurance, so they bid more than they would if they were risk neutral.
4.4.2 Budget Constrained Bidders

Let us relax the assumption that bidders are not constrained in budgeting their bids. An alternative assumption, among others, would be that bidders face a fixed budget equal to \( B \), so they cannot make any bid higher than \( B \). Letting \( \beta() \) denote the equilibrium strategy in the first-price auction with no budget constraint, we assume \( B < \beta(1) \) to avoid a trivial case.

- In the second-price auction, each player has a dominant strategy:

\[
\beta_S(\theta) = \min\{\theta, B\}.
\]

- The proof follows the same line as when there is no budget constraint. (Try for yourself.)

- In the first-price auction, a candidate equilibrium is the following:

\[
\beta_F(\theta) := \begin{cases} 
\beta(\theta) & \text{if } \theta \leq \hat{\theta} \\
B & \text{if } \theta > \hat{\theta},
\end{cases}
\]

where \( \hat{\theta} \) satisfies

\[
(\hat{\theta} - \beta(\hat{\theta}))F^{n-1}(\hat{\theta}) = (\hat{\theta} - B)\sum_{k=0}^{n-1} \frac{1}{k+1} \binom{n-1}{k} (1 - F(\hat{\theta}))^k F^{n-k-1}(\hat{\theta}).
\]

- Here, \( \theta > \hat{\theta} \) bids \( B \) while \( \theta < \hat{\theta} \) bids \( \beta(\theta) \) as if the budget were absent. And \( \hat{\theta} \) is indifferent between \( B \) and \( \beta(\hat{\theta}) \).

- Note that \( B > \beta(\hat{\theta}) \) so the equilibrium strategy jumps at \( \hat{\theta} \).

- Given all other bidders employ \( \beta_F() \), it is optimal for bidder \( i \) of type \( \theta_i \) to bid \( \beta_F(\theta_i) \). (Why?)
We have the following proposition.

**Proposition 4.2.** *With the same budget $B$, the expected revenue is higher in the first-price auction than in the second-price auction.*

*Proof.* The equilibrium allocation in the first-price auction is the same as the second-price auction with a larger budget equal to $\hat{\theta}$. Thus, by the revenue equivalence theorem, the first-price auction with budget $B$ and the second-price auction with budget $\hat{\theta} > B$ generate the same revenue, which is higher than the revenue generated by the second-price auction with budget $B$. 

A rough intuition behind this result is as follows: Without budget constraint, any given type of bidder makes the same interim (or average) payment in two auction formats but the payment in the SPA is random and thus sometimes exceeds the average. This makes the budget constraint more likely to bind in the SPA, which translates into a higher revenue of the FPA.

### 4.4.3 Optimal Auction with Correlated Values

We now relax the assumption that values are independently distributed. Relaxing this assumption leads to a striking result in terms of the design of optimal auction: The seller can achieve the first-best allocation and extract all the surplus from the bidders. Here, this result is shown in a simple $2 \times 2$ model, though it is much more general.

- Assume that there are 2 bidders and 2 possible valuations, $\theta_H$ and $\theta_L < \theta_H$, for each bidder.
Let $p_{ij}$ denote the probability that bidder 1 has $\theta_i$ and bidder 2 has $\theta_j$.

Suppose that valuations are correlated: $p_{LL}p_{HH} - p_{LH}p_{HL} \neq 0$: Letting $\psi := \frac{p_{HH}p_{LL}}{p_{HL}p_{LH}}$, if $\psi > 1$ ($< 1$), then values are positively (negatively) correlated.

Consider the general auction rule: $x_{ij} =$ the probability that bidder 1 gets the object and $t_{ij} =$ bidder 1’s payment, when his value is $i$ and bidder 2’s value is $j$.

The full extraction means that (i) the allocation must be efficient

$$x_{HH} = x_{LL} = \frac{1}{2}, \ x_{HL} = 1, \text{ and } x_{LH} = 0.$$  

(ii) the individual rationality for both types must be bidding:

$$p_{LL} \left( \frac{\theta_L}{2} - t_{LL} \right) - p_{LH}t_{LH} = 0 $$

$$p_{HH} \left( \frac{\theta_H}{2} - t_{HH} \right) + p_{HL}(\theta_H - t_{HL}) = 0 ,$$

from which we have

$$t_{LH} = \frac{p_{LL}}{p_{LH}} \left( \frac{\theta_L}{2} - t_{LL} \right)$$ (4.8)

$$t_{HL} = \theta_H + \frac{p_{HH}}{p_{HL}} \left( \frac{\theta_H}{2} - t_{HH} \right) .$$ (4.9)

Also, we need to satisfy the incentive compatibility condition:

$$0 = p_{LL} \left( \frac{\theta_L}{2} - t_{LL} \right) - p_{LH}t_{LH} \geq p_{LL}(\theta_L - t_{HL}) + p_{LH} \left( \frac{\theta_L}{2} - t_{HH} \right)$$ (IC$_L$)

$$0 = p_{HH} \left( \frac{\theta_H}{2} - t_{HH} \right) + p_{HL}(\theta_H - t_{HL}) \geq -p_{HH}t_{LH} + p_{HL}\left( \frac{\theta_H}{2} - t_{LL} \right).$$ (IC$_H$)

Substituting (4.8) and (4.9) into this yields

$$0 \geq \frac{p_{LH}}{2}(\theta_L - \psi\theta_H) - p_{LL}(\theta_H - \theta_L) + p_{LH}(\psi - 1)t_{HH}$$ (4.10)

$$0 \geq \frac{p_{LH}}{2}(\theta_H - \psi\theta_L) + p_{HL}(\psi - 1)t_{LL} .$$ (4.11)
Thus, we are done by choosing $t_{HH}$ and $t_{LL}$ that satisfy (4.10) and (4.11).

- Some remarks are in order.

  - If $\psi = 1$ or values are independent, then it is impossible to satisfy the above inequalities, which is what we already know from the analysis of independent types.

  - As $\psi$ gets larger beyond 1 or values become more positively correlated, (4.10) and (4.11) can be satisfied by setting $t_{HH}$ and $t_{LL}$ lower while setting $t_{HL}$ and $t_{LH}$ relatively higher: This is effective in giving bidders an incentive to tell truthfully when types are highly positively correlated.

  - If $\psi > 1$ but $\psi \simeq 1$, then $t_{HH}$ and $t_{LL}$ have to be very large, which may cause some problem for a budget-constrained bidder.

4.5 Appendix: Envelope Theorem

Let $X$ be the choice set and $[\underline{t}, \overline{t}]$ the parameter set. Consider the following problem:

$$\max_{x \in X} f(x, t).$$

Let $V(t) := \max_{x \in X} f(x, t)$, called value function, and $x(t) := \arg \max_{x \in X} f(x, t)$, called the set of maximizers. Assume for simplicity that $x(t)$ is a singleton (though it need not be). We have the following theorem.

**Theorem 4.3 (Envelope Theorem).** Take $t \in [\underline{t}, \overline{t}]$ and $x \in x(t)$ and suppose that $f_2(x, t)$ exists, where $f_2(x, t) = \frac{\partial f}{\partial t}$. Then, $V'(t) = f_2(x(t), t)$ whenever $V'(t)$ exists.
Example: Consider \( f(x, t) = (3t - 1)x - \frac{1}{2}x^2 \) and \( X = [0, 1] \) and \( t \in R \). Then,

\[
  x(t) = \begin{cases} 
  0 & \text{if } t \leq \frac{1}{3} \\
  3t - 1 & \text{if } \frac{1}{3} < t < \frac{2}{3} \\
  1 & \text{if } t \geq \frac{2}{3} 
  \end{cases} \quad \text{and} \quad V(t) = \begin{cases} 
  0 & \text{if } t \leq \frac{1}{3} \\
  \frac{1}{2}(3t - 1)^2 & \text{if } \frac{1}{3} < t < \frac{2}{3} \\
  3t - \frac{2}{3} & \text{if } t \geq \frac{2}{3} 
  \end{cases}
\]

One can verify

\[
  V'(t) = f_2(x(t), t) = \begin{cases} 
  0 & \text{if } t \leq \frac{1}{3} \\
  3(3t - 1) & \text{if } \frac{1}{3} < t < \frac{2}{3} \\
  3 & \text{if } t \geq \frac{2}{3}, 
  \end{cases}
\]

as desired by Theorem 4.3.

In addition to the assumptions in Theorem 4.3, let us assume that \( V'(t) \) exists almost everywhere.\(^2\) Then, we have the envelope theorem in the integral form:

\[
  V(t) = V(t) + \int_{\frac{1}{3}}^{t} f_2(x(s), s) ds. \quad (4.12)
\]

\(^2\)This is equivalent to assuming that the function \( V(\cdot) \) is *absolute continuous*. 

82
Chapter 5

Matching Theory

In this part, we will study the very basics of matching theory, which is becoming more popular among economists both theoretically and practically. A typical matching problem is described by a situation in which there are agents on two sides of a market who each wish to match with agent(s) on the other side.

- The examples include matchings between

  - Men and women in a ‘marriage market’
  - Firms and workers in a labor market
  - Medical interns and Hospitals
  - Students and colleges/high schools
  - Tenants and houses
  - Donators and patients in organ exchanges
A matching mechanism is a rule that specifies a matching for each preference profile of all involved agents.

- A desirable matching rule is expected to satisfy the following properties:
  - Pareto efficiency: Ex-ante or ex-post
  - Stability: No incentive for agents to break up their matches
  - Strategy-proofness: No incentive for agents to misreport their preferences

A matching rule that satisfies the above properties is hard to come by since they need to be satisfied by many matchings that can arise from all possible preference profiles of agents on two sides. It is an important task of the matching theory to find such a rule.
5.1 Stable Matchings

Our benchmark model is the marriage problem in which there are two finite and disjoint sets $M$ and $W$: $M = \{m_1, m_2, \ldots, m_n\}$ is the set of men, and $W = \{w_1, w_2, \ldots, w_p\}$ is the set of women. The preference of each man $m$ is represented by an ordered list, $P(m)$, on the set $W \cup \{m\}$: For example, $P(m) = w_3, w_1, m, w_2, \ldots, w_p$ indicates that $w_3$ is $m$’s first choice, $w_1$ is his second choice, remaining single is his third choice, and so on. Sometimes, we just write $P(m) = w_3, w_1$, ignoring the part of list that is worse than remaining single. For another example, $P(m') = w_2, [w_1, w_3], m', \ldots, w_k$ indicates that for man $m'$, $w_2$ is his first choice, $w_1$ and $w_3$ are his second choice and indifferent, and so on. We write $w >_m w'$ to mean that $m$ prefers $w$ to $w'$, and $w \geq_m w'$ to mean that $m$ likes $w$ at least as well as $w'$. We say that $w$ is acceptable to $m$ if $m$ likes $w$ at least as well as remaining single, that is $w \geq_m m$, and otherwise $w$ is unacceptable to $m$. The same notions apply to women as well.

- A matching $\mu$ is a one-to-one mapping from the set $M \cup W$ onto itself such that for each $m \in M$ and $w \in W$, $\mu(m) \in W \cup \{m\}$, $\mu(w) \in M \cup \{w\}$, and $\mu(m) = w$ if and only if $\mu(w) = m$.

- We say that the matching $\mu$ is blocked by an individual, say $m$, if $m > \mu(m)$, or blocked by a pair, say $(m, w)$, if $w >_m \mu(m)$ and $m >_w \mu(w)$.

- A matching $\mu$ is individually rational if it is not blocked by any individual.

- A matching is stable if it is not blocked by any individual or any pair of agents.

- What is the justification about requiring matching to be stable? → An unstable matching would not sustain if agents are well informed of one another’s preference and
can easily access each other.

- One can show that any stable matching is Pareto-optimal in the sense that there is no other matching that makes some agents better off without making anyone worse off. (Try to prove this for yourself.)

- Let us consider the following example with three men and three women:

  \begin{align*}
  P(m_1) &= w_2, w_1, w_3 \\
  P(w_1) &= m_1, m_3, m_2 \\
  P(m_2) &= w_1, w_3, w_2 \\
  P(w_2) &= m_3, m_1, m_2 \\
  P(m_3) &= w_1, w_2, w_3 \\
  P(w_3) &= m_1, m_3, m_2.
  \end{align*}

- The matching \( \mu \) given by

  \[
  \mu = \begin{pmatrix}
  w_1 & w_2 & w_3 \\
  m_1 & m_2 & m_3
  \end{pmatrix}
  \]

  is unstable since \((m_1, w_2)\) is a blocking pair.

- The matching given by

  \[
  \mu' = \begin{pmatrix}
  w_1 & w_2 & w_3 \\
  m_1 & m_3 & m_2
  \end{pmatrix}
  \]

  is stable.

- There is another stable matching

  \[
  \mu'' = \begin{pmatrix}
  w_1 & w_2 & w_3 \\
  m_3 & m_1 & m_2
  \end{pmatrix}
  \]

- The above example shows that there can be multiple stable matchings. Some inspection reveals that every man likes \( \mu'' \) as well as \( \mu' \) while every woman likes \( \mu' \) as well as \( \mu'' \).
- A matching is *man-optimal* if it is stable and every man likes it as well as any other stable matching. Also, a matching is *woman-optimal* if it is stable and every woman likes it as well as any other stable matching.

### 5.1.1 Gale-Shapley Algorithm

We ask if there is a matching rule or algorithm that always yields a stable matching whatever preferences agents have.

- Gale and Shapley provided one such algorithm, called *deferred acceptance procedure*.

- The man-proposing deferred acceptance procedure consists of the following steps:

  *Step 1:* Each man proposes to his first choice among the acceptable women on his preference list; Each woman only holds her most preferred among the acceptable men who have proposed to her, rejecting all others;

  In general, at

  *Step k:* Each man who was rejected in the previous step proposes to his next choice among the acceptable women on his list; Each woman only holds her most preferred among the group consisting of the new acceptable proposers and any man she has held from the previous step, rejecting all others;

  *Termination:* The process terminates when a step is reached in which no new man is rejected; After termination, each woman is married to the man she is holding.

- Let $\mu_M$ denote the matching that results from this procedure.

- There is also the woman-proposing deferred acceptance procedure in which the roles of man and woman are just switched.
Consider the following example with 5 men and 4 women:

\[
\begin{align*}
P(m_1) &= w_1, w_2, w_3, w_4 & P(w_1) &= m_2, m_3, m_1, m_4, m_5 \\
P(m_2) &= w_4, w_2, w_3, w_1 & P(w_2) &= m_3, m_1, m_2, m_4, m_5 \\
P(m_3) &= w_4, w_3, w_1, w_2 & P(w_3) &= m_5, m_4, m_1, m_2, m_3 \\
P(m_4) &= w_1, w_4, w_3, w_2 & P(w_4) &= m_1, m_4, m_5, m_2, m_3 \\
P(m_5) &= w_1, w_2, w_4
\end{align*}
\]

**Step 1:** \(m_1, m_4, \text{ and } m_5\) propose to \(w_1\), who only holds \(m_1\); \(m_2\) and \(m_3\) propose to \(w_4\), who only holds \(m_2\). This yields a tentative matching

\[
\begin{align*}
&\text{\hspace{1cm} } w_1 \ w_2 \ w_3 \ w_4 \\
&\hspace{1cm} \\
&\hspace{1cm} \text{\hspace{1cm} } m_1 \ m_2
\end{align*}
\]

**Step 2:** \(m_3, m_4, \text{ and } m_5\) propose to \(w_3, w_4, \text{ and } w_2\), respectively; \(w_4\) rejects \(m_2\) and holds \(m_4\). The tentative matching is thus revised to

\[
\begin{align*}
&\text{\hspace{1cm} } w_1 \ w_2 \ w_3 \ w_4 \\
&\hspace{1cm} \\
&\hspace{1cm} \text{\hspace{1cm} } m_1 \ m_5 \ m_3 \ m_4
\end{align*}
\]

**Step 3:** \(m_2\) proposes to \(w_2\), who rejects \(m_5\) and holds \(m_2\). The tentative matching is thus revised to

\[
\begin{align*}
&\text{\hspace{1cm} } w_1 \ w_2 \ w_3 \ w_4 \\
&\hspace{1cm} \\
&\hspace{1cm} \text{\hspace{1cm} } m_1 \ m_2 \ m_3 \ m_4
\end{align*}
\]

**Step 4:** \(m_5\) proposes to \(w_4\), who rejects \(m_5\) and holds \(m_4\), which stops the procedure and results in

\[
\mu_M = \begin{pmatrix}
& w_1 & w_2 & w_3 & w_4 & (m_5) \\
& m_1 & m_2 & m_3 & m_4 & m_5
\end{pmatrix}
\]

88
In our previous example with three men and three women, the man-proposing algorithm yields \( \mu'' \) while the woman-proposing algorithm yields \( \mu' \).

- A matching arising from the man-proposing (resp. woman-proposing) deferred acceptance procedure is man-optimal (resp. woman-optimal):

**Theorem 5.1.** Suppose that every agent has a strict preference list. Then, the matching \( \mu_M \) is man-optimal.

- **Proof of stability:** Suppose that man \( m \) and woman \( w \) are matched in \( \mu_M \) but \( m \) prefers another woman \( w' \). Then, it must be that at some step of the algorithm, \( m \) proposes to \( w' \) and she rejects \( m \) in favor of a man, say \( m' \), whom she prefers. This, given how the algorithm works, means that \( w' \) is matched in \( \mu_M \) with a man as least as high-ranked as \( m' \) on her list, which implies that \( (m, w') \) is not a blocking pair.

- **Proof of man-optimality:** Let us say that \( w \) is possible for \( m \) if there is some stable match at which these two are matched. We show by induction that there is no round at which a women rejects a man for whom she is possible.

**Induction hypothesis:** Up to some step \( t \), no woman has rejected a man for whom she is possible.

We now show that if at step \( t + 1 \), some woman \( w \) rejects some man \( m \) in favor of some other man \( m' \), then \( w \) is not possible for \( m \). To this end, we consider any matching \( \mu \) such that \( \mu(m) = w \) and show that it is not stable. Suppose to the contrary that \( \mu \) is stable. Then, \( w' := \mu(m') \) is possible for \( m' \) so, by the above hypothesis, \( m' \) has yet to propose to \( w' \) through step \( t \) (why?), which implies that \( m' \) prefers \( w \) to \( w' \). Thus, a pair \( (m', w) \) blocks \( \mu \), a contradiction.
– Symmetrically, the woman-proposing deferred acceptance algorithm yields a stable matching that is woman-optimal.

– One can also show that when all agents have strict preferences, the interests of two sides are opposed on the set of stable matchings: If \( \mu \) and \( \mu' \) are stable, then all men like \( \mu \) at least as well as \( \mu' \) if and only if all women like \( \mu' \) at least as well as \( \mu \).

**Proof.** Let \( \mu \) and \( \mu' \) stable and \( \mu(m) \geq_m \mu'(m) \) for all \( m \in M \). Suppose to the contrary that for some \( w \in W \), \( \mu(w) \succ_w \mu'(w) \). We show that \( \mu' \) cannot be stable. Let \( m = \mu(w) \).

Then, \( m \neq \mu'(w) \), so \( \mu'(m) \neq w \). By strict preference, \( m \) prefers \( w \) to \( \mu'(w) \). Hence, \( (m, w) \) blocks \( \mu' \).

– Moreover, the set of stable matchings has an algebraic structure, called **lattice**.

### 5.1.2 Incentive Problem

In a situation where agents’ preferences are unknown, a stable matching can be implemented by asking each agent to report his/her preference list and then running, for instance, the deferred acceptance procedure according to the reported preference profile. With no measures to check the truthfulness of each agent’s report, however, there arises an incentive problem, i.e. whether each agent is willing to report his/her true preference, given the specified matching algorithm. To investigate this problem, we do not confine ourselves to Gale-Shapley algorithm but consider all possible matching **mechanisms**:

**Definition 5.1.** A matching mechanism is a mechanism in which agents report their (possibly false) preferences and which maps the reported preference profile to a matching outcome.
A matching mechanism defined this way will be without an incentive problem if it is \textit{strategy-proof} in the following sense:

\textbf{Definition 5.2.} A matching mechanism is \textit{strategy-proof} if it is a dominant strategy for each agent to report his/her true preference.

Then, we obtain an impossibility result as follows.

\textbf{Proposition 5.1.} \textit{There does exist no stable matching mechanism that is strategy-proof.}

\textit{Proof.} The proof is done with an example with two men and two women:

\begin{align*}
P(m_1) &= w_1, w_2 & P(w_1) &= m_2, m_1 \\
P(m_2) &= w_2, w_1 & P(w_2) &= m_1, m_2.
\end{align*}

There are only two stable matchings, $\mu$ and $\nu$: $\mu(m_i) = w_i$ for $i = 1, 2$ and $\nu(m_i) = w_j$ for $i = 1, 2$ and $j \neq i$. Consider any stable mechanism that chooses $\mu$. Observe that if $w_2$ reports a preference $Q(w_2) = m_1$ while all others report their true preferences, then $\nu$ is the only stable matching with respect to the reported preference profile $P' = (P(m_1), P(m_2), P(w_1), Q(w_2))$ so any stable mechanism must select $\nu$ when $P'$ is reported. Thus, $w_2$ is better off being matching with $m_1$ after reporting $Q(w_2)$ instead of her true preference. 

Despite the negative result in Proposition 5.1, Gale-Shapley algorithm works well with respect to the incentives of proposers:

- Given the man-proposing algorithm, it is a dominant strategy for each man to report his true preference.
– For a proof, refer to Roth and Sotomayor (1992).

– Combined with this observation, Proposition 5.1 implies that women do have incentive to misrepresent their preference, as was showcased in the proof of Proposition 5.1.

– Let us go back to the previous example with three men and three women. If the man-proposing algorithm is run so that men are reporting truthfully, then women can report false preferences as follows:

\[
Q(w_1) = m_1 \quad Q(w_2) = m_3 \quad Q(w_3) = m_2,
\]

which will result in \( \mu' \), woman-optimal matching. Thus, it is a Nash equilibrium for agents to report \( P' = (P(m_1), P(m_2), P(m_3), Q(w_1), Q(w_2), Q(w_3)) \). (Why?)

### 5.1.3 Extension to Many-to-One Matching

In many matching situations such as college admission problem or internship assignment problem, an agent (college or hospital) on one side is matched to more than one agent (student or intern) on the other side. Our analysis in the previous section can be easily extended to this many-to-one matching problem. A prime example of the many-to-one matching is the college admission problem in which there are a set of colleges, \( C = \{C_1, \ldots, C_n\} \), and a set of students, \( S = \{s_1, \ldots, s_m\} \). Each college \( C \in C \) has a quota of \( q_C \geq 1 \) and has preference over individual students, denoted \( P(C) \) while each student can only attend one school and has preference over \( C \). Each college also has preference over \( 2^S \), i.e. sets of students. We assume that the preferences of colleges over \( 2^S \) are responsive in the following sense: For any college \( C \) and two sets of students \( A \) and \( A' \) that are identical with \( |A| = |A'| \leq q_C \) except that \( s \in A \setminus A' \) and \( s' \in A' \setminus A \) for some \( s, s' \in S \), we have \( A >_C A' \).
if and only if $s >_C s'$.

- A matching is a mapping $\mu$ that satisfies the followings: (i) $\mu(s) \in C \cup \{s\}$; (ii) $\mu(C) \subset S$ with $|\mu(C)| \leq q_C$; (iii) $\mu(s) = C$ if and only if $s \in \mu(C)$.

- As before, a matching $\mu$ is (pairwise) stable if it is not blocked by any individual agent or any college-student pair, that is if there is no pair $(C, s)$ for which $C >_s \mu(s)$ and $s >_C s'$ for some $s' \in \mu(C)$.

- There is a stronger notion of stability, called group stability, that requires a matching to be immune to blocking by a coalition of students and colleges. The group stability, however, is equivalent to the pairwise stability if colleges have responsive preferences.

There is an easy way to translate the college admission model into a related marriage problem. For a college $C$, treated as a men, create $q_C$ “men”, $c_1, c_2, \cdots, c_{q_C}$, each of whom has the same preference over students as $C$. The preference list of each student, treated as a woman, is then modified by replacing each entry $C$, wherever it appear on her list, by the string $c_1, c_2, \cdots, c_{q_C}$.

- Some, but not all, results from the marriage problem generalize to college admission problem.

- A matching of the college admission problem is stable if and only if the corresponding matchings of the related marriage problem are stable.

- When the colleges’ preferences over individual students are strict, the matching resulting from the student-proposing (college-proposing, resp.) mechanism is student-optimal (college-optimal, resp.).
In a matching mechanism that yields the student-optimal stable matchings, it is a dominant strategy for students to report their true preferences.

However, there does not exist any matching mechanism in which it is a dominant strategy for colleges to report their true preferences. This is proved with the following example in which there are four students and three colleges with $q_1 = 2$ and $q_2 = q_3 = 1$:

\[
P(s_1) = C_3, C_1, C_2 \quad P(C_1) = s_1, s_2, s_3, s_4
\]
\[
P(s_2) = C_2, C_1, C_3 \quad P(C_2) = s_1, s_2, s_3, s_4
\]
\[
P(s_3) = C_1, C_3, C_2 \quad P(C_3) = s_3, s_1, s_2, s_4
\]
\[
P(s_4) = C_1, C_2, C_3.
\]

There is a unique stable matching,

\[
\mu = C_1 \quad C_2 \quad C_3
\]
\[
\{s_3, s_4\} \quad \{s_2\} \quad \{s_1\}
\]

Suppose now that $C_1$ reports a false preference $P'(C_1) = s_1, s_4, C_1$ while all other agents report their true preferences, so that the reported preference profile is $P' = (P'(C_1), P(C_2), P(C_3), P(s_1), \ldots, P(s_4))$. Given $P'$, a unique stable matching is

\[
\mu' = C_1 \quad C_2 \quad C_3
\]
\[
\{s_1, s_4\} \quad \{s_2\} \quad \{s_3\}
\]

which is preferred by $C_1$ to $\mu(C_3)$.

The above matching $\mu$ has another problem: There is a (unstable) matching that is preferred by every college,

\[
\mu'' = C_1 \quad C_2 \quad C_3
\]
\[
\{s_2, s_4\} \quad \{s_1\} \quad \{s_3\}
\]
This kind of problem does not arise in the one-to-one matching.

Consider another example in which there are three colleges with \( q_i = 1 \), \( i = 1, 2, 3 \) and three students:

\[
\begin{align*}
P(s_1) &= C_2, C_1, C_3 & P(C_1) &= s_1, s_3, s_2 \\
P(s_2) &= C_1, C_2, C_3 & P(C_2) &= s_2, s_1, s_3 \\
P(s_3) &= C_1, C_2, C_3 & P(C_3) &= s_2, s_1, s_3.
\end{align*}
\]

It is easy to check there is only one stable matching,

\[
\begin{array}{ccc}
C_1 & C_2 & C_3 \\
\{s_1\} & \{s_2\} & \{s_3\}
\end{array}
\]

This matching, however, is Pareto-dominated for the students by a (unstable) matching,

\[
\begin{array}{ccc}
C_1 & C_2 & C_3 \\
\{s_2\} & \{s_1\} & \{s_3\}
\end{array}
\]

The last example shows that if one considers that only students’ preferences are important in evaluating the efficiency of matchings, then the stability may conflict with the Pareto efficiency.

5.2 Efficient Matching

In this section, we study the matching problem in which the preferences of agents on one side are not as important as those of agents on the other side. As shown in the last example of Subsection 5.1.3, the stability and Pareto efficiency may not stand together as far as the preferences of students are concerned. Here, we ask whether by forgoing the stability
requirement, we can find a matching mechanism that always yields the Pareto efficiency for the students while maintaining the strategy-proofness.

### 5.2.1 Top Trading Cycles Algorithm

We adopt the same setup as in Subsection 5.1.3 except that colleges are called (high) schools and colleges’ preferences are treated as schools’ priority rankings. For instance, a priority ranking of a student for a given school can be determined according to: (i) whether (s)he has a sibling in that school; (ii) how close the school is to the student’s residence; and so on. Students in the same priority group are randomly ordered by a lottery. We assume that the prior rankings of schools are publicly known before a matching is implemented while the preferences of students are only known to themselves.

- The following matching mechanism, called *top trading cycles*, yields a Pareto efficient matching from the students’ perspectives.

  - Given the reported preference profile, the top trading cycles mechanism is run in the following steps:

    **Step 1:** Each student points to his/her first choice among the schools; Each school points to the student who has the highest priority for the school; There must be at least one circle (why?), say \((s_1, C_1, s_2, \cdots, s_k, C_k)\), in which \(s_1\) points to \(C_1\), \(C_1\) points to \(s_2\), \(\cdots\), \(s_k\) points to \(C_k\), and \(C_k\) points to \(s_1\). Every student in a circle is assigned a seat in the school (s)he points to and is removed. Also, each school removed as soon as its quota is filled.

      In general, at
Step $k$: Each remaining student points to his/her first choice among the remaining schools; Each remaining school points to the student who has the highest priority for the school; Whatever circle is found, every student in that circle is assigned a seat in the school she points to and is removed. Also, each school is removed as soon as its quota is filled.

Termination: The algorithm terminates when all students are assigned a seat or all schools fill their quotas.

– Consider the last example in subsection 5.1.3 and apply the top trading cycle. In the fist step, there forms a circle $(s_1, C_2, s_2, C_1)$. Once these are removed, in the second step, there forms a circle $(s_3, C_3)$. The resulting matching is

\[
C_1 \quad C_2 \quad C_3
\]

\[
\{s_2\} \quad \{s_1\} \quad \{s_3\}
\]

which is Pareto-efficient for the students.

– The top trade cycles mechanism is Pareto efficient for the students.

– The top trading cycles mechanism is strategy-proof for the students.