Lecture Notes in Information Economics

Juuso Valimaki

February, 2014

Abstract

These lecture notes are written for a first-year Ph.D. course in Microeconomic Theory. They are based on teaching material from courses given on the topic in Finnish Doctoral Programme in Economics over a number of years. It is a special pleasure to thank Paula Mäkelä and Emilia Oljemark for their work as TA's for this course and also for their careful reading of these notes and their suggestions that improved the presentation.

1 Introduction

Microeconomic Theory analyzes individual behavior in various institutional settings. The first step in the analysis asks how a rational economic agent chooses between different options available to her. Consumer Theory, Producer Theory and Choice Under Uncertainty are examples of this type of analysis. More generally, this could be called Decision Theory encompassing e.g. Statistical Decision Theory. The key here is that the institutional setting within which the decision maker operates is left completely abstract: it is only reflected in the set of available choices and preferences over the alternatives.

Competitive Equilibrium Theory is the first institution that is explicitly modeled in typical courses in Microeconomic Theory. A price system is the only coordinating device between various interacting agents. Prices together with parameters of the model such as initial endowments of goods and ownership of productive assets determine the set of available choices for all the economic actors. Prices are assumed independent of individual choices, i.e. all agents are price takers. Furthermore, it is assumed that price is the only variable connecting the decision problems of the various agents. Hence externalities and public goods are ruled out by assumption.

The solution in a decision problem is represented by a map from the primitives of the model (preferences and choice sets) to optimal choices. The solution concept for competitive economies is called competitive equilibrium. A competitive equilibrium is a price and an allocation (consumption and production choices) such that the aggregate excess demand in all markets is zero.

Arrow, Debreu and McKenzie showed in the 1950's that a competitive equilibrium exists and is Pareto efficient under the assumptions of competitive behavior. Debreu, Mas-Colell, Mantel and Sonnenschein gave descriptions of the equilibrium correspondence, i.e. the map from the primitives (endowments, preference, technologies) of the model to equilibrium endogenous variables (prices and allocations).

Game Theory provides a flexible mathematical framework for the analysis of institutions where the agents' decision problems are interdependent. The choices of agent i affect the preferences of other agents in the economy. This may happen through the impact that i's decision has on the price formation (i.e. agents are no longer assumed price takers) or through direct external effects as in the case of congestion.

A Game consists of a set of players $\{1, ..., N\}$, a set of available choices A_i for each player $i \in \{1, ..., N\}$ and a preference relation $\succeq_i$ for each player over the set of choice vectors $a = (a_1, ..., a_N) \in A := \times_{i=1}^{N} A_i$. In the first part of this course, a number of solution concepts for games were given: Dominant Strategy Equilibrium, Rationalizability, Nash Equilibrium, Subgame Perfect Equilibrium, Bayes-Nash Equilibrium depending on the nature of interactions (simultaneous play vs. sequential play, complete information vs. incomplete information etc.).

This part of the course concentrates on simple allocation problems. A group of agents $i \in \{0, 1, ..., N\}$ must decide collectively on an allocation $a \in A$ that affects the payoffs of all players $u_i(a, \theta, y)$ where $\theta \in \Theta$ is a vector of information types of the players (their Harsanyi types in the sense of Bayesian games) at the beginning of the game and $y \in Y$ is a random variable whose distribution depends on (a, θ) and that is realized after a was chosen. I have written the payoffs in a general form to encompass all cases covered in the literature, but we shall concentrate on models that do not include all three variables simultaneously.

The allocation problems are assumed simple enough so that if a, θ and y were all commonly observed, the parties could reach an efficient allocation (in some game form). In the next chapter of these notes, we present the model of Adverse Selection. In most cases, it is useful to write the allocation $a = (x, t)$ where x represents a physical allocation and t is a monetary transfer. In Chapter 2, ignore the role of y, i.e. we assume that $Y = \{y_0\}$ and ask what allocation rules $\phi : \Theta \to X \times T$ are possible if we allow individual agents to use their information to their own best advantage. The idea is simple: since the payoffs of all agents depend only on (x, t, θ) let's allow the agents to communicate their private information through messages and let's decide the allocation based on these messages. The canonical model that you might have in mind here is one where an uninformed seller sells goods to informed agents (and information is interpreted as their valuation for the goods). Chapter 2 contains results on what rules ϕ can be realized in such communication. It covers optimal mechanism in principal-agent settings with applications to price discrimination and Myerson's results on revenue maximizing auctions. It also constructs the Vickrey-Clarke-Groves mechanism for efficient allocation in private values models with quasi-linear models.

Chapter 3 presents the basic model of Moral Hazard where information is symmetric between the contracting parties in the sense that there is no type θ in the model. The informational problem between contracting parties arises from the fact that the actions in the model (corresponding to the physical allocation x earlier) are not directly observable to all parties but affect the payoffs through their influence on y. We let $x = (x_0, x_1, ..., x_N)$ be a vector of action choices by the agents where each i chooses x_i and $t = (t_0, t_1, ..., t_N)$ is a vector of transfers (often wages) to the agents. We assume that y is publicly observable but that the choice of x_i is known by i only. Therefore we write the transfers as $t : Y \to \mathbb{R}^N$, and the payoff functions as $u_i : X \times Y$. We ask what choices of (x, t) are possible in equilibrium where each agent maximizes her own utility when choosing x_i. The canonical model here is that of an employer incentivizing a team of workers through incentive contracts where compensation depends on the observable output $y \in Y$. We cover the principal-agent model of Moral Hazard (i.e. a single agent or worker in the canonical model) where we display the incentive-insurance trade-off in optimal contracts. We cover also the model of Moral Hazard in teams, and we show that efficient effort in teams can be achieved only when a budget-breaker is present.

Chapter 4 covers two-player games where the players move sequentially and the first mover, often called the sender, has private information θ that affects the payoffs of both players in the game. The second mover, called the receiver, observes the sender's move often called the message sent m, but not the sender's private information θ. The receiver then chooses her action x. The two actions together with θ determine the payoffs of both players, i.e. we take the utility functions to be $u_R(\theta, m, x)$ and $u_S(\theta, m, x)$ respectively.. The key question is how the receiver's perceptions about θ change as a function of m. We cover the model of cheap-talk communication where m has no direct payoff implication to the players. We determine limits of possible communication for the case where the two players have conflicting interests. We also cover the model of costly signaling.

Chapter 5 outlines some recent further developments in the literature. We cover the following topics in the last Section: i) Information based trading in (financial) markets, ii) Rational Expectations Equilibrium as the solution concept for competitive economies with private information, iii) Competition in markets with contracts, iv) The role of commitment in Information Economics. The last topic covers simple examples of hold-up, soft budget constraints, and Coasian bargaining.

2 Adverse Selection, Screening and Mechanism Design

A set of agents $i \in \{0, 1, ..., N\}$ must decide upon an allocation $a \in A$. Some of the agents have private information, and we denote the (Harsanyi) type of agent i by $\theta_i \in \Theta_i$. Each player i has a payoff function that depends on the allocation and on the vector of types $\theta = (\theta_0, \theta_1, ..., \theta_N)$.

$$u_i : A \times \Theta \to \mathbb{R}.$$

In these notes, we only consider the quasi-linear case where the allocations consist of two components: a physical allocation x and a vector of monetary transfers $t = (t_0, ..., t_N)$. Furthermore, all agents' payoffs are assumed linear in t. Hence we write

$$u_i(a, t) = v_i(x, \theta) - t_i.$$

In some applications such as optimal income taxation and models of insurance markets, this is not a good assumption. Luckily enough, much of the theory goes through in models without quasi-linear utilities as well.

We say that he model has *private values* if for all i, θ_i and for all type vectors $\theta_{-i}, \theta'_{-i}$, we have

$$v_i(x, \theta_i, \theta_{-i}) = v_i(x, \theta_i, \theta'_{-i}).$$

If this condition is violated, we say that the model has interdependent values. This Chapter deals mostly with the private values case, but at the end, we will see some examples with interdependent values too.

The goal in this chapter is to see what types of allocation rules that assign an allocation to each type vector are reasonable predictions for the model. We start with the simplest setting with only two agents in the game and only one of them with private information.

2.1 Principal-Agent Model

In this first subsection on Adverse Selection, we introduce the main concepts that are used throughout these notes. In particular, we introduce Revelation Principle, Incentive Compatibility and Individual Rationality in the case with only one informed agent. These concepts generalize in a relatively straightforward manner to the case of multiple informed parties The hope in these notes is that by encountering these concepts repeatedly, students will become familiar with them by the end of this course.

There are two players $i \in P, A$, and only one of them, A has private information. We call the player with private information the *agent* and the player without private information the *principal.* We assume that the uninformed player, the principal, has bargaining power in the game. If the agent's private information was known in advance, the principal would simply make a take-it-or-leave-it offer to the agent. In the offer, the principal specifies a number of possible allocations or trades (x, t) from which the agent chooses the best deal for herself. The agent's ranking of the available options depends on her type θ_A. Since only the agent has private information, we will omit the subscript in the type θ_A for the rest of this section.

The principal is committed to her offers. Once the agent makes her choice, the principal can in most cases infer the true type of the agent. At this stage, the principal would like to renege on the initial offer and make a new proposal. The commitment assumption rules this out. Once an offer is on the table, the agent may pick it and keep it.

2.1.1 Example: Price Discrimination

Consider for concreteness a model where x denotes the quality of a product that the principal sells to the agent. The quality is observable to all parties and the different types of agents value quality differentially:

$$v_A(x, \theta) = \theta x.$$

The principal knows that different types of θ exist. If she can tell the true θ of each agent, she can engage in first degree price discrimination and set individual prices for all buyer types to extract their full surplus. If she knows the distribution of the θ's but not the realized value of θ for the current buyer, she must engage in second degree price discrimination or screening where the agents self select themselves into purchasing different qualities of the good as designed by the seller. This is the main focus in this subsection.

The principal produces quality with a cost function that is independent of the tastes of the agent.

$$v_P(x) = -\frac{1}{2}x^2.$$

Assume furthermore that $t_P = -t_A$, i.e. the price paid by the agent is received by the principal. From now on we let t denote the transfer from the agent to the principal. Hence we have:

$$u_A = \theta x - t \text{ and } u_P = t - \frac{1}{2}x^2.$$

Let me stress here one aspect of the transfer t. It denotes the total transfer paid by the agent and not a 'price per unit of x' as is typical in the competitive model. In models of price discrimination, the per unit price is not a constant independent of the level x of provision. As a result, second degree price discrimination is sometimes called non-linear pricing.

Assume that the agent may refuse to trade with the principal. Denote the payoff to the agent after such a refusal by $\underline{U}$. Let's start the analysis by considering what the principal would do if she knew the type of the agent (first degree price discrimination). In that case, the principal would find for each θ a pair $(x(\theta), t(\theta))$ to solve the following problem:

$$\begin{array}{ll} \max\limits_{(x,t)} & t - \frac{1}{2}x^2 \\ \text{subject to} & \theta x - t \geq \underline{U} \end{array} .$$

Since the objective function is strictly increasing in t, one sees immediately that the constraint must hold as an equality, and therefore the optimum solution is:

$$x(\theta) = \theta, t(\theta) = \theta^2 - \underline{U},$$

whenever

$$\frac{1}{2}\theta^2 \geq \underline{U}.$$

Here we have expressed the pairs $(x(\theta), t(\theta))$ that are traded parametrically as functions of θ. It may be more natural to consider the entire set

$$X := \{(x,t)|(x(\theta), t(\theta)) = (x,t) \text{ for some } \theta \in \Theta\}.$$

We may even write elements of this set as $(x, t(x))$ to emphasize that what is being offered to the agents in the market is a menu of qualities together with prices associated with those qualities. In the current example, the prices associated with traded qualities take the form

$$t(x) = x^2 - \underline{U} \text{ for } x \geq \sqrt{2\underline{U}}.$$

Returning to the case where the principal does not know the agent's type θ, consider how the agent will choose from a menu $\{(x, t(x))\} := X$ provided by the principal.

$$\max_{(x,t(x))\in X} \quad \theta x - t(x).$$

Assuming that $t(x)$ is differentiable, we get a necessary condition for interior quality choice from the condition

$$\theta = t'(x).$$

Notice that if the principal offered the menu of qualities and prices that maximize her revenue under complete information, the agent of type θ would choose pair $(x,t) = (\frac{1}{2}\theta, \frac{1}{4}\theta^2 - \underline{U})$. Hence the agent will not choose the pair $(x,t) = \left(\theta, \theta^2 - \underline{U}\right)$ intended for her under complete information. This should come as no surprise given that under complete information, each type of the agent is kept to her outside option level of payoff.

2.1.2 Revelation Principle

In this subsection, we discuss a conceptual simplification in the mechanism design problem. In general, the rules of communication between the principal and the agent could be quite complicated. A key result from late 1970's observes that only a simple form of communication is needed: the principal just asks the agent to announce her type and commits to an allocation conditional on the announced type.

To give the general statement of the Revelation Principle, we consider an arbitrary communication scheme between the principal and the agent. The principal chooses a set of possible messages M for the agent and commits to an allocation rule $\phi(m) = (x(m), t(m))$ for each $m \in M$ depending on the message m sent by the agent. We take the allocation rule to be deterministic in these notes, but everything generalizes to randomized allocations. We call the pair (M, ϕ) a *mechanism* and we let Γ stand for an arbitrary mechanism. A strategy for the agent in mechanism $\Gamma = (M, \phi)$ is denoted by $\sigma_\Gamma : \Theta \to M$, and an optimal strategy is denoted by σ_Γ^*. Hence for σ_Γ^*, we have

$$\sigma_\Gamma^*(\theta) \in \arg\max_m v_A(x(m), \theta) - t(m).$$

We call a mechanism *direct mechanism* if $M = \Theta$, i.e. if the messages are reported types. Finally, we say that a direct mechanism (Θ, ϕ) is *incentive compatible* if for all θ,

$$\theta \in \arg\max_m v_A(x(m), \theta) - t(m).$$

Incentive compatibility is a very important concept for what follows. Anther way of saying that a direct mechanism is incentive compatible is to say that truthful reporting is optimal in that mechanism.

The outcome in mechanism Γ under σ^* is given by the rule associating to each θ the allocation resulting from reporting according to σ^*.We are now ready to state formally the Revelation Principle.

Theorem 1 (Revelation Principle) *For any mechanism $\Gamma = (M, \phi)$ and optimal strategy σ_Γ^* in Γ, there is an incentive compatible direct mechanism $\widehat{\Gamma} = \left(\Theta, \widehat{\phi}\right)$ with the same outcome as in mechanism Γ.*

Proof. Let $\widehat{\phi}(m) := \phi(\sigma^*(m))$. The first claim is that $\left(\Theta, \widehat{\phi}\right)$ is incentive compatible. Once this is shown, the outcomes in the two mechanism are seen to be identical by construction.

Write $\phi(m) = (x(m), t(m))$ and $\widehat{\phi}(m) = (x(\sigma^*(m)), t(\sigma^*(m)))$. Incentive compatibility requires that

$$\theta \in \arg\max_m v_A(x(\sigma^*(m)), \theta) - t(\sigma^*(m)).$$

Suppose incentive compatibility does not hold. Then there is a $\theta \in \Theta$ and a message $m \in \Theta$ such that

$$v_A\left(x\left(\sigma^*\left(m\right)\right), \theta\right) - t\left(\sigma^*\left(m\right)\right) > v_A\left(x\left(\sigma^*\left(\theta\right)\right), \theta\right) - t\left(\sigma^*\left(\theta\right)\right).$$

Since $\sigma^*\left(m\right) \in M$, this inequality contradicts the optimality of σ^* in $\left(M, \phi\right)$.
■

At this stage, the Revelation Principle probably looks like a useless abstract result. The best way to view it is a very useful tool for solving problems in Information Economics. It is much more convenient to analyze only incentive compatible direct mechanisms rather than the set of all possible mechanisms.

For the price discrimination example, an indirect mechanism would have the set of allowed allocations X as the message space and ϕ would be the identity function. From an arbitrary set X, the optimal choice for type θ would the one yielding the maximal utility:

$$\max_{(x,t(x))\in X} \theta x - t(x)$$

Denote the optimal choices by $(x^H, t^H), (x^L, t^L)$ for types θ^H and θ^L respectively. In the corresponding direct revelation mechanism, the agent reports her type. If she is assigned allocation $\sigma^*\left(m\right)$, cognitional on her announcement, then it is clearly optimal for the agent to report truthfully.

For another example, consider a selling mechanism where the seller asks a buyer to submit a bid and trade is executed if and only if the bid exceeds a secret reservation price Z drawn from the distribution $F\left(\cdot\right)$. The buyer is in the role of the agent, and her willingness to pay for the good θ is her type. In the indirect mechanism, the message of the buyer is her (positive) bid $m \in \mathbb{R}_+$. The allocation is binary: the buyer gets the good at price m if $m \geq z$, where z is the realization of the r.v. Z, and the buyer does not get the good and does not pay a price if $m < z$. Denote these allocations (x, t) by $(1, m)$ and $(0, 0)$ respectively, where x denotes the probability of getting the good. The buyer's valuation is given by

$$v_A = \theta x - t.$$

So the buyer chooses m to maximize

$$\Pr\left(Z \leq m\right)\left(\theta - m\right).$$

Let $m^*\left(\theta\right)$ be the optimal message in this indirect mechanism. The outcome function in the direct mechanism is given by the family of lotteries $\begin{cases} \left(1, m^*\left(\theta\right)\right) \text{ with probability } F\left(m^*\left(\theta\right)\right), \\ \left(0,0\right) \text{ with probability } 1 - F\left(m^*\left(\theta\right)\right). \end{cases}$

2.1.3 Taxation Principle

Sometimes people feel uneasy about the use of Revelation Principle because direct revelation games do not seem natural. Have you ever been asked about your type upon entering a store? It seems much more natural to take the allocations and prices associated with them as the basic object of choice. Luckily enough we have the following result by Guesnerie and Laffont :

Theorem 2 *For every direct mechanism $(x(\theta), t(\theta))$, there is an equivalent mechanism where the agent chooses a physical allocation x and its associated transfer $T(x)$ from a set $\{(x, T(x))\}_{x \in X}$ to solve*

$$\max_{x \in X} v(x, \theta) - T(x),$$

and the outcome of $\{(x, T(x))\}_{x \in X}$ coincides with the outcome of $(x(\theta), t(\theta))$.

Proof. Just take $X = \{x(\theta)\}_{\theta \in \Theta}$ and let $T(\widehat{x}) = t(x^{-1}(\widehat{x}))$. This last expression is well-defined even in the case where $x^{-1}(\widehat{x})$ is not a singleton since incentive compatibility implies that the payments of two types receiving the same x must also be equal. ■

By using taxation principle, some arguments can be made more easily. For example with sufficient differentiability, taxation principle implies that

$$v_x(x, \theta) = T'(\theta).$$

2.1.4 Optimal Mechanisms: Two Types

There are two distinct senses in which one can talk about optimality: i) In the sense of maximizing social surplus between all agents. ii) In the sense of generating the maximal expected payoff to the uninformed principal in the model. With a single informed agent, the first sense is trivial regardless of what the set of possible types might be. If we just define $(x(\theta), t(\theta))$ by

$$\begin{aligned} x(\theta) &\in \arg\max_x v_A(x, \theta) + v_P(x), \\ t_A(\theta) &= -v_P(x(\theta)), \\ t_p &= -t_A, \end{aligned}$$

then the mechanism is clearly incentive compatible and efficient. Hence we turn attention to the case where we find a mechanism to maximize the principal's payoff in the case where $\theta \in \{\theta^H, \theta^L\}$ with $\theta^H > \theta^L$. We assume here that the principal receives the transfers that the agent makes, i.e. $t_P = -t_A$. In this way, the mechanism has a balanced budget in the sense that it does not rely on positive monetary flows from the outside. We also denote $t_A = t$.

The plan for the rest of this subsection is as follows: First, we formalize the constraints of incentive compatibility and individual rationality. These constraints delineate the set of allocations that the principal can offer successfully to the agent. In the second step, we optimize the Principal's payoff over all these allocations.

Incentive Compatibility and Monotonicity

With two types, the menu $\{(x^H, t^H), (x^L, t^L)\}$ consists of (at most) two different (x, t) -pairs. Incentive compatibility then reduces simply to

$$\begin{aligned} v_A(x^H, \theta^H) - t^H &\geq v_A(x^L, \theta^H) - t^L, \\ v_A(x^L, \theta^L) - t^L &\geq v_A(x^H, \theta^L) - t^H. \end{aligned}$$

We say that the mechanism is *individually rational* if it is incentive compatible and the equilibrium payoff to each type is at least as high as the outside option payoff $\underline{U}$. We normalize $\underline{U} = 0$. Hence individual rationality is satisfied if

$$\begin{aligned} v_A(x^H, \theta^H) - t^H &\geq 0, \\ v_A(x^L, \theta^L) - t^L &\geq 0. \end{aligned}$$

By summing the incentive compatibility constraints, we get a necessary condition for incentive compatibility:

$$v_A\left(x^H,\theta^H\right)-v_A\left(x^L,\theta^H\right)\geq v_A\left(x^H,\theta^L\right)-v_A\left(x^L,\theta^L\right). \tag{1}$$

We pause here for a moment to introduce a concept that is important in monotone comparative statics in general and for solving adverse selection models, in particular. You may recall from your Mathematical Methods course that the sign of comparative statics in a parameter of the model depends crucially the sign of the cross partial derivatives of the endogenous variable and the parameter. Strictly Increasing Differences is a slight generalization of such conditions on signs of cross partial derivatives.

Definition 1 *A function f has* strictly increasing differences *(or is strictly supermodular) if for all $x'>x$ and $\theta'>\theta$ we have:*

$$f\left(x',\theta'\right)-f\left(x,\theta'\right)>f\left(x',\theta\right)-f\left(x,\theta\right).$$

A twice differentiable function is supermodular if its cross partial derivative is positive; $\frac{\partial^2 f(x,\theta)}{\partial x\partial\theta}>0$ for all (x,θ).

Hence we have as an immediate result that if v_A has strictly increasing differences, then

$$\theta^H>\theta^L\Rightarrow x^H\geq x^L.$$

It is a good exercise to provide a formal proof of this.

Which Constraints Bind (Eliminating Transfers)?

Consider next the expected payoff to the principal from an incentive compatible mechanism. Let p denote the prior probability that the agent is of type θ (and hence $(1-p)$ is the probability of type θ^L). We have

$$Ev_P(x)+t=p\left(v_P\left(x^H\right)+t^H\right)+(1-p)\left(v_P\left(x^L\right)+t^L\right).$$

This objective function is increasing in $\left(t^H,t^L\right)$. This leads to the first observation.

Lemma 1 *Suppose that v_A has strictly increasing differences and that v_A is increasing in θ. Then at any optimal incentive compatible, individually rational mechanism $\left\{\left(x^H,t^H\right),\left(x^L,t^L\right)\right\}$,*

1.
$$v_A\left(x^L,\theta^L\right)-t^L=0.$$

2.
$$v_A\left(x^H,\theta^H\right)-t^H=v_A\left(x^L,\theta^H\right)-t^L.$$

Proof. 1. Suppose to the contrary that $\left\{\left(x^H,t^H\right),\left(x^L,t^L\right)\right\}$ is IC and IR and $v_A\left(x^L,\theta^L\right)-t^L>0$. Then $v_A\left(x^L,\theta^H\right)-t^L>0$ since v_A is increasing in θ. By IC for θ^H, we have that

$$v_A\left(x^H,\theta^H\right)-t^H>0.$$

Hence both IR constraints are slack. But this contradicts optimality of $\left\{\left(x^H,t^H\right),\left(x^L,t^L\right)\right\}$ since the value of the objective function can be raised without violating the constraints by adding a small enough ε to both t^H and t^L.Notice that we have also argued that IR is always satisfied for θ^H if IR holds for θ^L and IC holds for θ^H.

2. If
$$v_A\left(x^H,\theta^H\right)-t^H>v_A\left(x^L,\theta^H\right)-t^L,$$
then IR for θ^H is slack since

$$v_A\left(x^L,\theta^H\right)-t^L\geq v_A\left(x^L,\theta^L\right)-t^L\geq 0.$$

But then the optimality of $\left\{\left(x^H,t^H\right),\left(x^L,t^L\right)\right\}$ is contradicted by raising t^H to $t^H+\varepsilon$ for a small enough ε. ■

Notice also that from equation (1), and binding IR for θ^L, we get

$$v_A\left(x^H,\theta^H\right)+t^L-v_A\left(x^H,\theta^L\right)\geq v_A\left(x^L,\theta^H\right).$$

Since IC for θ^H binds,

$$v_A\left(x^H,\theta^H\right)=v_A\left(x^L,\theta^H\right)-t^L+t^H.$$

Therefore
$$v_A\left(x^L,\theta^L\right)-t^L=0\geq v_A\left(x^H,\theta^L\right)-t^H,$$
so that IC is also satisfied for θ^L.

Optimizing over Quantities

Hence when looking for the optimal menu $\left\{\left(x^H, t^H\right), \left(x^L, t^L\right)\right\}$, the relevant constraints are just IC for θ^H and IR for θ^L. Solving the problem is then easy:

$$\max_{\{(x^H,t^H),(x^L,t^L)\}} p\left(v_P\left(x^H\right) + t^H\right) + (1-p)\left(v_P\left(x^L\right) + t^L\right) \tag{2}$$
$$\text{subject to } v_A\left(x^L, \theta^L\right) - t^L = 0,$$
$$v_A\left(x^H, \theta^H\right) - t^H = v_A\left(x^L, \theta^H\right) - t^L.$$

Substituting the constraints into the objective function by eliminating $\left(t^H, t^L\right)$, we get the problem:

$$\max_{(x^H,x^L)} p\left(v_P\left(x^H\right) + v_A\left(x^H, \theta^H\right) - v_A\left(x^L, \theta^H\right) + v_A\left(x^L, \theta^L\right)\right)$$
$$+ (1-p)\left(v_P\left(x^L\right) + v_A\left(x^L, \theta^L\right)\right).$$

The first order conditions of the problem are

$$\begin{aligned} \frac{\partial}{\partial x^H}\left(v_P\left(x^H\right) + v_A\left(x^H, \theta^H\right)\right) &= 0, \\ \frac{\partial}{\partial x^L}\left(v_P\left(x^L\right) + v_A\left(x^L, \theta^L\right)\right) &= \frac{p}{1-p}\frac{\partial}{\partial x^L}\left(v_A\left(x^L, \theta^H\right) - v_A\left(x^L, \theta^L\right)\right). \end{aligned}$$

If we let $S(x,\theta) := v_P(x) + v_A(x,\theta)$ stand for the social surplus in the model, then we can rewrite these conditions as

$$\begin{aligned} \frac{\partial}{\partial x} S\left(x^H, \theta^H\right) &= 0, \\ \frac{\partial}{\partial x} S\left(x^L, \theta^L\right) &= \frac{p}{1-p}\frac{\partial}{\partial x^L}\left(v_A\left(x^L, \theta^H\right) - v_A\left(x^L, \theta^L\right)\right). \end{aligned}$$

Hence we see that x^H is chosen at its socially optimal level $x^*\left(\theta^H\right)$. For x^L, we note that $\frac{\partial}{\partial x^L}\left(v_A\left(x^L, \theta^H\right) - v_A\left(x^L, \theta^L\right)\right) > 0$ by strictly increasing differences. Hence for concave (in x) social surplus functions, we have $x^L < x^*\left(\theta^L\right)$.

Interpreting the Results

What is the reason for this result? From IC for θ^H, and IR for θ^L, we get:

$$t^H = v_A\left(x^H, \theta^H\right) - \left(v_A\left(x^L, \theta^H\right) - v_A\left(x^L, \theta^L\right)\right).$$

By strictly increasing differences the term in brackets is increasing. Hence to increase the transfer t^H, x^L must be lowered if everything else is kept equal. The losses from such reductions around $x = x^*\left(\theta^L\right)$ are small by the definition of optimality, Hence a reduction in x^L must be optimal. Notice that there is no similar effect for x^H. Since IC for θ^L does not bind, x^L matters only for t^H and therefore should be chosen socially optimally.

We call the term $\left(v_A\left(x^L, \theta^H\right) - v_A\left(x^L, \theta^L\right)\right)$ the information rent to the high type agent. It simply measures how much the high type would gain relative to the low type by lying about her type. Since the IC for θ^H is binding, the information rent measures exactly the equilibrium payoff of θ^H. The argument above shows that in the principal-agent framework, the principal faces a tradeoff between efficient provision of x^L to θ^L and the information rent extraction from θ^H resulting from x^L.

We conclude by summarizing the main points that we got here. If v_A satisfies strictly increasing differences, then at the optimal mechanism IR binds for θ^L, IC binds for θ^H. The allocation is increasing in the sense that $x^H \geq x^L$, x^H is chosen efficiently and x^L is distorted downwards.

I leave it for you to find the solutions to special cases such as $v_A(x, \theta) = x\theta$, $v_P(x) = \frac{-1}{2}x^2$ or alternatively $v_A(x, \theta) = \theta v(x)$, $v_P(x) = -cx$, where $v(x)$ is some concave increasing function.

2.1.5 Optimal Mechanisms: Continuum of Types

Revenue Equivalence Theorem We consider also the case where the set of possible types for the agent is an interval on the real line: $\Theta = \left[\underline{\theta}, \overline{\theta}\right]$. The principal has a prior distribution $F(\theta)$ on Θ with density function $f(\theta)$. The reason for solving this case is that when generalizing to multi-agent settings, having continuous type sets often allows us to concentrate on pure strategy equilibria whereas with discrete types, one would have to result to mixed strategy equilibria. A second reason is that the most fundamental result in Information Economics, the Revenue Equivalence Theorem needs a connected set of types to be valid.

For the remainder in this subsection, we concentrate on the incentives of the agent, and therefore we drop the superscripts in the payoff functions. The key result for characterizing the (equilibrium) maximized payoff to the agent is the envelope theorem.

Envelope theorems consider a family of parametrized optimization problems where the objective functions are given by $\{g(x,\theta)\}_{\theta\in\Theta}$. The choice variable is $x \in X$ and we take X to be a compact set of the real line, and we assume that $g(x,\theta)$ is differentiable in θ with uniformly bounded derivatives in θ.(i.e. uniformly as x varies).

Let

$$V(\theta) = \max_{x\in X} g(x,\theta),$$

and $X^*(\theta) := \arg\max_{x\in X} g(x,\theta)$. A function $x^*(\theta)$ is a *selection from* $X^*(\theta)$ if for all $\theta, x^*(\theta) \in X^*(\theta)$.

Theorem 3 (Envelope Theorem) *Assume that X is compact, and $\Theta = [\underline{\theta}, \overline{\theta}] \subset \mathbb{R}$ and $g : X \times \Theta \to \mathbb{R}$ is differentiable in θ with uniformly bounded derivatives in θ Then for any selection $x^*(\theta)$ from $X^*(\theta)$, $V'(\theta) = g_\theta(x^*,\theta)$ for almost every θ and furthermore*

$$V(\theta) = V(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} g_\theta(x^*, s)\, ds. \tag{3}$$

Proof. Equation (3) follows immediately from the Fundamental Theorem of Calculus if we prove that $V'(\theta) = g_\theta(x^*,\theta)$ for almost all θ.

Fix any selection $x^*(\theta)$. Then by definition of $V(\theta)$:

$$\begin{aligned} V(\theta) &= g(x^*(\theta), \theta) \geq g(x^*(\theta'), \theta), \\ V(\theta') &= g(x^*(\theta'), \theta') \geq g(x^*(\theta), \theta'), \end{aligned}$$

and therefore:

$$V(\theta) - V(\theta') \leq g(x^*(\theta), \theta) - g(x^*(\theta), \theta').$$

Consider the dividing this inequality by $(\theta - \theta')$ and taking the limits as $\theta' \uparrow \theta$ and $\theta' \downarrow \theta$ respectively, we get

$$\begin{aligned} V'(\theta-) &\leq g_\theta(x^*(\theta), \theta), \\ V'(\theta+) &\geq g_\theta(x^*(\theta), \theta). \end{aligned}$$

Hence if $V'(\theta)$ exists, we have

$$V'(\theta) = g_\theta(x^*(\theta), \theta)$$

whenever $g_\theta(x^*(\theta), \theta)$ exists. Hence

$$V(\theta) = V(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} g_\theta(x, s)\, ds.$$

■

To see how this connects to the contracting problem, consider the family of functions $\{v(x(m), \theta) - t(m)\}_{\theta \in \Theta}$. Let

$$V(\theta) = \max_{m \in \Theta} v(x(m), \theta) - t(m).$$

Incentive compatibility requires that $\theta \in m^*(\theta)$ for all θ. Hence we can compute using the envelope theorem:

$$V(\theta) = V(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} v_\theta(x(s), s)\, ds.$$

By noting that $V(\theta) = v(x(\theta), \theta) - t(\theta)$, we see that

$$t(\theta) = v(x(\theta), \theta) - V(\underline{\theta}) - \int_{\underline{\theta}}^{\theta} v_\theta(x(s), s)\, ds.$$

Hence apart from the additive constant $V(\underline{\theta})$, the transfer from the agent in any incentive compatible mechanism is uniquely pinned down by the physical allocation function $x(\theta)$. This is the Revenue Equivalence Theorem. You can view this as a generalization of the step of eliminating the transfers in the two-type model. Sometimes this envelope theorem characterization is also referred to as Local Incentive Compatibility since it is a direct consequence of the first-order condition for optimal announcements in the direct mechanism.

Theorem 4 (Revenue Equivalence Theorem) *Suppose X, Θ, v satisfy the conditions of the envelope theorem. Then the equilibrium payoff to the agent $V(\theta)$ in an incentive compatible mechanism $(x(\theta), t(\theta))$ can be written as*

$$V(\theta) = V(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} v_\theta(x(s), s)\, ds,$$

and the transfer $t(\theta)$ *must satisfy:*

$$t(\theta) = v(x(\theta), \theta) - V(\underline{\theta}) - \int_{\underline{\theta}}^{\theta} v_\theta(x(s), s)\, ds.$$

This remarkable result dates back to Holmstrom and Green&Laffont in late 1970's. To see that no such result is possible when the type sets are discrete, it is sufficient to consider the problem of selling an indivisible good (i.e. $x \in \{0,1\}$) to an agent with $\theta \in \{0,1\}$.Let $v(x,\theta) = x\theta$. Consider mechanisms where the agent gets the good if and only if $\theta = 1$. Then any $(x(\theta), t(\theta))$ with $x(\theta) = \theta$, $t(\theta) = \theta p$ for $p \in [0,1]$ is incentive compatible and has the same allocation function. The utility of $\theta = 0$ is zero for all p, but the transfer (and expected payoff) of $\theta = 1$ depends on p. Hence Revenue Equivalence Theorem fails.

Consider next the case where v satisfies strictly increasing differences and is twice differentiable with uniformly bounded second derivatives. We show that if the conditions for the Envelope Theorem (and hence for Revenue Equivalence Theorem) hold, then incentive compatibility is equivalent to Revenue Equivalence Theorem, i.e.

$$V(\theta) = V(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} v_\theta(x(s), s)\, ds. \tag{4}$$

and monotonicity of $x(\theta)$.

Theorem 5 *Suppose* $\frac{\partial^2 v(x,\theta)}{\partial x \partial \theta} > 0$. *Then* $(x(\theta), t(\theta))$ *is incentive compatible if and only if* $x(\cdot)$ *is non-decreasing and*

$$t(\theta) = v(x(\theta), \theta) - V(\underline{\theta}) - \int_{\underline{\theta}}^{\theta} v_\theta(x(s), s)\, ds.$$

Proof. Take any $m, \theta \in \Theta$.

$$\begin{aligned}
& v(x(\theta), \theta) - t(\theta) - (v(x(m), \theta) - t(m)) \\
= & \int_{\underline{\theta}}^{\theta} v_\theta(x(s), s)\, ds - \left(v(x(m), \theta) - v(x(m), m) + \int_{\underline{\theta}}^{m} v_\theta(x(s), s)\, ds \right) \\
= & \int_{m}^{\theta} \left(v_\theta(x(s), s) - v_\theta(x(m), s) \right) ds \\
= & \int_{m}^{\theta} \int_{x(m)}^{x(s)} v_{x\theta}(z, s)\, dz ds.
\end{aligned}$$

Since $v_{x\theta} > 0$, the double integral is nonnegative for all m, θ if and only if

$$(\theta - m)(x(\theta) - x(m)) \geq 0 \text{ for all } m, \theta.$$

■

The first equality uses Revenue Equivalence Theorem, the second is housekeeping and the third uses the Fundamental Theorem of Calculus. This result lies behind virtually all applications of mechanism design and adverse selections. For the case of strictly increasing differences, incentive compatibility is equivalent to envelope theorem and monotonicity of $x(\cdot)$. Surprisingly little is known about the implications of incentive compatibility without strictly increasing differences.

Optimizing over Incentive Compatible Mechanisms Rewriting the Objective Function

We are now in the position to find the optimal incentive compatible mechanism for the principal. Any mechanism $(x(\theta), t(\theta))$ is incentive compatible if $x(\cdot)$ is non-decreasing and $t(\theta)$ is calculated from $x(\theta)$ using the envelope formula. The expected payoff to the principal from an incentive compatible mechanism $(x(\theta), t(\theta))$ is:

$$\mathbb{E}_\theta \left[t(\theta) + v_P(x(\theta)) \right].$$

Using the envelope formula, we can eliminate the transfers and write this as:

$$\mathbb{E}_\theta \left[v(x(\theta), \theta) - V(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} v_\theta(x(s), s)\, ds + v_P(x(\theta)) \right].$$

Letting again

$$S(x, \theta) := v_P(x) + v(x, \theta),$$

The objective function simplifies to:

$$\mathbb{E}_\theta \left[S(x(\theta), \theta) - V(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} v_\theta(x(s), s)\, ds \right].$$

By interchanging the order of integration or by integrating by parts, we get

$$\begin{aligned}
\mathbb{E}_\theta \int_{\underline{\theta}}^{\theta} v_\theta\left(x\left(s\right), s\right) ds &= \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\theta}}^{\theta} v_\theta\left(x\left(s\right), s\right) ds f\left(\theta\right) d\theta \\
&= \int_{\underline{\theta}}^{\overline{\theta}} \int_{s}^{\overline{\theta}} f\left(\theta\right) d\theta v_\theta\left(x\left(s\right), s\right) ds \\
&= \int_{\underline{\theta}}^{\overline{\theta}} (1 - F\left(s\right)) v_\theta\left(x\left(s\right), s\right) ds.
\end{aligned}$$

Hence by substituting this expression into the objective function, we have:

$$\mathbb{E}_\theta \left[S\left(x\left(\theta\right), \theta\right) - V\left(\underline{\theta}\right) - \frac{\left(1 - F\left(\theta\right)\right)}{f\left(\theta\right)} v_\theta\left(x\left(\theta\right), \theta\right) \right].$$

The problem of the principal is thus to:

$$\begin{aligned}
&\max_{\left(x\left(\theta\right)\right)} \mathbb{E}_\theta \left[S\left(x\left(\theta\right), \theta\right) - V\left(\underline{\theta}\right) - \frac{\left(1 - F\left(\theta\right)\right)}{f\left(\theta\right)} v_\theta\left(x\left(\theta\right), \theta\right) \right] \\
&\text{subject to } x\left(\theta\right) \text{ nondecreasing.}
\end{aligned}$$

It is clear that it is optimal for the principal to set $V(\underline{\theta})$ equal to the outside option that we normalize to 0 from now on.

Finding a Solution

There are two ways of solving such problems, The first approach is easy: just ignore the monotonicity constraint and see what happens if you maximize the expectation separately for each θ. Recall that

$$\max_x \mathbb{E}_\theta u\left(x, \theta\right) \leq \mathbb{E}_\theta \max_x u\left(x, \theta\right).$$

If the resulting x is nondecreasing in θ, then the solution thus obtained is optimal.

Assuming an interior solution for x in the problem

$$\max_x S\left(x, \theta\right) - V\left(\underline{\theta}\right) - \frac{\left(1 - F\left(\theta\right)\right)}{f\left(\theta\right)} v_\theta\left(x, \theta\right)$$

yields the first order condition:

$$S_x(x,\theta) - \frac{(1-F(\theta))}{f(\theta)} v_{x\theta}(x,\theta) = 0. \tag{5}$$

Notice the similarity in this formula to the result derived for the two-type case. At $\overline{\theta}$, $(1 - F(\theta)) = 0$, and therefore $x^*(\overline{\theta})$ coincides with the socially optimal solution. Hence as in the two-type model the allocation of the highest type is not distorted. Similarly also the allocation of all types below the highest is distorted downwards. Clearly also the IR constraint binds only for the lowest type agent.

To proceed with the solution, differentiating with respect to x and θ gives:

$$\left(S_{x\theta} - \frac{d}{d\theta}\left(\frac{(1-F(\theta))}{f(\theta)}\right) v_{x\theta} - \left(\frac{(1-F(\theta))}{f(\theta)}\right) v_{x\theta\theta}\right) d\theta$$
$$+ \left(S_{xx} - \frac{(1-F(\theta))}{f(\theta)} v_{xx\theta}\right) dx = 0$$

The second term is non-positive by the second order condition for maximum. The first term is hard to sign in general, but much of the literature deals with the linear case, where $v(x,\theta)$ is linear in θ. In this case, the first term reduces to

$$\left(1 - \frac{d}{d\theta}\left(\frac{(1-F(\theta))}{f(\theta)}\right)\right) v_{x\theta}$$

since $S_{x\theta} = v_{x\theta}$ by the private values assumption. Hence a sufficient condition for $x^*(\theta)$ to be increasing is that

$$\frac{d}{d\theta}\frac{f(\theta)}{1-F(\theta)} \geq 0$$

Since $\frac{f(\theta)}{1-F(\theta)}$ is the hazard rate of the distribution $F(\theta)$, this condition is known in the literature as the monotone hazard rate condition.

If the resulting $x^*(\theta)$ fails to be increasing, then a more sophisticated approach using optimal control theory is needed. We shall not cover this but we shall rather look at some applications.

Models of Price Discrimination **Mussa-Rosen Model** A monopolist firm (the principal) chooses price-quality pairs for differentiated buyers (the agent) to maximize her profit. The agent's payoff function is given by $v(x,\theta) = \theta x$. The prior on agent types is uniform on $[0,1]$. The principal's payoff function $v_P(x) = -\frac{1}{2}x^2$ represents a quadratic cost of producing quality. The main question is the shape of the optimal pricing function for different quality levels.

We apply formula 5 to determine a candidate optimal quality provision rule:

$$S_x(x,\theta) - \frac{(1-F(\theta))}{f(\theta)} v_{x\theta}(x,\theta) = \theta - x - (1-\theta) = 0.$$

Therefore the optimal quality provision rule is given by:

$$x^*(\theta) = \max\{2\theta - 1, 0\}.$$

Clearly this is non-decreasing in θ. The associated transfer is computed from Revenue Equivalence Theorem as:

$$t(\theta) = \theta(2\theta - 1) - \int_{\frac{1}{2}}^{\theta} (2s-1)ds$$

$$= 2\theta^2 - \theta - \theta^2 + \theta - \frac{1}{4} = \theta^2 - \frac{1}{4}.$$

Since $\theta = \frac{1+x}{2}$, we can write

$$T(x) = (\frac{1+x}{2})^2 - \frac{1}{4}.$$

Hence we see that $T'(x) = (1+x)$ is increasing. The marginal unit of quality is supplied at an increasing price. Furthermore, the price grows more quickly that the marginal cost of providing quality $v_P'(x) = x$. The economics message from this model can be summarized as follows: A price-discriminating monopolist sells quality at a premium. It is easy to verify that the general properties discussed in the previous subsection hold for this model. The analysis goes through for other convex cost functions for providing quality.

Maskin-Riley Model

This model assumes a linear cost of provision $v_p(x) = cx$ for the monopolist and a concave utility for the agent in x, i.e. $v(x,\theta) = \theta v(x)$ for some concave function $v(x)$. The physical allocation x is here best interpreted as the quantity sold. Again, we seek to determine the optimal schedule $x(\theta)$. We continue with the assumption that θ is uniformly distributed on $[0,1]$.

An application of formula (5) gives now:

$$\theta v'(x) - c - (1-\theta)v'(x) = 0.$$

Therefore we have:

$$v'(x^*) = \frac{c}{2\theta - 1}.$$

By taxation principle, each x in the optimal schedule $(x^*(\theta), t(\theta))$ is associated with a unique transfer $T(x)$. The agent's optimization problem yields immediately:

$$\theta v'(x) = T'(x).$$

Taking together with the optimality formula for the principal, we get:

$$2T'(x) = c + v'(x).$$

Therefore we have: $T''(x) = \frac{1}{2}v''(x) \leq 0$. Therefore the marginal price is decreasing in the amount sold, and we have a model of quantity discounts.

2.2 Optimal Mechanism Design with Many Agents

A canonical example of mechanism design problems is the organization of a sales procedure. A seller want to sell one or more goods to a number of interested parties, the buyers. Again, the assumption is that the seller may make a take-it-or-leave-it offer and the buyers are privately informed.

The key difference to the single buyer case covered in the previous section is that each individual agent is now uncertain about the eventual terms of trade. Since the allocation and transfers depend on the private information of all buyers, each buyer must anticipate the behavior of others in the mechanism. In other words, each buyer os faced with a Bayesian game whose outcomes depend on the action of all buyers.

We cover only the private values case with quasilinear preferences. Gibbard-Satterthwaite Theorem tells us that in the general private values case, elicitation of preferences is impossible. In the first subsection, we set up the general mechanism problem. We recall that Revelation Principle remains valid exactly as in the previous section. Hence we focus on incentive compatible or truthful direct revelation mechanisms.

After covering notions of incentive compatibility and individual rationality for multi-agent mechanisms, we specialize to linear environments where the payoff of each buyer is linear in her own type. We cover the optimal auction result of Roger Myerson. Throughout this section, we emphasize the Revenue Equivalence Theorem derived in the previous section and use it to derive equilibrium bidding strategies for different auction formats.

The remainder of this section is spent on efficient mechanisms. The objective here is to explore the possibility if efficient trade under incomplete information in the spirit of the classical Welfare Theorems of the competitive model. We start with the Vickrey-Clarke-Groves (VCG) mechanism that is the most canonical of all multi-agent mechanisms. In a sense the VCG mechanisms provide a similar benchmark for economies with incomplete information as the competitive model is for the complete information case.

We end this section by discussing the requirement of budget balance. The last result is the celebrated Myerson-Satterthwaite Theorem that demonstrates the impossibility of efficient trade if incentive compatibility, individual rationality and budget balance are required. This gives a good reason for studying the various trade-offs arising in models of incomplete information.

2.2.1 Set-up

The Mechanism Designer or Principal has the bargaining power and N agents $i \in \{1, 2, ..., N\}$ have private information. For each agent i, her type is $\theta_i \in \Theta_i$ and is known only to agent i. All agents and the principal share the same common prior on the types given by the distribution function $F(\theta)$, where $\theta = (\theta_1, ..., \theta_N) \in \Theta = \times_{i=1}^{N} \Theta_i$. An agent i of type θ_i has interim beliefs $F(\theta_{-i} | \theta_i)$ on the type vector θ_{-i} of the other players.

There is a set of potential allocation A with a typical element $a \in A$. The agents and the principal have payoff function (von Neumann - Morgenstern) over the allocations and type vectors. Agent i has a payoff function $u_i(a, \theta)$ and the principal has a payoff function $u_0(a, \theta)$.

The principal chooses message spaces M_i for each of the agents, and commits to an allocation rule (possibly random) $\phi : M \to \Delta(A)$, where $M = \times_{i=1}^{N} M_i$ with a typical element $m \in M$. The agents announce their messages simultaneously, in other words they engage in a Bayesian game with (pure) strategies

$$\sigma_i : \Theta_i \to M_i,$$

and payoffs (for pure allocation rules) given by $u_i(\phi(m_i, m_{-i}), \theta_i, \theta_{-i})$. Again we call the pair (M, ϕ) a mechanism.

We have now different possible solution concepts for the message game between the agents: Dominant strategy equilibrium, Bayes-Nash equilibrium, Ex-post equilibrium.

1. A vector $\sigma = (\sigma_1, ..., \sigma_N)$ is a *dominant strategy equilibrium* in pure strategies if for all i, θ_i and for all θ_{-i}, m_{-i},

$$u_i(\phi(\sigma_i(\theta_i), m_{-i}), \theta_i, \theta_{-i}) \geq u_i(\phi(m_i, m_{-i}), \theta_i, \theta_{-i}) \text{ for all } m_i.$$

2. A vector $\sigma = (\sigma_1, ..., \sigma_N)$ is an *ex-post equilibrium* in pure strategies if for all i, θ_i and for all θ_{-i},

$$u_i(\phi(\sigma_i(\theta_i), \sigma_{-i}(\theta_{-i})), \theta_i, \theta_{-i}) \geq u_i(\phi(m_i, \sigma_{-i}(\theta_{-i})), \theta_i, \theta_{-i}) \text{ for all } m_i.$$

3. A vector $\sigma = (\sigma_1, ..., \sigma_N)$ is an *Bayes-Nash equilibrium* in pure strategies if for all i, θ_i and for all m_i,

$$\mathbb{E}_{\theta_i} u_i \left(\phi\left(\sigma_i\left(\theta_i\right), \sigma_{-i}\left(\theta_{-i}\right)\right), \theta_i, \theta_{-i}\right) \geq \mathbb{E}_{\theta_i} u_i \left(\phi\left(m_i, \sigma_{-i}\left(\theta_{-i}\right)\right), \theta_i, \theta_{-i}\right)$$

Clearly a dominant strategy equilibrium is an ex-post equilibrium and an ex-post equilibrium is a Bayes-Nash equilibrium. For all three equilibrium concepts, one can prove that the revelation principle holds: For any mechanism (M, ϕ) and an equilibrium σ in the Bayesian game induced by (M, ϕ), there is another mechanism (M', ϕ') such that truthful reporting, i.e. $\sigma_i(\theta_i) = \theta_i$ for all i, θ_i is an equilibrium in the game induced by (M', ϕ') and the outcomes in the two games coincide, i.e. $\phi(\sigma(\theta)) = \phi'(\theta)$.The formal proof for the three equilibrium concepts is left as an exercise.

2.2.2 Notions of Incentive Compatibility

We define here the corresponding three notions for incentive compatibility. A direct mechanism is said to be incentive compatible if truthful reporting is optimal for every type of every agent. Since what follows in these lectures is in the quasilinear setting, I write direct mechanism as $(x(\theta), t(\theta))$ where x denotes again a physical allocation and t denotes a vector of transfers. We write $u_i(x, t, \theta) = v_i(x, \theta) - t_i$ for $i \in \{1, ..., N\}$ and $u_0(x, t, \theta) = v_0(x, \theta) + t_0$.

1. A direct mechanism $(x(m), t(m))$ is dominant strategy incentive compatible if for all $i, \theta_i, \theta_{-i}, m_{-i}$, we have:

$$\theta_i \in \arg\max_{m_i \in \Theta_i} v_i(x(m_i, m_{-i}), \theta_i, \theta_{-i}) - t_i(m_i, m_{-i}),$$

   i.e. truthtelling is optimal for each type of each player regardless of what the type and announcement of the other players.

2. The mechanism $(x(m), t(m))$ is ex post incentive compatible if for all i, θ_i, θ_{-i}, we have

$$\theta_i \in \arg\max_{m_i \in \Theta_i} v_i(x(m_i, \theta_{-i}), \theta_i, \theta_{-i}) - t_i(m_i, \theta_{-i}),$$

i.e. truthtelling is optimal for all θ_i given that $m_{-i} = \theta_{-i}$. In other words, as long as the other players are truthful, their type is not important for a players' incentives for truthful reporting. Notice also that for private values environments, the two notions coincide in the static case. For dynamic settings, they will differ also for private values.

3. The mechanism $(x(m), t(m))$ is Bayes-Nash incentive compatible if for all i, θ_i, we have

$$\theta_i \in \arg\max_{m_i \in \Theta_i} \mathbb{E}_{\theta_{-i}} \left[v_i(\theta_i, \theta_{-i}, q(m_i, \theta_{-i})) - t_i(m_i, \theta_{-i}) \left| \theta_i \right. \right],$$

i.e. truthtelling is optimal given that $m_{-i} = \theta_{-i}$ and given the distribution $F(\theta_{-i} \left| \theta_i \right.)$ for the types of other players.

Clearly, a dominant strategy incentive compatible mechanism is also ex post incentive compatible and an ex post IC mechanism is Bayes-Nash IC.

2.2.3 Individual Rationality

Sometimes it makes sense to consider settings where the agents have an option of staying out from the mechanism. In general, the payoff from this option could depend on the type, but for these notes, I will assume that this outside option value is constant on type and normalized to 0. There are different notions of participation constraints depending on when the participation decision is taken. We consider here incentive compatible mechanisms and assume truthful reports by the other players.

1. A mechanism $(x(\cdot), t(\cdot))$ is *ex-post individually rational* if for all i and all θ,

$$v_i(x(\theta), \theta) - t(\theta) \geq 0.$$

2. A mechanism $(x(\cdot), t(\cdot))$ is *interim individually rational* if for all i and all θ_i,

$$\mathbb{E}_{\theta_{-i}} \left[v_i(x(\theta), \theta) - t(\theta) \left| \theta_i \right| \right] \geq 0.$$

3. A mechanism $(x(\cdot), t(\cdot))$ is *ex-ante individually rational* if for all i and all θ_i,

$$\mathbb{E}_{\theta} \left[v_i(x(\theta), \theta) - t(\theta) \left| \theta_i \right| \right] \geq 0.$$

Clearly an ex-post IR mechanism is interim IR and ex-ante IR. For the ex-post IR, it is sometimes useful to consider also the case where the other players are not truthful and then one would have the requirement that for all i and all m_{-i}, θ,

$$v_i\left(x\left(\theta_i, m_{-i}\right), \theta\right) - t\left(\theta_i, m_{-i}\right) \geq 0.$$

2.2.4 Linear Environments

Assume next the simple linear form of preferences:

$$u_i\left(x, t, \theta\right) = \theta_i v_i\left(x\right) - t_i,$$

where v_i is an increasing function of the physical allocation x. We also assume that for all i, $\theta_i \in \left[\underline{\theta}_i, \overline{\theta}_i\right]$ and that the distribution admits a well defined density function with full support. Furthermore, we assume that for $i \neq j$, $\theta_i \perp \theta_j$. We look for a characterization of the truthful equilibria of direct mechanisms. Denote an arbitrary announcement by player i by m_i, and denote by $T_i\left(m_i, \theta_i\right)$ the expected transfer given true type θ_i and announcement m_i. Notice that by our assumption of independence of types, we can write

$$T_i\left(m_i, \theta_i\right) = T_i\left(m_i\right).$$

Similarly, let

$$X_i\left(m_i\right) = E_{\theta_{-i}}\left[v_i\left(x\left(m_i, \theta_{-i}\right)\right)\right].$$

The key here is that agent i's expected payment conditional on report m_i is independent of θ_i by the independence of the valuations. This is clearly not the case when types are correlated and this is the main reason why revenue equivalence fails in that setting.

Because of the linearity of the utilities, we may now write the expected payoff from announcing m_i while of type θ_i.

$$E_{\theta_{-i}}\left[u_i\left(x\left(m_i, \theta_{-i}\right), \theta\right)\right] = \theta_i X_i\left(m_i\right) + T_i\left(m_i\right),$$

Define also the following functions for $i = 1, ..., N$:

$$\widehat{V}_i\left(\theta_i, m_i\right) = \theta_i X_i\left(m_i\right) + T_i\left(m_i\right).$$

Now we are in a position to characterize the outcomes that are achievable in Bayesian Nash equilibria of mechanism design problems. Let

$$V_i := \widehat{V}_i(\theta_i, \theta_i).$$

Since the payoff function $\theta_i X_i$ satisfies strictly increasing differences for all i, we have from the results on single agent implementability the following theorem.

Proposition 1 *(Bayesian incentive compatibility of outcomes)*
The mechanism $(x(\cdot), t(\cdot))$ is Bayesian incentive compatible if and only if for all $i \in \{1, ..., N\}$,

1. *$X_i(m_i)$ is nondecreasing.*
2. *$V_i(\theta_i) = V_i(\underline{\theta}_i) + \int_{\underline{\theta}_i}^{\theta_i} X_i(s)\, ds$ for all θ_i.*

2.2.5 Optimal Myerson Auctions

We continue with the IPV case of auctions. There are N bidders, each with a type θ_i drawn according to the distribution function $F_i(\theta_i)$. We assume that each F_i has a continuous density function f_i and that the types are statistically independent. For each allocation decision $x = (x_1, ..., x_N)$, we have the payoff of bidder i as

$$u_i(\theta, q, t) = \theta_i x_i - t_i.$$

We also require that $x \in \mathbb{R}_+^N$ and $\sum_i q_i \leq 1$. Let

$$X_i(\theta_i) = \mathbb{E}_{\theta_{-i}} x_i(\theta_i, \theta_{-i}) = \int_{\Theta_{-i}} x_i(\theta_i, \theta_{-i}) \prod_{j \neq i} f_j(\theta_j)\, d\theta_{-i},$$

and

$$T_i(\theta_i) = \mathbb{E}_{\theta_{-i}} t_i(\theta_i, \theta_{-i}) = \int_{\Theta_{-i}} t_i(\theta_i, \theta_{-i}) \prod_{j \neq i} f_j(\theta_j)\, d\theta_{-i}.$$

The expected payoff for bidder i of type θ_i from announcing type m_i is

$$\widehat{V}_i(\theta_i, m_i) = \theta_i X_i(m_i) - T_i(m_i).$$

By the previous lecture, an allocation $X_i(\theta_i)$ is incentive compatible if and only if we have:

1. $V_i(\theta_i) = \widehat{V}_i(\theta_i, \theta_i) = V_i(\underline{\theta}_i) + \int_{\underline{\theta}_i}^{\theta_i} X_i(s)\,ds.$

2. $X_i'(\theta_i)$ non-decreasing.

The expected payment going to the auctioneer is

$$\mathbb{E}\sum_i T_i(\theta_i) = \int_\Theta \sum_i t_i(\theta_i, \theta_{-i}) \prod_i f_i(\theta_i)\,d\theta.$$

Using the envelope result, we have

$$\mathbb{E}\sum_i T_i(\theta_i) = \int_\Theta \sum_i [\theta_i x_i(\theta) - V_i(\underline{\theta}_i) - \int_{\underline{\theta}_i}^{\theta_i} x_i(s)\,ds] \prod_i f_i(\theta_i)\,d\theta.$$

If the outside option for the bidders is fixed at 0, it is optimal to set $V_i(\theta_i) = 0$. Integrating by parts or interchanging the order of integration we get:

$$\int_{\Theta_i}\int_{\underline{\theta}_i}^{\theta_i} x_i(s)\,ds \prod_i f_i(\theta_i)\,d\theta = \int_\Theta x_i(s)\frac{(1 - F_i(\theta_i))}{f_i(\theta_i)} \prod_i f_i(\theta_i)\,d\theta.$$

Therefore

$$\mathbb{E}\sum_i T_i(\theta_i) = \int_\Theta \sum_i [\theta_i q_i(\theta) - x_i(s)\frac{(1 - F_i(\theta_i))}{f_i(\theta_i)}] \prod_i f_i(\theta_i)\,d\theta.$$

The seller's revenue is maximized if $q(\theta)$ is chosen to maximize the above integral. A good strategy is to first ignore the constraint that $X_i(\theta_i)$ must be non-decreasing in $\theta_{i.}$Since the expected revenue is linear in x, $x^*(\theta)$ maximizes the integral for each θ if:

$$x_i^*(\theta) = \begin{cases} 1 \text{ if } \theta_i - \frac{(1-F_i(\theta_i))}{f_i(\theta_i)} \geq \theta_j - \frac{(1-F_i(\theta_i))}{f_j(\theta_j)} \text{ and } \theta_i - \frac{(1-F_i(\theta_i))}{f_i(\theta_i)} \geq 0. \\ 0 \text{ otherwise.} \end{cases}$$

We call

$$\theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)}$$

the virtual type of player i. If the virtual surplus is nondecreasing in θ_i, then we have arrived at a full solution of the problem. Otherwise we must take the constraint $X_i'(\theta_i) \geq 0$ into account and use optimal control theory to find the optimal feasible solution. A sufficient condition for having a nondecreasing virtual surplus is that

$$\frac{f_i(\theta_i)}{1 - F_i(\theta_i)}$$

be nondecreasing in θ_i. In this case, the optimal auction thus allocates the good to the buyer with the maximal virtual type if this type is non-negative for some bidder and leaves the good unallocated otherwise.

Comments on Optimal Auction

1. Symmetric F_i : optimal reserve prices that are independent of N.
2. Asymmetric F_i : optimal auction favors the disadvantaged bidder.
3. Notice the trade-off between efficiency and information rent
4. A good exercise is to use the revenue equivalence theorem to conclude the following: the expected revenue from an optimal auction with n symmetric bidders is dominated by the expected revenue from a second price auction without reserve prices with $n+1$ bidders from the same distribution. Hence the amount that can be gained by optimally tailoring the auction rules (reservation prices) is quite limited.

Application: Symmetric Equilibria Through Revenue Equivalence

FPA

From revenue equivalence theorem, we have:

$$V(\theta_i) = V_i(\underline{\theta}_i) + \int_{\underline{\theta}_i}^{\theta_i} X_i(s)\, ds.$$

If the first price auction has an equilibrium with strictly increasing strategies $b_i(\theta_i)$, then we have:

$$X_i(\theta_i) = F(\theta)^{N-1}$$

and

$$V_i(\underline{\theta}_i) = 0$$

On the other hand, in first price auction, the payoff to bidder i is:

$$(\theta_i - b_i(\theta_i))X_i(\theta_i)$$

or

$$(\theta_i - b_i(\theta_i))X_i(\theta_i) = \int_{\underline{\theta}_i}^{\theta_i} X_i(s)\, ds$$

or

$$(\theta_i - b_i(\theta_i))F(\theta)^{N-1} = \int_{\underline{\theta}_i}^{\theta_i} F(\theta)^{N-1} ds$$

or

$$b_i(\theta_i) = \theta_i - \frac{\int_{\underline{\theta}_i}^{\theta_i} F(\theta)^{N-1} ds}{F(\theta)^{N-1}}$$

War of attrition

In this auction format the highest bidder wins the object, all losers pay their own bid and the winner pays the highest losing bid. This can be seen as a model where the bidders compete for a price by expending resources (as expressed in their bids). The contest ends when all but one bidders give up. At that point, the winner collect the prize.

We look for a symmetric equilibrium in increasing strategies. Any such equilibrium allocates the prize efficiently (i.e. the highest bidder is the bidder with the highest valuation. Hence the allocation is the same as in the second price auction. By revenue equivalence theorem, the expected payment must also coincide be the same as above Denote the bid of bidder i in the symmetric equilibrium of the war of attrition by $b^W(\theta_i)$ and the expected payment by Π^W. We have:

$$\Pi^W = (1 - F^{N-1}(\theta_i))b^W(\theta_i) + \int_0^{\theta_i} b^W(z)dF^{N-1}(z) = \int_0^{\theta_i} z dF^{N-1}(z) = \Pi^{SPA}.$$

Since this holds for all θ_i we may differentiate to get:

$$\frac{db^W(\theta_i)}{d\theta_i}(1 - F^{N-1}(\theta_i)) = \theta_i dF^{N-1}(\theta_i)$$

and therefore

$$b^W(\theta_i) = \int_0^{\theta_i} \frac{\theta_i dF^{N-1}(\theta_i)}{1 - F^{N-1}(\theta_i)}.$$

We turn next to the issue of efficient mechanisms. We shall see how the revenue equivalence theorem turns out to be useful for deducing properties of social surplus maximizing allocation schemes.

2.3 Efficient Mechanisms: VCG Mechanisms

In this subsection, we address the incentive compatibility of socially efficient decision rules in models with private values and quasi-linear payoffs.

Definition 2 *The model has* private values *if for all* $i \in \{0, 1, ..., N\}$,

$$v_i(x, \theta_i, \theta_{-i}) = v_i(x, \theta_i, \theta'_{-i}) \text{ for all } \theta_{-i}, \theta'_{-i} \in \Theta_{-i}.$$

We start with the simplest setting where $v_0 = 0$, and $t_0 = \sum_{i=1}^{N} t_i$. In other words, the mechanism designer collects the payments made by the agents. With this specification, Pareto-efficiency reduces to surplus maximization, and efficiency is unambiguous.

Let

$$x^*(\theta) \in \arg\max_{x \in X} \sum_{i=1}^{N} v_i(x, \theta) = \arg\max_{x \in X} \sum_{i=1}^{N} v_i(x, \theta_i),$$

where the last equality follows from the private values assumption.

Consider the direct mechanism with choice rule $x^*(m)$ and with transfer function

$$t_i^*(m) = -\sum_{j \neq i} v_j(x^*(m), m_j) + \tau_i(m_{-i}). \tag{6}$$

It is easy to see that for an arbitrary announcement m_{-i} by the other players, $m_i = \theta_i$ is the optimal report for an arbitrary function τ_i. The mechanism $(x^*(\theta), t^*(\theta))$ as above is called a *Vickrey-Clarke-Groves* mechanism (VCG mechanism). Green and Laffont (1977, 1979) show that if there are no restrictions on the domain of preferences for the players, then VCG mechanisms are the only mechanisms that make truthful revelation a dominant strategy for the efficient transfer rule $x^*(\theta)$. Basically this observation is a direct consequence of the Revenue Equivalence Theorem (separately for each announcement m_{-i} by the other agents), and therefore the set of parameters Θ must be sufficiently regular for the uniqueness result (e,g, a convex subset of $\mathbb{R}^k$ will do). As equation (6) makes clear, the transfers are pinned down by the efficient allocation rule $x^*(\cdot)$ up to the constant $\tau_i(m_{-i})$.

A particularly useful VCG mechanism is the pivot mechanism or the externality mechanism, where

$$\tau_i(m_{-i}) = \max_{x \in X} \sum_{j \neq i} v_j(x, m_j).$$

If we denote the maximized surplus for a coalition $\mathcal{S}$ of players by

$$W_{\mathcal{S}}(\theta) = \max_{x \in X} \sum_{i \in \mathcal{S}} v_i(x, \theta_i),$$

then we see that the utility in the pivotal mechanism going to player i is given by

$$\max_{x \in X} \sum_{i=1}^{N} v_i(x, \theta) - \max_{x \in X} \sum_{j \neq i} v_j(\theta_j, q) = W_{\{1,\ldots,N\}}(\theta) - W_{\{1,\ldots,N\}\setminus\{i\}}(\theta) := M_i(\theta),$$

i.e. in the pivotal mechanism, each player gets her marginal contribution to the social welfare as her payoff.

Notice that since $(x^*(\cdot), t^*(\cdot))$ is dominant strategy incentive compatible, correlation in the agents' types makes no difference for the incentive compatibility of the mechanism. It works for all private values settings with quasi-linear payoffs.

2.3.1 Budget Balance

An unfortunate feature of the VCG mechanism is that it is not budget balanced (BB) between the agents. In some problems this is not problematic: the mechanism designer may be a real seller that does not know the valuations of potential buyers. In other economic situations, positive sums of transfers from the agents represent wasted money, and negative sums of transfers feature financing originating outside the model.

We say that the mechanism $(x(\cdot), t(\cdot))$ satisfies *ex-post budget balance* if for all θ,

$$\sum_{i=1}^{N} t_i(\theta) = 0.$$

It satisfies *ex-ante budget balance* if

$$\mathbb{E}_\theta \sum_{i=1}^{N} t_i(\theta) = 0.$$

A less demanding requirement for mechanisms would be that they have no deficits (again either in the ex-post and ex-ante senses), i.e. $\Sigma_i t_i \geq 0$.

2.3.2 BB in BNE

The previous chapter raises the question of what more can be done with BNE as the solution concept since this is less restrictive that dominant strategy. As long as the types are independent, we know from the characterization of incentive compatibility that all BNE implementing the efficient decision rule have the same expected payoffs apart from the utility of the lowest type player. Hence the expected payments etc. can be calculated using VCG mechanisms.

2.3.3 IC, BB and Participation

Assume now that the type sets of the agents are intervals on the real line $\Theta_i = [\underline{\theta}_i, \overline{\theta}_i]$. Assume also that $v_i(x, \theta_i)$ satisfies strictly increasing differences. Then we know from the revenue equivalence theorem that in any VCG mechanism, type $\underline{\theta}_i$ of player i gets the lowest expected equilibrium payoff. If the outside option for each i is given by payoff 0, then interim Individual Rationality is satisfied for all players if and only if $V_i(\underline{\theta}_i) \geq 0$.

Amongst interim individually rational VCG mechanisms, the following transfer payment rule collects the largest expected payment from the agents:

$$t_i^P(\theta) = W(\underline{\theta}_i, \theta_{-i}) - W_{-i}(\theta).$$

We call this the *pivotal mechanism*. It is immediately clear that with these transfers, the lowest type of each agent has a zero expected payoff. Hence a very useful observation is that if the VCG mechanism with payment function t^P runs a deficit, then there is no efficient IR mechanism that satisfies budget balance.

It is slightly less trivial (and not so important) to show that if $(x^*(\cdot), t^P(\cdot))$ results in a surplus and the types are independent, then there exists another efficient, IR mechanism that satisfies budget balance.

We discuss next the issue of individual rationality or participation in a particular mechanism. To fix ideas, we consider a game of public goods provision. A project can be completed at cost $c > 0$ or alternatively it may remain uncompleted at zero cost. Therefore $x \in \{0, 1\}$. Assume that $v_i(0, \theta) = 0$ for all i and $v_i(1, \theta) = \theta_i$ for all i, and θ_i. Then the efficient decision rule is to have

$$x^*(m) = \begin{cases} 1 \text{ if } \Sigma_i m_i \geq c. \\ 0 \text{ otherwise.} \end{cases}$$

Consider the pivot mechanism for this problem:

$$t_i(m) = \begin{cases} \underline{\theta} x^*(\underline{\theta}, m_{-i}) + c - \Sigma_{j \neq i} m_j \text{ if } \Sigma_i m_i \geq c \geq \Sigma_{j \neq i} m_j, \\ 0 \text{ otherwise.} \end{cases}$$

It is a good exercise to verify that all players have a dominant strategy to report $m_i = \theta_i$. Suppose next that all player have the opportunity to opt out from the mechanism and get a utility of 0. Then it is immediate from the revenue equivalence and the transfer function that the largest transfers consistent with participation by all types are given by the pivot mechanism transfers. If we understand budget balance to mean that the sum of transfers covers the cost c, then we can simply evaluate the expected payments in the pivot mechanism and ask whether the cost c is covered. I leave it as an exercise to show that cost is covered only in cases where $N\underline{\theta} > c$ or $c < N\overline{\theta}$. In all other cases, there is a deficit. The computation of this is left as an exercise.

The same line of argument can be used in e.g. the bilateral trade model. Here a buyer has a valuation $v \in [\underline{v}, \overline{v}]$, and a seller has a cost of producing the good given by $c \in [\underline{c}, \overline{c}]$. The cost and the valuation are private information to the seller and the buyer, respectively. Assume that $\overline{v} > \underline{c}$ and $\underline{v} < \overline{c}$ The buyer's expected utility from the good is $xv - t$, where x is the probability of trade, The seller's gross utility is $t - xc$. Construct the pivot mechanism. Argue that it has the largest transfers, and show that the pivot mechanism transfers result in a deficit. By doing this example, you will have demonstrated an instance of the famous Myerson-Satterthwaite Theorem:

Theorem 6 (Myerson-Satterthwaite) *If $\underline{v} < \overline{c}$ and $\underline{c} < \overline{v}$, then there does not exist any NE of any bargaining game that is ex post efficient, expected budget balanced, and interim individually rational.*

Why is this theorem so important? Recall an equally famous theorem by one of the leaders of Chicago School, Ronald Coase:

Theorem 7 (Coase Theorem) *If economic agents can bargain without any cost, then the economic outcome will be Pareto efficient regardless of the initial allocation of property rights.*

It does not take much to see the political consequences of Coase Theorem. Myerson-Satterthwaite Theorem gives a convincing and simple example of a setting where the assumption of costless bargaining makes no sense. Consider a modification of the buyer-seller example above where individual 1 has value v_1 from the use of the good and individual 2 has value v_2, If these values are privately known, it is immediate that the model is isomorphic to the buyer seller model above (with v_1 being the (opportunity) cost in the new model). Hence there is no efficient allocation mechanism that balances the budget and respects the property rights of both bargainers (1 must give up the good voluntarily and 2 must agree to make any payment).

Nevertheless, it is obvious that in the absence of any property rights, an efficient mechanism exists: Just auction the good without reserve prices between the two agents. Hence the initial assignment of property rights matters for efficient final allocation and the conclusion of Coase Theorem fails.

There is a lengthy literature on bargaining under incomplete information, and there are numerous ways in which the inefficiency in Myerson Satterthwaite can manifest: Delays in bargaining is one possibility if the agents discount future payoffs, lack of mutually profitable trade is a more immediate way to resolve this.

2.3.4 Correlated Values and Interdependent Values (Optional)

We do not have time here to go into any details of mechanism design in the cases where the types of the agents are correlated or private values assumption is violated. In the notes 'Mechanism Design: Review of Basic Concepts' (MD), I discuss these topics to some extent. The main observations can be summarized in the list below that I provide for your convenience.

1. With correlated values, Revenue Equivalence Theorem fails. Furthermore, almost any allocation rule $x(\theta)$ can be made incentive compatible

with a suitable transfer scheme $t(\theta)$ without changing interim expected payoffs. the reasons why this apparently fantastic result is not so great are discussed in (MD). The fact that correlation makes almost everything possible is mostly viewed as a theoretical nuisance. As a result, much of the mechanism design literature has been confined to the independent values case.

2. Under interdependent values efficient allocation rules may fail to be incentive compatible even when types are independent. There are two reasons for this. First, even in the case of single-dimensional types, efficient allocation rules may fail to be monotonic. A simple example of this is the following:

    - How to allocate a single indivisible object between two players? Only player 1 has private information $\theta_1 \in [0, 2]$. Both players' payoffs depend on this information:

    - $$v_1 = \theta_1, v_2 = 2\theta_1 - 1.$$

      Efficient allocation of the object (letting $x(\theta_1)$ be the probability of assigning the good to 1)

      $$\theta_1 > 1 \Rightarrow x(\theta_1) = 0, \qquad \theta_1 \leq 1 \Rightarrow x(\theta_1) = 1.$$

    - Is there a transfer function that could make this allocation incentive compatible? At most two transfers can be used (transfers must be constant conditional on the allocation). (Why?) Let $t(0)$ be the transfer payment when 1 gets the object, and $t(1)$ the payment that 1 makes when 2 gets the object.

    - Then we have from incentive compatibility:

      $$\theta_1 - t(0) \geq -t(1) \text{ for all } \theta_1 \leq 1,$$

      and

      $$-t(1) \geq \theta_1 - t(0) \text{ for all } \theta_1 > 1.$$

- These inequalities are seen to be incompatible by e.g. summing them up. The efficient decision rule fails to be implementable in this example because monotonicity of the decision rule fails. The reason for this failure can be traced to the fact that player 1's type has a larger effect on the valuation of player 2 than on the valuation of player 1 herself.

3. With multi-dimensional independent but interdependent valuations, efficient allocation rules are almost never incentive compatible.

4. With correlated values (with or without private values), attention is typically focused on the comparison of different indirect mechanisms, e.g. the performance of various types of auctions.

3 Moral Hazard

In contractual arrangements in which the principal offers the contract, we distinguish between cases where informational asymmetries are present at the moment of contracting (adverse selection models) and those where contracting is under symmetric information. The latter case we call moral hazard, and this includes settings where the agent may take an unobservable action after contracting or where the agent may learn new information after signing the contract. In general, complete information at the moment of contracting means that the agent will be kept to reservation utility and the efficiency-extraction of adverse selection models is not present unless the model also includes additional restrictions such as limited liability.

In this chapter, we discuss the basic moral hazard setting. At the contracting stage, neither player has any private information. Hence θ does not appear in the payoff functions. The key informational issue is now that the actions of the agent x (corresponding to the allocations a in the previous chapter) are not observable. Hence it is not possible to commit to x in advance. We assume that the parties observe publicly a signal $y \in Y$ that is statistically related to x. We assume that y is contractible in the sense that there exists a legal enforcement mechanism guaranteeing to contracts based on y. The principal has von Neumann - Morgenstern payoff function $u_P(x, y)$ and the agent has payoff function $u_A(x, y)$. Since x is not observable to the principal, $u_P(x, y)$ cannot depend directly on x. Let $v_P(y)$ be the Bernoulli payoff function to the principal when signal y is observed (including any contractual obligations at y). We have then

$$u_P(x, y) = \int_Y v_P(y) f(y|x)\, dy,$$

Where $f(y|x)$ denotes the density of y when $x \in X$ was chosen by the agent (the idea is the same for finite Y except that we need then a conditional probability mass distribution $\pi(y\,|x)$). Similarly for the agent, we have

$$u_A(x, y) = \int_Y v_A(x, y) f(y|x)\, dy.$$

Since the agent knows her choice of x, her Bernoulli utility function $v_A(x, y)$ may obviously depend on x directly.

By far the most popular application of the Moral Hazard framework is that of the employment relation. The employer, i.e. the principal, needs a costly effort x from a worker, the agent. There may be a conflict of interest between the two parties in terms of the level of effort or direction of attention etc. This is often modeled in crude terms by assuming that the employer cares only about the level output y of the worker whereas the worker does not care about y but wants to minimize effort cost of x. I stress that nothing in what follows depends on assumptions of workers as lazy, employers as greedy etc. The key point is the conflict in the desired effort x in the absence of any contracts.

To align the incentives of the two parties, the principal may promise to pay a wage $t(y)$ based on the publicly observable output y. Ee could have for example the net profit after wages $y - t(y)$ as the variable of interest to the principal, and the vector $(x.t(y))$ determining the payoff of the worker:

$$\begin{aligned} u_P(x,y) &= \int_Y v_P(y - t(y)) f(y|x)\, dy, \\ u_A(x,y) &= \int_Y v_A(x, t(y)) f(y|x)\, dy. \end{aligned}$$

To specialize further, we could have a risk-neutral principal and a risk-averse agent with separable utility in effort and wage. We revert now to conventional notation and write $w(y)$ for the wage (in stead of $t(y)\, m$).

$$\begin{aligned} u_P(x,y) &= \int_Y (y - w(y)) f(y|x)\, dy, \\ u_A(x,y) &= \int_Y u(w(y)) f(y|x)\, dy - c(x). \end{aligned}$$

This is the most common setting in the literature.

Simple Example: Binary Action Choice, Binary Signals

The agent can take action $x \in \{x_l, x_h\}$ at cost $c \in \{c_l, c_h\}$.The outcomes $y \in \{y_l, y_h\}$ occur randomly, where the probabilities are governed by the action as follows.

$$p_l = \Pr(y_h | x_l) < p_h = \Pr(y_h | x_h)$$

The principal offers a wage $w \in \{w_l, w_h\}$ to the agent contingent on whether the outcome is $y = y_l$ or $y = y_h$, and the utility of the agent is

$$u(w_i) - c_i$$

and of the principal it is

$$v_P(y_i - w_i)$$

where we assume that u and v are strictly increasing and weakly concave.

First Best

We consider initially the optimal allocation of risk between the agents in the presence of the risk and with observable actions. If the actions are observable, then the principal can induce the agent to choose the preferred action a^* by

$$w(y_i, x) = -\infty$$

if $x \neq x^*$ for all y_i.

As the outcome is random and the agents have risk-averse preference, the optimal allocation will involve some risk sharing. The optimal solution is characterized by

$$\max_{\{w_l, w_h\}} \{p_i v(y_h - w_h) + (1 - p_i) v(y_l - w_l)\}$$

subject to

$$p_i u(w_h) + (1 - p_i) u(w_l) - c_i \geq U, \ (\lambda)$$

which is the individual rationality constraint. Here we define

$$w(y_i, a^*) \triangleq w_i$$

This is a constrained optimization problem, with Lagrangian:

$$L(w_l, w_h, \lambda) =$$

$$p_i v(y_h - w_h) + (1 - p_i) v(y_l - w_l) + \lambda\{p_i u(w_h) +$$

$$(1 - p_i) u(w_l) - c_i.\}$$

This gives:

$$\frac{V'(y_i - w_i)}{U'(w_i)} = \lambda.$$

This relation is called Borch's rule. It is just the efficient rule for risk sharing between two partners. You should check what happens if one of the parties is risk-neutral.

Second Best

Consider now the case in which the action is unobservable and therefore

$$w(y_i, x) = w(y_i)$$

for all x. Suppose the principal wants to induce high effort, then the *incentive constraint* is:

$$p_h u(w_h) + (1 - p_h) u(w_l) - c_h \geq p_l u(w_h) + (1 - p_l) u(w_l) - c_l$$

or

$$(p_h - p_l)(u(w_h) - u(w_l)) \geq c_h - c_l \tag{7}$$

as $p_l \to p_h$ $w_h - w_l$ must increase, incentives become more high-powered.

The principal also has to respect a *participation constraint* (or *individual rationality* constraint)

$$p_h u(w_h) + (1 - p_h) u(w_l) - c_h \geq U \tag{8}$$

We first show that both constraints are binding at the optimum of the principal's maximization problem

$$\max_{\{w_h, w_l\}} p_h (y_h - w_h) + (1 - p_h)(y_l - w_l)$$

subject to (7) and (8).

For the participation constraint, principal could lower both payments while satisfying the incentive constraint. To see that the incentive constraint holds as an equality suppose to the contrary. Then it is possible to subtract

$$\frac{(1 - p_h)\varepsilon}{u'(w_h)}$$

from w_h and add

$$\frac{p_h \varepsilon}{u'(w_l)}$$

to w_l.

The incentive constraint would still hold for ε sufficiently small. This change in wages subtracts

$$(1 - p_h)\,\varepsilon$$

from $u(w_h)$ if y_h is realized and adds

$$p_h \varepsilon$$

to $u(w_l)$ if y_l is realized so that the expected utility remains constant and hence individual rationality constraint still holds.

The expected wage bill is reduced by

$$\varepsilon p_h (1 - p_h) \left(\frac{1}{u'(w_h)} - \frac{1}{u'(w_l)} \right) > 0,$$

since $w_h > w_l$ and since u is concave contradicting optimality. Hence we conclude that both the incentive constraint and the individual rationality constraint must bind at optimum.

Using the incentive constraint and the individual rationality constraint, we get:

$$u(w_l) = U - \frac{c_h p_l - p_h c_l}{p_h - p_l}$$

and

$$u(w_h) = U - \frac{c_h p_l - p_h c_l}{p_h - p_l} + \frac{c_h - c_l}{p_h - p_l}$$

Since the utility function $u(\cdot)$ is strictly increasing, you can back out the wage payments w_h and w_l from these expressions.

More Generally

Suppose next that effort levels and signals take on a finite number of values. In other words, for some $I, J < \infty$, we have:

$$y_i \in \{y_1, ..., y_I\},$$

and

$$x_j \in \{x_1, ... x_J\},$$

and the probability of observing output y_i if x_j was chosen by the agent is given by:

$$p_{ij} = \Pr\left(y_i \left| x_j \right.\right).$$

The utility function of the agent is $u(w) - x$, and $y - w$ is the payoff of the principal (i.e. output net of wage payments). The principal's problem is then to

$$\max_{\{w_i\}_{i=1}^{I}, x_j} \left\{ \sum_{i=1}^{I} (y_i - w_i)\, p_{ij} \right\}.$$

Given the wage contract the agent selects a_j if and only if

$$\sum_{i=1}^{I} u(w_i)\, p_{ij} - x_j \geq \sum_{i=1}^{I} u(w_i)\, p_{ik} - x_k \quad (\mu_k)$$

and

$$\sum_{i=1}^{I} u(w_i)\, p_{ij} - x_j \geq \underline{U} \ \ (\lambda)$$

Fix x_j, then the Lagrangian is

$$\begin{aligned} \mathcal{L}(w, \lambda, \mu) &= \left\{ \sum_{i=1}^{I} (y_i - w_i)\, p_{ij} \right\} \\ &+ \sum_{k \neq j} \mu_k \left\{ \sum_{i=1}^{I} u(w_i)(p_{ij} - p_{ik}) - (x_j - x_k) \right\} \\ &+ \lambda \left\{ \sum_{i=1}^{I} u(w_i)\, p_{ij} - x_j \right\} \end{aligned}$$

Differentiating with respect to w_i yields

$$\frac{1}{u'(w_i)} = \lambda + \sum_{k \neq j} \mu_k \left(1 - \frac{p_{ik}}{p_{ij}}\right) \text{ for all } i. \tag{9}$$

With a risk-averse principal the condition (9) would simply by modified to:

$$\frac{v'(y_i - w_i)}{u'(w_i)} = \lambda + \sum_{k \neq j} \mu_k \left(1 - \frac{p_{ik}}{p_{ij}}\right), \ \forall i \tag{10}$$

In the absence of an incentive problem, or $\mu_k = 0$, (10) reproduces the Borch rule of optimal risk sharing:

$$\frac{v'(y_i - w_i)}{u'(w_i)} = \lambda, \forall i.$$

The problem with this model is that it is not at all easy to see a priori which incentive constraints are binding. A good guess might be that local incentive compatibility yields global incentive compatibility as in the case of adverse selection. Unfortunately, the results are much weaker in the case of moral hazard models.

As is often the case, having a continuum of outputs and effort levels yields (at least a notational) simplification. The fundamental problems obviously remain the same: when do we know that nearby (local) incentive constraints are the only ones that bind? The conditions (i.e. monotone likelihood ratio property and convexity of the c.d.f.'s of the outputs as a function of x) are sufficient also in the finite case.

3.1 First Order Approach

In this subsection, we outline the model for the case where the agent chooses her action $x \in [\underline{x}, \overline{x}] \subset \mathbb{R}$, and the action results in a random output $y \in [\underline{y}, \overline{y}] \subset \mathbb{R}$. Denote the conditional density of y given effort x by $f(y\,|x)$. Let $F(y\,|x)$ denote the corresponding cumulative distribution function. We assume that the density is strictly positive and differentiable in x at all $y \in [\underline{y}, \overline{y}]$ for all $x \in [\underline{x}, \overline{x}]$. The output is publicly observable, but the chosen action is known only to the agent. Hence the contract offered to the agent can depend only on y. As before, the wage contingent on output is denoted by $w(y)$.

The agent is assumed to have a payoff function that depends on the wage and chosen effort in a separable fashion, i.e. the payoff can be written as $u(w(y)) - c(x)$. The agent has the possibility of refusing the contract altogether which yields her the payoff U_0. The key assumption is that the agent is risk averse, and to this effect, we assume that u is twice continuously differentiable with a non-positive second derivative. The principal is most often assumed to be risk neutral, and hence her payoff is simply $y - w(y)$.

(See Rogerson (1985) for the analysis of the slightly more complicated case where the principal is also risk averse).

The principal's problem is then to find the contract that maximizes her own expected payoff. Alternatively, we may take the principal to choose the wage schedule and the action for the agent subject to the constraint that the chosen action must indeed give the agent the highest possible expected payoff given the contract. In other words, the problem is to

$$\max_{w(y),x} \int_{\underline{y}}^{\overline{y}} (y - w(y))\, f(y\,|x)\, dy$$

subject to

$$x \in \arg\max \int u(w(y)) f(y\,|x)\, dy - c(x)$$

and

$$\int u(w(y)) f(y\,|x)\, dy - c(x) \geq U_0.$$

The first constraint is the incentive compatibility constraint, and the second is the participation constraint or the individual rationality constraint. The first of these is the harder to deal with. The reason for the difficulties is that it is actually not a single constraint, but rather a continuum of inequality constraints. These types of families of constraints are typically not very easy to deal with. As a result much of research has been devoted to study alternative formulations to that constraint. The most obvious candidate would be to replace the incentive compatibility constraint by the first order condition:

$$\int u(w(y)) f_x(y\,|x)\, dy - c'(x) = 0.$$

A Standard method called the *First Order Approach* tackles the problem by solving the principal's problem subject to the individual rationality and the above first order constraint. We get the Lagrangean:

$$\int_{\underline{y}}^{\overline{y}} \left(y - w(y) + \left(\lambda + \mu \frac{f_x(y\,|x)}{f(y\,|x)} \right) u\left(w\left(y\right)\right) \right) f(y\,|x)\, dy$$
$$-\lambda \left(c\left(x\right) + U_0\right) - \mu c'\left(x\right).$$

Maximizing w for each fixed y yields the first order condition of an interior optimal wage $w(y)$ for the principal:

$$\frac{1}{u'(w(y))} = \lambda + \mu \frac{f_x(y|x)}{f(y|x)}.$$

The first assumption that is needed to make any progress is called the *Monotone Likelihood Property* on the distribution of signals conditional states that $\frac{f_x(y|x)}{f(y|x)}$ is increasing in y. This simply requires that larger efforts x induce stochastically larger outputs y. Under this assumption, and under the assumption that u is strictly concave, we see that $w(y)$ must be increasing in y. You can see this by differentiating the above first order condition with respect to w and y.

The question remains though if the first order approach is valid. A recent paper by Rene Kirkegaard (2013) re-establishes old results by Rogerson and Jewitt in a much easier setting. If $F(y|x)$ is convex in x and in the monotone likelihood property holds, then FOA is valid. Furthermore the paper presents additional assumptions on u and the likelihood ratio $\frac{f_x(y|x)}{f(y|x)}$, and $F(y|x)$ that guarantee the validity of FOA. It makes a nice connection to the theory of stochastic dominance that you encountered in the fall part of this course.

Linear contracts with normally distributed performance and exponential utility

This constitutes another "natural" simple case. Performance is assumed to satisfy $y = x + \varepsilon$, where ε is normally distributed with zero mean and variance σ^2. The principal is assumed to be risk neutral, while the agent has a utility function:

$$U(w, x) = -e^{-r(w-c(x))}$$

where r is the (constant) degree of absolute risk aversion ($r = -U''/U'$), and $c(a) = \frac{1}{2}cx^2$.

We restrict attention to linear contracts:

$$w = \phi y + \beta.$$

A principal trying to maximize his expected payoff will solve :

$$\max_{x,\phi,\beta} \; E(y-w)$$

subject to:

$$E(-e^{-r(w-c(x))}) \geq U(\overline{w})$$

and

$$x \in \arg\max_{x} \; E(-e^{-r(w-c(x))})$$

where $U(\overline{w})$ is the default utility level of the agent, and $\overline{w}$ is thus its certain monetary equivalent.

Even though we only cover the simplest one-shot single task problem in these notes, it should be pointed out that the main reason for adopting these functional assumptions is the ease with which the setup generalizes. The two papers by Milgrom and Holmström apply this framework to the study of optimal contracts in a dynamic setting and in a model with multiple tasks. These articles are essential reading for anybody that wishes to study contract theory further.

Certainty Equivalent

The certainty equivalent w of random variable y is defined as follows

$$u(w) = \mathbb{E}\left[u(y)\right]$$

The certainty equivalent of a normally distributed random variable x under CARA preferences, i.e. the w which solves

$$-e^{-rw} = \mathbb{E}\left(-e^{-ry}\right)$$

has a particularly simple form, namely[1]

$$w = \mathbb{E}\left[y\right] - \frac{1}{2} r\sigma^2. \tag{11}$$

The difference between the mean of random variable and its certain equivalent is referred to as the risk premium:

$$\frac{1}{2} r\sigma^2 = \mathbb{E}\left[y\right] - w$$

Rewriting Incentive and Participation Constraints

Maximizing expected utility with respect to a is equivalent to maximizing the certainty equivalent wealth $\widehat{w}(a)$ with respect to a, where $\widehat{w}(a)$ is defined by

$$-e^{-r\widehat{w}(a)} = \mathbb{E}(-e^{-r(w-c(x))}).$$

Hence, the optimization problem of the agent is equivalent to:

$$\begin{aligned} x \in \arg\max \left\{\widehat{w}(x)\right\} = \\ \arg\max \left\{\phi x + \beta - \frac{1}{2} c x^2 - \frac{r}{2}\phi^2\sigma^2\right\}, \end{aligned}$$

which yields

$$x^* = \frac{\phi}{c}.$$

Inserting x^* into the participation constraint

$$\phi\frac{\phi}{c} + \beta - \frac{1}{2} c\left(\frac{\phi}{c}\right)^2 - \frac{r}{2}\phi^2\sigma^2 = \bar{w}$$

[1] To see this recall that a Normal variable with mean μ and variance σ^2 has density $k\left(\sigma^2\right)\exp\left\{-\left(\frac{x-\mu}{\sigma}\right)^2\right\}$. Hence the expected utility of a random variable $\widetilde{w}$ with normal distribution is simply $k\left(\sigma^2\right)\int \exp\left\{-\gamma x\right\}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}dx$

$= k\left(\sigma^2\right)\int \exp\left\{-\gamma x - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}dx$

$= k\left(\sigma^2\right)\int \exp\left\{-\left(\frac{x-\mu+\gamma\sigma^2}{\sigma}\right)^2 - \mu\gamma + \frac{1}{2}\gamma^2\sigma^2\right\}dx$

$= \exp\left\{-\gamma\left(\mu - \frac{1}{2}\gamma\sigma^2\right)\right\}.$

yields an expression for β,

$$\beta = \bar{w} + \frac{r}{2}\phi^2\sigma^2 - \frac{1}{2}\frac{\phi^2}{c}.$$

This gives us the agent's effort for any performance incentive ϕ.

The principal then solves :

$$\max_{\phi} \left\{ \frac{\phi}{c} - (\bar{w} + \frac{r}{2}\phi^2\sigma^2 + \frac{1}{2}\frac{\phi^2}{c}) \right\}.$$

The first order conditions are

$$\frac{1}{c} - (r\phi\sigma^2 + \frac{\phi}{c}) = 0,$$

which yields :

$$\phi^* = \frac{1}{1 + rc\sigma^2}.$$

Effort and the variable compensation component thus go down when c (cost of effort), r (degree of risk aversion), and σ^2 (randomness of performance) go up, which is intuitive. The constant part of the compensation will eventually decrease as well as r, c or σ^2 become large, as

$$\beta = \bar{w} + \left(\frac{1}{2} \frac{r\sigma^2 - \frac{1}{c}}{(1 + rc\sigma^2)^2} \right).$$

3.1.1 Risk-Neutrality and Limited Liability

Here is a variant of the basic Moral Hazard problem, but now with risk-neutrality for all parties. In order to make the problem interesting, we assume limited liability, i.e. that $w(y) \geq 0$ for all y. In the absence of this requirement, all incentive problems could be solved by letting the principal sell the project to the agent at the outset (i.e. set $w(y) = y - \mathbb{E}y$ for all y).

The agent chooses $x \in \mathbb{R}_+$, and $y \in \{0, v\}$. Let $x = \Pr\{y = v|x\}$. Let $c(x)$ denote the increasing convex cost of effort, The outside option value to the agent is normalized to zero. Contracts between the parties take the form of wage payments (positive) contingent on y from principal to agent. It is left as an exercise to show that one of the two possible payments must be zero, and in fact the wage contingent on $\{y = v\}$ must be positive. Let this wage be denoted by w.

The expected payoff to the agent from (x, w) is $xw - c(x)$ and the principal gets $x(v - w)$. The agent chooses $\widehat{x}$ to solve

$$w = c'(\widehat{x}).$$

Since $c(x)$ is assumed convex, the first order condition is sufficient here. The principal's problem can then be written as:

$$\max_x x(v - c'(x)).$$

The first order condition for this problem is

$$v - c'(x^*) - xc''(x^*) = 0..$$

The objective function is concave if $xc'(x)$ is convex (holds for almost any cost function you can think of) and for that case, x^* is the optimum. Socially optimal level of x^{**} solves:

$$v = c'(x^{**}),$$

indicating that $x^* < x^{**}$.

Notice that in this problem individual rationality is not binding because of limited liability, The principal ends up in a tradeoff between rent going to the agent (from the imperfect observability of x) and efficiency of actions much like in the adverse selection model.

3.2 Multi Agent Moral Hazard (Optional)

3.2.1 Deterministic Team Models

One of the best studied forms of collective moral hazard problems comes under the name of partnership problems or team problems. The basic model there is due to early work by Radner and Marschak, and the modern treatment emphasizing the role of incentives is due to Holmström (1982).

N players must take an unobservable effort x_i at private cost $c_i(x_i)$ in order to produce output y. The output is observable and verifiable, but depends on the collective efforts of all the players according to the production function:

$$y = f(x_1, ..., x_N).$$

We assume that f is increasing, concave and differentiable. We will also assume that c_i are twice continuously differentiable, strictly increasing and convex for all i. The partners (or the team members) must design a sharing rule, i.e. the output y must be divided into $(y_1(y), ..., y_N(y))$ in such a way that budget is balanced.

$$y = \Sigma_{i=1}^{N} y_{i.} \tag{12}$$

The preferences of the players are then given by:

$$u_i(y_i, x_i) = y_i - c_i(x_i). \tag{13}$$

The first best level of effort is given by the vector $x^* = (x_1^*, ..., x_N^*)$, where

$$c_i'(x_i^*) = f_i(x^*) \text{ for all } i. \tag{14}$$

The basic question is then if it is possible to design the sharing rule in such a way that all players i have the right incentives to supply effort at the first best level. If the first best level of effort for each player is strictly positive, then this question can be answered in the negative.

To see the basic intuition, observe that at the first best level of effort as determined in 14, the marginal cost of effort is positive. In order to provide player i with the proper incentives, it must be the case that y_i increases by $f_i(x^*)$ units for each marginal increase of a single unit in a_i.

But since

$$\Sigma_j y_j(y(x)) = y(x) \text{ for all } x,$$

we have by differentiation w.r.t. x_i and evaluating at x^* :

$$\Sigma_j y_j'(y) f_i(x^*) = f_i(x^*).$$

Hence if the incentives for i are right, it must be the case that $y_j'(y) = 0$ for all $j \neq i$.But then it is not optimal for any $j \neq i$ to choose x_j^*. Hence it is impossible to provide the right incentives for the players.

This argument gives only the intuition rather than an actual proof since it assumes differentiability of the sharing schemes. Differentiability may play a role in the theory since it is well known that for discontinuous production functions, first best may be achievable. To see how this can make a difference, consider an example where the effort of player 1 is nonproductive, i.e. $f_1(x) = 0$ for all x, and the dependence of player $j's$ effort for $j > 1$ is symmetric.

Then the following scheme achieves the first best. Let $y^* = f(x^*)$.

$$\begin{aligned} y_1(y) &= \begin{cases} 0 \text{ if } y \geq y^*, \\ y \text{ if } y < y^*. \end{cases}, \\ y_j(y) &= \begin{cases} \frac{y}{N-1} \text{ if } y \geq y^*, \\ 0 \text{ if } y < y^*. \end{cases} \text{ for } j > 1. \end{aligned}$$

In this example player 1 becomes effectively the principal for the model, and the incentive scheme is a milder form of the shoot the agents scheme where a collective punishment is imposed when a deviation is observed. Since there is no need to provide 1 with any incentives, this scheme overcomes the inherent difficulties in team problems.

The balanced budget requirement in 12 can be circumvented easily, and in a sense this scheme should be regarded as selling the partnership to an outsider. For the complete argument in the case where $x_i^* > 0$, consider a point $y' < y^*$, where y^* denotes again the first best level of production. Since $f_i(x^*) > 0$ for all i,

$$y_i(y') \leq y_i(y^*) - c_i'(x_i^*) \frac{(y^* - y')}{f_i(x^*)} \text{ for all } i,$$

if y' is sufficiently close to y^*. (Can you supply the proof?)

But then the argument for the differentiable case can be applied to consider the changes from y' to y^*. Observe that the key difference to the example above is that it would be impossible to reward one of the players in the case of a fall in the output since all of the players can induce this fall through their own actions.

3.2.2 Principal and Many Agents

In this section, we maintain much of the machinery from the previous section, but we switch attention back to the principal agent formulation for moral hazard models. Assume now that y depends on x in a stochastic manner. In other words, the vector of efforts x parametrizes the distribution of the output:

$$y \sim F(y\,|x).$$

We assume that all of the distribution functions are sufficiently many times differentiable. Denote by z the vector of stochastic variables whose realizations are contractible and depend on the effort choice x according to:

$$z \sim G(z\,|x).$$

In many models, it will be convenient to assume that y is a subvector of z, but z may also contain other information that is correlated with the effort levels. Since we no longer require budget balance, we switch back to the usual notation on wages $w_i(z)$, and we also reintroduce the possibility for risk aversion for the agents. The problem of the principal is then to solve the following maximization problem:

$$\max_{x,w} \int_x \left(E(y\,|x,z) - \Sigma_{i=1}^N w_i(z)\right) g(z\,|x)\, dz$$

subject to

$$\int_x u_i(w_i(z))\, g(z\,|x)\, dz - c_i(x_i) \geq U_i \text{ for all } i,$$

$$x_i \in \arg\max_{x_i' \in X_i} \int_x u_i(w_i(z))\, g(z\,|x_{-i}, x_i')\, dz - c_i(x_i') \text{ for all } i.$$

The constraints are again the IR and IC constraints for the agents, but it should be noted that each agent takes the effort choice of other agents as given. Hence the assumption is that the principal gets to choose the equilibrium that the agents play, and the analysis is thus in accordance with our analysis in the adverse selection case. It is important to observe that the efforts chosen by one player may make information about other players' choices more accurate. Thus a modification of the single person sufficiency notion is required.

Definition 3 *A statistic $T_i(z)$ is sufficient for z with respect to x_i if there are positive functions ϕ and ζ such that*

$$g(z|x) = \zeta_i(z, x_{-i})\,\phi_i(T_i(z), x) \text{ for all } z \text{ and } x.$$

The following theorem is proved exactly as the corresponding result in the single agent case.

Theorem 8 *If $T_i(z)$ is sufficient for z with respect to x_i for all i, then there for each wage profile $\{w_i(z)\}_{i=1}^N$, there is another wage profile $\{\widetilde{w}_i(z)\}_{i=1}^N$ that Pareto dominates $\{w_i(z)\}_{i=1}^N$.*

The intuition of this result runs as follows. From Blackwell's theorem, we know that a random variable is sufficient for another if and only if the latter is a garbling of the first. In other words, the second must be the first plus noise. As long as the agents are risk averse adding this noise to the compensation package makes the agent strictly worse off. A converse to this theorem is also possible.

Definition 4 *A statistic $T_i(x)$ is globally insufficient for i if for all x, T_i*

$$\frac{g_{x_i}(z|x)}{g(z|x)} \neq \frac{g_{x_i}(z'|x)}{g(z'|x)} \text{ for almost all } z, z' \in \{z\,|T_i(z) = T_i\}.$$

Holmström (1982) proves the following theorem.

Theorem 9 *If $\{w_i(T_i(z))\}$ is a wage profile where $T_i(z)$ is globally insufficient for some i, then there exist wage schedules $\{\widehat{w}_i(z)\}$ that yield a strict Pareto improvement on $\{w_i(T_i(z))\}$ and induce the same action.*

These two theorems are useful in analyzing optimal compensation schemes in environments with many agents. For concreteness, we may think about many salesmen selling a new product in a number of different geographic areas. Suppose that the effort of each salesman results in a level of sales given by

$$y_i(x_i, \theta_i),$$

where θ_i is a random variable that represents the business conditions in the area of salesman i. Assume also that z contains $(y_1, ..., y_N)$. In this case, it is possible for the principal to base individual contracts on individual sales (since the efforts of others do not affect the sales). This will, however, be in general suboptimal.

The reason for this is that if the θ_i are correlated (as is presumably the case for this particular example), y_j is statistically dependent on y_i even if the effort choices are known (as is the case in equilibrium analysis). Hence y_i is globally insufficient for i and the theorem above shows that the compensation scheme can be improved. This observation is generally used as the justification for compensations based on relative performances within organizations.

4 Signaling

The final model to be covered in these notes is the model of signaling. The informational imperfection is similar to the adverse selection model. Two parties are deciding on an economic allocation and one of them has private information. Let $\theta \in \Theta$ denote again the private information. The uninformed party has prior probability (measure) π^0 on Θ. We assume that the two players play a fixed extensive form game. In the first stage, the informed player chooses a message $m \in M$ and in the second stage, the uninformed party chooses and action $x \in X$ after observing m but not θ.

We call the informed party the *sender* since she sends a message to the *receiver*, i.e. the uninformed party. Signaling games are also called *sender-receiver games*. A pure strategy for the sender is a function $\mu : \Theta \to M$, and a pure strategy for the receiver is $\xi : M \to X$. Let $\pi(m)$ denote the posterior probability (measure) of the receiver on Θ after receiving message m. Assuming pure strategies, after any message m, we derive the receiver's beliefs $\pi(m)$ on Θ by using Bayes' rule:

$$\pi(m)(\theta) = \frac{\pi^0(\theta)}{\sum\limits_{\{\theta' \in \Theta : \mu(\theta) = m\}} \pi^0(\theta')}, \text{ if } \sum_{\{\theta' \in \Theta : \mu(\theta) = m\}} \pi^0(\theta') > 0,$$

and $\pi(m)(\theta)$ is arbitrary if $\sum\limits_{\{\theta' \in \Theta : \mu(\theta) = m\}} \pi^0(\theta') = 0$. Here $\pi(m)(\theta)$ denotes the probability that the receiver attaches to type θ of the sender after observing message m.

We allow the utilities of the players to depend on all the variables in the model and write $u_S(m, x, \theta)$ and $u_R(m, x, \theta)$ for the sender and the receiver respectively. We solve for equilibria in the game using the concept of Perfect Bayesian Equilibrium. Hence we require that the receiver acts optimally given her beliefs, and therefore she has no commitment power as in the adverse selection part.

4.1 Cheap Talk

We start by considering a special case of the model where the payoffs do not depend on m. Messages are then best thought of as talk between the parties, hence the name cheap-talk models (so no feelings of guilt etc. from broken promises). Let's start with the easiest possible case where the payoffs of the two parties are perfectly aligned, i.e.

$$u_S(m, x, \theta) = u_R(m, x, \theta) = u(x, \theta).$$

Let's assume that the set of possible messages is rich enough so that a one-to-one function $\zeta : \Theta \to M$ exists. Obviously the easiest way to satisfy this is by having $\Theta \subset M$. Under this assumption, perfect communication is possible between the players If the sender adopts ζ as her strategy in the game, then the receiver can deduce the type θ in an unambiguous manner from each message in the image of Θ under ζ.

Define

$$X^*(\theta) := \arg\max_x u(x, \theta).$$

Also let

$$X^0 = \arg\max_x \int_\Theta u(x, \theta)\, d\pi^0(\theta)$$

It is clear that the following pair of strategies$(\mu(\theta), \xi(m))$ and beliefs $\pi(m)$ forms a Perfect Bayesian Equilibrium $\{\mu(\theta), \xi(m), \pi(m)\}$ of the sender-receiver game for any selection $\xi(\cdot)$ from $X^*(\cdot), X^0$:

$$\begin{aligned}
\mu(\theta) &= \theta,\ \xi(m) \in X^*(m) \text{ for } m \in \Theta,\ \xi(m) \in X^0 \text{ for } m \notin \Theta, \\
\pi(m) &= \delta_m \text{ if } m \in \Theta, \\
\pi(m) &= \pi^0 \text{otherwise,}
\end{aligned}$$

where δ_m denotes point mass on type $\theta = m$. Optimality of all actions on the equilibrium is an immediate consequence of the fact that the receiver's action maximizes both players' full-information payoffs. Clearly this equilibrium satisfies all normative criteria that one might wish for in an equilibrium. Equilibria where different types of senders send different messages are called separating equilibria. Hence the informative equilibrium is an example of separating equilibria.

It is rather annoying that the game has other equilibria as well. The following pair of strategy vectors and beliefs is also a Perfect Bayesian Equilibrium for any fixed message $m^0 \in M$:

$$\begin{aligned} \mu(\theta) &= m^0, \xi(m) \in X^0 \text{ for all } \theta \text{ and all } m, \\ \pi(m) &= \pi^0 \text{ for all } m. \end{aligned}$$

No information is transmitted in this equilibrium and hence it is called the babbling equilibrium. You should prove as an exercise that a babbling equilibrium exists for all cheap talk games, and not just ones with aligned interests. Equilibria where all senders send the same message are called pooling equilibria. Hence the babbling equilibrium is an example of pooling equilibria.

The more interesting results start when we move away from this assumption of aligned interests. In these notes, we do so in a simple parametric manner. We assume that $\Theta = X = [0, 1]$, the prior is uniform on Θ, and that the payoff functions take the particularly simple form:

$$u_S(x, \theta) = -(x - \theta - b)^2, \; u_R(x, \theta) = -(x - \theta)^2.$$

Hence the receiver would always choose $x = \theta$ and we refer to b as the bias of the sender: she would want $x = \theta + b$. The size of b measures the conflict of interest between the two parties.

It is immediately clear that the fully informative equilibrium described above is not an equilibrium in the game with conflicting interests. If the receiver chooses

$$\xi(m) = m \text{ for all } m \in \Theta,$$

then the best response of the sender would be

$$\mu(\theta) = \min\{\theta + b, 1\}.$$

In fact, only a finite number of differentially informative messages are sent in any equilibrium of the game. The statement means that only a finite number of different posteriors are realized on the equilibrium path of the game. For simplicity, we shall denote the messages sent in an equilibrium with k different posteriors by $m_1, m_2, ..., m_k$. We order them in such a way that $x_1 = \xi(m_1) < x_2 \xi(m_2) < ... < x_k \xi = (m_k)$. We start the analysis by showing that

$$\theta < \theta' \Rightarrow \mu(\theta) \leq \mu(\theta')$$

in the order defined above. This implies that the set of types that send message m_i in equilibrium is an interval for all i. Let

$$\mu^*(\theta) = \arg\max_{m \in \{m_1, \ldots, m_k\}} -(\xi(m) - \theta - b)^2.$$

Lemma 2 *The best response by the sender $\mu^*(\theta)$ is increasing in θ in strict set order. Hence any selection from $\mu(\theta)$ is an increasing function.*

Proof. Since $u_S = -(\xi(m) - \theta - b)^2$ satisfies strictly increasing differences in (m, θ), the clam follows by Milgrom Shannon theorem (see chapter 2 of these notes). ■

Hence all equilibria can be characterized by a finite number of cutoffs $\theta_0 = 0,\ \theta_k = 1,\ \theta_l < \theta_{l+1}$. All $\theta \in [\theta_{l-1}, \theta_l)$ send message m_l for $l \in \{1, ..., k\}$ and $\theta = 1$ also sends message m_k. For $l \in \{1, ..., k-1\}$, θ_l must be indifferent between messages m_l and m_{l+1}.Furthermore, x_k must solve

$$\max_x \mathbb{E}_\theta \left[-(x - \theta)^2 \,|\theta_{k-1} \leq \theta \leq \theta_k \right].$$

It is left as an exercise to show that

$$x_k = \frac{\theta_{k-1} + \theta_k}{2}.$$

Using this in the indifference condition for θ_k, we have

$$-\left(\frac{\theta_{k-1} + \theta_k}{2} - \theta_k - b\right)^2 = -\left(\frac{\theta_k + \theta_{k+1}}{2} - \theta_k - b\right)^2.$$

The nontrivial solution to this occurs when

$$\frac{\theta_{k-1} + \theta_k}{2} - \theta_k - b = -\left(\frac{\theta_k + \theta_{k+1}}{2} - \theta_k - b\right)$$

or

$$\theta_{k+1} - \theta_k = \theta_k - \theta_{k-1} + 4b.$$

This is a nice linear difference equation that can be solved with boundary conditions $\theta_k = 1,\ \theta_0 = 0$. Since the distance between consecutive cutoffs grows by $4b$ after each interval, there can be at most finitely many such cutpoints in the unit interval. Hence b determines the extent to which information can be credibly transmitted in this equilibrium. We summarize the findings in the following theorem due to Crawford and Sobel (1982).

Theorem 10 (Crawford and Sobel) *If $b > \frac{1}{4}$ in the quadratic cheap-talk game, then the unique PBE of the game is the babbling equilibrium. The maximal number of differentially informaitove messages is bonded by $\frac{1}{4b}$ and hence fully separating informative equilibria do not exist if $b > 0$.*

This model of cheap-talk has served as the launching pad for a huge literature in i) organization theory: when to delegate and when to communicate? ii) political economy: how can experts influence decision makers.

4.2 Costly Signaling

We cover here a variant of the Spence job market signaling model. The idea is that now messages are costly for the sender but useless for the receiver. To fix ideas, we call the sender the worker, the messages schooling and the receiver is the firm. The firm must make a binary decision: either hire the worker $x = 1$ and pay union wage w (this is a parameter, not a choice variable in the model) or not hire $x = 0$ and pay nothing. The productivity of the worker θ is private information to the worker, and the firm has a uniform prior belief for $\theta \in \{0, 1, 2, ..., k\}$.

Prior to contacting the firm, the worker chooses her level of education $m \in \mathbb{R}_+$. The cost of schooling depends on productivity $c(m, \theta) = \frac{m}{1+\theta}$. The firm sees m but not θ. The firm wants to hire the worker if and only if her expected productivity is above w. Expected productivity is computed using the posterior belief induced by the observed education choice m in a Perfect Bayesian Equilibrium of the game.

The payoff to the worker when hired is

$$u_S(m, x, \theta) = wx - \frac{m}{1+\theta}.$$

Notice that this function satisfies increasing differences in m, θ. This is often stated in the literature by saying that the model satisfies single crossing. The seller's payoff is:

$$u_R(m, x, \theta) = x(\theta - w).$$

Since x is assumed binary, at most two different levels of education will be chosen in equilibrium (show this!). There is always a pooling equilibrium where $m = 0$ is chosen by all worker types. Here is an example of a semi-separating equilibrium In semi-separating (or partially pooling) equilibria some types of senders pool on the same message, but some also send separate messages. All workers with productivity above w choose $m^H = (1 + w)\, w$, all types below w choose $m^L = 0$

It is left as an exercise for you to consider other semi-separating equilibria and other pooling equilibria. Finally for a more challenging exercise, recall the definition of Intuitive Criterion. Consider any out-of-equilibrium message m, i.e. a message that is not sent with positive probability by any θ in a Perfect Bayesian Equilibrium $\{\mu(\theta), \xi(m), \pi(m)\}$. Denote the expected equilibrium payoff of the sender of type θ *in this equilibrium* by $v(\theta)$. Then $\{\mu(\theta), \xi(m), \pi(m)\}$ satisfies the Intuitive Criterion if for all out-of-equilibrium messages m':

$$\begin{aligned} u_S(m', x, \theta) &< v(\theta) \text{ for all } x \in X \text{ and,} \\ u_S(m', x, \theta') &\geq v(\theta') \text{ for some } x \in X, \\ &\Longrightarrow \pi(m')(\theta) = 0. \end{aligned}$$

In words, Intuitive Criterion requires that if an out of equilibrium message m' is worse for some type θ than her equilibrium move regardless of how the receiver acts, and if m' is (weakly) better than the equilibrium message for type θ' for some reaction by the receiver, then the receiver should conclude that the sender is not of type θ. The idea is simply that it would have been better to stick with the original equilibrium message.

The exercise is to determine whether the pooling equilibria survive IC. Suppose that w is so low that the workers are hired in the pooling equilibrium. Do separating equilibria still exist?

Let me end this section by noting that signaling is an example of a concept developed largely in the economic theory that has taken on a life in another field. In biology, signaling gives a nice explanation to initially puzzling phenomena such as the tail of the peacock. The story runs something like this. By having a plush long tail, the peacock is telling the peahen: 'If I can get away from the predators with this big tail, I must be really strong' and hence the mating success of those with long tails that survive to maturity.

5 Further Topics

Any first-year course will miss on some topics that could reasonably be included. I end these notes by discussing briefly a selection of what I would have included next if time had permitted this.

5.1 No Trade Theorem, Market Microstructure

One of the main reasons for having stock exchanges is that trading in equity reveals information that is useful for allocating scarce financial resources to their most productive uses. It comes then as an unpleasant surprise to students of financial markets that trading will not take place for informational reasons.

To state the result somewhat more formally, let's define an exchange economy formally. The economy consists of agents $i \in \{1, ..., I\}$, a finite set of possible types $\Theta = \times_i^I \Theta_i$ with typical element $\theta \in \Theta$, for each i an initial endowment e_i, and payoff function $u_i(x_i(\theta), \theta)$, where $x_i(\theta)$ is the consumption of agent i that depends on the information of all agents in the economy. At the prior stage all this is common knowledge and so is a common prior over the set of type vectors Θ. To rule out trades that are not driven by information flows, assume that the initial allocation of endowments are Pareto efficient given the prior information.

We need to distinguish between various notions of Pareto efficiency. An allocation $x = (x_1, ..., x_I)$ is Pareto efficient if it maximizes the sum of utilities in the economy for some vector of non-negative welfare weights $\lambda_i(\theta)$ on the agents i and their types θ:

$$\sum_{i=1}^{I} \sum_{\theta \in \Theta} \lambda_i(\theta) p(\theta) u_i(x_i(t), \theta).$$

We distinguish between ex ante and interim Pareto efficiency by placing restrictions on the measurability of the weights $\lambda_i(\theta)$. For ex ante efficiency, we restrict $\lambda_i(\theta) = \lambda_i(\theta')$. For interim efficiency, $\lambda_i(\theta_i, \theta_{-i}) = \lambda_i(\theta_i, \theta'_{-i})$. For ex post efficiency we would let $\lambda_i(\theta)$ be completely arbitrary.

From this it is obvious that ex ante efficient allocations are interim efficient. Hence if the initial allocation is ex ante efficient, it is also interim efficient and there cannot be a trade to another $x(t)$ that would be agreed to by all types at the interim stage. This is called the no-trade theorem in the literature. In the form presented here, it is due to Holmstrom and Myerson (1983) 'Efficient and Durable Decision Rules under Incomplete Information', but the same result is contained in earlier work by Milgrom and Stokey (1982) 'Information, Trade, and Common Knowledge'.

In order to have trading in equilibrium, something has to give. The alternatives are essentially the following: Perhaps the initial allocation is not Pareto efficient. A nice model of this is Duffie, Garleanu and Pedersen (2005) 'Over-the-Counter Markets'. Perhaps there are liquidity traders or traders that trade regardless of their information. Sometimes these are called noise traders. Prominent papers on this are Glosten and Milgrom (1985) 'Bid, Ask and Transaction Prices in a specialist Market with Heterogeneously Informed Traders' in a price taking setting and Kyle (1985) 'Continuous Auctions and Insider Trading' is an example of a strategic model where informed traders consider their impact on prices. Perhaps trades are not common knowledge. The original paper by Milgrom and Stokey (1982) already contains an analysis of possible trades when there is only mutual knowledge of order k about trades . Perhaps there is common knowledge of different priors. A prominent example of this is Morris (1994) 'Trade with Heterogeneous Prior Beliefs and Asymmetric Information'. All these routes have been explored for modeling financial trading.

5.2 Rational Expectations Equilibrium and Game Theory of Large Markets

One of the more successful approaches to general equilibrium under incomplete information is the study or Rational Expectations Equilibria. The starting point is an exchange economy where the agents have private information. A Rational Expectations Equilibrium is a price and an allocation such that given the price markets clear, and the price is measurable with respect to the information generated by individual trades. Agents trades are measurable with respect to their own information and the information contained in the price function.

Notice the funny chicken and egg quality in the definition of REE. Price function reflects information contained in the net trade of agent i. Suppose agent i changes her trading strategy. Since agents are price takers by assumption, price remains the same. It was noticed early on that this can lead to problems with existence. Suppose that agents trade in a revealing manner for price functions that do not reveal agent i's information, but in a non-revealing manner if the price reveals her information. This problem centers on the main shortcoming of general equilibrium theory in general. The process of price formation is not explained, and hence confusions such as the above arise. Nevertheless REE seems to work well in most applications in finance and macroeconomics. Hence a major research has been focused on finding foundations for REE or alternative price setting models for financial markets.

There is a relatively complete literature on private values double auctions (i.e. auctions with potential sellers who possess the good and potential buyers who do not have the good). One of seminal papers demonstrating the nice convergence of double auctions to Walrasian outcomes as the number of bidders gets large is Rustichini, Satterthwaite and Williams (1994) 'Convergence to Efficiency in a Simple Market with Incomplete Information'. 'Efficiency of Large Double Auctions' by Cripps and Swinkels (2006) allows for correlated valuations, multiple unit demands and supplies. The crucial maintained assumption is that of private values.

There is also a rather well developed theory for common value auctions with fixed supply. Kremer (2002) 'Information Aggregation in Common Value Auctions' is the best source on this. The theory of double auctions with common values is much harder. See (if you dare) 'Toward a Strategic Foundation for Rational Expectations Equilibrium' by Reny and Perry (2006) for first steps in this direction.

5.3 Competition in Contracts

The first Chapter on Adverse Selection and Mechanism Design was conducted under the assumption that there is a single mechanism designer or principal in the model. A natural thought would be to relax this monopoly assumption and allow for multiple principals. A first question then is what the offered contracts should look like. For each mechanism designer, the space of relevant details regarding the agents has become larger. The preferences of the agent are formed partially by the contract offers made by the other principals. Shouldn't each principal k then be allowed to offer contracts conditioning on the offers of principal l? But since l's offers condition on k's offers, we get into a bizarre circular situation. To get a glimpse of the difficulties encountered, take a look at Epstein and Peters (1999) 'A Revelation Principle for Competing Mechanisms' and 'Definable and Contractible Contracts' by Peters and Szentes (2012).

Another route to take is to step back from the Revelation Principle and require that the principals compete in menus of price-allocation pairs. This is the approach in the seminal article by Rothschild and Stiglitz (1976) 'Equilibrium in Competitive Insurance Markets: An Essay in the Economics of Incomplete Information' that demonstrated non-existence of equilibrium when the contracts are exclusive in the sense that each agent can choose at most one offer.

Here is a quick idea of how the model works. It is given in the context of an insurance market. Two (or more) firms are selling insurance contracts to a single buyer with private information about her type. She can be either high risk θ^H or low risk θ^L. In case of no accident, the agent has endowment y, and in case of accident, her endowment is $y-d$. We let θ^i be the probability of accident for $i \in \{H, L\}$. The prior probability of a high-risk type is p.

The firms offer contracts that promise to pay x in case of an accident in exchange for an up-front fee t. Each firm can offer as many of such contracts as they like without any cost. By usual undercutting arguments, this leads to the zero profit for all firms in the market in any equilibrium. The agent is risk averse:

$$u(x, t, \theta) = \theta v(y - d + x - t) + (1 - \theta) v(y - d - t),$$

where v is a strictly concave function. The firms are assumed risk-neutral. If θ was public information, the equilibrium contracts would feature full information for both types at actuarily fair prices.

$$\left(x^i, t^i\right) = \left(d, \theta^i d\right).$$

Obviously this is not incentive compatible if θ is private information. The high-risk agent would claim to be low-risk to reduce the premium t. Consider then a pooling contract where the same amount of insurance x is sold to both types at the actuarily fair pooling price

$$t = \left(p\theta^H + (1-p)\,\theta^L\right) x.$$

Could this be an equilibrium? If this is the only contract offered in the market (or the best available contract for all agents), a firm could try to come up with another contract that is attractive to the low-risk agents, but not the high-risk ones. Since the marginal rates of substitution between the cases of no accident and accident are different for the two different types of agents at any pooling contract, it is always possible to make an offer that attracts only the low-risk agent and increases the firm's profit (show this as an exercise).

In any separating equilibrium, full insurance is sold to the high-risk type. Low risk type gets partial insurance $x^L < d$ at price $\theta^L x^L$. The level x^L is fixed from the high-risk types incentive compatibility condition. The most striking observation that Rothschild and Stiglitz made was that in some cases, this separating contract leads to a profitable pooling offer. The details are left as a somewhat challenging exercise.

Current literature on the issue has taken two routes. Attar, Mariotti and Salanie (2012,2013) have relaxed the assumption that contracts are exclusive. They allow instead that agents buy partial insurance from multiple sources. This leads to nicer characterizations of the equilibria. Another route that has been proposed is a switch towards informed principal models. Maskin and Tirole (1992) 'The Principal-Agent Relationship with an Informed Principal, II: Common Values' is a good source for this.

5.4 Lack of Commitment and Hold-Up

We conclude these notes by demonstrating two types of problems that arise when contracting abilities are limited. The first example shows what happens when some publicly observable variables are not contractible in the sense that contracts based on those variables cannot be enforced. The second two examples demonstrate problems that arise if commitment to future contracts is not possible. Even though the first two examples in their simplest forms do not have informational problems, many of their generalizations do. All of the examples covered here have given rise to a sizable literature.

Example 1 (Hold-Up) *A buyer and a seller contract on the term of a sale. The quality of the good for the agent is* x *where* x *is an investment into quality made by the seller at private cost* $c(x) = \frac{1}{2}x^2$*. The investment level is fully observable, but it is not possible to write contracts of the form* $t(x)$ *since the level of* x *is not verifiable in court. The cost of investment is sunk.*

After the investment, the two parties bargain over the sales price. The key assumption is that the seller does not have full bargaining power. Assume for concreteness that both parties have equal bargaining power. The resulting transfer from the buyer to the seller is then $t(x) = \frac{1}{2}x$*, The seller foresees this outcome at the investment stage and hence invests* $x^{eq} = \frac{1}{2}$ *even though the socially efficient level would be* $x^* = 1$*.*

Example 2 (Soft Budget Constraint) *The issue here is commitment in moral hazard settings. Suppose an Entrepreneur needs* 1 *unit of funds to operate. A Bank may either grant this funding or not. After getting the money, the Entrepreneur may either exert costly effort that results in a return* $R > 1$ *to the Bank or not exert effort which results in* $r < 1$*. After observing the effort, the Bank may either allow the project to continue until the return is collected or liquidate the project early resulting in* λr *or* λR *with* $\lambda < 1$ *depending on effort choice to the Entrepreneur. The payoff to the entrepreneur is* $v - c$ *if the project continues until the end and* $-c$ *if it is liquidated. Assume that* $v > c$*.*

If B *could commit to a contract, she would grant the funding and continue if and only if* E *exerted effort. In equilibrium, this would result in granting the funding, effort exerted and continuation of the project, i.e. the efficient outcome.*

Without commitment, the game is solved by backward induction. Continuing is a dominant strategy for B *therefore* E *never exerts effort and as a result, no funding is granted.*

To see how problems of commitment manifest in the standard Moral Hazard model, it is a good exercise to consider a two-period extension of the model with risk-neutral principal and agent with say quadratic cost of effort in each period and the assumption that the game ends when the first success arrives. You should show that best contracts (with or without commitment) specify a wage conditional on success for each period. You should then solve for the optimal (w^0, w^1) when w^1 is set optimally in period 1 i.e. without commitment. Then you should solve for $(\widehat{w}^0, \widehat{w}^1)$ under commitment to $\widehat{w}^1$. How do w^1 and $\widehat{w}^1$ compare?

The next example highlights the importance of commitment in adverse selection problems.

Example 3 (Coase Conjecture) *Consider now the adverse selection model where a monopolist seller sells an indivisible good to a single buyer. Assume that the buyers valuation for the good θ is uniformly distributed on $[0, 1]$. The optimal sales mechanism is to sell at fixed price $\frac{1}{2}$.*

An obvious question is what happens if the buyer refuses to buy at that price. In the Myerson optimal mechanism the answer is: nothing. There is commitment to the mechanism, and it is a feature, not bug of the mechanism that it is sometimes inefficient. One might nevertheless question the commitment power in some applications and hence it is a useful exercise to see what happens if after any rejection, the seller makes a new price offer. (Actually it takes some work to show that this is the form that the optimal sequential mechanism will take). It is a good exercise to compute the equilibrium with two periods paying particular attention to the fact that the buyer may now wait in order to get a lower price tomorrow.

Coase conjecture states that as the number of periods becomes large, and as the parties become arbitrarily patient (or time interval between the periods shrinks), in the limit of the unique stationary equilibrium of the model, all prices converge to the lowest valuation, i.e. 0 here.

6 Bibliography

Adverse Selection and Mechanism Design

Mas-Colell, Whinston and Green (1995): Chapter 13, 23.

Salanie (1997): The Economics of Contracts. MIT Press. Chapter 2-3.

Paul Milgrom (2004) Putting Auction Theory to Work. Cambridge University Press.

Vijay Krishna (2010) Auction Theory. Academic Press.

Tilman Borgers (2014) An Introduction to the Theory of Mechanism Design. Book Manuscript. Available at: www-personal.umich.edu/~tborgers/TheoryOfMechanismDesign.pdf

Juuso Välimäki: Lecture Notes on Mechanism Design.

Moral Hazard

Mas-Colell, Whinston and Green (1995): Chapter 14.

Salanie (1997): The Economics of Contracts. MIT Press. Chapter 4.

William Rogerson (1985): “The First-Order Approach to Principal-Agent Problems” Econometrica , Vol. 53, No. 6 (Nov., 1985) , pp. 1357-1367.

Rene Kirkegaard (2013): “Local Incentive Compatibility in Moral Hazard Problems: A Unifying Approach”, May 2013. Available at: http://www.uoguelph.ca/~rkirkega/LocalIC.pdf

Signaling

Mas-Colell, Whinston and Green (1995): Chapter 13.

Salanie (1997): The Economics of Contracts. MIT Press. Chapter 5.

Crawford and Sobel (1982): “Strategic Information Transmission”, Econometrica, Vol. 50, No. 6 (Nov., 1982), pp. 1431-1451.