2 Adverse Selection: General Theory

We use the term "adverse selection" when a characteristic of the Agent is imperfectly observed by the Principal.[1] This term comes from a phenomenon well known to insurers: If a company offers a rate tailored only to the average-risk population, this rate will attract only the high risk population, and the company will therefore lose money. This effect may even induce the insurer to deny insurance to some risk groups. Other terms sometimes used are "self-selection" and "screening." The general idea of adverse selection can be grasped from the following example, which will be analyzed fully in section 2.2.

Suppose that the Principal is a wine seller and the Agent a buyer. The Agent may have cultivated tastes for good wines or have more modest tastes. We will say there are two "types": the sophisticated Agent who is ready to pay a high price for good vintage and the frugal Agent whose tastes—or means—may be less developed.

We can assume that the Principal cannot observe the type of any given Agent, or at least that the law (as is often the case) forbids him to use nonanonymous prices that discriminate between the two types.[2]

The key to the solution of the adverse selection problem is the following observation: if the sophisticated Agent is willing to pay more

1. This chapter and the next chapter develop the Principal–Agent paradigm introduced in section 1.2.

2. In Pigou's terms, first-degree price discrimination is infeasible besides being illegal.

than the frugal Agent for a given increase in the quality of the wine, then the Principal can segment the market by offering two different wine bottles:

$$\begin{cases} \text{a wine of high quality for a high price} \\ \text{a wine of lower quality for a lower price} \end{cases}$$

We will see in section 2.2 how these qualities and prices can be chosen optimally.

If all goes according to plan, the sophisticated type will choose the top high-priced wine, while the frugal type will pick a lower quality bottle. Thus the two types of Agent "reveal themselves" through their choices of wine. As we will see, this implies that the frugal type buys a lower quality than might be socially optimal. The whole point of adverse selection problems is to make the Agents reveal their type without incurring too high a social distortion.

Let us briefly consider a few other relevant examples of adverse selection.

- In life insurance, the insured's state of health (and therefore risk of dying soon) is not known to the insurer, even if the insured has had a medical checkup. As a result the insurer is better off offering several insurance packages, each tailored to a specific risk class. (This situation will be studied in section 3.1.3.)

- In banking, the borrowers' default risk can be only imperfectly assessed, in particular, where entrepreneurs request financing for risky projects. A natural idea is to use interest rates to discriminate among entrepreneurs. However, this may induce credit rationing, unless banks also vary collateral levels.[3]

- In labor markets, potential workers have an informational advantage over employers in that they know their innate abilities better.

3. This is an admittedly very brief summary of a body of literature that started with Stiglitz-Weiss (1981).

Because of this firms must screen workers to select the promising candidate and reject all others.

- In government-regulated firms (state-owned or not), the regulated firm has better information on its costs or productivity than the regulator. The obvious implication is that it can manipulate the way it discloses information to the regulator to maximize its profits (see section 3.1.1).

*2.1 Mechanism Design

Mechanism design is at the root of the study of adverse selection. Mechanism design is so important to adverse selection models that some authors also call these models mechanism design problems. I will not attempt here to give a self-contained presentation of mechanism design. I will assume that the reader has already been exposed to this theory. My sole aim will be to remind the reader of the general formalistic properties and the results that will be needed later in the book.[4] The reader who finds this section too abstract can skip it without losing the thread of the chapter.

The object of mechanism design theory is to explore the means of implementing a given allocation of available resources when the relevant information is dispersed in the economy. Take, for instance, a social choice problem where each agent $i = 1, \ldots, n$ has some relevant private information θ_i. Assume that despite all the reservations exemplified by Arrow's theorem, society has decided that the optimal allocation is

$$y(\theta) = (y_1(\theta_1, \ldots, \theta_n), \ldots, y_n(\theta_1, \ldots, \theta_n))$$

Presumably it is be easy to implement the allocation if the government knows all the θ_i's. However, if only i knows his θ_i and, say, his

4. See Laffont (1989) or Moore (1992) for a more complete exposition.

optimal allocation $y_i(\theta)$ increases with θ_i, he is likely to overstate his θ_i so as to obtain a larger allocation. This can make it very difficult for the government to implement $y(\theta)$.

The provision of public goods is another example. Everyone benefits from a bridge, but no one particularly cares to contribute to its building costs. The optimal financing scheme presumably depends on each agent's potential use of the bridge: for example, commuters heavily using the bridge might be asked to pay more than infrequent commuter types. In the absence of a reliable way to differentiate between these individuals, the government will have to rely on voluntary declarations. Naturally, to avoid bearing a large portion of the cost, the heavy user type of Agent will understate the utility he derives from the bridge. As a result the bridge may not be built, as its cost may exceed the reported benefits.

As a final example, consider the implementation of a Walrasian equilibrium in an exchange economy. We all know that this has good properties under the usual assumptions. However, it is not clear how the economy can move to a Walrasian equilibrium. If information were publicly available, the government could just compute the equilibrium and give all consumers their equilibrium allocations.[5] In practice, the agents' utility functions (or their true demand functions) are their private information, and they can be expected to lie so as to maximize their utility. As information is dispersed throughout the economy, implementable allocations are subject to a large number of incentive constraints.

In all these examples, two related questions arise:

Can $y(\theta)$ be implemented? In other words, is it incentive compatible (some authors say "feasible")? What is the optimal choice among incentive compatible allocations?

In more abstract terms we consider a situation where

5. This was the original vision of the proponents of market socialism.

- there are n agents $i = 1, \ldots, n$ characterized by parameters $\theta_i \in \Theta_i$, which are their private information and are often called their "types";
- agents are facing a "Center" whose aim is to implement a given allocation of resources, and generally (which is the more interesting case) this allocation will depend on the agents' private characteristics θ_i.

Think of the Center as government, or as some economic agent who has been given the responsibility of implementing an allocation, or even as an abstract entity such as the Walrasian auctioneer. The Center needn't be a benevolent dictator; he may be, for instance, the seller of a good who wants to extract as much surplus as possible from agents whose valuations for the good he cannot observe.

2.1.1 General Mechanisms

The problem facing the Center is an incentive problem. The Center must try to extract information from the Agents so that he can implement the right allocation. To do this, he may resort to very complicated procedures, using bribes to urge the Agents to reveal some of their private information. This process, however complicated, can be summed up by a *mechanism* $(y(.), M_1, \ldots, M_n)$. This consists of a message space M_i for each Agent i and a function $y(.)$ from $M_1 \times \ldots \times M_n$ to the set of feasible allocations. The allocation rule $y(.) = (y_1(.), \ldots, y_n(.))$ determines the allocations of all n Agents as a function of the messages they send to the Center.[6] Note that generally these allocations are vectors.

Given an allocation rule $y(.)$, the Agents play a message game in which the message spaces M_i are their strategy sets and the allocation rule $y(.)$ determines their allocations and therefore their utility

6. In general, the mechanism involves stochastic allocation rules. Here we will assume that they are deterministic.

levels. Agent i then chooses a message m_i in M and sends it to the Center, who imposes the allocation $y(m_1, \ldots, m_n)$.

Note that in general, the message chosen by Agent i will depend on his information I_i, which contains his characteristic θ_i. The Agent's information may in fact be richer, as is the case where each Agent knows the characteristic of some of his neighbors. Equilibrium messages thus will be functions $m_i^*(I_i)$ and the implemented allocation will be

$$y^*(I_1, \ldots, I_n) = y(m_1^*(I_1), \ldots, m_n^*(I_n))$$

Assume, for instance, that the Center is the proverbial Walrasian auctioneer and tries to implement a Walrasian equilibrium in a context where he does not know the Agents' preferences. Then one way for him to proceed is to ask the agents for their demand functions, to compute the corresponding equilibrium, and to give each agent his equilibrium allocation. If he is the builder of a bridge, he might announce a rule stating under which conditions he will decide to build the bridge and how it will be financed; then he would ask each Agent for his willingness to pay.

2.1.2 Application to Adverse Selection Models

The models we are concerned with in this chapter are very special and simple instances of mechanism design. The Principal here is the Center, and only one Agent is involved. Thus $n = 1$, and the information I of the Agent boils down to his type θ. Given a mechanism $(y(.), M)$, the Agent chooses the message he sends so as to maximize his utility $u(y, \theta)$:

$$m^*(\theta) \in \arg\max_{m \in M} u(y(m), \theta)$$

and he obtains the corresponding allocation

$$y^*(\theta) = y(m^*(\theta))$$

The revelation principle below[7] implies that one can confine attention to mechanisms that are both *direct* (where the Agent reports his information) and *truthful* (so that the Agent finds it optimal to announce the true value of his information).

Revelation Principle.
If the allocation $y^*(\theta)$ can be implemented through some mechanism, then it can also be implemented through a direct truthful mechanism where the Agent reveals his information θ.

The proof of this result is elementary. Let $(y(.), M)$ be a mechanism that implements the allocation y^*, and let $m^*(\theta)$ be the equilibrium message, so that $y^* = y \circ m^*$. Now consider the direct mechanism $(y^*(.), \Theta)$. If it were not truthful, then an Agent would prefer to announce some θ' rather than his true type θ. So we would have

$$u(y^*(\theta), \theta) < u(y^*(\theta'), \theta)$$

But, by the definition of y^*, this would imply that

$$u(y(m^*(\theta)), \theta) < u(y(m^*(\theta')), \theta)$$

Consequently m^* cannot be an equilibrium in a game generated by the mechanism $(y(.), M)$, since the Agent of type θ prefers to announce $m^*(\theta')$ rather than $m^*(\theta)$. Thus the direct mechanism (y^*, Θ) must be truthful, and by construction, it implements the allocation y^*.

Note that in a direct mechanism the message space of the Agent coincides with his type space. Thus in the example of the bridge, the Agent needs only to announce his willingness to pay.

Assume that as is often the case, the allocation y consists of an allocation q and a monetary transfer p. The revelation principle states that to implement the quantity allocation $q(\theta)$ using transfers

7. I only state this principle for the case where $n = 1$. It is valid more generally, but the shape it takes depends on the equilibrium concept used for the message-sending game among the n agents. These complications do not concern us here.

$p(\theta)$, it is enough to offer the Agent a menu of contracts. If the Agent announces that his type is θ, he will receive the allocation $q(\theta)$ and will pay the transfer $p(\theta)$.

Direct truthful mechanisms are very simple but rely on messages that are not explicit. In the example of the wine seller, one can hardly expect the buyer to come into the shop and declare "I am sophisticated" or "I am frugal." A second result sometimes called the *taxation principle* comes to our aid in showing that these mechanisms are equivalent to a nonlinear tariff $\tau(.)$ that lets the Agent choose an allocation q and pay a corresponding transfer $p = \tau(q)$. The proof of this principle again is simple. Let there be two types θ and θ' such that $q(\theta) = q(\theta')$; if $p(\theta)$ is larger than $p(\theta')$, then the Agent of type θ can pretend to be of type θ', and the mechanism will not be truthful. Therefore we must have $p(\theta) = p(\theta')$, and the function $\tau(.)$ is defined unambiguously by

$$\text{if} \quad q = q(\theta), \quad \text{then} \quad \tau(q) = p(\theta)$$

In our earlier example the wine seller only needs to offer the buyer two wine bottles that are differentiated by their quality and price. This is, of course, more realistic; although most retailers do not post a nonlinear tariff on their doors, they often use a system of rebates that approximates a nonlinear tariff.

2.2 A Discrete Model of Price Discrimination

In section 2.3, we will obtain the general solution for the standard adverse selection model with a continuous set of types. Here we learn first to derive the optimum in a simple two-type model by way of heavily graphical techniques and very simple arguments.

To simplify things, we will reuse the example of a wine seller who offers wines of different qualities (and at different prices) in order to segment a market in which consumers' tastes differ. This is therefore a model that exhibits both vertical differentiation and second-degree price discrimination.[8]

2.2.1 The Consumer

Let the Agent be a moderate drinker who plans to buy at most one bottle of wine within the period we study. His utility is $U = \theta q - t$, where q is the quality he buys and θ is a positive parameter that indexes his taste for quality. If he decides not to buy any wine, his utility is just 0.

Note that with this specification,

$$\forall \theta' > \theta, \quad u(q, \theta') - u(q, \theta) \quad \text{increases in } q$$

This is the discrete form of what I call the Spence-Mirrlees condition in section 2.3. For now, just note its economic significance: At any given quality level, the more sophisticated consumers are willing to pay more than the frugal consumers for the same increase in quality. This is what gives us the hope that we will be able to segment the market on quality.

There are two possible values for θ: $\theta_1 < \theta_2$; the prior probability that the Agent is of type 1 (or the proportion of types 1 in the population) is π. In the following, I will call "sophisticated" the consumers of type 2 and "frugal" the consumers of type 1.

2.2.2 The Seller

The Principal is a local monopolist in the wine market. He can produce wine of any quality $q \in (0, \infty)$; the production of a bottle of good quality q costs him $C(q)$. I will assume that C is twice differentiable and strictly convex, that $C'(0) = 0$ and $C'(\infty) = \infty$.

---

8. The classic reference for this model is Mussa-Rosen (1978), who use a continuous set of types.

The utility of the Principal is just the difference between his receipts and his costs, or $t - C(q)$.

2.2.3 The First-Best: Perfect Discrimination

If the producer can observe the type θ_i of the consumer, he will solve the following program:

$$\max_{q_i, t_i} (t_i - C(q_i))$$

$$\theta_i q_i - t_i \geq 0$$

The producer will therefore offer $q_i = q_i^*$ such that $C'(q_i^*) = \theta_i$ and $t_i^* = \theta_i q_i^*$ to the consumer of type θ_i, thus extracting all his surplus; the consumer will be left with zero utility.

Figure 2.1 represents the two first-best contracts in the plane (q, t). The two lines shown are the indifference lines corresponding to zero utility for the two types of Agent. The curves tangent to them are isoprofit curves, with equation $t = C(q) + K$. Their convexity is a consequence of our assumptions on the function C. Note that the utility of the Agent increases when going southeast, while the profit of the Principal increases when going northwest.

Both q_1^* and q_2^* are the "efficient qualities." Since $\theta_1 < \theta_2$ and C' is increasing, we get $q_2^* > q_1^*$, and the sophisticated consumer buys a higher quality wine than the frugal consumer. This type of discrimination, called first-degree price discrimination, is generally forbidden by the law, according to which the sale should be anonymous: You cannot refuse a consumer the same deal you prepared for another consumer.[9] However, we are interested in the case

9. As we will see shortly, the sophisticated consumer envies the frugal consumer's deal.

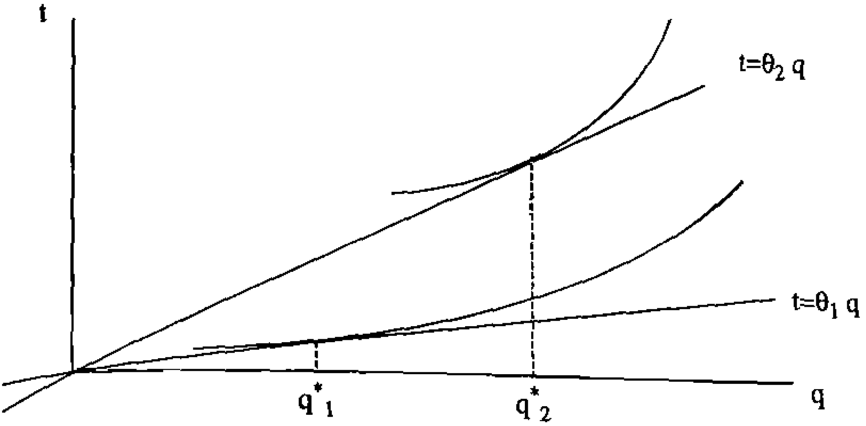

Figure 2.1
The first-best contracts

where the seller cannot observe directly the consumer's type. In this case perfect discrimination is infeasible no matter what is its legal status.

2.2.4 Imperfect Information

Now in the second-best situation in information is asymmetric. The producer now only knows that the proportion of frugal consumers is π. If he proposes the first-best contracts (q_1^*, t_1^*), (q_2^*, t_2^*), the sophisticated consumers will not choose (q_2^*, t_2^*) but (q_1^*, t_1^*), since

$$\theta_2 q_1^* - t_1^* = (\theta_2 - \theta_1) q_1^* > 0 = \theta_2 q_2^* - t_2^*$$

The two types cannot be treated separately any more. Both will choose the low quality deal (q_1^*, t_1^*).

Of course, the producer can get higher profits by proposing (q_1^*, t_1^*) the point designated A in figure 2.2, since A will be chosen only by the sophisticates and only by them. Note that A is located on a higher isoprofit curve than (q_1^*, t_1^*), and therefore it gives a higher profit to the seller.

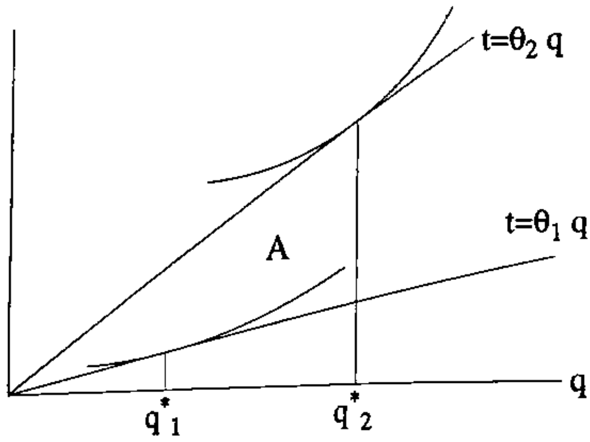

Figure 2.2
A potentially improving contract

A number of other contracts are better than A. Our interest is in the best pair of contracts (the second-best optimum). This is obtained by solving the following program:

$$\max_{t_1,q_1,t_2,q_2} \{\pi[t_1 - C(q_1)] + (1-\pi)[t_2 - C(q_2)]\}$$

subject to

$$\begin{cases} \theta_1 q_1 - t_1 \geq \theta_1 q_2 - t_2 & (IC_1) \\ \theta_2 q_2 - t_2 \geq \theta_2 q_1 - t_1 & (IC_2) \\ \theta_1 q_1 - t_1 \geq 0 & (IR_1) \\ \theta_2 q_2 - t_2 \geq 0 & (IR_2) \end{cases}$$

The constraints in this program are identified as follows:

• The two (IC) constraints are the *incentive compatibility* constraints; they state that each consumer prefers the contract that was designed for him.

• The two (IR) constraints are the *individual rationality*, or *participation* constraints; they guarantee that each type of consumer accepts his designated contract.

We will prove that at the optimum:

1. (IR_1) is active, so $t_1 = \theta_1 q_1$.
2. (IC_2) is active, whence

$$t_2 - t_1 = \theta_2(q_2 - q_1).$$

3. $q_2 \geq q_1$.
4. (IC_1) and (IR_2) can be neglected.
5. Sophisticated consumers buy the efficient quality

$$q_2 = q_2^*.$$

Proofs We use (IC_2) to prove property 1:

$$\theta_2 q_2 - t_2 \geq \theta_2 q_1 - t_1 \geq \theta_1 q_1 - t_1$$

since $q_1 \geq 0$ and $\theta_2 > \theta_1$. If (IR_1) was inactive, so would be (IR_2), and we could increase t_1 and t_2 by the same amount. This would increase the Principal's profit without any effect on incentive compatibility.

Property 2 is proved by assuming that (IC_2) is inactive. Then

$$\theta_2 q_2 - t_2 > \theta_2 q_1 - t_1 \geq \theta_1 q_1 - t_1 = 0$$

We can therefore augment t_2 without breaking incentive compatibility or the individual rationality constraint (IR_2). This obviously increases the Principal's profit, and therefore the original mechanism cannot be optimal.

To prove property 3, let us add (IC_1) and (IC_2). The transfers t_i cancel out, and we get

$$\theta_2(q_2 - q_1) \geq \theta_1(q_2 - q_1)$$

and

$$q_2 - q_1 \geq 0$$

since $\theta_2 > \theta_1$.

By property 4, the (IC_1) can be neglected, since (IC_2) is active. By property 3,

$$t_2 - t_1 = \theta_2(q_2 - q_1) \geq \theta_1(q_2 - q_1)$$

The proof of assertion 1 shows that (IR_2) can be neglected.

Finally, by property 5, we can prove that $C'(q_2) = \theta_2$. If $C'(q_2) < \theta_2$, for instance, let ε be a small positive number, and consider the new mechanism (q_1, t_1), $(q_2' = q_2 + \varepsilon, t_2' = t_2 + \varepsilon\theta_2)$. It is easily seen that

$$\theta_2 q_2' - t_2' = \theta_2 q_2 - t_2 \text{ and } \theta_1 q_2' - t_2' = \theta_1 q_2 - t_2 - \varepsilon(\theta_2 - \theta_1)$$

so the new mechanism satisfies all four constraints. Moreover

$$t_2' - C(q_2') = t_2 - C(q_2) + \varepsilon(\theta_2 - C'(q_2))$$

This tells us that the new mechanism yields higher profits than the original one, which is absurd. We can prove in the same way that $C'(q_2) > \theta_2$ is impossible (just change the sign of ε).

It is an easy and useful exercise to obtain graphical proofs of these five points. The optimal pair of contracts appears to be located as shown in figure 2.3. (q_1, t_1) is on the zero utility indifference line of the Agent of type 1, and (q_2, t_2) is the tangency point between an iso-

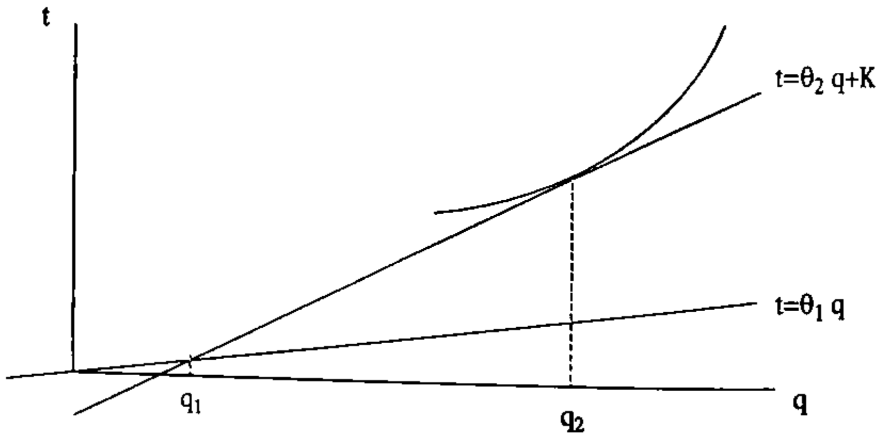

Figure 2.3
The second-best optimum

profit curve of the seller and the indifference line of the Agent of type 2 that goes through (q_1, t_1).

To fully characterize the optimal pair of contracts, we just have to let (q_1, t_1) in figure 2.3 slide on the line $t_1 = \theta_1 q_1$. Formally the optimum is obtained by replacing q_2 with q_2^* and expressing the values of t_1 and t_2 as functions of q_1, using

$$\begin{cases} t_1 = \theta_1 q_1 \\ t_2 - t_1 = \theta_2(q_2 - q_1) \end{cases}$$

This gives

$$\begin{cases} q_2 = q_2^* \\ t_1 = \theta_1 q_1 \\ t_2 = \theta_1 q_1 + \theta_2(q_2^* - q_1) \end{cases}$$

We can substitute these values in the expression of the Principal's profit and solve

$$\max_{q_1} \left(\pi(\theta_1 q_1 - C(q_1)) - (1 - \pi)(\theta_2 - \theta_1)q_1\right)$$

Note that the objective of this program consists of two terms. The first term is proportional to the social surplus[10] on type 1 and the second represents the effect on incentive constraints on the seller's objective. Dividing by π, we see that the Principal should maximize

$$(\theta_1 q_1 - C(q_1)) - \frac{1 - \pi}{\pi}(\theta_2 - \theta_1)q_1$$

which we can call the *virtual surplus*. We will see a similar formula in section 2.3. The difference between the social surplus and the virtual surplus comes from the fact that when the Principal increases q_1, he makes the type 1 package more alluring to type 2. To prevent type 2

10. The social surplus is the sum of the objectives of the Principal and the type 1 Agent. We do not have to worry about the social surplus derived from selling to Agent 2, since we know that we implement the first-best $q_2 = q_2^*$.

from choosing the contract designated for type 1, he must therefore reduce t_2, which decreases his own profits.

We finally get

$$C'(q_1) = \theta_1 - \frac{1-\pi}{\pi}(\theta_2 - \theta_1) < \theta_1$$

so that $q_1 < q_1^*$: the quality sold to the frugal consumers is subefficient.[11]

The optimal mechanism has five properties that are common to all discrete-type models and can usually be taken for granted, thus making the resolution of the model much easier:

- The highest type gets an efficient allocation.
- Each type but the lowest is indifferent between his contract and that of the immediately lower type.
- All types but the lowest type get a positive surplus: their *informational rent,* which increases with their type.
- All types but the highest type get a subefficient allocation.
- The lowest type gets zero surplus.

Informational rent is a central concept in adverse selection models. The Agent of type 2 gets it because he can always pretend his type is 1, consume quality q_1, pay the price t_1, and thus get utility

$$\theta_2 q_1 - t_1$$

which is positive. However, type 1 cannot gain anything by pretending to be type 2, since this nets him utility

$$\theta_1 q_2 - t_2$$

11. If the number of frugal consumers π is low, the formula will give a negative $C'(q_1)$. Then it is optimal for the seller to propose a single contract designed for the sophisticated consumers. A more general treatment should take this possibility into account from the start. Here this *exclusion* phenomenon can be prevented by assuming that π is high enough. We will see in section 3.2.6 that this is not possible when the Agent's characteristic is multidimensional.

which is negative. For n types of consumers $\theta_1 < \ldots < \theta_n$, each type $\theta_2, \ldots, \theta_n$ can get informational rent, and this rent will increase from θ_2 to θ_n. Only the lowest type, θ_1, will receive no rent.

Remark By the taxation principle, there is a nonlinear tariff that is equivalent to the optimal mechanism. It is simply

$$\begin{cases} t = t_1 & \text{if } q = q_1 \\ t = t_2 & \text{if } q = q_2 \\ t = \infty & \text{otherwise} \end{cases}$$

So the seller needs only to propose the two qualities that will segment the market.[12]

2.3 The Standard Model

The model we study in this section sums up reasonably well the general features of standard adverse selection models. It introduces a Principal and an Agent who exchange a vector of goods q and a monetary transfer p. The Agent has a characteristic θ that constitutes his private information. The utilities of both parties are given by

$$\begin{cases} W(q, t) & \text{for the Principal} \\ U(q, t, \theta) & \text{for the Agent of type } \theta \end{cases}$$

Note that we do not make the Principal's utility function depend on the type θ of the Agent. This is because the model involves "private values" as opposed to "common values." This distinction will be used again in chapter 3. When the contract is signed, the Agent knows his

12. Such an extremely nonlinear tariff is less reasonable when the variable q is a quantity index, as it is in the price discrimination problem studied by Maskin-Riley (1984). Then it is sometimes possible to implement the optimum mechanism by using a menu of linear tariffs. Rogerson (1987) proves that a necessary and sufficient condition is that the optimal nonlinear schedule $t = T(q)$ be convex.

type θ.[13] The Principal entertains an a priori belief about the Agent's type. This belief is embodied in a probability distribution f with cumulative distribution function F on Θ, which we will call the Principal's *prior*. Because the Agent has a continuous set of possible types to choose from, the graphical analysis we used in section 2.2 no longer meets our needs, so we must use differential techniques.

From the revelation principle we already know that the Principal just has to offer the Agent a menu of contracts $(q(.), t(.))$ indexed by an announcement of the Agent's type θ that must be truthful at the equilibrium. We need to characterize the menus of contracts such that

(*IC*) Agent θ chooses the $(q(\theta), t(\theta))$ that the Principal designed for him,

(*IR*) Agent θ thus obtains a utility level at least as large as his reservation utility, meaning the utility he could obtain by trading elsewhere (his second-best opportunity).

The menu of contracts $(q(.), t(.))$ maximizes the expected utility of the Principal among all menus that satisfy (*IR*) and (*IC*).

Remarks

- As in section 2.2, the acronyms (*IR*) and (*IC*) come from the terms *individual rationality* and *incentive compatibility*.

- As in section 2.2.4, it may be optimal for the Principal to exclude some types θ from the exchange by denying them a contract (or at least falling back on a prior "no trade" contract). We, however, neglect this possibility in the following analysis.

- We can neglect the possibility that the optimal mechanism is random; exercise 2.5 gives a sufficient condition for the optimal mechanism to be deterministic.

13. It may be more reasonable to expect the Agent to learn his type only at some point after the contract has been signed but before its provisions are executed. I discuss this variant of the standard model in section 3.2.5.

- We can assume that the Principal faces a population of Agents whose types are drawn from the cumulative distribution function F. This case is isomorphic to that we study here, with a single Agent whose type is random in the Principal's view. Many papers vacillate between the two interpretations, and so will I here.

2.3.1 Analysis of the Incentive Constraints

Let $V(\theta, \hat{\theta})$ be the utility achieved by an Agent of type θ who announces his type as $\hat{\theta}$ and therefore receives utility

$$V(\theta, \hat{\theta}) = U(q(\hat{\theta}), t(\hat{\theta}), \theta)$$

The mechanism (q, t) satisfies the incentive constraints if, and only if, being truthful brings every type of Agent at least as much utility as any kind of lie:

$$\forall (\theta, \hat{\theta}) \in \Theta^2, \quad V(\theta, \theta) \geq V(\theta, \hat{\theta}) \qquad (IC)$$

To simplify notation, we can assume that q is one-dimensional. More important, we can take Θ to be a real interval[14] $[\underline{\theta}, \overline{\theta}]$ and let the Agent's utility function take the following form:

$$U(q, t, \theta) = u(q, \theta) - t$$

This presumes a quasi-linearity that implies that the Agent's marginal utility for money is constant; it simplifies some technical points but primarily allows us to use surplus analysis.

We can further assume that the mechanism (q, t) is differentiable enough. It is sometimes possible to justify this assumption rigorously by proving that the optimal mechanism indeed is at least piecewise differentiable.

---

14. The problem becomes more complicated, and the solution takes a very different form when θ is multidimensional; see section 3.2.6.

For (q, t) to be incentive compatible, it must be that the following first- and second-order necessary conditions hold:[15]

$$\forall \theta \in \Theta, \quad \begin{cases} \dfrac{\partial V}{\partial \hat{\theta}}(\theta, \theta) = 0 \\ \dfrac{\partial^2 V}{\partial \hat{\theta}^2}(\theta, \theta) \leq 0 \end{cases}$$

The first-order condition boils down to

$$\frac{dt}{d\theta}(\theta) = \frac{\partial u}{\partial q}(q(\theta), \theta)\frac{dq}{d\theta}(\theta) \qquad (IC_1)$$

As to the second-order condition, that is,

$$\frac{d^2 t}{d\theta^2}(\theta) \geq \frac{\partial^2 u}{\partial q^2}(q(\theta), \theta)\left(\frac{dq}{d\theta}(\theta)\right)^2 + \frac{\partial u}{\partial q}(q(\theta), \theta)\frac{d^2 q}{d\theta^2}(\theta) \qquad (IC_2)$$

it can be simplified by differentiating (IC_1), which gives

$$\frac{d^2 t}{d\theta^2}(\theta) = \frac{\partial^2 u}{\partial q^2}(q(\theta), \theta)\left(\frac{dq}{d\theta}(\theta)\right)^2 + \frac{\partial u}{\partial q \partial \theta}(q(\theta), \theta)\frac{dq}{d\theta}(\theta)$$
$$+ \frac{\partial u}{\partial q}(q(\theta), \theta)\frac{d^2 q}{d\theta^2}(\theta)$$

whence by substituting into (IC_2),

$$\frac{\partial^2 u}{\partial q \partial \theta}(q(\theta), \theta)\frac{dq}{d\theta}(\theta) \geq 0$$

The first- and second-order necessary incentive conditions thus can be written as

15. These conditions are clearly not sufficient in general; however, we will soon see that they are sufficient in some circumstances.

$$\forall \theta \in \Theta, \quad \begin{cases} \dfrac{dt}{d\theta}(\theta) = \dfrac{\partial u}{\partial q}(q(\theta), \theta)\dfrac{dq}{d\theta}(\theta) & (IC_1) \\ \dfrac{\partial^2 u}{\partial q \partial \theta}(q(\theta), \theta)\dfrac{dq}{d\theta}(\theta) \geq 0 & (IC_2) \end{cases}$$

Most models used in the literature simplify the analysis by assuming that the cross-derivative $\partial^2 u/\partial q \partial \theta$ has a constant sign. This is called the Spence-Mirrlees condition. I will assume that this derivative is positive:

$$\forall \theta, \forall q, \quad \frac{\partial^2 u}{\partial q \partial \theta}(q, \theta) > 0$$

This condition is also called the *single-crossing condition*; it indeed implies that the indifference curves of two different types can only cross once,[16] as is shown in figure 2.4 (where, for the sake of concreteness, I take u to be increasing and concave in q).

The Spence-Mirrlees condition has an economic content; it means that higher types (those Agents with a higher θ) are willing to pay more for a given increase in q than lower types. We may thus hope that we will be able to separate the different types of Agents by offering larger allocations q to higher types and making them pay for the privilege. This explains why the Spence-Mirrlees condition is also called the *sorting condition*, as it allows us to sort through the different types of Agent.

Let us now prove that if q belongs to a direct truthful mechanism (q, t) if, and only if, q is nondecreasing.[17] To see this, consider

$$\frac{\partial V}{\partial \hat{\theta}}(\theta, \hat{\theta}) = \frac{\partial u}{\partial q}(q(\hat{\theta}), \theta)\frac{dq}{d\theta}(\hat{\theta}) - \frac{dt}{d\theta}(\hat{\theta})$$

16. The simplest way to see this is to note that for a given q where they cross, the indifference curves of different types are ordered. Higher types have steeper indifference curves because the slopes $\partial u/\partial q$ increase with θ.

17. If we had assumed the Spence-Mirrlees condition with $\partial^2 u/\partial q \partial \theta < 0$, then q would be nonincreasing.

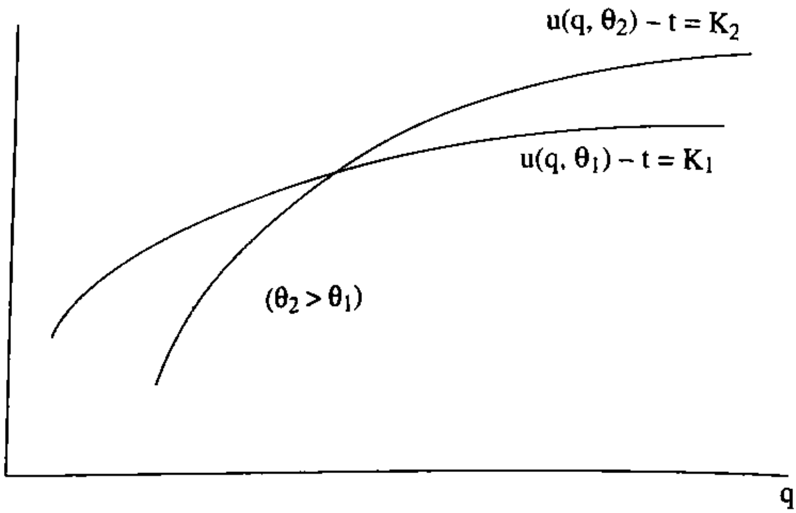

Figure 2.4
The Spence-Mirrlees condition

By writing (IC_1) in $\hat{\theta}$, we get

$$\frac{\partial u}{\partial q}(q(\hat{\theta}), \hat{\theta}) \frac{dq}{d\theta}(\hat{\theta}) = \frac{dt}{d\theta}(\hat{\theta})$$

whence

$$\frac{\partial V}{\partial \hat{\theta}}(\theta, \hat{\theta}) = \left(\frac{\partial u}{\partial q}(q(\hat{\theta}), \theta) - \frac{\partial u}{\partial q}(q(\hat{\theta}), \hat{\theta}) \right) \frac{dq}{d\theta}(\hat{\theta})$$

But the sign of the right-hand side is that of

$$\frac{\partial^2 u}{\partial q \partial \theta}(q(\hat{\theta}), \theta^*) \, (\theta - \hat{\theta}) \frac{dq}{d\theta}(\hat{\theta})$$

for some θ^* that lies between θ and $\hat{\theta}$. Given the Spence-Mirrlees condition, this term has the same sign as $\theta - \hat{\theta}$ if q is nondecreasing. That is, the function $\hat{\theta} \rightarrow V(\theta, \hat{\theta})$ increases until $\hat{\theta} = \theta$ and then decreases. Therefore $\hat{\theta} = \theta$ is the global maximizer of $V(\theta, \hat{\theta})$.

This is a remarkable result. We started with the doubly infinite (in number) global incentive constraints (IC) and the Spence-Mirrlees condition allowed us to transform the constraints into the much simpler local conditions, (IC_1) and (IC_2), without any loss of generality. Note how the problem separates nicely: (IC_2) requires that q be nondecreasing and (IC_1) gives us the associated t. This will be very useful in solving the model. If the Spence-Mirrlees condition did not hold, the study of the incentive problem would be global and therefore much more complex.[18]

2.3.2 Solving the Model

Let us go on analyzing this model with a continuous set of types. We will neglect technicalities in the following. In particular, we assume that all differential equations can safely be integrated.[19] We also assume that the Principal's utility function is quasi-separable and is

$$t - C(q)$$

We further assume that

$$\forall q, \forall \theta, \quad \frac{\partial u}{\partial \theta}(q, \theta) > 0$$

meaning that a given allocation gives the higher types a higher utility level. Finally, we assume that the Spence-Mirrlees condition holds:

$$\forall \theta, \forall q, \quad \frac{\partial^2 u}{\partial q \partial \theta}(q, \theta) > 0$$

18. In the few papers (e.g., Moore 1988) that adopt a "nonlocal" approach that does not rely on the Spence-Mirrlees condition, typically assumes that only the downward incentive constraints are assumed to bind. Milgrom-Shannon (1994) establish a connection between the Spence-Mirrlees condition and the theory of supermodular functions.

19. Readers interested in a more full and rigorous analysis should turn to Guesnerie-Laffont (1984).

Let $v(\theta)$ denote the utility the Agent of type θ gets at the optimum of his program. As the optimal mechanism is truthful, we get

$$v(\theta) = V(\theta, \theta) = u(q(\theta), \theta) - t(\theta)$$

and IC_1 implies that

$$\frac{dv}{d\theta}(\theta) = \frac{\partial u}{\partial \theta}(q(\theta), \theta)$$

which we have assumed is positive. The utility $v(\theta)$ represents the *informational rent* of the Agent; the equation above shows that this rent is an increasing function of his type. Higher types thus benefit more from their private information. That is, if type θ can always pretend his type is $\hat{\theta} < \theta$, he will obtain a utility

$$u(q(\hat{\theta}), \theta) - t(\hat{\theta}) = v(\hat{\theta}) + u(q(\hat{\theta}), \theta) - u(q(\hat{\theta}), \hat{\theta})$$

which is larger than $v(\hat{\theta})$ since u increases in θ. The ability of higher types to "hide behind" lower types is responsible for their informational rent.[20] This rent is the price that the Principal has to pay for higher types to reveal their information.

In most applications the individual rationality constraint is taken to be independent of the Agent's type.[21] This amounts to assuming that the Agent's private information is only relevant in his relationship with the Principal. Under this assumption, which is not innocuous,[22] we can normalize the Agent's reservation utility to 0 and write his individual rationality constraint as

$$\forall \theta, \quad v(\theta) \geq 0 \qquad (IR)$$

Given that v is increasing, the individual rationality constraint (IR) boils down to

20. Note, however, that lower types have no incentive to hide behind higher types.
21. We will make an important exception in section 3.1.3.
22. See section 3.2.8 for a general analysis of the adverse selection problem in which reservation utilities are allowed to depend on types in a nonrestricted way.

$$v(\underline{\theta}) \geq 0$$

which must actually be an equality, since transfers are costly for the Principal.

These preliminary computations allow us to eliminate the transfers $t(\theta)$ from the problem; so we have

$$v(\theta) = \int_{\underline{\theta}}^{\theta} \frac{\partial u}{\partial \theta}(q(\tau), \tau)d\tau$$

whence

$$\begin{aligned} t(\theta) &= u(q(\theta), \theta) - v(\theta) \\ &= u(q(\theta), \theta) - \int_{\underline{\theta}}^{\theta} \frac{\partial u}{\partial \theta}(q(\tau), \tau)d\tau \end{aligned}$$

Let us now return to the Principal's objective[23]

$$\int_{\underline{\theta}}^{\bar{\theta}} \Big(t(\theta) - C(q(\theta))\Big) f(\theta)d\theta$$

Substituting for t, it can be rewritten as

$$\int_{\underline{\theta}}^{\bar{\theta}} \left(u(q(\theta), \theta) - \int_{\underline{\theta}}^{\theta} \frac{\partial u}{\partial \theta}(q(\tau), \tau)d\tau - C(q(\theta)) \right) f(\theta)d\theta$$

Let us define the *hazard rate*

$$h(\theta) = \frac{f(\theta)}{1 - F(\theta)}$$

This definition is borrowed from the statistical literature on duration data:[24] if $F(\theta)$ is the probability of dying before age θ, then $h(\theta)$ represents the instantaneous probability of dying at age θ provided that one has survived until then.

23. Recall that f is the probability distribution function and F the cumulative distribution function of the Principal's prior on Θ.

24. Some economists improperly define the hazard rate as $1/h(\theta)$.

Now applying Fubini's theorem[25] or simply integrating by parts, the Principal's objective becomes

$$I = \int_{\underline{\theta}}^{\bar{\theta}} H(q(\theta), \theta) f(\theta) d\theta$$

where

$$H(q, \theta) = u(q, \theta) - C(q) - \frac{\partial u}{\partial \theta}(q, \theta)\frac{1}{h(\theta)}$$

The function $H(q(\theta), \theta)$ is the *virtual surplus*. It consists of two terms. The first term,

$$u(q(\theta), \theta) - C(q(\theta))$$

is the first-best social surplus,[26] namely the sum of the utilities of the Principal and the type θ Agent. The second term, $-v'(\theta)/h(\theta)$, therefore measures the impact of the incentive problem on the social surplus. This term originates in the necessity of keeping the informational rent $v(\theta)$ increasing. That is, type θ's allocation is increased, then so is his informational rent, and to maintain incentive compatibility, the Principal must also increase the rents of all types $\theta' > \theta$ who are in proportion $1 - F(\theta)$.

We still need to take into account the second-order incentive constraint

$$\frac{dq}{d\theta}(\theta) \geq 0$$

The simplest way to proceed is to neglect this constraint in a first attempt. The (presumed) solution then is obtained by maximizing the integrand of I in every point, whence

25. Fubini's theorem states that if f is integrable on $[a, b] \times [c, d]$, then

$$\int_a^b \int_c^d f(x, y) dx dy = \int_a^b \left(\int_c^d f(x, y) dy\right) dx = \int_c^d \left(\int_a^b f(x, y) dx\right) dy$$

26. It is appropriate to speak of surplus here because the transfers have a constant marginal utility equal to one for both Principal and Agent.

$$\frac{\partial H}{\partial q}(q^*(\theta), \theta) = 0$$

Writing this equation in full, we have

$$\frac{\partial u}{\partial q}(q^*(\theta), \theta) = C'(q^*(\theta)) + \frac{\partial^2 u}{\partial q \partial \theta}(q^*(\theta), \theta)\frac{1}{h(\theta)}$$

Note that the left-hand side of this equation has the dimension of a price; it is in fact just the inverse demand function of Agent θ. Since we have assumed that the cross-derivative is positive, this equation tells us that price is greater than marginal cost. The difference between them is the source of the informational rent, and this difference represents the deviation from the first-best.

The Separating Optimum.

If the function q^* is nondecreasing, it is an optimum. We can say that types are separated and that revelation is then perfect, as shown in figure 2.5.

Higher types θ have a larger allocation q, and they pay more for it. Note that it is often possible to make assumptions that guarantee the separation result. If, for instance, $u(q, \theta) = \theta q$ and C is convex, then it is easily varified that assuming the hazard rate h to be nondecreasing is sufficient to imply that q^* is increasing. The literature often resorts to such an assumption because it is satisfied by many classic probability distributions.

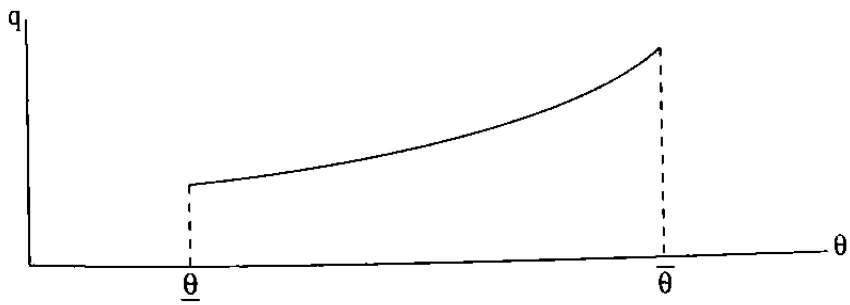

Figure 2.5
A separating optimum

It is hard to say much about the shape of the tariff $t(q)$ in general. The reader is referred to exercise 2.4 to prove that $t(q)$ is convex if $u(q, \theta)$ is linear. As Rogerson (1987) has shown, such a convex $t(q)$ can be approximated by a menu of linear tariffs.

**The Bunching Optimum.*

If the function q^* happens to be decreasing on a subinterval, it cannot be the solution. It is then necessary to take into account the constraint that q should be nondecreasing, which means resorting to optimal control theory. Since I do not expect optimal control theory to be a prerequisite to understanding the discussions in this book, I give a self-contained analysis below, using only elementary concepts. Readers who prefer a more direct treatment should consult Laffont (1989, 10) and Kamien-Schwartz (1981), for example, for the basics of optimal control theory.

First, note that the solution will consist of subintervals in which q is increasing and subintervals in which it is constant. Take a subinterval $[\theta_1, \theta_2]$ in which q is increasing and $\partial H/\partial q$ is positive. We then add a positive infinitesimal function $dq(\theta)$ to $q(\theta)$ in that subinterval so that $dq(\theta_1) = dq(\theta_2) = 0$ and $q + dq$ stays increasing. This clearly increases H on $[\theta_1, \theta_2]$ and so improves the objective of the Principal. A similar argument applies when $\partial H/\partial q$ is negative on a subinterval where q is increasing. Thus, whenever q is increasing, the solution must satisfy $\partial H/\partial q = 0$, which is just to say that it must coincide with q^*.

The determination of the subintervals where q is constant is trickier. We take such a (maximal) subinterval $[\theta_1, \theta_2]$. On this subinterval the solution must equal a constant $\tilde{q}$ such that $q^*(\theta_1) = q^*(\theta_2) = \tilde{q}$. This defines two functions $\theta_1(\tilde{q})$ and $\theta_2(\tilde{q})$. We just have to determine the value of $\tilde{q}$. We let

$$F(q) = \int_{\theta_1(q)}^{\theta_2(q)} \frac{\partial H}{\partial q}(q, \theta)d\theta$$

and assume that $F(\tilde{q}) > 0$. Then we add to the solution an infinitesimal positive constant on $[\theta_1, \theta_2]$ (and afterward, a smaller, decreasing amount on $[\theta_2, \theta_2 + \varepsilon]$, where $q^*(\theta_2 + \varepsilon) = \tilde{q} + dq$). The Principal's objective will be unchanged on $[\theta_2, \theta_2 + \varepsilon]$, since $\partial H/\partial q = 0$ there by assumption. However, the objective will increase by $F(\tilde{q})dq$ on $[\theta_1, \theta_2]$. This, and a similar reasoning when $F(\tilde{q}) < 0$, prove that we must have $F(\tilde{q}) = 0$. Because $\partial H/\partial q = 0$ in θ_1 and θ_2, we can easily write the derivative of F as

$$F'(q) = \int_{\theta_1(q)}^{\theta_2(q)} \frac{\partial^2 H}{\partial q^2}(q, \theta)d\theta$$

Thus, if we make the reasonable assumption that the virtual surplus is concave in q,[27] $\partial^2 H/\partial q^2$ will be negative and therefore F will be decreasing. This implies that if there is a $\tilde{q}$ such that $F(\tilde{q}) = 0$, then it is unique, and this completes our characterization of the solution.

The solution in this more complicated case is depicted in figure 2.6. In sum, we speak of *bunching* or pooling of types on the subintervals where q is constant, and there is less than perfect revelation. Obviously all the types $\theta \in [\theta_1, \theta_2]$ pay the same transfer t for their constant allocation.

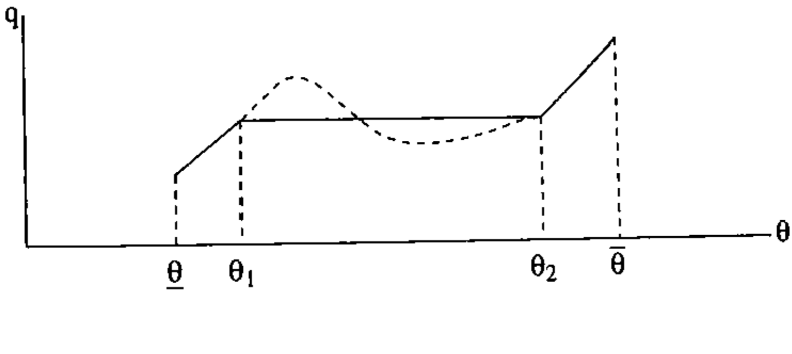

Figure 2.6
An optimum with bunching

27. We assume, for instance, that u is concave in q, C is convex and $\partial^2 u/\partial q^2$ increases in θ.

Exercises

Exercise 2.1

Assume that there are n types of consumers in the wine-selling example of section 2.2 and that $\theta_1 < \ldots < \theta_n$. Their respective prior probabilities are $\pi_1, \ldots, \pi_n$, with $\sum_{i=1}^{n} \pi_i = 1$. Show that the only binding constraints are the downward adjacent incentive constraints

$$\theta_i q_i - t_i \geq \theta_i q_{i-1} - t_{i-1}$$

for $i = 2, \ldots, n$ and the individual rationality constraint of the lowest type

$$\theta_1 q_1 - t_1 \geq 0$$

Exercise 2.2

In the context of section 2.3.2, assume that $u(q, \theta) = \theta q$ and C is convex.

1. Show that a necessary and sufficient condition for q^* to be increasing is that $\theta - 1/h(\theta)$ be increasing.
2. A function g is *log-concave* iff $\log g$ is concave. Show that all concave functions are log-concave. Show that if $(1 - F)$ is log-concave, then q^* is increasing.
3. Show that q^* is increasing if θ is uniformly distributed.
4. A bit more tricky: Show that if f is log-concave, then so is $(1 - F)$.
5. Conclude that q^* is increasing if θ is normally distributed.

Exercise 2.3 (difficult)

My characterization of the bunching optimum in section 2.3.2 implies a hidden assumption: bunching does not occur "at the bottom" (on some interval $[\underline{\theta}, \theta_1]$) nor "at the top" (on some interval $[\theta_2, \bar{\theta}]$). Modify the proof so that it covers these two cases as well.

Exercise 2.4

Denote $t(q)$ the optimal tariff in the continuous-type model of section 2.3 and $\theta(q)$ the inverse function to the optimal $q(\theta)$.

1. Prove that $t'(q) = \dfrac{\partial u}{\partial q}(q, \theta(q))$

2. Assume that $u(q, \theta)$ is linear in q; prove that $t(q)$ is convex.

Exercise 2.5

Let us study the sufficient conditions for the optimal mechanism to be deterministic in the continuous-type model of section 2.3. Let the Agent's utility function be $u(q, \theta) - t$ and the Principal's utility function be $t - C(q)$. We assume that u is increasing in θ and has a positive cross-derivative, and that C is increasing and convex in q.

Denote by $(Q(\theta), T(\theta))$ a stochastic mechanism that is a lottery from which the (q, t) pair is drawn after the Agent announces his type.

1. Rewrite the arguments of section 2.3 to show that the $Q(\theta)$ in the optimal stochastic mechanism maximizes

$$I = \int_{\underline{\theta}}^{\bar{\theta}} EH(q(\theta), \theta) f(\theta) d\theta$$

under the second-order incentive constraint that

$$E \frac{\partial^2 u}{\partial q \partial \theta}(Q(\theta), \theta) \frac{dQ}{d\theta}(\theta) \geq 0$$

2. Assume that $EQ'(\theta) \geq 0$ everywhere. Let $q^e = EQ$. Use Jensen's inequality to show that if $\dfrac{\partial u}{\partial \theta}$ is concave in q, then the deterministic

mechanism schedule q^{ϵ} satisfies the incentive constraint and improves the objective I.

3. Assume that $u(q, \theta) = q\theta$. Show that the optimal mechanism is deterministic.

References

Guesnerie, R., and J.-J. Laffont. 1984. A complete solution to a class of principal-agent problems with an application to the control of a self-managed firm. *Journal of Public Economics.* 25:329–69.

Kamien, M., and N. Schwartz. 1981. *Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management.* Amsterdam: North-Holland.

Laffont, J.-J. 1989. *The Economics of Uncertainty and Information.* Cambridge: MIT Press.

Maskin, E., and J. Riley. 1984. Monopoly with incomplete information. *Rand Journal of Economics* 15:171–96.

Milgrom, P., and C. Shannon. 1994. Monotone comparative statics. *Econometrica* 62:157–80.

Moore, J. 1988. Contracts between two parties with private information. *Review of Economic Studies* 55:49–70.

Mussa, M., and S. Rosen. 1978. Monopoly and product quality. *Journal of Economic Theory* 18:301–17.

Palfrey, T. 2002. Implementation theory. In *Handbook of Game Theory,* vol. 3, R. Aumann and S. Hart, eds. Amsterdam: North-Holland.

Rogerson, W. 1987. On the optimality of menus of linear contracts. Mimeo. Northwestern University.

Salanié, B. 2000. *The Microeconomics of Market Failures.* Cambridge: MIT Press.

Stiglitz, J., and A. Weiss. 1981. Credit rationing in markets with imperfect information. *American Economic Review* 71:393–410.