bookboon.com

Essentials of Microeconomics

Krister Ahlersten

Krister Ahlersten

Microeconomics

Microeconomics

ISBN 978-87-7681-410-6

Contents

1 Introduction

Economics is often defined as something along the lines of "the study of how society manages its scarce resources." The starting point of most such studies is that individuals allocate their resources such that they themselves will get the highest possible level of utility. An individual has an idea of what the consequences of different actions will be, and she chooses that action she believes will produce the best result for her. She is, in other words, selfish and rational. Note that she is also forward-looking. She acts so that she in the future will get the highest possible level of utility, independently of what she has already done. That she is selfish does not have to mean that she is an egoist. However, it does mean that she will only voluntarily share with others if she believes that she thereby will maximize her own utility. We often call this simplification of human beings **Homo Economicus**.

Economics: The study of how society manages its scarce resources

The **resources** that we are talking about here could be labor, capital (such as machines), and raw materials. That they are scarce means there are not enough resources to produce everything we want. That, in turn, means that one has to weight different things against each other. To get more of one thing, one has to give up something else. If you, e.g., want to sleep an extra hour, it is impossible to do so without giving up something else, such as an hour of studying. There is, consequently, a sort of a hidden cost to sleeping longer. This type of cost is called **opportunity cost** (or **alternative cost**). A classical saying in economics is that "there is no such thing as a free lunch." This means that, even if you do not actually pay for the lunch, you always have to give up at least the time when you could have done something else. That is, you always have to pay the opportunity cost.

Homo Economicus: A model of human beings. She is assumed to maximize her own utility.
Resources: Labor, capital and raw materials. The things we use to produce goods and services.

Opportunity/alternative cost: The (hidden) cost of choosing one alternative instead of another.

When we study **microeconomics**, it is primarily individual human beings and individual firms, **agents**, that we study. This is in contrast to **macroeconomics**, where one studies whole economies, and questions such as unemployment and inflation.

Microeconomics: The study of the economic behavior of individual human beings and firms.
Agent: An entity that is capable of making a decision, e.g. a human being or a firm.
Macroeconomics: The study of whole economies.

Roughly speaking, there are three types of decisions that need to be made in an economy: Which goods and services to produce, how to produce them, and who should get them. Often in economic models, the prices of goods (or services, labor, capital, etc.) automatically coordinate these decisions in a market. A **market** is any mechanism where buyers and sellers meet. That could be, for example, a market square, a stock exchange, or a computer network where one can buy and sell things.

Market: Meeting place where buyers and sellers are able to trade with each other.

Microeconomics is often based on **models**. We try to describe a real phenomenon as simply as possible by only highlighting a few central features. Many economic models can be used for predictions and can therefore be tested against reality. Such models are called **positive**. The opposite kind of models, models that are about values, is called **normative**. For example, to decide about an economic policy one would first use positive economics to make assessments about the consequences of different alternatives. Then one would use one's opinions about what is desirable and what is not to choose between the different alternatives. That is then a normative decision.

Model: A simplified description of reality.

Positive economics: A testable economic model.
Normative economics: An economic model that includes values (and therefore is not testable).

1.1 Plan

Before we begin, it is probably wise to make it clear where we are trying to go. We want to develop a number of models that together can describe how an economy works. They should be able to produce clear and testable predictions and be as simple as possible.

- In a market, products and/or services are being bought and sold (or traded). We begin by looking at consumers and producers, and their respective demand and supply in a market. That way, we will see an example of how the market price of a good is determined.
- Consumers and producers, however, have difficult problems to solve before they arrive at their respective demand and supply. First, we look at a consumer's problem in a very simple case: She has to choose between two different goods for which she has different preferences. We show how it is possible to go from her preferences and income to her demand for one of the goods. Then we show how one can derive the demand for the whole market.
- Then we change perspectives and study a producer's problem. We will then discover that the model looks very similar to that of the consumer. The producer has to produce the good with the help of labor and capital, and different combinations of the two will lead to different quantities of the good. She also has to think about the fact that, different combinations will have different costs. The results will help us to show how the market supply is determined.
- There are usually quite many consumers but substantially fewer producers. This has a large impact on how the market operates, and we therefore continue to study different market forms. We will differentiate between cases where there are one, two, some, and many producers. We also study the welfare effects of different market forms.
- The producers have a demand for labor and the workers supply it. The labor market has some odd features that we will treat separately.
- Equilibrium is a central concept in economics. We show how consumer and producer markets, as well as the market for goods, simultaneously reach equilibrium in a simple and stylized economy.
- Lastly, we relax some of the assumptions we have made so far. We show how undesirable results can arise because of so-called market failures, e.g. because different agents have different amounts of information about a good, or because it is difficult to keep out users who do not pay.

2 Supply, Demand, and Market Equilibrium

We begin our study of microeconomics by looking at a market with many buyers and sellers, i.e. a market where there is a large amount of competition. We will study such a market in more depth in Chapter 9, as well as other market types, but starting here makes it easy to get a feel for how the subject works.

2.1 Demand

2.1.1 The Demand Curve

The **demand curve** shows what quantities of a good buyers are willing to buy at different prices. Note the expression “are willing.” It is not about how much they actually buy, but about how much they would want to buy if a certain price was offered.

Demand curve: Shows how much the buyers are willing to buy at different prices of a good.

A demand curve is only valid if all other relevant factors are held constant (***ceteris paribus***: with other things the same). The most important other factors that can affect demand are:

Ceteris paribus: Latin for “with other things the same”.

- The buyers' income.
- Prices and price changes on other goods. We will make a distinction between **complementary goods** and **substitute goods**. An example of complementary goods is right and left shoes. If the price of right shoes rises then the demand for right shoes will typically decrease. However, the demand for left shoes will also typically decrease. Consequently, the demand for left shoes partly depends on the price of another good: right shoes.

  Substitute goods work in the opposite way. An example could be blue and green pens: If one cannot use blue, one can often use green instead. If the price of green pens rises, the demand for green pens typically decreases. However, if the price of blue pens is unchanged one can use these instead of the green ones, and then the demand for blue pens increases. Consequently, the demand for blue pens depends on the price of another good: green pens.

  Note that for substitute goods, a rise in the price of the other good leads to an increase in the demand for the good we are analyzing, whereas for complementary goods it is the other way around; a rise in the price of the other good leads to a decrease in the demand for the good analyzed.

Complementary goods: Goods that are typically consumed together
Substitute goods: Goods that can be used instead of each other.

- **Preferences**. What consumers demand is largely a matter of taste. If there is a change in taste, there is usually also a change in demand. Taste can change for many different, underlying, reasons. For example, changes in moral perception or in fashion.

Preferences: What an individual prefers; her taste.

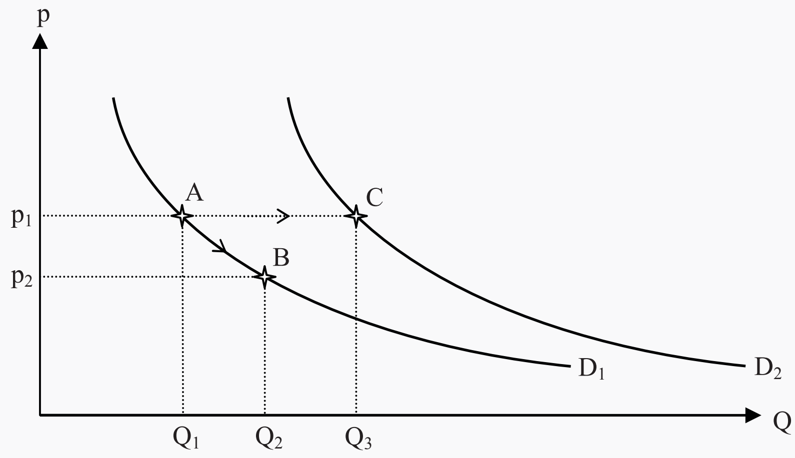
If these factors are held constant, then the demand curve is valid and it usually slopes downwards. In other words, the lower the price is the higher is the demand, and vice versa. Demand is defined for a certain period. One can for example think of it as defined over a month, corresponding to a monthly salary.

When drawing a demand curve in a diagram, the quantity demanded is on the X-axis and the price is on the Y-axis. This is slightly odd, since we often think of the quantity demanded as a function of the price, not the other way around. There are historical reasons for drawing it this way.

2.1.2 When do We Move along the Demand Curve, and When Does It Shift?

The relation between price and quantity that is described by the demand curve is valid *only* if it is the price of the good itself that changes. Look at Figure 2.1 and the demand curve D_1. If, in the beginning, the price is p_1, then the quantity demanded is Q_1 (point A). If the price of the good falls to p_2, then the quantity demanded changes to Q_2 (point B). We, consequently, *move along* the demand curve when the price of the good changes.

Figure 2.1: The Demand Curve



If, instead, something else changes (e.g. income, the prices of other goods, consumer preferences, or anything that affects the demand on the good), then the demand curve *shifts*. Assume again that the price is p_1 so that the quantity demanded is Q_1 (point A). If the consumer's income increases, she can buy more of the good than she could before. Consequently, the whole demand curve shifts from D_1 to D_2. If the price is still p_1, the quantity demanded increases to Q_3 (point C).

2.2 Supply

2.2.1 The Supply Curve

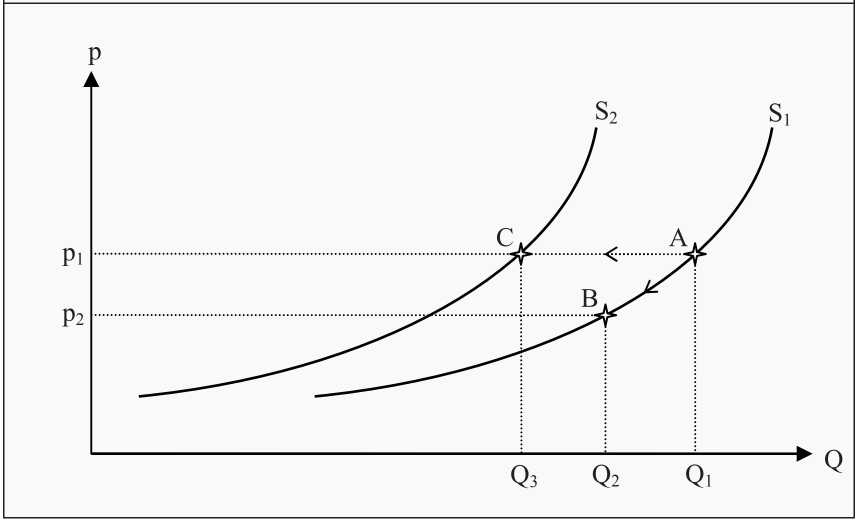
The producer counterpart to the demand curve is the **supply curve**. It shows how large quantities the producers are willing to sell at different prices, given that other factors that can affects supply are held constant. The supply curve is typically upward sloping or horizontal (but it could also be downward sloping). The demand curve is also valid over a certain period. Later, we will distinguish between two time periods: short and long horizons.

The most important factors, beside the price, that affect supply are:

- **Factor prices**, i.e. wages, prices of machines and compensation to owners and lenders. In other words, changes in the cost of production.
- Laws and regulations that apply to the production.
- Prices of other goods the firm produces or could potentially produce. Perhaps the producer is producing blue and green pens. If the price of green pens rises, she is likely to shift over resources (workers and machines) to that production and there is less left with which to produce blue pens. Consequently, the supply of blue pens decreases, even though the price of blue pens is unchanged.

Factor prices: The prices of the factors of production.

Figure 2.2: The Supply Curve



Supply curve: Shows how much the sellers are prepared to sell at different prices of a good.

The supply curve behaves in a way that is similar to that of the demand curve. Look at Figure 2.2 and the supply curve S_1. If the price is p_1, then the producers are willing to sell the quantity Q_1 (point A). If the price of the good falls to p_2, we move along S_1 to point B, where the quantity is Q_2. If, instead, some other factor changes, e.g. if wages increase so that it becomes more expensive to produce the good, the whole supply curve shifts. For instance from S_1 to S_2. If the price is still p_1, then the quantity supplied changes from Q_1 to Q_3 (point C).

2.3 Equilibrium

A market is in **equilibrium** when both of these conditions are fulfilled:

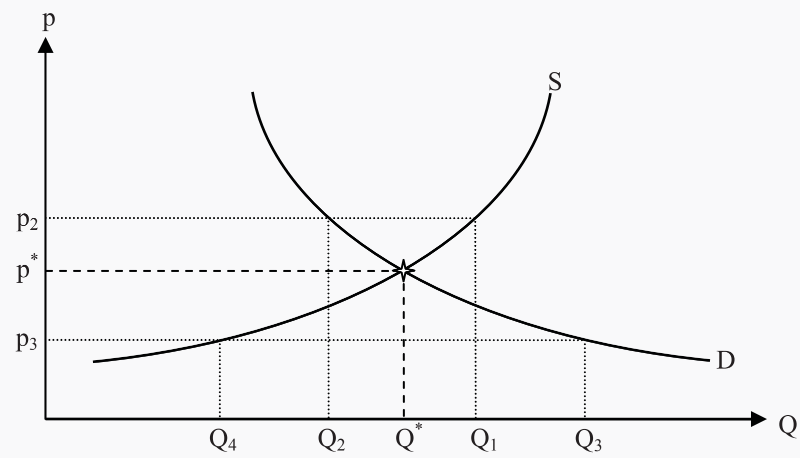
1. No agent wants to change her decision or strategy.
2. The decisions of all agents are compatible with each other, so that they can all be carried out simultaneously.

Equilibrium: A situation in which no agent wants to change her decision and all decisions are compatible.

If we join the supply and demand curves in one diagram, we get an equilibrium point where the two curves intersect. At this point, the price the consumers are willing to pay is the same as the price the producers demand. In Figure 2.3, the **equilibrium price** (market-clearing price) is p^* and the **equilibrium quantity** is Q^*.

Equilibrium price: The price that arises when there is an equilibrium in the market.
Equilibrium quantity: The quantity that is bought and sold when there is an equilibrium in the market.

Figure 2.3: Equilibrium



The equilibrium point has two important properties in that it is most often (but not always) stable and self-correcting. That it is stable means that, if the market is in equilibrium there is no tendency to move away from it. That it is self-correcting means that, if the market is not in equilibrium then there is a tendency to move towards it.

To see more clearly what this means, suppose the price is higher than in equilibrium, e.g. that it is p_2. At that price, producers are willing to supply the quantity Q_1 whereas the consumers are only willing to buy the quantity Q_2. Therefore, there is an excess supply of the good. To get rid of the extra units the producers are prepared to lower the price. This will push the price downwards, closer to p^*. At p^*, there is no excess supply and the downward push on the price ends.

Then assume, instead, that the price is lower than p^*, e.g. that it is p_3. At this price, the consumers demand the quantity Q_3 whereas the producers are only willing to supply the quantity Q_4. Consequently, there will be a shortage of the good, and the consumers will be prepared to bid up the price to get more units. This will tend to push the price upwards, closer to p^* where, again, the push will end.

2.3.1 How to Find the Equilibrium Point Mathematically

Supply and demand can be written as mathematical functions, and in simple examples, they are often straight lines. They could, for instance, be:

$$\begin{cases} Q_S = 85 + 30p \\ Q_D = 185 - 20p. \end{cases}$$

Here, Q_D is the quantity demanded, Q_S is the quantity supplied, and p is the price.

We now want to find the price, p^*, that makes $Q_D = Q_S$. If the left-hand sides above are equal, the right-hand sides must also be so. Therefore, substitute p^* for p and set the right-hand sides equal to each other:

$$85 + 30p^* = 185 - 20p^*$$

To get p^* alone on the left-hand side, we add 20 p^* on both sides and subtract 85 from both sides. Then we have that

$$50p^* = 100.$$

Dividing by 50 on both sides yields the result that

$$p^* = 2.$$

If we then want to know the equilibrium quantity, Q^*, we substitute the result we got for p^* into either the supply or the demand function above. (Note that they must yield the same quantity, since p^*, by definition, is the price that makes $Q_D = Q_S$.)

$$Q^* = \begin{cases} Q_S^* = 85 + 30p^* \quad = 85 + 30 * 2 \quad = 145 \\ Q_D^* = 185 - 20p^* = 185 - 20 * 2 = 145 \end{cases}$$

Consequently, we have the equilibrium price, $p^* = 2$, and the equilibrium quantity, $Q^* = 145$.

2.4 Price and Quantity Regulations

Many markets are, for a number of reasons, **regulated**. The government could for instance decide about prices that the market is not allowed to go above or below, or about maximum quantities. Such regulations will benefit certain groups of people, but often have unintended negative side effects. These are often called **secondary effects**.

Regulation: Laws that influence prices and/or quantities in a market.

Secondary effect: An unintended side effect of, for instance, a law.

2.4.1 Minimum Prices

Minimum prices (also called **price floors**) are often used for wages (the price of labor) and for certain types of goods such as agricultural goods. The minimum price is usually chosen above the equilibrium price, as in the opposite case it would not have any effect. (The market participants would then choose p^* instead.) Consumers and producers are consequently prevented from reaching the equilibrium price p^*.

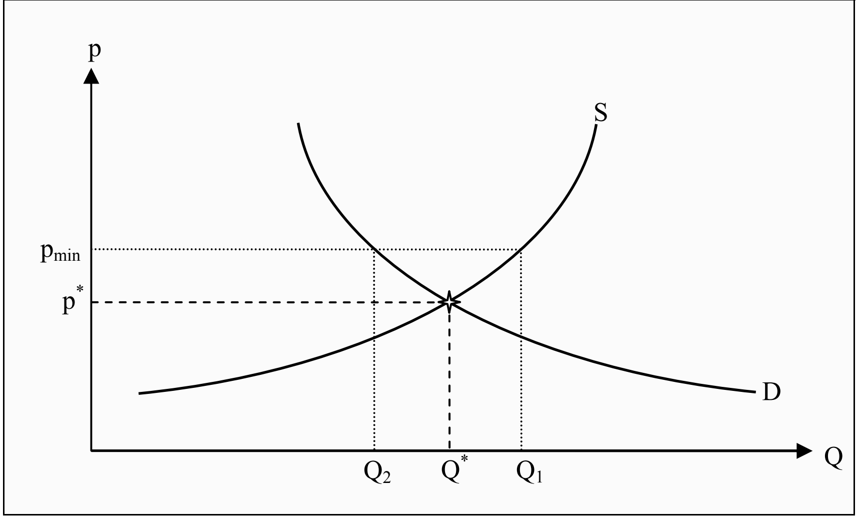
Minimum price/price floor: The lowest price a regulation allows.

Look at Figure 2.4. The effect of the minimum price is that the consumers only demand the quantity Q_2 whereas the producers supply the quantity Q_1. Therefore, we get an excess supply of the good.

Note that consumers and producers are allowed to buy and sell at any price above the minimum price. A price higher then p_{min} will however result in an even larger excess supply, so typically the minimum price is chosen.

The situation described is not an equilibrium. To see that, note that point 2 in the definition of an equilibrium (see Section 2.3) is not satisfied: Given the price p_{min} producers want to sell the quantity Q_1, but that is not possible since the consumers only want to buy the quantity Q_2.

Figure 2.4: The Effect of a Minimum Price

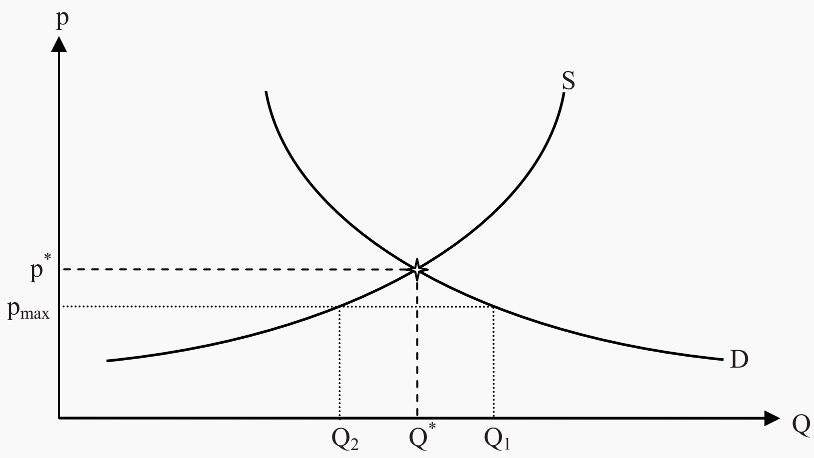


2.4.2 Maximum Prices

Maximum prices (also called **price ceilings**) are in several countries used for apartment rentals. For a maximum price to have any effect, it has to be below the equilibrium price, and the effects are the opposite to those of a minimum price. In Figure 2.5, p_{max} is the maximum price. It causes the consumers to demand the quantity Q_1 whereas the producers only want to supply Q_2, and, consequently, there is a shortage. A typical consequence of a maximum price is that the search time to find an appropriate good is increased since the supply is too small to meet the demand.

Maximum price/price ceiling: The highest price a regulation allows.

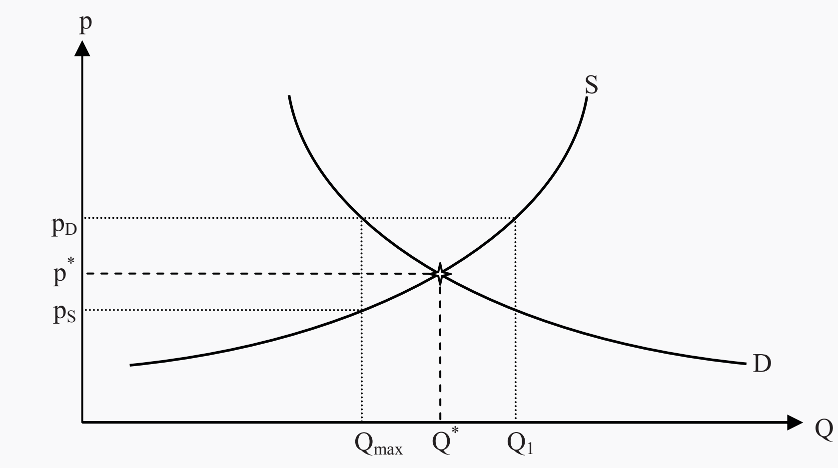
Figure 2.5: The Effect of a Maximum Price



2.4.3 Quantity Regulations

The effects of quantity regulations are similar to those of price regulations. Assume for instance that there is a restriction stating that one may only import the quantity Q_{max} of a certain good, say Asian textiles.

Figure 2.6: The Effect of a Quantity Regulation



Producers would have been willing to supply the quantity Q_{max} at a price of p_S, whereas the consumers would have been willing to buy that quantity at a price of p_D. Since the quantity is not allowed to increase, there is excess demand at all prices other than p_D. When there is excess demand, consumers are likely to bid up the price, so the price that this market is likely to arrive at is p_D.

Note that at the price p_D, producers are willing to supply a much larger quantity, Q_1, but that they are prevented from doing so by the regulation. The consumers have to pay a price that is larger than the equilibrium price (p_D instead of p^*) and they get fewer units of the good, so they typically are made worse off by a quantity regulation.

3 Consumer Theory

Where does the demand curve come from? In order to explain why individuals choose different quantities at different prices, we will use a model with three components:

- Consumers have certain restrictions on how they can choose. Most importantly, they have a **budget**, but there can also be other restrictions.
- Individual preferences (or tastes) determine how satisfied an individual will be with different combinations of goods and/or services. We measure the level of satisfaction in terms of **utility**.
- Given preferences and restrictions, the individual **maximizes** her utility of consumption.

We will now discuss these three components.

Budget: The amount of money, or wealth, a consumer has access to.

Utility: A measure of how satisfied a consumer is.

Maximize: Choose in such a way that one gets as much as possible of something else.

3.1 Baskets of Goods and the Budget Line

As a consumer, one can choose between several different goods and services. A certain combination of goods and services is called a **basket of goods** (a **bundle of goods**, or a **market basket**). The consumer's problem can therefore be described as having to choose between different baskets, given the restrictions she faces, such that she maximizes her utility.

Basket / bundle of goods: A combination of goods and services.

We begin by looking at a simple case where we have just two goods, good 1 and 2, with prices p_1 and p_2. A basket that consists of the quantity q_1 of good 1 and q_2 of good 2 is written (q_1,q_2). For example, (4,3) means that we have 4 units (or kilos, liters, etc) of good 1 and 3 units of good 2. The price of the basket (q_1,q_2) is then:

$$p_1 * q_1 + p_2 * q_2 .$$

If we have a limited amount of money to buy these goods for, this will impose a restriction on how much we can buy of each good. Letting m denote the amount of money available, the price of the basket chosen must not exceed m. The different combinations of good 1 and 2 that cost exactly m can be written

$$p_1 * q_1 + p_2 * q_2 = m.$$

Solving this expression for q_2, we get the function of the **budget line**:

Budget line: A graphical description of the baskets a consumer can buy, given a certain budget.

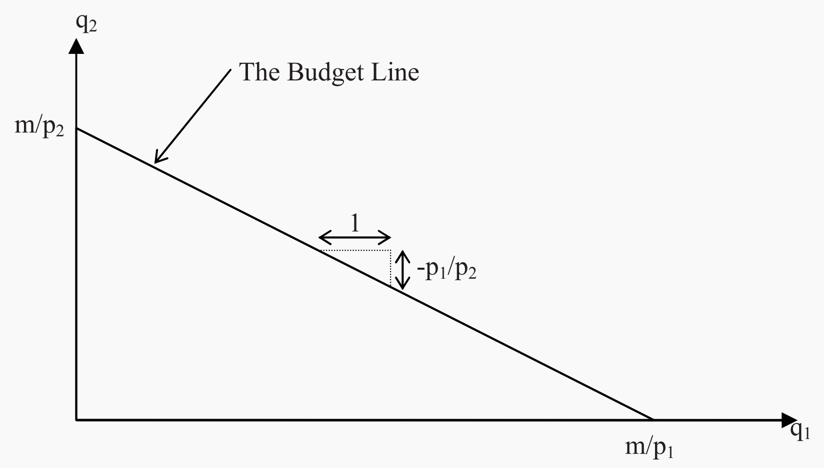
$$q_2 = \frac{m - p_1 * q_1}{p_2} = \frac{m}{p_2} - \frac{p_1}{p_2} * q_1$$

This function is a straight line that intercepts the Y-axis at m/p_2 and has the slope p_1/p_2 (see Figure 3.1).

All the points on the budget line cost exactly m. The points in the grey area below the budget line cost less than m whereas the points above cost more than *m*. The baskets that a consumer with wealth m can buy are, consequently, the ones on and below the budget line.

There is a simple strategy for finding the budget line: If we only buy good 2, the maximum quantity that we can buy is m/p_2, whereas if we buy only good 1, the maximum quantity that we can buy is m/p_1. Indicate the first point on the Y-axis and the second on the X-axis, and then draw a straight line between them. The line you have drawn is the budget line, and it will automatically have the slope $-p_1/p_2$.

Figure 3.1: The Budget Line



The slope of the budget line is called the **marginal rate of transformation** (*MRT*). We consequently have that

Marginal rate of transformation: The slope of the budget line.

$$MRT = -\frac{p_1}{p_2}$$

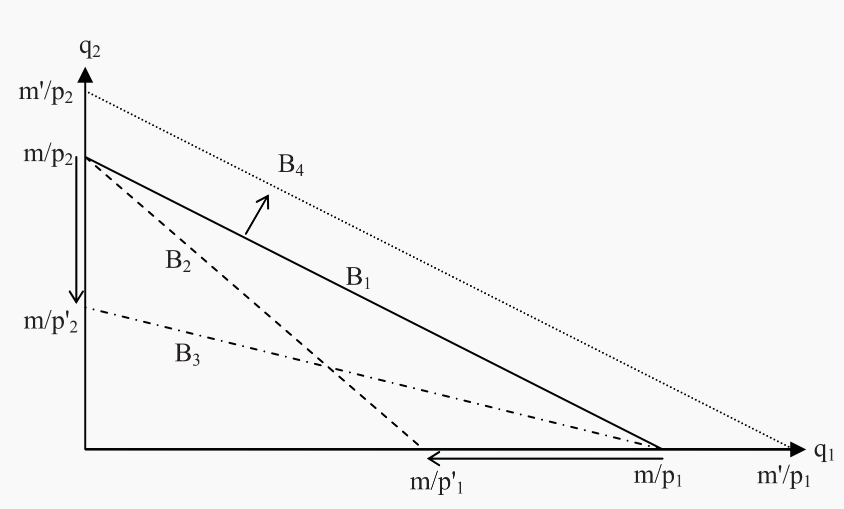
Suppose, for instance, that the two goods are ice cream (price 10) and pizza (price 20). *MRT* will then be -10/20 = -0.5. We can interpret this such that you have to give up half a pizza if you want to have one more ice cream (or, vice versa, that you have to give up two ice creams, -20/10, to get one more pizza). To *transform* your basket into another basket with one more ice cream, you have to give up half a pizza. Note that this means that the price of ice cream measured in pizzas (instead of money) is half a pizza.

If income or prices change, the budget line will also change. Look at Figure 3.2 and the budget line B_1. If the price of good 1 rises from p_1 to p'_1, we can only buy a maximum of m/p'_1 of that good, but we can still buy m/p_2 of good 2. Consequently, the budget line rotates about the intercept with the Y-axis to B_2. If, instead, the price of good 2 rises from p_2 to p'_2, then B_1 rotates about the intercept with the X-axis to B_3.

When a price changes, *MRT* also changes since the slope of the budget line changes. If the price of ice cream rises from 10 to 20, *MRT* will be -20/20 = -1. Now, you have to give up a whole pizza to get one more ice cream. Note that this also means that the pizza has become cheaper, *relatively speaking*: You can now get one more pizza for just one ice cream, even though the price of pizza is unchanged.

Assume now that the prices are p_1 and p_2, as they were originally, but that the income increases to m'. We can then buy a maximum of m'/p_2 of good 2 and a maximum of m'/p_1 of good 1. B_1 consequently shifts to B_4. Note that the slope of B_4 is exactly the same as the slope of B_1, since the prices are unchanged: You have more money, but you still have to give up half a pizza if you want to have one more ice cream.

Figure 3.2: Changes in the Budget Line



If prices rise or if income falls, the area under the budget line becomes smaller. In the opposite cases, it becomes larger. The larger the area is, the more choices of consumption you have.

3.2 Preferences

The theory of preferences belongs to the most difficult parts of basic microeconomics, so take your time with this section. It is very important to both understand and be able to use preference-theory in the rest of the material.

You have probably heard the expression that "one should not compare apples and oranges," or something similar. The point here is precisely that one should do that, and even to compare anything with anything else. This is done through a **preference order**. We will assume that an individual always knows what she prefers: she prefers basket A to basket B, she prefers B to A, or she is **indifferent** between them. If all baskets are ordered accordingly, we have a preference order and such an order is valid for a certain individual.

Usually, the following four assumptions are made about preference orders:

- *Complete*. The individual can order all conceivable baskets of goods.
- *Transitive*. If the individual prefers A to B, and B to C, she also prefers A to C. In other words, there are no "circles" in preferences.
- *Non-satiation*. An individual always prefers more of a good to less. This assumption is a bit tricky. Suppose we think of pollution as a good. Is more pollution usually preferred to less pollution? No, obviously not. To get around this type of problem, we have to define the good in the opposite way: Instead of pollution, we define clean air to be the good. More clean air is better than less.
- *Convexity*. Suppose we have two baskets that an individual is indifferent between, A and B. She will then always prefer (or at least be indifferent between) baskets that lie between these two baskets. Say that she is indifferent between a basket consisting of (2 apples, 4 bananas) and (4 apples, 2 bananas). She will then, according to the assumption, prefer a basket of (3 apples, 3 bananas) to the other two (or, at least, be indifferent between all of them).

Are these assumptions true? Many people have debated the reasonableness of them. Are you, for instance, non-satiable? Which do you prefer: 2 liters of milk or 10,000 liters? Probably 2 liters. The rest will not fit into the refrigerator and will soon start to smell. It will also require a lot of work to get rid of them.

In many models, however, it is also assumed that there are no **transaction costs**. This means that, there are no costs to trading, except for the price of the goods. Examples of transaction costs are the cost of a stamp if you mail in an order, the effort it takes to go to the market where you can buy things, or the cost to hire a lawyer to go through a contract before you sign it. Models that include transaction costs become much more complicated, but, on the other hand, they also become more realistic. In the example, you would probably prefer 10,000 liters of milk if it would not cost you anything to sell them and immediately get rid of them. In a worst-case scenario, you would sell then at a price of 0, which should make you indifferent between 2 and 10,000 liters of milk.

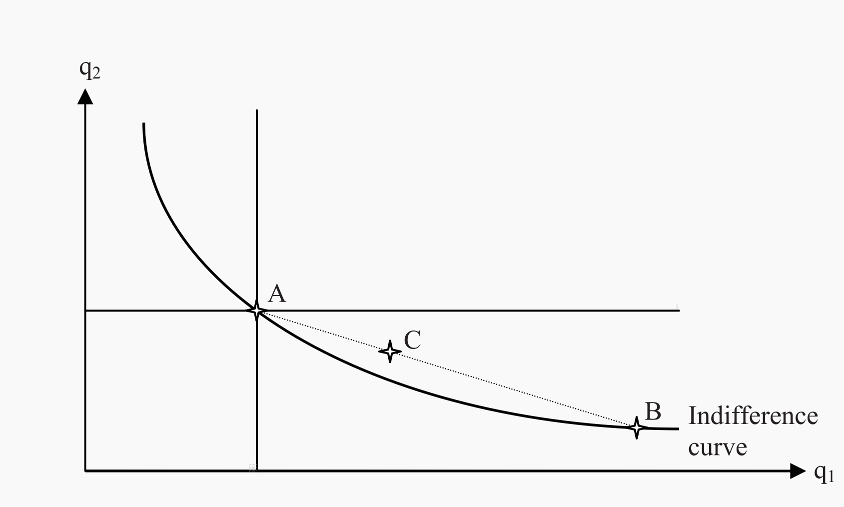
Transaction costs: Any cost, apart from the price of the good, that is associated with buying or selling it.

3.3 Indifference Curves

If we only have two goods, we can illustrate different baskets that the individual if indifferent between with **indifference curves**. All points on an indifference curve are baskets that the individual perceives are equally good. She is, in other words, indifferent between them. An example of a typical indifference curve is shown in Figure 3.3.

Indifference curve: A curve showing different combinations of two goods between which a consumer is indifferent. Similar to elevation contours on a map.

Figure 3.3: An Indifference Curve



After having made the four assumptions above, we can say a lot about what an indifference curve must look like. All points in the diagram (i.e. all possible combinations of good 1 and 2) correspond to a basket. Since the preferences are complete, there must be some preference curve that runs through any point in the diagram. Another way to say the same thing: Pick any point in the diagram; whichever point you picked, there is an indifference curve running through that point.

We now randomly select a point in the diagram, say point A. Since the individual is non-satiable, all points where she gets more of either good 1, or good 2, or of both are better for her. This corresponds to the grey area northeast of point A. Similarly, all points where she gets less, the grey area southwest to A, must be worse for her. Consequently, she cannot be indifferent between basket A and any point in the grey areas. Therefore, a preference curve that runs through A cannot also run through any point in the two grey areas. This means that an indifference curve will slope downwards. (See, however, the case of complementary goods in Section 3.6)

The assumption of convexity implies that the slope will become smaller and smaller as we move to the right. Convexity means that, if we randomly choose any other point on the indifference curve that runs through A, say point B, and then choose a point in between them, say point C, then point C must be better than (or at least as good as) A and B. C must therefore lie on a higher indifference curve than the one that runs through A and B. If this is true for any choices of A, B, and C, then the curve must slope less and less the farther to the right we get.

An economic interpretation of this criterion is that, the less one has of a certain good, e.g. the lower q_1 is, the less inclined one is to give up one more unit of that good. If that is so, then one will demand more of the other good to compensate for the loss of that one unit. We, consequently, have to increase q_2 more and more, the lower q_1 is, to ensure that the individual has the same utility. And as we need larger and larger amounts of good 2 to keep the individual indifferent after having lost one more unit of good 1, the slope of the indifference curve will increase as we move to the left (i.e. as we reduce good 1), and vice versa when we move to the right.

3.4 Indifference Maps

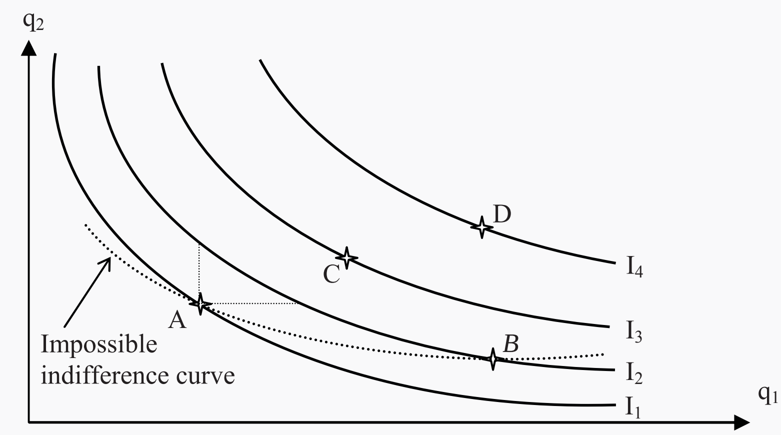
Since the preferences are complete, some indifference curve must run through each point, i.e. each basket. If we randomly choose four baskets, A, B, C, and D, there will be some indifference curve that runs through each point (see Figure 3.4).

If we move to the northeast in the diagram, the level of utility increases. Labeling the indifference curves I_1, I_2, I_3, and I_4, they must therefore represent higher and higher levels of utility. A collection of several indifference curves in one figure is called an **indifference map**. It is common to compare indifference maps to elevation contours on a regular map: It is like walking up or down a hill when one moves from one indifference curve to another.

Indifference map: A collection of indifference curves in a diagram.

After we have drawn the indifference curves, we can also compare points that do not lie to the northeast or southwest of each other. In the figure, point B is not to the northeast of point A, but it does lie on an indifference curve that is "higher" than the one that runs through A. Consequently, point B represents a basket that is better than the one represented by point A. We can also see this in the following way: Note that there are points on I_2 that lie to the northeast of point A (between the two dotted lines that originate at A). Those points must therefore be better than A. Moreover, all points on I_2 are equally good for the individual (since she, by definition, is indifferent between all of them). Consequently, point B represents a level of utility that is exactly the same as the points on I_2 that are to the northeast of A. Therefore, B must be better than A is. Note that if we argue that way, we have used the assumption of transitivity.

Figure 3.4: An Indifference Map



The indifference curves have the following four important properties:

- Baskets that are further away from the origin (the point (0,0) in the graph) are better than the ones closer to the origin.
- Every point has an indifference curve that runs through it, since the preferences are complete.
- Indifference curves cannot cross each other. This follows from the assumptions of transitivity and non-satiation.
- The indifference curves slope downwards. If they would slope upwards, we would violate the assumption of non-satiation.

3.5 The Marginal Rate of Substitution

Look at one of the indifference curves in Figure 3.4. The slope of the curves is of central importance. Think about what the slope means: If you choose some basket on one of the curves, how much would you be willing to give up of good 2 to get one more unit of good 1? If you would be willing to give up only a small quantity of good 2, the magnitude of the slope would be small, whereas if you were willing to give up a lot, it would be large.

Imagine that we have two individuals who each have 5 apples (good 1) and 5 bananas (good 2). To get one more apple, the first is willing to give up one banana, whereas the other is willing to give up two bananas. The first individual's indifference curve running through the point (5,5) will then slope less than the second individual's indifference curve. These two individuals have different tastes regarding apples and bananas.

The numerical value of the slope of an indifference curve, the magnitude of the slope, is called the **marginal rate of substitution** (*MRS*), and it can *approximately* be calculated as

Marginal rate of substitution: How much an individual is willing to pay for an additional unit of a good in terms of another good (rather than money). It corresponds to the slope of an indifference curve.

$$MRS = \frac{\Delta q_2}{\Delta q_1}$$

Here, Δq_1 and Δq_2 are the changes in quantity for good 1 and good 2, respectively. Individual 2 above was willing to give up 2 bananas to get one more apple. Then $\Delta q_2 = -2$, $\Delta q_1 = 1$, and $MRS = -2/1 = -2$. The fact that the indifference curves slope less and less to the right implies that MRS is decreasing.

Often, one does not keep the minus sign in *MRS*. It is then implicitly understood that one gets less (minus) of one good to get more (plus) of the other. Note that, if one leaves out the minus in *MRS*, one typically does so for *MRT* as well (see Section 3.1).

The expression for *MRS* above is only approximate. The smaller one chooses Δq_1, the better the approximation will be. For it to become completely exact, Δq_1 must be chosen infinitely small. This, in turn, makes it necessary to use derivatives. That, however, lies outside the scope of this book. Note that the word "**marginal**" means "infinitely small." You will hear that word many times in economics.

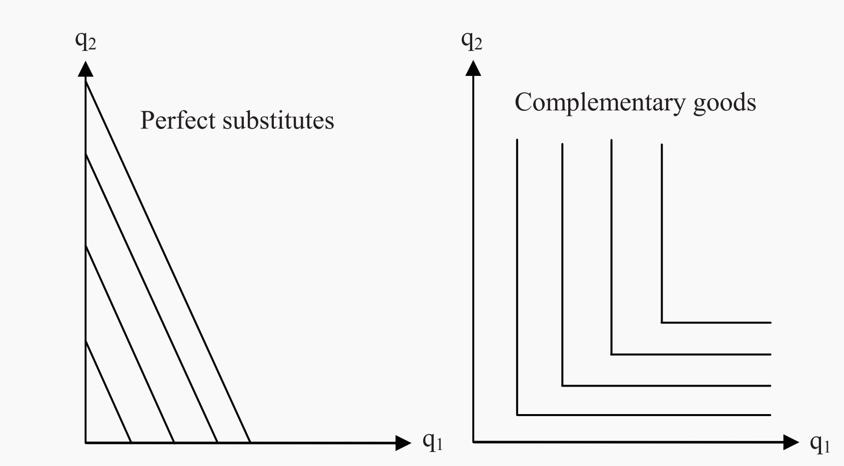
Marginal: An infinitely small change. Usually, one speaks of a change of one unit.

3.6 Indifference Curves for Perfect Substitutes and Complementary Goods

An example of (almost) perfect substitutes we have already seen is green and blue pens. **Perfect substitutes** have the property that, instead of decreasing MRS, they have constant *MRS*. This means that they have the same slope everywhere, i.e. they are straight lines sloping downwards to the right (see Figure 3.5). In the case of the pens we have that $MRS = 1/1 = 1$ (where we have dropped the minus sign), but *MRS* could be any number. The defining criterion for perfect substitutes is that *MRS* is constant.

Perfect substitutes: Two goods that are possible to use interchangeably, so that a consumer is indifferent between them. This causes the indifference curves to be straight lines.

Figure 3.5: Perfect Substitutes and Complementary Goods



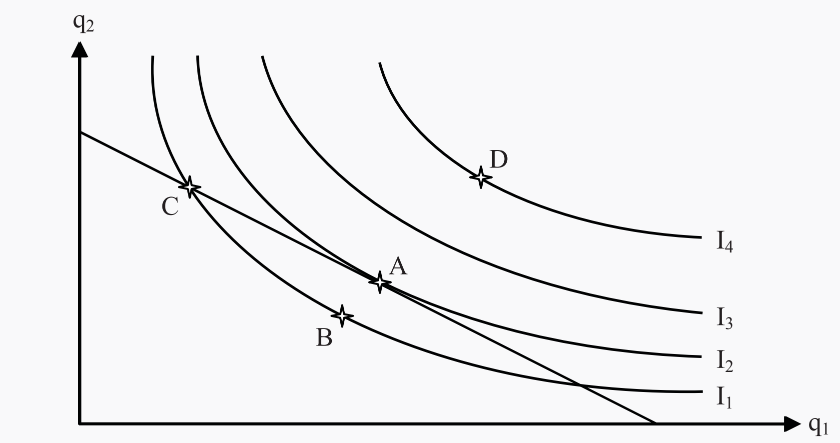
The example of **complementary goods** we saw before was right and left shoes. One has no use for one without the other. This fact causes the indifference curves to become L-shaped (see Figure 3.5). Assume we have two left shoes and two right shoes. Even if we get many more right shoes, we will still have the same utility as before. The indifference curves are therefore vertical along q_2 and horizontal along q_1, and the only way to reach a higher level of utility is to get more of *both* good 1 and good 2.

Complementary goods: Goods that go together, so that a consumer needs them both to have any use of them. This causes the indifference curves to be L-shaped.

3.7 Utility Maximization: Optimal Consumer Choice

So far, we have described two of the three parts we need to explain how consumers choose goods. First, we described their limitations (scarceness; income; the budget line), and then we described their preferences (desires, taste). Now, we put these two parts together. Moreover, if we add the assumption that the consumer will maximize her utility, we will be able to predict which basket of goods she will choose: She will choose a point on an indifference curve that she can afford and that gives her maximum utility. This usually, but not always, singles out one point.

Figure 3.6: Utility Maximization



In Figure 3.6, we see the indifference curves from Figure 3.4 and the budget line from Figure 3.1 combined. Which of the points A – D is an optimal, utility maximizing, choice?

- Is, for instance, point B optimal? No, A is better than B since A is on a higher indifference curve. The consumer can also afford A, since A is on the budget line.
- Is C optimal? No, C is on the same indifference curve as B, and is therefore as good as B. However, A is better than B and, consequently, A must be better than C.
- Is D optimal? D is on a higher indifference curve than any of the other baskets, A – C. It therefore produces the highest level of utility. However, the consumer cannot afford D since it lies outside the budget line. Therefore, D is not an optimal choice.
- Is A optimal? Yes, A is the only basket that, given the consumers indifference curves and budget line, produces a maximum level of utility. All other points that lie on or below the budget line produce lower levels of utility. At point A, an indifference curve just touches the budget line (i.e. the budget line is a tangent to the indifference curve).

Point A has an interesting property. In that point, the budget line and the indifference curve have exactly the same slope. Remember that the slope of the budget line is (minus) the quotient between the prices $-p_2/p_1$, which we called the marginal rate of transformation (*MRT*), and that the slope of the indifference curves is the marginal rate of substitution (*MRS*). A criterion for being exactly at the point where we maximize utility is then that

$$MRT\left(=-\frac{p_1}{p_2}\right)=MRS$$

However, there are cases when the point of utility maximization does not fulfill this criterion. Look for instance at the indifference curves for perfect substitutes and complementary goods. If you fit a budget line into any of those graphs, you will find that the criterion *MRT* = *MRS* usually is not fulfilled. For perfect substitutes, the consumer will usually maximize her utility at either the X-axis or at the Y-axis, where she only consumes one of the goods (this is called a **corner solution**; the opposite is called an **interior solution**). If the budget line is parallel to an indifference curve, the consumer can choose any point on the line. She can afford them all, and she is indifferent between all of them.

Corner solution: The consumer chooses to consume only one of the goods, so that she ends up in a corner in the graph.
Interior solution: She chooses to consume at a point in the graph where there are no particular restrictions (as there are in a corner solution)

For complementary goods, she will maximize her utility at a point where an indifference curve has a corner. In such a point, the curve has no defined slope (since it has different slopes to the left and to the right) and, hence, *MRS* does not exist.

Use the following strategy to find the point of utility maximization:

- Draw the budget line.
- Find the indifference curve that just barely touches the budget line (i.e. an indifference curve that the budget line is a tangent to). In most cases, there is only one such indifference curve. All other indifference curves either crosses the budget line or does not touch it at all. Be careful, however, to check if there exists a corner solution.
- The point of utility maximization is the point of tangency (or the corner solution).

3.8 More than Two Goods

The method we have described uses only two goods. So, what do we do if we have more goods? One method we can use, if we want to use graphs in the same spirit as before, is to define a sort of composite good as "everything else," alternatively as "money" (since money represents possibilities to consume something else). Then we can draw a graph where good 1 is the good we want to analyze and good 2 is "everything else."

Another strategy that is used in more advanced textbooks is the so-called **utility function**. This mathematical function assigns a numerical value to the utility level of a certain consumption choice. For two goods, the utility of consuming a certain combination of them could be:

Utility function: A mathematical function that gives a numerical value that corresponds to the level of utility a consumer attains.

$$U(q_1,q_2)=q_1*q_2$$

The utility, *U*, of consuming, for instance, 2 units of good 1 and 3 units of good 2 will then be 2*3 = 6. The number 6 does not mean much more than that it is better than, for instance, 4 but worse than, for instance, 14. The analysis is then carried out such that one maximizes the value of *U*, given that the cost of buying must not exceed the budget. If you continue to study microeconomics, the analyses will become increasingly more concentrated on utility functions and less on graphical descriptions. In Chapter 6, we will briefly use a utility function in the analysis of attitudes towards risk.

4 Demand

4.1 Individual Demand

We will now show how to use the theory of preferences from last chapter, to derive an individual's demand curve. Remember that the consumer's budget line can change because of changes in prices or because of changes in income. Here, we will assume that the preferences themselves do not change. This makes us able to derive both the demand curve that we used in Chapter 2, and the so-called Engel curve, which shows how demand depends on income.

4.1.1 The Individual Demand Curve

As we showed in Chapter 3, it is possible to find the point of utility maximization if one knows a consumer's preferences, the prices of the goods, and her budget. Let us now do that, but vary the price of good 1 and see what effect that has on, q_1, the quantity demanded.

Suppose we hold the price of good 2 (which you can think of as "all other goods") constant. Then the effect of varying the price of good 1 will be that the budget line rotates about the intercept on the Y-axis and intersects the X-axis at different points m/p_{1i}, where p_{1i} is the price one has chosen for good 1.

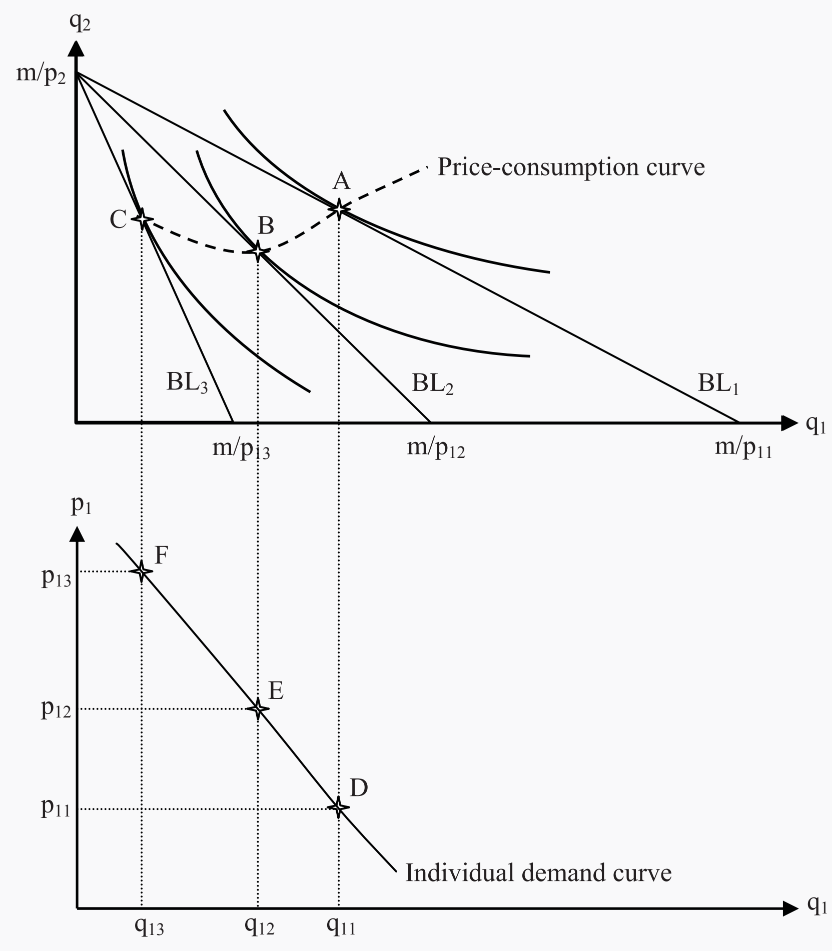
Look at the upper part of Figure 4.1. Suppose the price of good 1 is initially p_{11}. Then the budget line is BL_1. We find the indifference curve that just touches that budget line and label the point where it does so, point A. If we would raise the price of good 1 to p_{12}, the possible choices become limited to BL_2 (that intersects the X-axis in m/p_{12}) and then the consumer maximizes her utility in point B. If we continue to raise the price to p_{13}, and repeat the maximization, we get point C. If we would repeat this procedure for all possible prices, we would get a curve that is called the **price-consumption curve**. It shows how the optimal choice of quantity of good 1 varies with the price of that good, given that preferences, other prices and the income are held constant.

Price-consumption curve: A curve that shows how a consumer chooses to consume at different prices.

As you can see in the figure, the consumer will usually buy less of the good when the price increases. This is, however, not necessary. To see that, imagine that the indifference curve that runs through point B had been steeper. If it had been steep enough, it would touch BL_2 so far to the right that it would also be to the right of point A.

Now we want to find the demand curve for good 1. To that end, we indicate the prices we used for good 1 on the Y-axis in the lower graph of the figure, i.e. p_{11}, p_{12}, and p_{13}. Then we check which are the corresponding quantities demanded in the upper graph, at points A, B, and C, and indicate them on the X-axis in the lower diagram. (Note that both diagrams have q_1 on the X-axis.) After that, we find the points where the quantities and the corresponding prices in the lower diagram intersect, the points labeled D, E, and F. Finally, we draw a line through those points and fill in for all those numerous points for which we have not done the analysis. This curve is the individual's demand curve for good 1.

Figure 4.1: Derivation of an Individual Demand Curve



4.1.2 The Engel Curve

In the previous section, we showed how it is possible to derive the relation between the price and the quantity demanded for a certain good. Now, we will instead show how to derive the relation between *income* and the quantity demanded. The resulting curve is called the **Engel curve**.

Engel curve: A curve that shows the relation between income and quantity demanded. Compare to the demand curve.

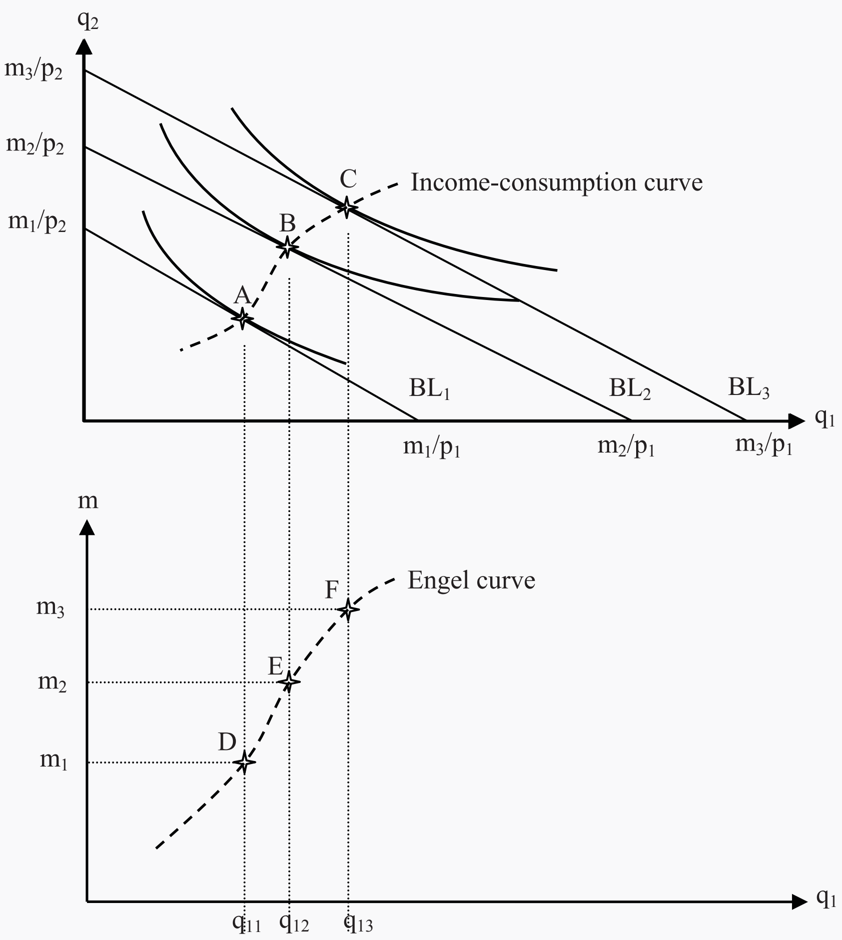
Look at Figure 4.2. Just as in the previous case, we start with the individual's maximization problem where she must choose quantities of good 1 and good 2. (Again, think of good 2 as "all other goods.") However, instead of varying the price, we now vary the income m. This means that the budget line will shift outwards for higher incomes and inwards for lower incomes. We assume that preferences and prices are unchanged. For the increasingly higher incomes m_1, m_2, and m_3, the budget lines become BL_1, BL_2, and BL_3.

In the same way as before, we find the utility maximization points for each budget line: points A, B, and C. If we would do that for all possible incomes, we would get the so-called **income-consumption curve**. That curve shows the optimal consumption of good 1 and good 2 at different incomes, given preferences and prices.

Income-consumption curve: A curve that shows the relation between income and consumption.

Similarly to before, we indicate the quantities that correspond to points A, B, and C, i.e. q_{11}, q_{12}, and q_{13} in the diagram below. Then we indicate the incomes m_1, m_2, and m_3 on the Y-axis, and the points where the incomes intersect the corresponding quantities: points D, E, and F. Thereafter, we draw a line through the points of intersection, as it would probably have looked if we had performed the same procedure for the points in between. The resulting curve is the so-called Engel curve, and it shows how the optimal consumption of good 1 varies with the income, given preferences and prices.

Figure 4.2: Derivation of the Engel curve



4.2 Market Demand

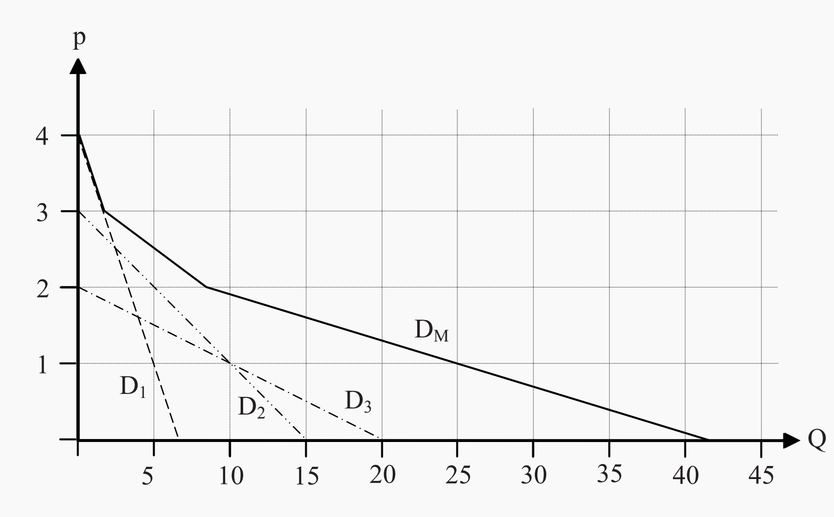
The market's demand consists of all individuals' demand. To find the market demand curve, we have to sum up the demand of all individuals for each price.

Suppose, for instance, that we have found demand curves for three different individuals, and that these three individuals together are the whole market. In Figure 4.3, their demand curves (for simplicity, they are all straight lines) are labeled D_1, D_2, and D_3.

If the price of the good is 4, all individuals demand a quantity of 0, but at a price of 3, the first individual demands 2 units. Since the others do not demand anything, the market's total demand is those 2 units. For prices between 3 and 4, the market's demand coincides with D_1, i.e. the demand curve of the first individual. (A straight line from point (0,4) to point (2,3))

When the price is 2, the first individual demands 4 units and the second demands 5 units. The market's total demand is then 9 units. For prices between 2 and 3, total demand is $D_1 + D_2$. It will then be a straight line beginning in point (2,3) and ending in point (9,2).

Figure 4.3: Market Demand



When the price is close to 0, all individuals demand it: The first demands 7 units, the second demands 15 and the third demands 20. Total demand is then 42 units. For prices between 0 and 2, total market demand is $D_1 + D_2 + D_3$. It will then be a straight line starting in point (9,2) and ending in point (42,0). Note that we should not really allow a price of 0. Demand would then be infinite as more of the good is, by assumption, always better.

The **market demand curve**, D_M, will consequently be the sum of the individual demand curves. If the individual demand curves are straight lines, the market demand curve will become a succession of straight lines, where a break signals that a new consumer starts demanding the good at that price.

Market demand curve: A curve that shows the demand for the whole market at different prices.

4.3 Elasticity

Suppose we want to study the effects a price change has on the demand of a good. It is practical to do that in terms of percentages: If the price rises by one percent, how many percentages will demand change?

More generally, one can study how many percent any one variable changes when another variable changes by one percent. This is called **elasticity**. The types of elasticity that are used the most are price elasticity, income elasticity and cross-price elasticity.

Elasticity: A measure of how sensitive a variable is to changes in another variable.

4.3.1 Price Elasticity

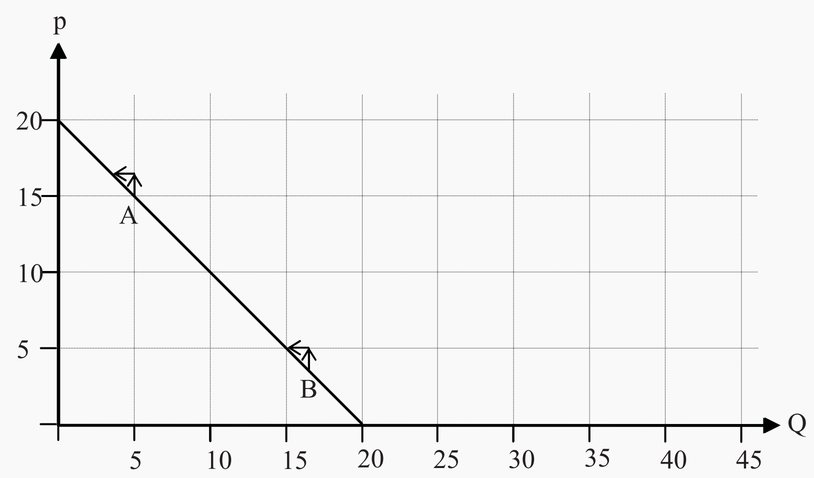
Price elasticity (of demand) is how many percent demand changes if the price changes one percent. We use the notation e_p for price elasticity, Q for quantity demanded, ΔQ for the change in quantity demanded, p for the price, and Δp for the change in price. The price elasticity can then be calculated as

$$e_p = \frac{\Delta Q / Q}{\Delta p / p}$$

Note that the expression in the numerator is the relative change in quantity (relative to the level) and the expression in the denominator is the relative change in price.

The elasticity is usually different depending on where on the demand curve it is calculated, even if the demand curve is a straight line. To see that, look at Figure 4.4. We start with point A where the price is 15 and the quantity demanded is 5. For simplicity, we choose Δp to be 1. If the price increases with 1, the quantity demanded decreases by 1 (ΔQ = -1), i.e. we move one step upwards and one step to the left following the arrows. The price elasticity at point A is consequently e_p = (-1/5)/(1/15) = -3. If we perform the same exercise at point B, we get that e_p = (-1/16)/(1/4) = -0.25.

Figure 4.4: Price Elasticity at Different Quantities Demanded



Income elasticity: How sensitive demand is to changes in income.

The price elasticity also depends on which type of good we study. Most importantly, one distinguishes between cases where the price elasticity is less than -1 or between -1 and 0. If it is less than -1 that means that the quantity decreases more (in percent) than the price increases (again, in percent), which is called **elastic demand**. If it is between -1 and 0 it means that the quantity decreases less than the price increases, which is called **inelastic demand**.

$0 < e_p$	Positive price elasticity
$-1 < e_p \leq 0$	Inelastic demand
$e_p = -1$	Unity elastic demand
$e_p < -1$	Elastic demand

Note that the good in Figure 4.4 has elastic demand at point A but inelastic at point B. Also note that $0 < e_p$ is *very* unusual (see, however, Section 5.2). Often, one does not include the minus sign. It is then implicitly understood that, for instance, a price elasticity of 3 means that demand *decreases* by 3 percent if the price increases by 1 percent.

4.3.2 Income Elasticity

Correspondingly, **income elasticity** (of demand) is the percentage change in demand if income changes one percent:

$$e_m = \frac{\Delta Q / Q}{\Delta m / m}$$

Here, e_m is income elasticity, and m and Δm are income and change in income, respectively. Similarly to price elasticity, goods are grouped depending on their income elasticity:

$e_m < 0$	Inferior goods
$0 < e_m$	Normal goods

$1 < e_m$	Luxury goods
$0 < e_m < 1$	Necessary goods

A **normal good** ($0 < e_m$) is a good one buys more of if income increases. An **inferior** good ($e_m < 0$) is a good one buys less of when income increases. These goods are typically of low quality, and one decreases one's consumption of them as one can afford better quality.

Normal goods are further divided into **necessary goods** and **luxury goods**. If income increases with one percent, one buys less than one percent more of a necessary good, but more than one percent more of a luxury good.

Normal good: A good one buys more of if income increases.
Inferior: A good one buys less of if income increases.
Necessary good: If income increases, one buys more of it, but not as many percentages more as the increase in income.
Luxury good: If income increases, one increases consumption by more percentages than the income.
Cross-price elasticity: How sensitive demand is to price changes in another good.

4.3.3 Cross-Price Elasticity

Cross-price elasticity is defined as the percentage change in demand on a good if the price of *another* good changes with one percent:

$$e_{12} = \frac{\Delta Q_1 / Q_1}{\Delta p_2 / p_2}$$

Here, e_{12} is the cross-price elasticity between good 1 and good 2; Q_1 and ΔQ_1 are quantity demanded and quantity change for good 1, whereas p_2 and Δp_2 are price and price change on good 2. Again, goods are grouped depending on their cross-price elasticity (compare with Figure 3.5):

$e_{12} < 0$	Complementary goods
$e_{12} = 0$	Independent goods
$0 < e_{12}$	Substitute goods

Suppose the price of good 2 rises by one percent. If that leads to a decrease in the demand for good 1 ($e_{12} < 0$) then good 1 and good 2 are probably goods that go together in some way: complements. If, instead, it leads to an increase in the demand for good 1 ($0 < e_{12}$) then good 1 is probably something one can buy instead of good 2: a substitute. (Also, compare to Section 2.1.1.)

5 Income and Substitution Effects

In Section 4.1.1, when we derived the individual demand curve, we saw how the quantity demanded changed when the price changed. We will now use consumer theory to perform a slightly more complicated analysis of a price change.

Suppose that we have a consumer, with a certain income, who has to choose between different quantities of good 1 and good 2 (which, again, can be thought of as "all other goods") in such a way that she maximize her utility. If the price of good 1 falls, we get two different effects:

- Since the price of good 1 falls, that good becomes cheaper relative the other good. This means that the marginal rate of transformation (*MRT*; the slope of the budget line) changes. Say that the prices of both goods initially are 1. The relative price is then 1/1 = 1. If the price of good 1 falls to 0.50, the relative price becomes 0.50/1 = 0.50. The consumer can now exchange one unit of good 2 for two units of good 1, and therefore good 1 becomes more attractive to her. As a result, she consumes more of the good. This effect is called the **substitution effect**.

Substitution effect: The effect on demand that depends of the change in relative prices.

- The purchasing power of the consumer becomes larger because of the drop in the price. She can now buy as much as she did before the price changed, and still have money left. That extra money she can spend on both good 1 and on good 2. This is called the **income effect**.

Income effect: The effect on demand that depends on the fact that one can afford more after a drop in the price, and vice versa.

In reality, we can only observe the total effect of the price change, i.e. how much more or less the consumer buys of the good. However, we will now see that it is possible to split up the total effect into the substitution- and income effects. Depending on whether good 1 is a normal or an inferior good, we get two different cases.

5.1 Normal Good

Assume we have the same case as we did earlier: A consumer chooses between good 1 and good 2. Giver her income, m, the prices of the goods, p_{11} and p_2, and her preferences, she chooses that basket of goods that maximizes her utility. In Figure 5.1, this means that she initially chooses point A.

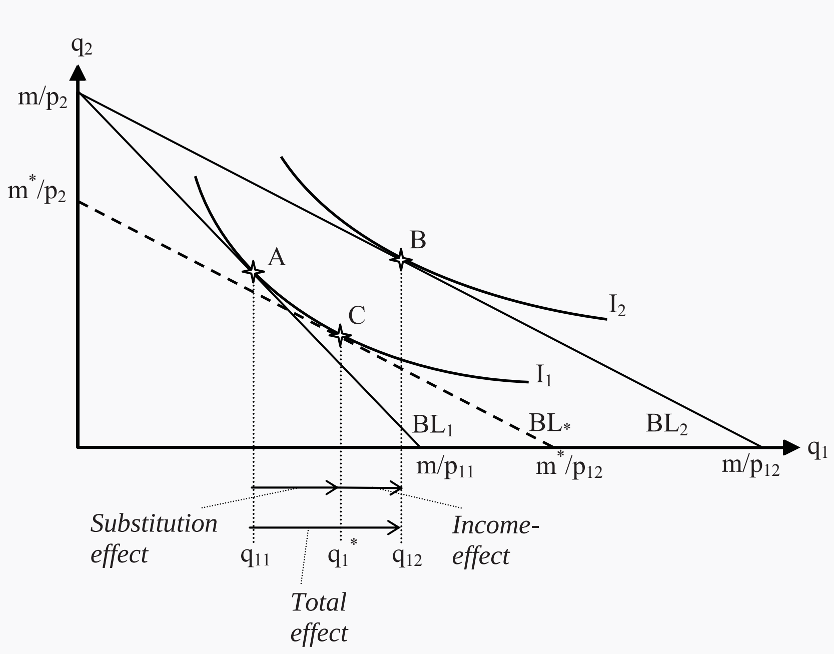
If the price of good 1 falls from p_{11} to p_{12}, the budget line rotates outwards from BL_1 to BL_2. When the consumer chooses a new basket, she ends up in point B. Her consumption of good 1 has consequently increased from q_{11} to q_{12}, which is the total effect.

We now ask ourselves how much of the change in quantity from q_{11} to q_{12} that depends on the income effect (i.e. on the increase in purchasing power) and how much that depends on the substitution effect (i.e. on the change in the slope of the budget line). To answer this question, we first ask another question: If only the relative prices had changed, without the consumer getting any increase in utility, what effect had we then seen.

If the relative prices change, the slope of the budget line changes. All budget lines that have the same relative prices as BL_2 must also have the same slopes as that budget line. Furthermore, for the consumer to have the same utility as before, she must consume on the same indifference curve as she did before, i.e. on I_1. We therefore construct an imaginary budget line, BL_*, that has the same slope as BL_2 and that, just as BL_1, is a tangent to I_1. (However, since it has a different slope than BL_1, it must touch I_1 at different point than that budget line does.) If this had been the real situation, the consumer would have chosen point C. She had then increased her consumption of good 1 from q_{11} to q_1^*. At the same time, she would have decreased her consumption of good 2. This substitution from good 1 to good 2 depends on the change in the relative price, but it does not result in any change in the level of utility. This part is the substitution effect.

The remaining change, from q_1^* to q_{12}, is the part that depends on the increase in the consumer's purchasing power. As she moves to a higher indifference curve, from I_1 to I_2, she increases her utility. This part is the income effect.

Figure 5.1: Income and Substitution Effects for a Normal Good



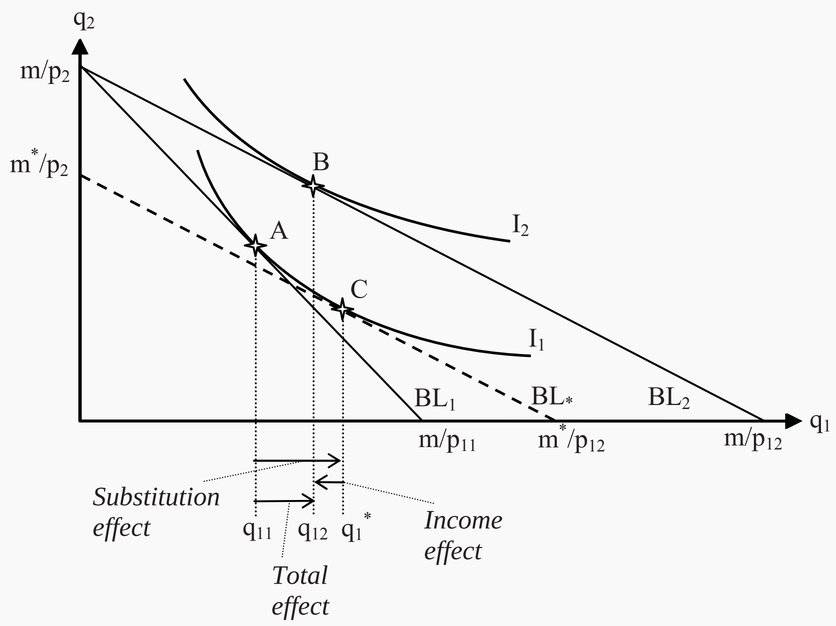
5.2 Inferior Good

The strategy to find the income- and substitution effects for an inferior good is exactly the same as for a normal good, but the result will look slightly different. As previously mentioned, an inferior good is a good one buys less of if one's income increases. The underlying reason for that is to be found in the preferences. As one becomes wealthier, one can afford to buy something of higher quality instead. This preference will have an effect on the shape of the indifference curves.

This time, when we split up the total effect into a substitution effect and an income effect, the income effect for the inferior good is negative. The substitution effect is always positive, which means that we get two cases depending on whether the negative income effect is smaller or larger in magnitude than the always-positive substitution effect. Goods that belong to the latter case are called **Giffen goods**, and these are a very rare kind of goods. Their distinguishing feature is that one buys more of them if the price rises. In Section 2.1.1, we said that the demand curve almost always slopes downwards. Giffen goods are consequently an exception from that rule.

Giffen good: A good one buys less of when the price falls, and vice versa.

Figure 5.2: Income and Substitution Effects for an Inferior Good

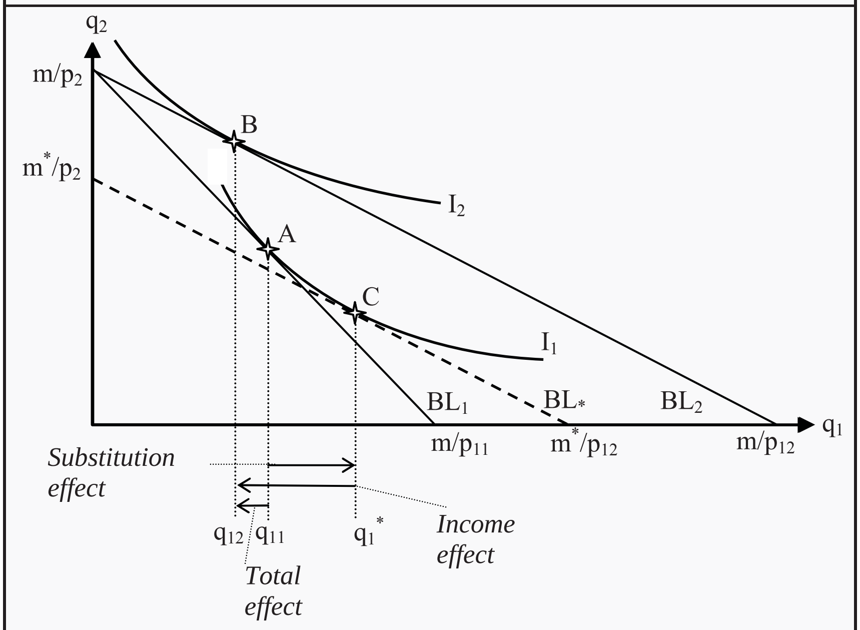


In Figure 5.2, we have almost the same situation as in 5.1. The difference is that the consumer's indifference curve I_2 has been changed so that it touches the budget line BL_2 at a point between points A and C. This change makes the income effect negative and the total effect is smaller than before.

In Figure 5.3, the indifference curve I_2 has been changed again, so that it touches BL_2 at a point to the left of point A. The income effect now becomes very negative, so negative that it dominates over the substitution effect. The total effect thereby also becomes negative and we have a Giffen good.

Note, however, that the consumer does increase her utility. This can seem strange, as the total effect is that she consumes less of the good analyzed (and we have assumed that more is always better). The drop in the price of the Giffen good means that the consumer can afford to buy more of other goods. Furthermore, these other goods function as substitutes for the Giffen good. Hence, the increase in utility. The increase in consumption of good 2 can be read off as the distance between A and B on the Y-axis.

Figure 5.3: Income and Substitution Effects for a Giffen Good



6 Choice under Uncertainty

The situations we have discussed up to this point have all lacked any elements of uncertainty. Individuals and firms have made their choices knowing what the outcomes would be. That is, of course, very unrealistic. Most of the time, we cannot be certain about which consequences our actions will have, even though we can perhaps know which consequences they will *probably* have.

A few examples of important decisions under uncertainty are:

- Buying a house or an apartment. You know what you pay for it, but what will it be worth when you sell it? What happens if the house burns down?
- Investments in an education. It is often easy to get statistics on today's salaries but in the future, they might change substantially.
- A firm invests in a new factory. Will the goods produced in the factory still be in demand in the future? Think, for instance, about the computer market where development is very rapid.

It is common in economics to view uncertainty as a sort of lottery. In a lottery, one often knows which outcomes are possible: There might, for instance be a list showing how much you can win. It is also possible to calculate the probabilities for the different outcomes. In the situations covered here, we will also assume that this is the case.

6.1 Expected Value

In statistics, "expected value" is a technical term. Suppose we toss a coin. If "heads" comes up, we win 5; if "tails" comes up we lose 5. The expected value of that lottery is then

$$\begin{aligned} EV &= \Pr(\text{"heads"}) * \text{value}(\text{"heads"}) + \Pr(\text{"tails"}) * \text{value}(\text{"tails"}) \\ &= 50\% * 5 + 50\% * (-5) \\ &= 0 \end{aligned}$$

Here, Pr(.) is the probability that the event within the parentheses will occur. If there are more than two possible outcomes, the expected value is the probability of each outcome multiplied with the value of that outcome, and then summed together. Note that the expected value need not to be something you would expect to occur. In the lottery above, we do not expect the outcome to be 0. We expect it to be either +5 or -5.

The expected value can be seen as a sort of average over the outcomes, where an outcome with high probability has a higher weight than one with lower probability. A lottery with an expected value of zero is called a **fair lottery**. Note that most real-world lotteries are not fair.

Fair lottery: A lottery with an expected value of zero.

6.2 Expected Utility

Consider a case where an agent has to choose between several alternatives, all of which will lead to an uncertain outcome. A naïve method could then be to analyze them as if they were lotteries, and then choose the one with the highest expected value. There are several reasons why such a method would not be a good one.

To begin, we need to define a utility function (compare to Section 3.8) over wealth. Usually, more wealth is better for an individual but as she becomes wealthier additional wealth matters less and less. The utility an individual has of wealth is often written as

$$U = U(W)$$

U stands for utility and W for wealth. The expression can then be read as "the utility level is a function of wealth." What form the utility function takes varies between individuals, but a function that is often used for illustrations is

$$U = \sqrt{W}$$

In Figure 6.1, we have drawn this function (sqrt = "the square root"). Note that the slope of the function becomes less and less steep. That means that the individual, as we just noted, receives less utility of extra wealth as she becomes wealthier. Note that how much extra utility the individual gets from a small increase in wealth corresponds to the slope of the utility function. The slope is called the **marginal utility**, MU.

Marginal utility: The additional amount of utility an agent receives from an additional amount of wealth.

We differentiate between three different kinds of utility functions

- *Diminishing marginal utility*. The slope decreases with increased wealth, such as the utility function in Figure 6.1.
- *Constant marginal utility*. The utility function is straight line, i.e. the slope is constant.
- *Increasing marginal utility*. The slope increases with increased wealth.

If we use the utility function together with expected value, we can calculate the *expected utility*. Suppose our wealth is 5. If we participate in the lottery above, we will either win 5 or lose 5, both with a 50% probability. The result will then be that we have either 0 or 10. The utility we would have of these outcomes, and the expected utility is then

$$\mathrm{U}(0) = \sqrt{0} = 0$$
$$\mathrm{U}(10) = \sqrt{10} \approx 3{,}2$$
$$\mathrm{E}(\mathrm{U}) = 50\% * (0) + 50\% * (3{,}2) = 1{,}6$$

E(U) stands for expected utility. Earlier, we calculated the expected value of the lottery to 0. Our expected wealth, as opposed to expected utility, is then 5 + 0 = 5 (the sum of what we have plus the expected value of the lottery). Note now that, the utility we have from a certain wealth of 5 is

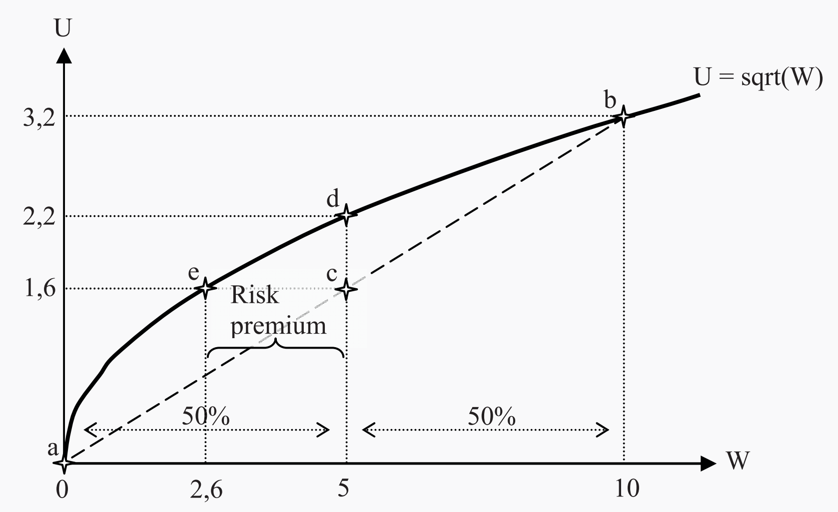
$$U(5) = \sqrt{5} = 2{,}2$$

The utility from getting the expected value with certainty (the utility of 5) is usually higher than the expected utility of participating in a fair lottery (the utility of either 0 or 10, both with a 50% probability). We will study this more closely in the next section.

6.3 Risk Preferences

Study Figure 6.1 again. We have a wealth of 5 plus an uncertain outcome of a lottery. Together, these give us an end wealth of either 0 or 10. We have indicated these values on the X-axis. The corresponding utilities we have indicated with points a and b. The expected utility of the wealth plus the lottery will be a point somewhere along a straight line from a to b, depending on the probability of each outcome. In the present case, the probability for each outcome is 50% and then the expected utility will be a point exactly half-way in between a and b, i.e. in point c where the utility is 1.6. (With other probabilities, we would have ended up in another point on the same straight line.) We have now illustrated the expected utility of an uncertain outcome of either 0 or 10.

Figure 6.1: Utility Function



What if we chose not participate in the lottery at all? In that case, we would keep a certain wealth of 5. The expected utility is unchanged: The expected value of a certain 5 is 5, and the expected value of a certain 5 plus the lottery is also 5. However, the utility of a certain wealth of 5 is 2.2, corresponding to point d in the figure. The outcome that, participating in the lottery gives less utility than not participating, depends on the fact that the utility function slopes less and less steep as wealth increases, i.e. that we have diminishing marginal utility. A person with such a utility function will always prefer not to participate in a fair lottery, and she is said to be risk averse.

Depending on which type of marginal utility an individual has (compare to above), we can classify her attitude towards risk:

- **Risk averse** (diminishing marginal utility). Prefers not to participate in a fair lottery. Most people, if not all, are risk averse. Note, however, that this theory (at least in the basic version presented here) is unable to explain why many people in real life participate in lotteries.
- Risk neutral (constant marginal utility). Is indifferent.
- Risk loving (increasing marginal utility). Prefers to participate in the lottery. A very unusual property!

Risk averse: An agent who dislikes risk.

6.4 Certainty Equivalence and the Risk Premium

Look at Figure 6.1 again. We have already seen that an ordinary person (i.e. a risk averse person) prefers not to participate in the lottery. One may then ask which level of certain wealth she would value as much as participating in the lottery.

As we saw before, her utility of participating is 1.6 (point c). The question is then which wealth would give her that same utility. Follow the line from 1.6 to the utility function and you will end up at point e, corresponding to a certain wealth of 2.6. This individual is, consequently, indifferent between participating in the lottery and having a certain wealth of 2.6. The value 2.6 is then said to be **certainty equivalent** to participating in this lottery.

Certainty equivalent: The amount of wealth an agent has the same utility of as another uncertain amount of wealth.

Since she now has a wealth of 5, she is, in other words, prepared to pay 5 - 2.6 = 2.4 to avoid the lottery. Alternatively, if her wealth had been 2.6, how much would we have had to pay her in order to make her willing to participate in the lottery? The answer is the same: 2.4. This amount is called the **risk premium**.

Risk premium: How much an agent is maximally willing to pay not to participate in a lottery (uncertain outcome).

6.5 Risk Reduction

Since most people are risk averse, they want to reduce risk. That is often achieved by pooling the risk and sharing it. This is, for instance, the idea behind insurance, where the risks are shared between many people.

7 Production

A producer uses raw materials, capital, and labor to produce goods and services. Here, we will present a simple model for how they decide how much to produce and which technology to use for the production.

A large part of producer theory is very similar to consumer theory. Basic assumptions for consumer theory are that consumers have a goal to maximize their utility, but that they have restrictions due to limited income and prices. Producers also have a goal. They wish to maximize their profit. They also have restrictions. These are, for instance, the costs of labor and capital; but they also have restrictions regarding the technology of production.

An aspect that will also prove important for a firm is the amount of competition they face: Do they have one, a couple, or many competitors? Alternatively, do they not face any competition at all? We will study different market forms in later chapters.

- The producers have certain restrictions. Primarily because different combinations of **inputs** (labor and capital) have different associated costs.
- The firm operates in a market that, in turn, has certain structures that the firm cannot influence.
- Technology: Different combinations of input produce different quantities of goods.
- We will distinguish between production in the short run and in the long run. In the short run, the quantity of available capital is fixed; in the long run, both labor and capital are variable.
- Given production and restrictions, the producer maximizes her profit.
- Another important question is how large a firm should be. The important concept here is returns to scale. How firm size affects how efficiently it can transform input to output.

Input: What a producer uses to produce goods: Labor, capital and raw materials.

7.1 The Profit Function

We will use a very simple model of a firm. It produces a single good, and the most important input factors are labor and capital (for instance machines). The producer has a certain cost, *C*, and a certain **revenue**, *R*. Her profit, π, can then be written as the difference between revenue and cost:

Revenue: What a firm is paid for goods and services.

$$\pi = R - C$$

The revenue, in turn, depends on the price of the good and how large quantity she sells,

$$R = p * q.$$

The costs are, of course, also dependent on how large quantity she produces, but usually in a more complicated way. The profit can therefore be written as

$$\pi = p * q - C(q),$$

where *C(q)* means that the cost, *C*, is a function of the quantity, *q*.

We will analyze each of the three variables, *p*, *q*, and *C*, in the profit function. The price is often set by the market. *C* depends both on the costs of the input factors and the quantity produced. The firm therefore chooses the quantity that maximizes profit. In this chapter, we will analyze production. In Chapter 8, we will look at the costs of production, and in later chapters, we will study different market forms.

7.2 The Production Function

The quantity the producer will produce of the single good, depends on the number of working hours, *L* (for Labor), and the amount of capital, *K*, that she uses. *q* is consequently a function of *L* and *K*:

$$q = f(L, K)$$

The letter *f* in the expression means that we have a *function* of *L* and *K*. That could mean 'L + K', 'L*K' or 'L^2 + 9*ln(K)' to just mention a few arbitrary examples. Which function that is appropriate depends on the technology, which good one produces, etc.

Note that we, of course, assume that the production of *q* units is done in the most efficient way possible. If it, for instance, is possible to produce 10 units of the good using a certain combination of *L* and *K*, it is also possible to produce only 9 units with the same combination. However, that production cannot be efficient, since one is wasting resources that could be used for the production of one more unit.

7.2.1 Average and Marginal Product

Before we begin the analysis, we need to define a few concepts that will be important later on. The **average product** is how much of the good that on average is produced by a certain input, *L* or *K*. We therefore have that the average products, *AP*, for labor and capital, respectively, are

Average product: The quantity of goods that, on average, is produced per hour worked or per unit of capital.

$$\begin{cases} AP_L = \dfrac{q}{L} \\ AP_K = \dfrac{q}{K} \end{cases}$$

The **marginal product**, *MP*, is how much extra quantity that can be produced if one increases the amount of either labor or capital with one unit, keeping the other one constant:

Marginal product: How much the quantity produced increses if either labor or capital is increased by one unit.

$$\begin{cases} MP_L = \dfrac{\Delta q}{\Delta L} \\ MP_K = \dfrac{\Delta q}{\Delta K} \end{cases}$$

The expression for *MP* is, just as in the case of *MRS*, only approximate. It will also become more and more exact the smaller one chooses ΔK or ΔL.

7.2.2 The Law of Diminishing Marginal Returns

Beside the conclusion that "supply = demand", the law of diminishing marginal returns (or the law of diminishing marginal product) is probably the most frequently cited concept from microeconomics. Suppose we keep everything constant, except for one single input factor, for instance *L*. If we increase the number of hours worked, we will probably produce more. The law of diminishing marginal returns states that the increase will eventually become smaller and smaller when the number of hours worked is large enough.

To use an example: Suppose you start a firm that produces photocopies. You buy a powerful copying machine and you are the only worker. Say that you work one hour per day. You probably will not be able to make many copies during that time. The machine needs fifteen minutes to warm up, you need to prepare things, etc. If you increase the number of hours worked by one hour, you will probably make more copies during the second hour than during the first. Consequently, the marginal product is higher during the second hour than during the first, and over that interval, we therefore have increasing marginal product.

However, if you continue to increase the number of hours, you will eventually not produce many more copies per additional hour. You will become too tired to work. The same thing will happen if you hire other people to work for you: eventually, additional hours or additional workers produce very little additional output. If you, for instance, hire several people, the space will become crowded and the workers will become less efficient.

Short run: In the short run, some costs are fixed.
Long run: In the long run, all costs are variable.

This "law" is based on experience and speculation, but is not considered particularly controversial. We will use it when we construct the product curve.

7.3 Production in the Short Run

It is common to distinguish between the **short run** and the **long run** regarding production. The short run is defined as the time during which (at least) one of the input factors is fixed, usually capital. If the firm, for instance, buys a factory, it may not be able to increase or decrease its size as fast as they would wish. During the time that the firm is stuck with the factory as it is, it amounts to a fixed cost. In the long run, all costs are variable.

Short run: In the short run, some costs are fixed.
Long run: In the long run, all costs are variable.

We will assume that in the short run, labor is variable but capital is fixed. To make it clear that the quantity of capital is fixed in the short run, one often adds a line above the *K* in the production function: $q = f\left(L, \overline{K}\right)$.

The relationship between total production and the number of hours worked can be drawn in a graph. Often, one combines that graph with another graph that shows the marginal product and the average product of labor. We will now show how to construct such a graph.

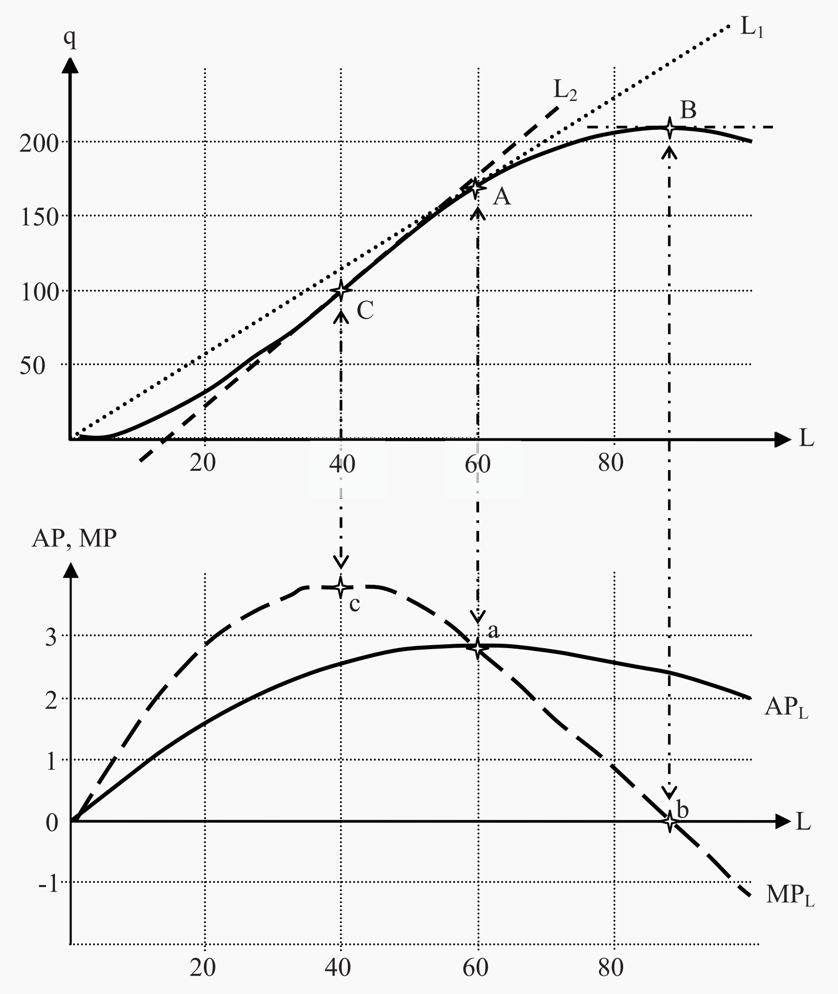
7.3.1 The Product Curve in the Short Run

If we keep the amount of capital constant, the quantity produced is a just function of the number of hours worked, *L*. In Figure 7.1, we see a typical product curve with associated average and marginal product curves.

The product curve has a few typical features: In the beginning, when the number of hours worked is low, production increases slowly, and later it becomes steeper and steeper. Eventually it reaches a maximum and thereafter it decreases.

After we have drawn the product curve, we want to construct the curves for the average and marginal product of labor. (The corresponding values for capital are not as interesting, since capital is a fixed cost in the short run.) To do that, we first observe that there is a simple method to find the value of the average product.

Figure 7.1: The Production Function with Average and Marginal Product



When you have drawn the product curve in the upper part of the graph, you draw a similar diagram below it with the same scale on the X-axis. Now, take a ruler and position it in the upper graph, with one point at the origin (0,0) and another at some point on the product curve, for instance as the line L_1 indicated in the graph. The slope of the ruler will now be equivalent to the average product, AP_L, at that point where the ruler touches the product curve. (That is, AP_L at point A is equivalent to the slope of line L_1.)

The maximum value of AP_L we get at point A, when L is 60. Indicate a point in the lower graph at $L = 60$, point a. To find the correct value on the Y-axis for point a, we calculate the slope of L_1 in the upper part of the diagram: Point A is at $L = 60$ and $q = 170$, so the slope is 170/60 = 2.83. Point a should then be at (60,2.83) in the lower diagram. At point a, AP_L reaches its highest value and must consequently slope downwards both to the left and to the right. Draw such a curve and label it "AP_L."

To construct MP_L, let instead the ruler glide along the product curve in the upper graph so that it indicates the slope of the curve at different points. That way we can see that at point B, when production is at its maximum, the slope must be zero. Consequently, $MP_L = 0$ in that point and we indicate the corresponding value for MP_L in the lower graph, point b.

Then, let the ruler glide along the product curve and note when the slope is as high as possible. In Figure 7.1, that is at point C (when the ruler looks like L_2). In that point, MP_L reaches its highest value. Indicate it in the lower graph. It is easy from the upper graph to see that, in this case, the slope of L_2 is higher than the slope of L_1. Consequently, MP_L (at point C) must be higher than AP_L (at point A) and point c in the lower graph must be higher than point a.

After that, draw the graph for MP_L: It must slope downwards both to the left and to the right of *c*. It must also pass through a (where $MP_L = AP_L$) and then through point b. Now, the graph is finished. Note that we have obeyed the law of diminishing marginal returns: To the right in the graph, MP_L becomes smaller and smaller (and eventually it becomes negative).

Note also that, the two graphs for MP_L and AP_L in the lower graph are closely related to each other. MP_L must intersect AP_L in the latter's maximum point. That fact has a purely mathematical reason: To the left of point a, $MP_L > AP_L$. That means that when we add one more unit of L in that region, we produce exactly MP_L more units of the good. Since that is more than the average so far, the average must increase. This is true as long as $MP_L > AP_L$. To the right of point a, we have that $MP_L < AP_L$. That means that if we add one more unit of L, we produce MP_L more units of the good, which here is less than the average so far. Consequently, the average must decrease and AP_L must slope downwards.

7.4 Production in the Long Run

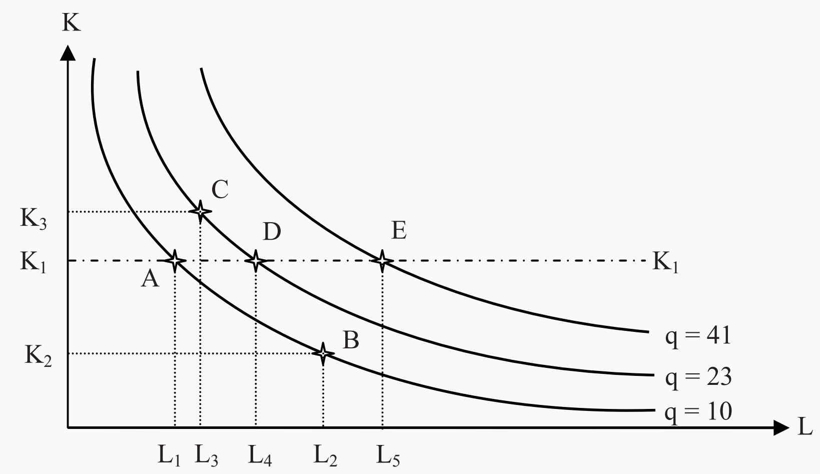
In the long run, both labor and capital are variable inputs. That means that the quantity produced is a function of both *L* and *K*, where either of them can be changed, i.e. $q = f(L, K)$ (and not $f(L, \overline{K})$).

It is usually the case that, the same quantity can be produced with different combination of labor and capital. Workers can, at least to some extent, be substituted for machines, and vice versa. This idea can be illustrated in a graph that is very similar to the indifference curves in consumer theory (see Section 3.3). An **isoquant** ("iso" = similar/same; "quant" = quantity) is a curve that shows different combinations of *L* and *K* that produce the same quantity of the single good (still assuming that the production is efficient). They are usually drawn in a way that is similar to indifference curves. Since we have diminishing marginal returns, they must slope less and less steep to the right. In Figure 7.2, you can see an example.

Start by looking at point A. With that combination of labor and capital (L_1,K_1), a maximum of $q = 10$ units of the good can be produced. The same quantity could also have been produced with the combination (L_2,K_2), point B. If one wants to increase the production to 23 units, one can choose, for instance, point C, where both the amount of capital and of labor has been increased. Consequently, one can move from A to C only in the long run.

In the short run, *K* is fixed. If one wants to increase production, one therefore has to choose the same *K*. If one wants to produce 23 units in the short run (assuming we start at point A), one has to choose the combination (L_4,K_1), point D, and if one wants to produce 41 units, one has to choose the combination (L_5,K_1), point E.

Figure 7.2: An Isoquant Map



Note that:

- Isoquants further away from the origin correspond to larger quantities.
- The isoquants cannot cross each other.

- The isoquants slope downwards.
- The extreme cases of isoquants are, just as for indifference curves, straight lines and L-shapes; compare to Figure 3.5.

7.4.1 The Marginal Rate of Technical Substitution

From the isoquants in the previous section, one can derive the **marginal rate of technical substitution**, *MRTS*. The *MRTS* corresponds to the marginal rate of substitution, *MRS*, from consumer theory (Section 3.5).

Marginal rate of technical substitution: The amount of capital one has to add to production if one reduces labor with one unit but still wishes to produce the same quantity. Corresponds to the slope of the isoquant curve.

MRTS can *approximately* be calculated as

$$MRTS = \frac{\Delta K}{\Delta L}$$

Since the curves slope downwards, if ΔK is positive then ΔL must be negative, and vice versa. That means that *MRTS* is a negative number. By convention, however, the minus sign is often omitted. (Compare to Section 3.5.)

7.4.2 The Marginal Rate of Technical Substitution and the Marginal Products

There is an important relation between *MRTS* and the marginal products of labor and capital. As we have seen, $MP_L = \Delta q/\Delta L$, which means that if we increase the amount of hours worked with ΔL, production will increase with $\Delta q = MP_L * \Delta L$. Similarly, for capital we have that $MP_K = \Delta q/\Delta K$, so if we use one unit less of capital, production decreases with $\Delta q = MP_K * \Delta K$. For instance, if the marginal product of labor, MP_L, is 3 and we add 1 more hour of labor, we will produce 3*1 = 3 units more of the good.

Let us combine these two observations in the following way: Suppose that you are on an isoquant, for instance in point B in Figure 7.2. If you increase labor with ΔL, production will increase with $\Delta q = MP_L * \Delta L$. However, suppose that you at the same time decrease your use of capital *exactly so much* that you still produce the same quantity as you did in point B. Then the total change in q must be zero. We can express this as

$$\begin{cases} MP_L * \Delta L = \Delta q \\ MP_K * \Delta K = -\Delta q \end{cases}$$
$$MP_L * \Delta L + MP_K * \Delta K = 0$$

If we rearrange the last expression, we can get an alternative expression for *MRTS* (see the previous section):

$$-\frac{MP_L}{MP_K} = \frac{\Delta K}{\Delta L} = MRTS$$

The marginal rate of technical substitution, *MRTS*, which, by definition, equals (minus) the change in capital divided by the change in labor, also equals one marginal product divided by the other. Note that on the left-hand side, we have the marginal product of labor in the numerator but in the middle, we have ΔL in the denominator.

7.4.3 Returns to Scale

Suppose that, using labor *L* and capital *K*, we produce the quantity *q* of a good. If we would double both *L* and *K*, we would probably increase the quantity produced as well, but by how much? If q is also doubled, we have constant **returns to scale** (we can think of this as that the *scale* is the same for *(L,K)* and *q*). If instead *q* increases by less than two times, we have decreasing returns to scale, and if it increases by more we have increasing returns to scale.

Returns to scale: By how much does the quantity change if one changes the inputs?

More generally, we increase *L* and *K* by a factor t, and then check if *q* increases by more, less, or by the same factor. We can express this as

f(t*L,t*K) < t*f(L,K)	Decreasing returns to scale
f(t*L,t*K) = t*f(L,K)	Constant returns to scale
f(t*L,t*K) > t*f(L,K)	Increasing returns to scale

Look again at the expressions above: *f(L,K)* is the quantity produced from the start. We introduced this same expression in Section 7.2. *f(t*L,t*K)* is the quantity produced if you increase both *K* and *L* by the factor *t*. Then you ask if you get more, less, or the same as t times the production you had from the beginning, i.e. *t*f(L,K).*

There can be different reasons why we get different returns to scale. For instance:

- *Constant returns to scale*. Suppose we have a factory that produces a certain quantity of a good. Then we build another factory that has the same size and that uses the same number of workers, so that we now have two factories. It seems reasonable to assume that the second factory produces the same quantity as the first one does. This means that, as we double the inputs we also double the output.
- *Decreasing returns to scale*. Suppose it becomes more and more difficult to coordinate the production as the size increases, so that we get higher and higher costs for administration. Then the costs will increase proportionally more than the production, and the production will grow by less than the inputs.
- *Increasing returns to scale*. Oftentimes, large firms are more efficient than small firms are. This is called *large-scale advantages*.

8 Costs

So far, we have only studied how different combinations of inputs produce different quantities of a good. Now, we will instead look at the cost of production. As before, we distinguish between the short and the long run.

When we study costs, it is important that we use the *opportunity cost* as measure, i.e. the cost of using our resources equals how much they had been worth if we had instead used them for the best alternative. (See also Chapter 1.)

One reason for studying the cost side is that we want to find the cheapest way of producing a good. However, there is also another reason. The relationship between output and cost play an important role for which type of market that will arise: how many firms will there be, and how high the price will be relative to the cost of production. That question we will return to in later chapters.

8.1 Production Costs in the Short Run

In the short run, not all input factors are variable. We therefore distinguish between fixed cost, *FC*, and variable cost, *VC*. Total cost, *TC*, is the sum of the two:

$$TC = VC + FC$$

We also need to define a few other central concepts. Regarding average cost, we will have use for the averages of all three of the above. If we divide each of them with q, we get average total cost, *ATC*, average variable cost, *AVC*, and average fixed cost, *AFC*:

$$\begin{cases} ATC = \dfrac{TC}{q} \\ AVC = \dfrac{VC}{q} \\ AFC = \dfrac{FC}{q} \end{cases}$$

Note that the following must hold:

$$ATC = AVC + AFC$$

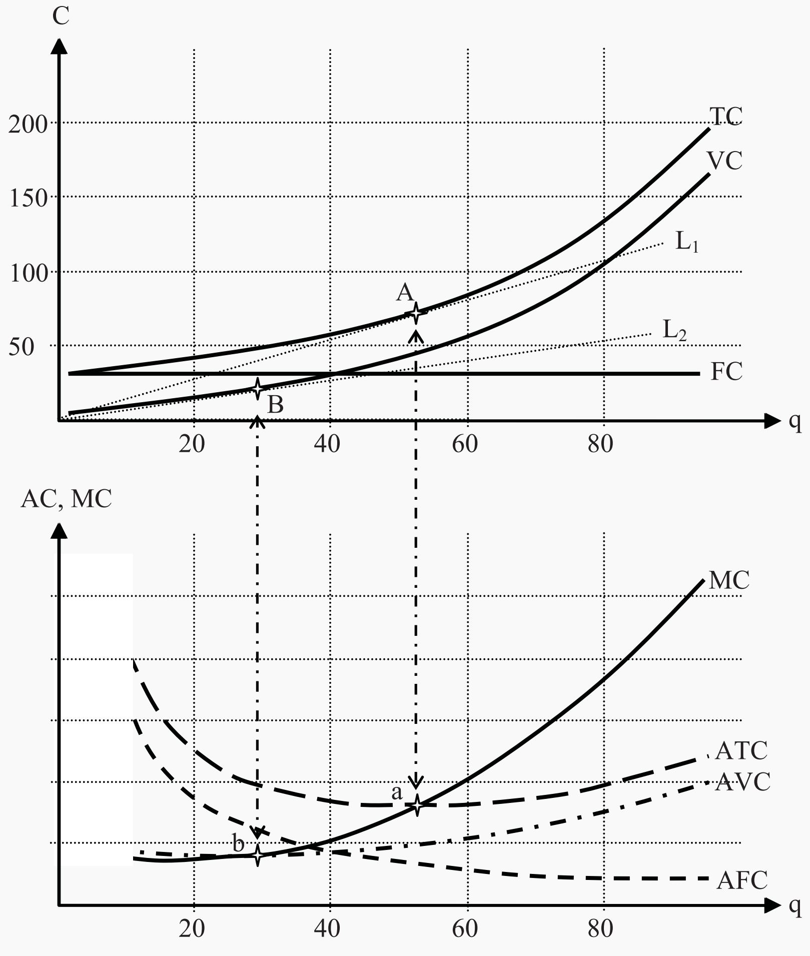
The marginal cost, *MC*, in turn, measures the cost of producing one more unit of the good:

$$MC = \frac{\Delta TC}{\Delta q} = \left[\frac{\Delta(VC + FC)}{\Delta q} = \right] \frac{\Delta VC}{\Delta q}$$

Note that we can use either the change in total cost or the change in variable cost. Both must give the same answer, since the fixed cost does not change ($\Delta FC = 0$). As before, the expression for marginal change is only an approximation.

Now, we will construct a graph to illustrate these different measures of costs (see Figure 8.1). The fixed cost, *FC*, is constant, independent of how many units we produce, so the curve illustrating *FC* must be a horizontal line. Total cost, *TC*, must always increase with production; else, the production is not efficient. Furthermore, since if we produce nothing *TC* must equal *FC*, the curve for *TC* must start in the same point as *FC* on the Y-axis. Since $TC = VC + FC$, the curve for variable cost, *VC*, must have the same shape as *TC*. Obviously, *VC* of producing nothing is zero, so the curve for *VC* must start at the origin.

Figure 8.1: The Cost Function with Average and Marginal Costs



Using the curves we have drawn in the upper part of the figure, we will now construct the ones in the lower part, i.e. *ATC*, *AVC*, *AFC*, and *MC*. First, we use the same technique as we did in Section 7.3.1. Lay a ruler in the upper part of the figure such that it has one point at the origin and another point on the *TC* curve. Find the point on *TC* where the ruler has the smallest slope. In the figure, this corresponds to line L_1 and point A. You have now found the smallest possible average cost, *ATC*. Proceed in the same way to find the point on *VC* where the ruler has the smallest slope: the line L_2 and point B. That point corresponds to the lowest possible average variable cost, *AVC*.

Draw the two curves for *ATC* and *AVC* in the lower part of the figure. Of course, *ATC* must lie above *AVC*. *ATC* should have its lowest point at a, and *AVC* at b. If you wish to find the numerical value for, for instance, point a, then read off the location of point A in the upper part of the figure: (53,75). Then calculate 75/53 = 1.4. Point a should then be at (53,1.4).

You can now also draw *MC* in the lower part of the figure. It must run through both point a and point b (for the same reasons as in Section 7.3.1) and it should correspond to the slope of *TC* (or *VC*, since it has exactly the same slope) in the upper part of the figure. Note that, since the slope of *TC* becomes higher and higher, MC should increase as we move to the right. Lastly, AFC must become smaller and smaller the more products we produce, since FC is a constant and we divide with an increasingly higher quantity, q.

8.2 Production Cost in the Long Run

In the long run, both labor and capital are variable. That allows us to, in a way similar to the budget line in consumer theory, construct the so-called **isocost lines** ("iso" = similar/same). If the price of one hour of work is w (for wage) and the price of one unit of capital is r (for rental rate), an isocost line is defined as all combinations of L and K that cost the same:

Isocost line: All combinations of inputs that cost the same.

$$w * L + r * K = C$$

Solving this expression for K, we get

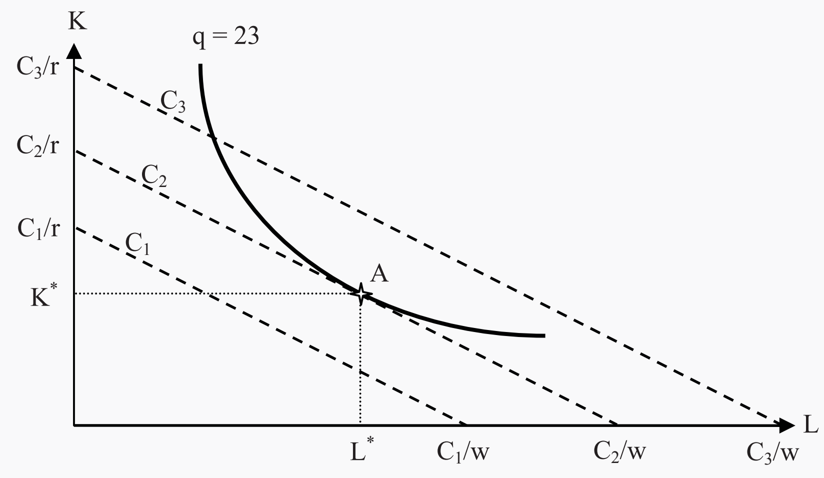
$$K = \frac{C}{r} - \frac{w}{r} * L.$$

For the time being, we assume that the prices of labor and capital are given. If we insert a few different values for C, we can draw a few isocost lines, for instance the lines C_1, C_2, and C_3 in Figure 8.2. Similar to the budget line, the slope of the isocost line depends on the value of w/r, and the easiest way to draw the budget line is to find the points on the X- and Y-axes where one buys only labor or capital. One can, for instance, calculate C_1/r, indicate that value on the Y-axis, and C_1/w, and indicate that value on the X-axis. Then draw a straight line between the two points to get all combinations of L and K that cost C_1.

Now, you can probably see where we are heading. In Figure 8.2 we have also drawn an isoquant (see Section 7.4), and in point A, the isoquant just barely touches one of the isocost lines, C_2. If we are prepared to pay the cost C_2, then $q = 23$ is the maximum quantity we can possibly produce. We do that by choosing the quantities L^* and K^* of labor and capital, respectively.

Note that we can also reason the other way around. If we want to produce 23 units, then the lowest cost we can possibly do that at is C_2. The optimal choices of inputs are then L^* and K^*. For a production of exactly 23 units of the good, all other combinations of L and K are either inefficient or do not produce enough of the good.

Figure 8.2: Isocost Lines



Just as we did in consumer theory, we can give a mathematical formulation of the result. At the point of tangency, the isoquant and the isocost line must have the same slope. The slope of the isocost line is -w/r and the slope of the iso-quant is the marginal rate of technical substitution, MRTS (see Section 7.4). At the point of tangency, we must then have that

$$-\frac{w}{r} = MRTS \quad \left[= -\frac{MP_L}{MP_K} = \frac{\Delta K}{\Delta L} \right]$$

Note that, if one uses the convention to omit the minus sign in front of *MRTS*, one must also do so in front of w/r. As a reminder, we have also included how we found *MRTS* earlier.

It is important to understand how to interpret this criterion. If we rearrange the expression, we can get

$$\frac{MP_K}{r} = \frac{MP_L}{w}$$

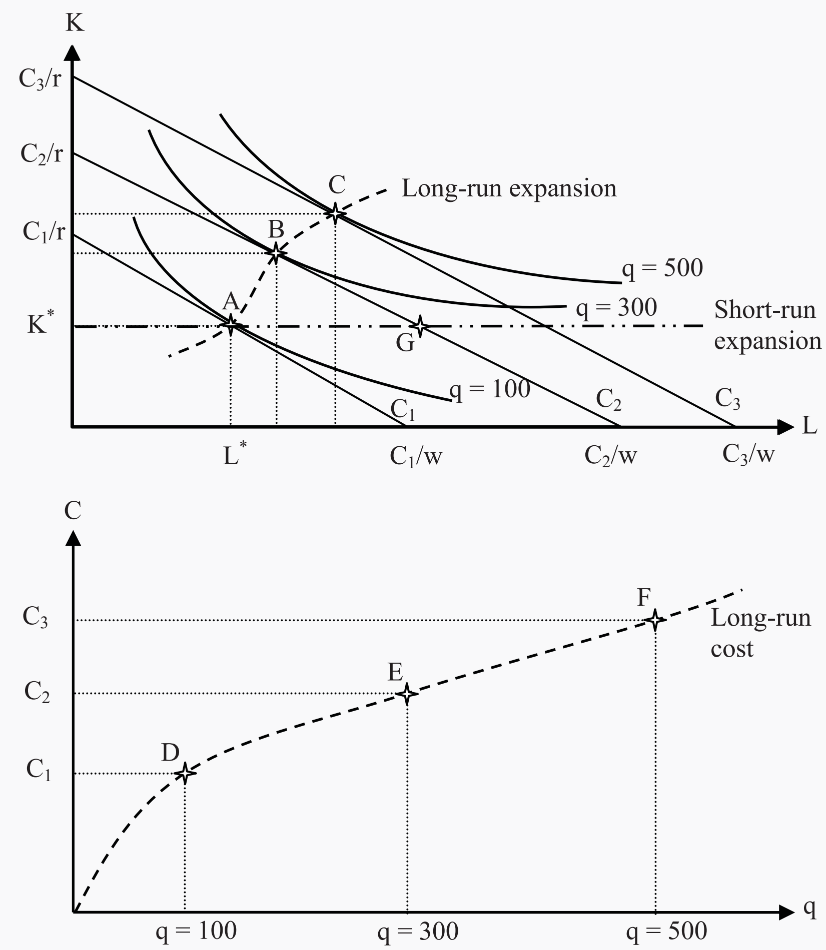
Remember that MP_K is the number of additional units we produce if we add one more unit of capital, while holding everything else constant. MP_L is the same for one more unit of labor. MP_K/r is then the number of additional units we get per dollar (or other currency), if we use one more unit of capital. MP_L/w is the number of additional units we get per dollar if we use one more unit of labor (work one more hour). At the optimal point, these two must be equal, and the producer is then indifferent between using labor and using capital.

If we repeat the procedure for finding the optimal point for many different iso-cost lines and isoquants, we will trace out a curve that shows all efficient combinations of labor and capital. This is the so-called **expansion path**. From that curve, it is possible to derive the long-run cost of production. (Compare to the income-consumption curve and the Engel curve in Section 4.1.2.)

Expansion path: How much one can produce at different costs.

Look at Figure 8.3. In the upper part of the figure, we have drawn three isocost lines, C_1, C_2, and C_3, and then found the points on each of them where an isoquant just about touches them. The three points are A, B, and C where the produced quantities are 100, 300, and 500 units of the good. If we had done this for all possible costs, we would have gotten the long-run expansion path. We see that, for a cost of C_1 we can produce a maximum of 100 units, for C_2 a maximum of 300 units, and for C_3 a maximum of 500 units. In the lower part of the figure, we have indicated those combinations at points D, E, and F, and then drawn a line through them. That line is the long-run cost of production. Note that the line must start at the origin, since the long-run cost of producing nothing is zero.

Figure 8.3: Derivation of the Long-Run Cost Curve and the Expansion Path



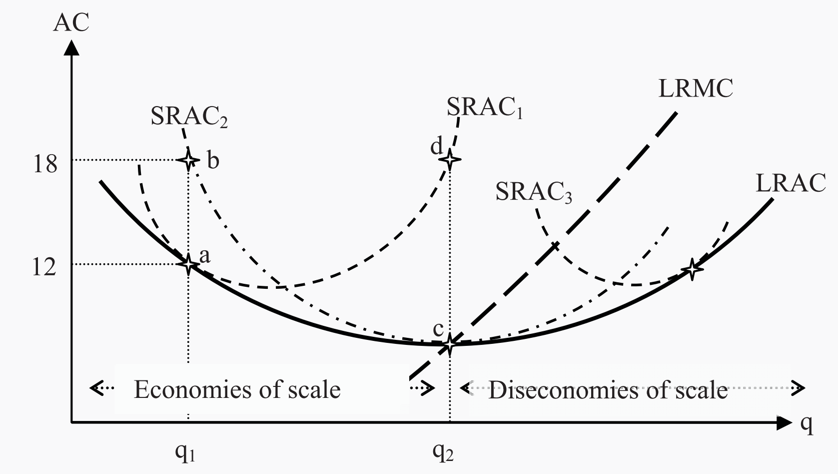
In Figure 8.3, we have also drawn the short run expansion path. In the short run, the quantity of capital is fixed, K^*. In the diagram, that amount of capital is optimal for a production of 100 units. That can be seen from the fact that the isocost line C_1 touches the isoquant $q = 100$ in point A where the amount of capital is K^*. If we want to produce more than 100 units in the short run, we must do that using only additional labor, i.e. the expansion must follow the short-run expansion path. At point G, the cost of production is as high as at point B, but the number of produced units must be smaller since the isoquant $q = 300$ is further from the origin than point G. When one chooses the long-run amount of capital to use, one does so under the assumption that the production in the short run will be optimal at precisely that amount. In the diagram, one has chosen K^* because one believed that, in the short run one will produce 100 units. For other quantities, K^* is not an efficient choice.

8.3 The Relation between Long-Run and Short-Run Average Costs

As we saw above, the short-run cost of production must always be higher than the long-run cost, except for at one certain point where they are the same. In Figure 8.3 the short-run cost and the long-run cost of producing 100 units is the same, given that one has chosen an amount of capital equal to K^*. For every other choice of capital, it is more expensive to produce 100 units in the short run than in the long run.

If this is true about the cost, it must also be true about the average cost, i.e. the long-run average cost must always be smaller than the short-run average cost, except for at one point where they are the same. If we draw a few different curves for short-run average costs, where each curve is valid for different long-run investments in capital, we get a picture as the one in Figure 8.4.

Figure 8.4: Long-Run and Short-Run Average Costs



We have three different short-run average cost curves, $SRAC_1$, $SRAC_2$, and $SRAC_3$ (Short Run Average Cost) and one long-run average cost curve, *LRAC* (Long Run Average Cost). Each *SRAC* curve has one point where it is optimal in the long run as well. For example, $SRAC_1$ is optimal at point a, where it touches *LRAC* and the average cost is 12. If one had instead chosen a different amount of capital, for instance, such that one would have been constrained to $SRAC_2$, one would have ended up at point b if one wanted to produce the quantity q_1. Average cost would then have been 18. $SRAC_2$ is instead long-run optimal if one wants to produce q_2.

If one takes all such points where the short-run average cost is optimal, i.e. for all possible *SRAC* curves, not just the ones drawn here, one will get the curve for the long-run average cost, *LRAC*. Just as the short-run marginal cost curves must intersect the *SRAC* curves at their lowest points, the long-run marginal cost curve, *LRMC*, must intersect *LRAC* at its lowest point.

The *LRAC* curve in Figure 8.4 has another interesting property. To the left of the quantity q_2, *LRAC* slopes downwards (towards q_2), but to the right it slopes upwards. That means that the lowest cost per unit of the good is achieved at the quantity q_2 and at point c in the diagram. To the left of q_2, we have **economies of scale** and to the right we have **diseconomies of scale**. Note that this can be very important if there is competition in the market. If the firm is at a point to the left of q_2, it can lower its cost per unit by increasing the scale of production. In the next step, it can undercut the price of its competitors.

Economies of scale: The more one produces, the lower is the cost per unit.
Diseconomies of scale: The more one produces, the higher is the cost per unit.

9 Perfect Competition

9.1 Introduction

So far, we have discussed how the consumers make their decisions, and what the producers' production possibilities and cost of production look like. The consumers often take prices as given and choose quantities based on the prices. The question is how prices arise. One factor is, of course, the cost of production. The price cannot be below the cost, at least not in the long run. The price is, however, very dependent on the structure of the market. Among the most important questions one can ask about the market structure are:

- The degree of **concentration** of buyers and sellers. Do we have many, a few, or one?
- The degree of **product differentiation**. Are the products identical to each other, or how different are they from each other? (See Chapter 15, on monopolistic competition.)
- Are there any **barriers to entry** in the market? (See Chapter 11 on monopoly)

Concentration: How many independent agents there are in a market.
Product differentiation: How big the differences there are between products.

Barriers to entry: Things that make it impossible or excessively costly to enter a market.

The answers to these questions largely determine which kind of market we get, and this, in turn, largely determines the price. In this chapter, we will look at one type of market, a perfectly competitive market. In later chapters, we will look at other market forms.

9.2 Conditions for Perfect Competition

For a market to be perfectly competitive, it has to fulfill the following conditions:

- All agents are **price takers**. No single buyer or seller can affect the price of the good. Everyone takes the price as given, and, depending on the price, decides about quantity. This condition will be true if there are many small buyers and sellers.
- **Homogenous products**. Each seller's products are identical to every other seller's products. Furthermore, there are no extra costs, such as transportation costs, for some sellers. The buyers are therefore neutral between different sellers.
- All factors of production are completely variable. There are no barriers to entry for new firms or barriers to leave for existing firms.
- All buyers and sellers have complete information about existing alternatives in the market.
- There are no agreements to collude in the market. For instance, the sellers cannot form a **cartel**.

Price takers: An agent that takes prices as given.

Homogenous products: Identical products.

Cartel: Several producers who cooperate on prices or quantities.

These conditions are hard to satisfy and few, if any, real markets do that. Even so, the model is very informative and delivers several interesting results. Furthermore, many economists are highly in favor of competition and some of the most important reasons for that will be revealed as we use this model.

9.3 Profit Maximizing Production in the Short Run

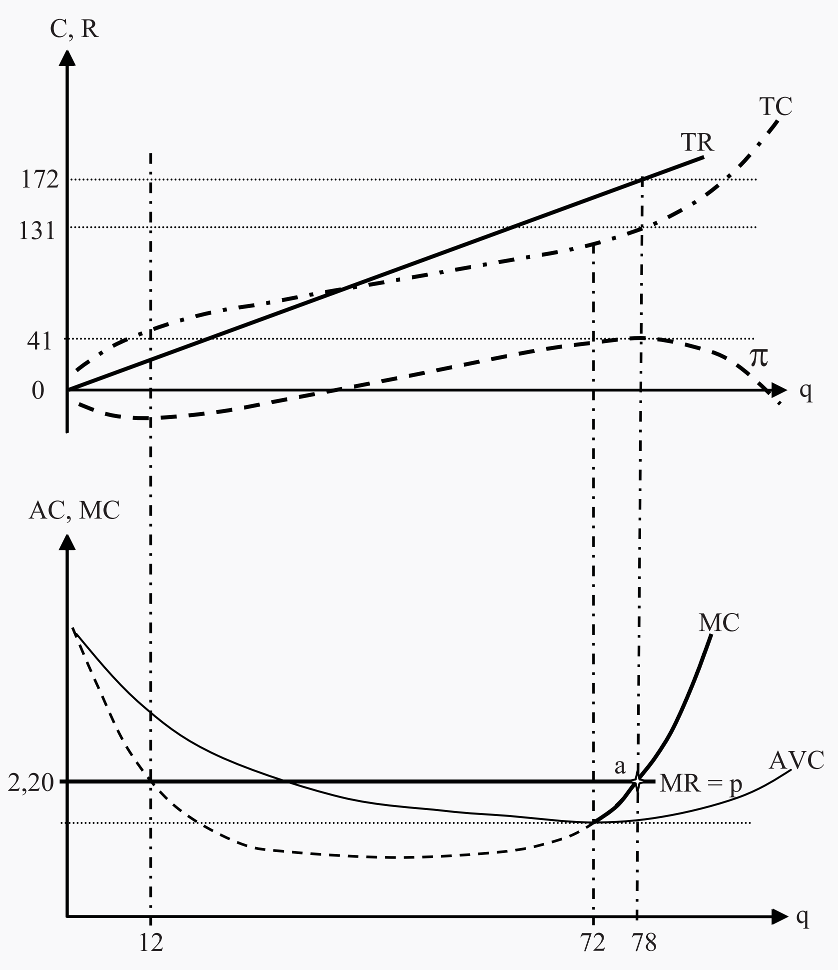
The goal of an individual firm is to maximize its profit, i.e. the difference between revenues and costs. In the short run, it does that under the restriction that it cannot change the amount of capital.

We will now study the short-run production in a diagram. In the upper part of the diagram in Figure 9.1, we have drawn the **total cost**, *TC*, **total revenue**, *TR*, and the **profit**, $\pi = TR - TC$. Since the firm cannot influence the price in a perfectly competitive market, *TR* will simply be a straight line with a slope equal to *p* (the price). This is since each additional unit of the good it sells will yield an income of *p*. The shape of *TC* is often more complicated (see Section 8.1).

Total cost: The total cost of producing a certain quantity of a good.
Total revenue: The total income from selling a certain quantity of a good.
Profit: The difference between revenue and cost.

The goal is to maximize the profit, which in the graph occurs at a quantity of $q = 78$ units, where the profit is 41. (This is the point where π reaches its maximum.) As profit is the difference between revenues and costs, the difference between *TR* and *TC* is at a maximum here: 172 - 131 = 41.

Figure 9.1: Profit Maximization under Perfect Competition



In the lower part of the diagram, we have drawn the **marginal revenue**, *MR*, the price, *p*, the **marginal cost**, *MC*, and the average variable cost, *AVC*. *MR* corresponds to the slope of *TR*, which we argued must be equal to *p*. In other words, if we sell one more unit of the good, we receive an additional income equal to the price, *p*. Therefore, *MR* is equal to *p*.

Marginal revenue: The additional income a firm receives if it sells one more unit.
Marginal cost: The additional cost a firm incurs if it produces one more unit.

In point a, the *MC* curve intersects the *MR* curve. This is a condition for profit maximization. To see why, think about what would happen if we sold one unit more or one unit less than 78. If we had produced and sold one unit more, we would have incurred a cost of *MC*, but we had only received an income of *MR*. And $MR < MC$, so we had reduced profit. Conversely, if we had produced one unit less, we would have saved the production cost of that unit, *MC*, but we would also have lost the revenue from selling it. Moreover, to the left of 78, $MR > MC$, so we had reduced profit that way as well. Therefore, 78 is indeed the best choice we can make. The condition for profit maximization is:

$$MR\left(q^*\right) = MC\left(q^*\right)$$

The firm, consequently, chooses the quantity, q^*, that makes $MR = MC$. Note that, at the quantity 78 in the figure, *TR* and *TC* have the same slope. That is the same thing as $MR = MC$.

9.3.1 Strategy to Find the Optimal Short-Run Quantity

We can summarize the strategy for finding the point where the firm maximizes its short-run profit in a few steps:

- Find the point where $MC = MR$ and where the *MC* curve is increasing.
- Is that point above (or equal to) *AVC*, i.e. is $p = MR \geq AVC$? In that case, choose to produce the corresponding quantity.
- In the opposite case, i.e. if $p = MR < AVC$, choose to produce nothing at all; $q = 0$. The condition $p = MR < AVC$ is called the **shut down condition**.

Shut down condition: The condition under which it is better to produce nothing rather than produce something: MR < AVC.

Note that (as in the first bullet point) the *MC* curve must be increasing. In Figure 9.1, we can see that the *MC* curve also intersects the *MR* curve at the quantity $q = 12$, but there the *MC* curve is decreasing. That point instead maximizes the loss!

Also, note for bullet points 2 and 3, the reasoning behind the condition $MR \geq AVC$: Since we are looking at the short run, the fixed cost, *FC*, cannot be changed. The firm can always choose to produce nothing. If it does so, it receives no revenues and incurs no variable costs, but it will still incur the total fixed cost, *FC*. Total profit will then be a loss of *-FC*. This means that the firm will choose to produce as long as it can at least recover some of that loss. Moreover, the firm will do so as long as the price, and therefore *MR*, is larger than or as large as *AVC*.

In the short run, the firm can consequently accept to produce at a (small) loss, since the loss will be smaller than if one chooses to shut down production completely. If instead $MR < AVC$, the revenues from additional units sold cannot even cover the average variable cost of producing them. Then it is better to shut down.

9.3.2 The Firm's Short-Run Supply Curve

What happens if the market price changes? Then *MR* changes, and the point of intersection between *MR* and *MC* also changes. The firm will then choose to produce the quantity that corresponds to the new point of intersection, so the quantity supplied follows the *MC* curve as the price changes.

However, this is only true as long as the price is higher than *AVC*. To see why, look at the shut down condition above again. Suppose the market price falls to the point where the MC curve intersects the *AVC* curve, i.e. at the quantity $q = 72$ in the figure. At that point, $MR = p = MC = AVC$ and the profit becomes $q^*(p - AVC) - FC = 72*0 - FC = -FC$. Note that $p - AVC$ is what the firm gets paid in excess of average variable cost for each unit it sells; $q^*(p - AVC)$ is then what it gets paid in excess of average variable cost for all units it sells; finally, subtracting *FC* yields what it gets paid in excess of all costs (= profit).

The loss is consequently as large as if we choose to produce nothing at all. If the price falls even more, the losses increase and it is better to produce nothing. The conclusion of this is that, the firm's short-run supply curve is the part of the *MC* curve that lies above *AVC*, i.e. the part that is drawn with a full line in the lower part of Figure 9.1.

9.3.3 The Market's Short-Run Supply Curve

The market is the sum of all individual firms. We get the market's supply curve by summing all individual firms' supply curves horizontally. Compare to the method used in Section 4.2.

9.4 Short-Run Equilibrium

In Figure 9.2, we have summarized the equilibrium in the market and the equilibrium for an individual representative firm. To the right in the figure, we have the individual firm's *MC*-, *ATC*-, and *AVC* curves. The short-run supply curve of the firm is the part of the *MC* curve that is above *AVC*. For prices below p_{min}, there is consequently no supply at all. If we sum all firms' supply curves, we get the market's supply curve, S, to the left in the figure. *(Σ(MC)* means "the sum of all *MC* curves.")

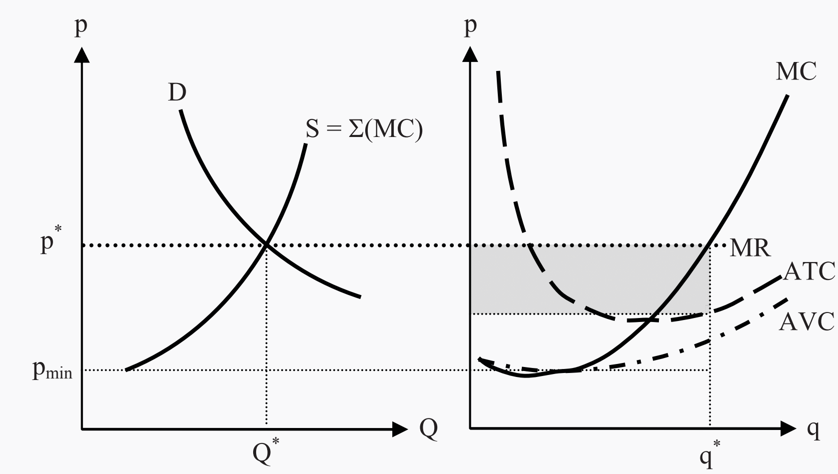
In the market, supply meets demand, *D*, and an equilibrium price, p^*, and an equilibrium quantity, Q^*, arise. p^* is the price that the individual firm receives for each unit of the good it sells. Since there are a large number of firms, no individual firm can charge a higher price than p^*. If some firm did, the consumers would choose one of its competitors instead. The *MR* curve of an individual firm is consequently horizontal and equal to the price, p^*. The firm chooses to produce the quantity q^*, as this quantity makes $MC = MR$ and, consequently, maximizes profit.

In the short run, a firm in a perfectly competitive market can make a profit. In Figure 9.2, the profit corresponds to the grey rectangle on the right-hand side. To see that this is the profit, note that in a perfectly competitive market *AR* (Average Revenue) is as large as *MR* is, since the firm is paid the same amount for each unit sold. Furthermore, $q*AR = TR$ and $q*ATC = TC$. Profit is then $\pi = TR - TC$. To summarize, we have that

$$\pi = TR - TC = q * AR - q * ATC = q * (MR - ATC)$$

The right-hand side of this expression corresponds to the grey rectangle in Figure 9.2. With the example we used before, we get that π = 78*(2.20 - 131/78) = 41.

Figure 9.2: Short-Run Equilibrium for the Market and an Individual Firm



9.5 Long-Run Production

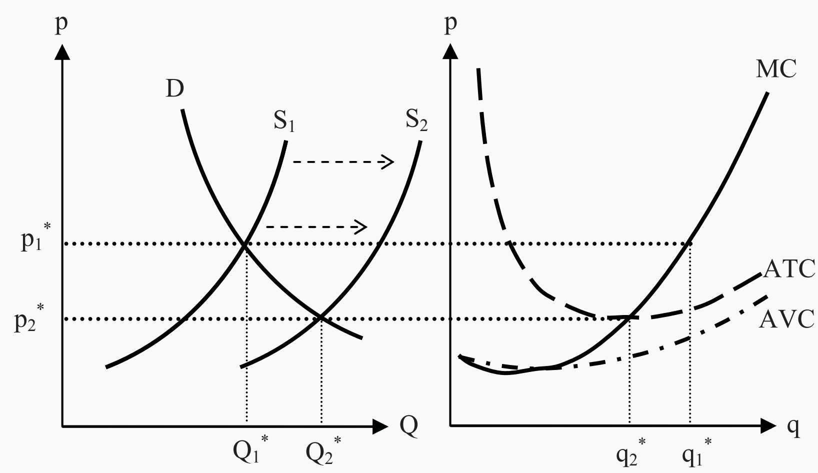
The firm in Figure 9.2 makes a short-run profit. However, what happens in the long run? There are three different possibilities: The firm could make a profit, make a loss, or break even. The first is called **excess profit** or **supernormal profit**, and the last is called **normal profit**. No firm can allow itself to make a long-run loss.

Excess profit / supernormal profit: The firm is paid more for the good than it costs to produce it.
Normal profit: The firm is paid what it costs to produce the good.

In the long run, a firm in a perfectly competitive market will make normal profits, i.e. break even. To see that, look at Figure 9.3. Here, we have the same initial situation as in Figure 9.2. Since the firm makes a profit, this market will attract new firms. Moreover, since there are no barriers to entry, new firms will establish themselves as soon as possible; i.e. in the long run. As more and more producers establish themselves in the market, larger and larger quantities will be supplied at each given price, and the market supply curve will shift to the right, from S_1 towards S_2. Consequently, the price will be pushed down, from p_1^* towards p_2^*.

For the individual firm, the decrease in the price means that it will reduce the quantity it produces, from q_1^* to q_2^*. Remember that the firm chooses to produce the quantity where $MR (= p) = MC$. This process, with new firms and reductions in prices, continues as long as there are any profits to be made in the market. When the price reaches p_2^*, and the firm produces the quantity q_2^*, we are exactly at the point where the *MC* curve intersects the *ATC* curve. In the previous section, we showed that $\pi = q^*(MR - ATC)$, and a quick look at that expression gives that if $MR = ATC$ then the profit must be zero, regardless of the quantity produced.

Figure 9.3: Long-Run Equilibrium for the Market and an Individual Firm



When the price has reached p_2^*, there are no longer any excess profits to be made in the market. No more firms establish themselves and the downward push on the price stops. The market is then in long-run equilibrium. Note that the individual firm *reduces* the quantity it produces, while the total quantity produced in the market *increases*. The reason for this is that enough many new firms establish themselves to counterweight the reduction in production for the individual firm.

In the short run, we can also have the opposite case: The individual firms can make short-run losses. That will cause some firms to leave the market in the long run and total supply decreases, which causes the equilibrium price to rise. That process will continue until no firm makes a loss, and the end point of that process is the same as before: p_2^* and Q_2^*.

Many people find the result that firms in a perfectly competitive market make zero profits, hard to accept. Remember, however, that by a cost in this context we mean *opportunity cost*. Therefore, the revenues are as large as the opportunity costs. In the opportunity costs, we include what the firm loses by not investing in the best alternative. If the best alternative is a very good one, the situation here will also be good for the firm. Also, remember that, there are very few real-life examples of a perfectly competitive market, so it might be difficult to have a good intuition for this result.

9.6 The Long-Run Supply Curve

For the individual firm, the long-run supply curve is found in exactly the same way as the short-run, only with long-run marginal cost, *LRMC*, and average cost, *LRAC*, instead. The supply curve is, consequently, that part of *LRMC* that lies above *LRAC*.

When it comes to the whole market's long-run supply curve, the situation is much more complicated. Remember that, for the short run we found the market's supply by summing up all existing firms' supply curves. That was possible because in the short run the number of producers is constant. This is not the case in the long run. If the price changes, then, in the long run, the number of producers also changes. The market long-run supply curve instead depends on how the cost of production changes with the size of production for the whole market.

- *Constant production cost*. If it is possible to establish new firms and they will have exactly the same cost of production as existing firms, then the industry has constant costs. If this is the case, the long-run supply curve will be a horizontal line. The reason for that is, if the price would be higher at some quantity, new firms would establish themselves. Moreover, the cost for the new firms is the same as for the old firms, so this will push down the price to the marginal cost.

- *Increasing production cost*. If it costs more and more to produce more units, the industry has increasing costs. To produce more, the firms have to increase the price per unit. The supply curve will therefore slope upwards (i.e. in the opposite direction of any of the supply curves in this book).
- *Decreasing production cost*. In the opposite case, if it costs less and less to produce more units, the supply curve will slope downwards.

9.7 Properties of the Equilibrium of a Perfectly Competitive Market

Several central properties of the perfectly competitive market equilibrium deserve to be pointed out. These properties are the primary reasons why many economists (normatively) view a high level of competition as a good thing.

- The equilibrium is efficient. The market price will be at the same level as the long-run average cost of production. There is consequently no other way to produce the same quantity of goods that is cheaper. There is, in other words, no waste of resources. (Compare to Sections 11.4 and 18.2.)
- All firms have normal profits, i.e. no profits. The consumers, consequently, pay only what the production costs.
- Total utility is maximized (see Section 10.1).

Note also that these results, that are positive for society, are achieved without any form of central planning or ruling. This phenomenon, that the resources automatically are allocated such that these results are achieved, is often called *the invisible hand*.

However, remember that to reach these results, we have assumed a perfectly competitive market, i.e. that all the assumptions in Section 9.2 are satisfied. We will soon look at other markets forms.

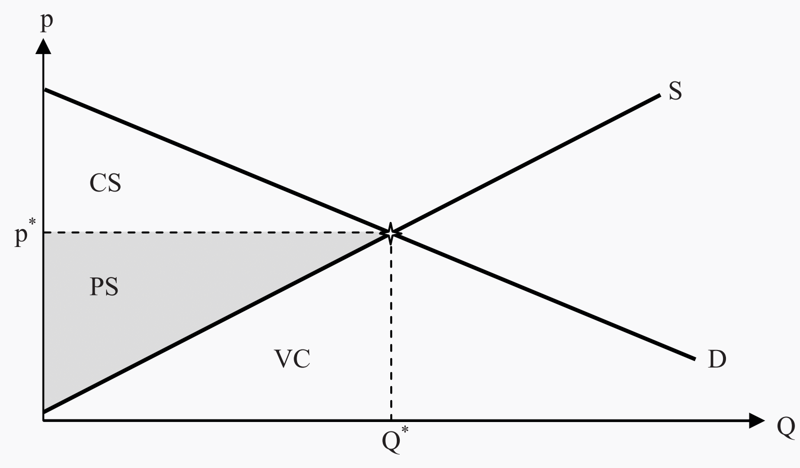
10 Market Interventions and Welfare Effects

There are many different opinions about what welfare is. When one talk of welfare effects in microeconomics, then it is most often about how much utility different groups of people get from different allocations of goods. Usually, welfare analyses separate between producers and consumers, and make the following distinctions:

- *Consumer surplus* (*CS*). The consumers have a certain valuation of a good and pay a certain price to get it. The consumer surplus consists of the difference between how high their valuation is and how much they pay. In Figure 10.1, this corresponds to the triangular area labeled *CS*, i.e. the area between the demand curve and the price.

  We can get an intuition of why this area is interesting in the following way. Suppose we are the customer with the highest valuation of the good. We would then be the customer that is willing to pay the highest price for it. In other words, we are the customer who defines the left-most point on the demand curve, *D*. If the price were so high that only one unit was sold, we would be the buyer. However, at the equilibrium we do not have to pay that high price. We only have to pay p^*. That means we get a surplus, as compared to what we are willing to pay, corresponding to the difference between the point on D and p^*. If we apply the same reasoning to the consumer with the second highest valuation, and so forth for all consumers who buy the good, we get the result that the total consumer surplus corresponds to the area *CS* in the figure.

Figure 10.1: Consumer and Producer Surplus



- *Producer surplus* (*PS*) is the difference between what the producers are paid for a good and the lowest price at which they would have supplied it (i.e. the marginal cost of production). In Figure 10.1, this corresponds to the triangular area labeled *PS*, i.e. the area between the price and the supply curve. The reasoning behind why this area is interesting parallels the one for the consumers.

- *Social surplus* is the sum of consumer and producer surplus: $CS + PS$.

One may note that PS is directly related to profit. The area below the supply curve, S, i.e. the triangle labeled VC, corresponds to the variable cost of production. The producers' total revenue, TR, is the price times the quantity sold. From the figure, we see that

$$p^* * Q^* = TR = PS + VC$$

Furthermore, we know that profit is revenues minus costs. We can therefore get an expression for the profit:

$$\begin{aligned} \pi &= TR - TC \\ &= TR - (VC + FC) \\ &= (PS + VC) - (VC + FC) \\ &= PS - FC \end{aligned}$$

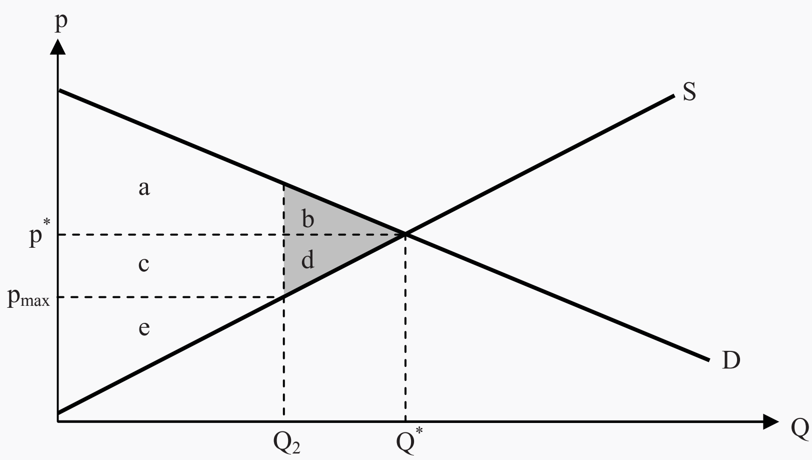
In the short run, profit consequently equals PS minus fixed costs. In the long run, there are no fixed costs, and PS and profit are equal to each other.

10.1 Welfare Analysis

CS, *PS*, and social surplus are often used to evaluate the effects of market interventions. Such an analysis is called a welfare analysis.

Let us use an earlier example. In Section 2.4.2, we studied the effect of a maximum price in a perfectly competitive market. With the help of Figure 10.2, we can now compare the social surplus with and without the maximum price.

Figure 10.2: Social Effects of a Maximum Price



The maximum price decreases the market price from p^* to p_{max} and the quantity from Q^* to Q_2. Before the maximum price was introduced, $CS = a + b$, $PS = c + d + e$, and the social surplus equaled $a + b + c + d + e$. After the maximum price is introduced, $CS = a + c$, $PS = e$, and the social surplus is $a + c + e$. The producers have consequently lost $c + d$, and the total welfare has decreased by $b + d$. The consumers have lost b but gained c.

Whether the consumers are better or worse off depends on if the area c is larger or smaller than b. However, social surplus is always diminished by introducing a maximum price in a perfectly competitive market. The amount of social surplus that is lost, $b + d$, is called the **deadweight loss**.

Deadweight loss: The welfare that is lost because of some imperfection in the market.

11 Monopoly

Monopoly can be viewed as the opposite of perfect competition. Instead of many firms, there is only one: the monopolist. This has important consequences for both price setting and the quantity produced.

Monopoly: A market with only one seller.

11.1 Barriers to Entry

Why do monopolies arise? There are many different reasons, but all of them have to do with barriers to entry in the market. The reasons for these barriers could be

- *Structural*. There are properties of the market that automatically shut competitors out:
  - *Economies of scale*. If there are economies of scale, large-scale advantages, the size of the firm is crucial for average cost. A situation can then arise in which only one firm can recover its costs. This is called a natural monopoly and an example of this is railroads.
  - *Cost advantages*. If the monopolist has access to a cheaper way of producing the good, for instance if she has a patent on a cheaper technology, she can push competitors out of the market.
- *Strategic limitations*. The monopolist can create barriers to entry. An example is **limit pricing**, where the monopolist sets the price so low that it becomes unattractive for competitors to enter.
- *Political*. The government may decide to grant a firm a monopoly in a certain market. A common example is for pharmaceutical goods.
- *Patents and exclusive rights*. If a firm has a patent on a certain good, other firms are shut out during the life span of the patent. It is also possible to have exclusive right to extracting, for instance, oil or metals.

Limit pricing: The monopolist sets a price lower than the monopoly price in order to keep competitors out.

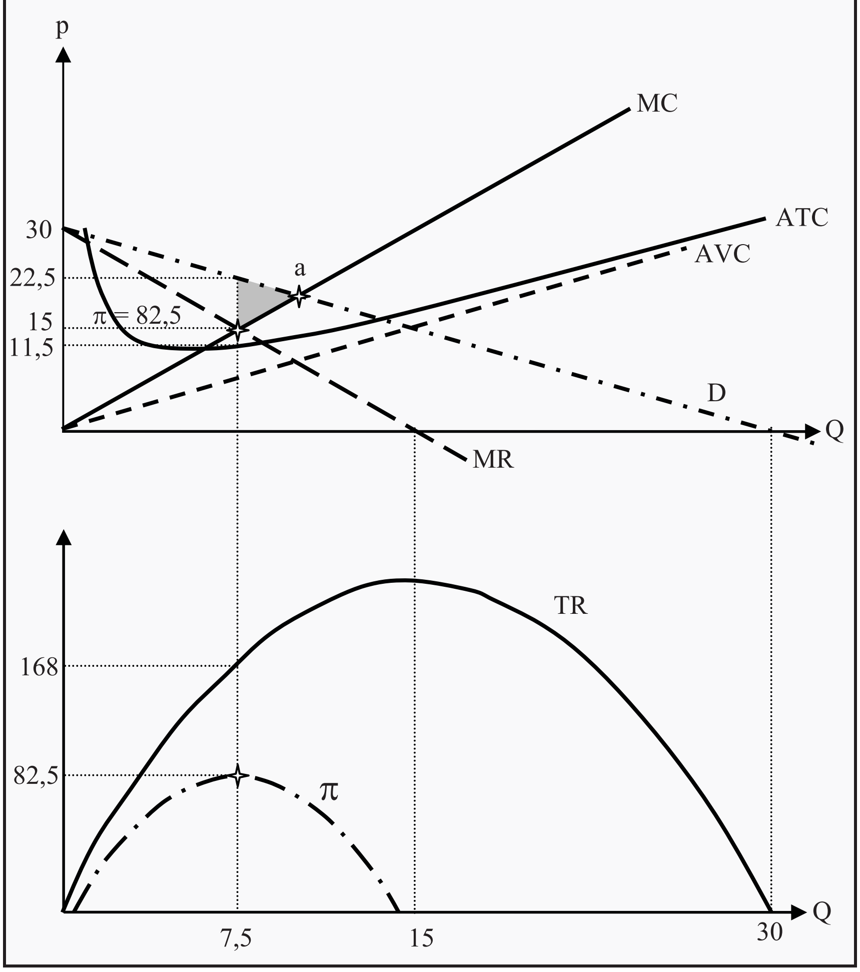
11.2 Demand and Marginal Revenue

In this chapter, we will assume that the monopolist charges all customers the same price. The monopolist faces the whole demand of the market. We can compare with a perfectly competitive market by looking again at Figure 9.2. The individual firm in a competitive market only faces a small part of the market. Therefore, it can be represented as in the right-hand side of the figure. A monopolist *is* the whole market. Therefore, it looks like the left-hand side of the figure. In order to sell more goods, the monopolist has do reduce the price, and the demand curve it faces will therefore slope downwards.

Now, note that the demand curve is decided by the consumers and not by the firm. It answers the question: *if* we would offer a certain price, how many units would we then be able to sell? In the perfectly competitive market, marginal revenue was equal to the price. That is not the case for a monopolist. For the monopolist to be able to sell an additional unit of the good, she must lower the price of *all* units. The total effect of selling one more unit then consists of both what she is paid for the last unit *and* of the reduction of revenue from all the other units that she now has to sell at the lower price. Consequently, the marginal revenue will be lower than the price.

Let us see what this means for a good with linear demand. If the demand curve is a straight line, the *MR* curve will also be a straight line with the same intercept on the Y-axis as the demand curve. However, it will have a slope with twice the magnitude. (To show that, we need to use the derivative, but this, again, is outside the scope of this book.)

Figure 11.1: The Profit Maximum for a Monopoly



We will use the demand curve $Q_D = 30 - p$, or if we solve for p: $p = 30 - Q_D$ (see Section 2.3.1). If the *MR* curve is to start in the same point and have a slope that is twice as large, its functional form must be $MR = 30 - 2*Q_D$. (The constant is the same, 30, and the slope is changed from -1 to -2). In Figure 11.1 the curves are drawn as *D* and *MR*. We have also drawn a marginal cost curve, *MC* (= 2**Q*), an average cost curve, *ATC*, and an average variable cost curve, *AVC* (= *Q*), and, in the lower part of the figure, total revenue, *TR*, and profit, π.

11.3 Profit Maximum

The monopolist wants to maximize her profit. She does that by producing the quantity, Q^*, at which $MC = MR$:

$$MC\left(Q^*\right) = MR\left(Q^*\right)$$

In Figure 11.1, this corresponds to the quantity 7.5, where both *MR* and *MC* equal 15. To see that this choice maximizes the profit, think of what would happen if she would produce more than that quantity. If she would produce one more unit, she would get paid *MR* but also incur a cost of *MC*. Moreover, since $MC > MR$, the cost is larger than the revenue and she would reduce profit; similarly if she would reduce the production.

The profit at a quantity of 7.5 is, according to the lower diagram, 82.5. The price the monopolist will charge is the one that the consumers, according to the demand curve, are prepared to pay when the total production is 7.5, i.e. 22.50. The corresponding *ATC* is 11.50. In other words, the monopolist makes 22.50 - 11.50 = 11 per unit sold, totaling to 11*7.5 = 82.5. This corresponds to the grey rectangle in the upper part of the figure.

Similarly to the firms in a perfectly competitive market (see Section 9.3.2), the price must also be above the average variable cost, *AVC*. If it is not, it is better to produce nothing at all. In the long run, the firm must also cover its fixed cost, and then the price must be above the average total cost, *ATC*.

In Figure 11.1, we have also indicated where total revenue is maximized. This occurs at the quantity $Q = 15$ and corresponds to the point in the upper part of the Figure where $MR = 0$. Note that this point does *not* maximize the profit. In the example, the firm makes a loss at that quantity.

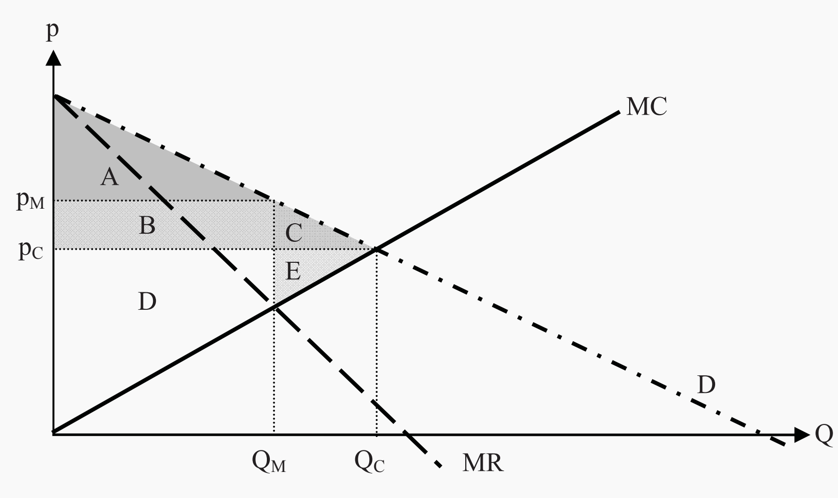
11.4 The Deadweight Loss of a Monopoly

To have a monopoly firm is often very beneficial for the monopolist, who can make a profit, but it is negative for society. To see why, look at point a in the upper part of Figure 11.1. In that point $MC = D$. The marginal cost is the cost of the last produced unit, and the demand is the consumers' corresponding valuation of that unit. At point a, the cost of producing the last unit and the valuation of that unit are the same. To the left of that point, the consumers' valuation is higher than the cost of producing additional units. Society therefore loses production of goods that cost less than what they are worth, according to the consumers' valuation. Then the quantity that the monopolist produces cannot be efficient. The deadweight loss that arises corresponds to the dark grey triangle in the figure. (Compare to Section 10.1.)

Let us look closer at the consequences. In Figure 11.2, we have drawn the supply curve, i.e. the *MC* curve. (Remember from Section 9.3.3 that the supply curve of a firm corresponds to the part of the *MC* curve that is above the average cost.) Furthermore, we have drawn the demand curve, *D*, and the marginal revenue curve, *MR*. Under perfect competition, equilibrium had been reached at the price p_C and the quantity Q_C, but in a monopoly market, equilibrium is reached at the price p_M and the quantity Q_M. The producer surplus (*PS*) in a monopoly market corresponds to the area $B + D$, and the consumer surplus (*CS*) corresponds to the area *A*. The social surplus will therefore be $A + B + D$. If the market had been perfectly competitive instead, *PS* had been $D + E$ while *CS* had been $A + B + C$ and the social surplus $A + B + C + D + E$.

In other words, society, particularly the consumers, loses utility: The deadweight loss is $C + E$. The monopolist, on the other hand, will lose if competition is increased. Note that the monopoly is not efficient (see Section 18.2): It would have been possible to produce the quantity Q_C to the price p_C and then compensate the firm by transferring the area *B* back from the consumers. The consumers would still get an increase in utility corresponding to the area *C* (they would now get $A + C$), and the producers would get an increase in utility corresponding to the area *E* (they would now get $B + D + E$). No group would lose anything. Therefore, the monopoly does not fulfill the Pareto criterion.

Figure 11.2: Deadweight Loss of a Monopoly



11.5 Ways to Reduce Market Power

To reduce the negative impact on society, governments often try to limit the **market power** of monopolists. Some popular measures include

Market power: Ability to affect the price.

- *Price regulations*. If it can be known what the market price would have been under perfect competition, or if the cost of production is known, the government can decide on a price ceiling at that price. Thereby, the equilibrium point is moved to the optimal point from society's viewpoint. It is, however, very difficult to estimate the optimal price.
- *Increase competition*. If the monopoly has been created through political decisions, the regulation can be changed.

12 Price Discrimination

There are more possibilities for a monopolist to take advantage of her situation: She can charge different prices from different customers. This is called **price discrimination**, and we distinguish between price discrimination of the first, second, and third degrees.

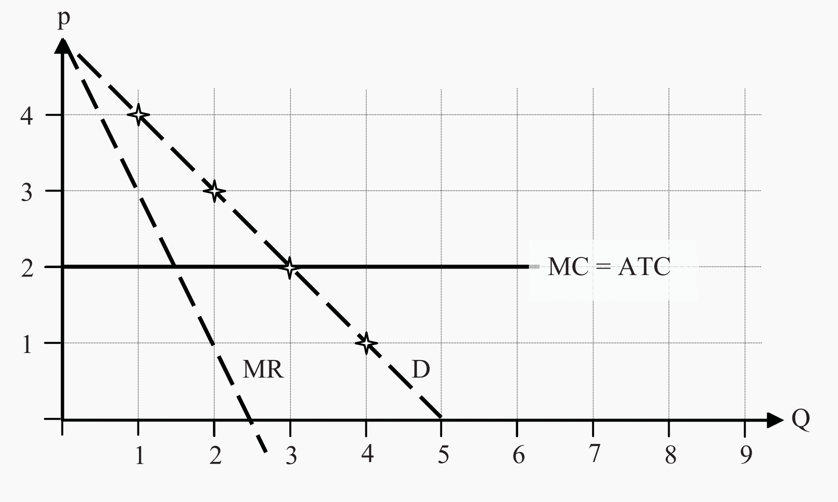
Price discrimination: Charging different prices from different consumers.

12.1 First Degree Price Discrimination

Remember that the demand curve corresponds to the consumers' valuation of different quantities of the good. Suppose, for example, that we have four presumptive consumers who want to buy a maximum of one unit of the good. The first is willing to pay 4 for one unit of the good, the others 3, 2, and 1, respectively. We then get a demand curve as D in Figure 12.1: If the price is 4, we sell one unit to the first customer, if it is 3 we sell one unit to each of the first two, and so on. We have indicated the **reservation prices** of each customer with a star and then joint them with a straight line.

Reservation price: The maximum price a buyer is willing to pay.

Figure 12.1: First Degree Price Discrimination



Furthermore, assume that the monopolist has a constant marginal cost, $MC = 2$, and no fixed cost. Then $AVC = ATC = MC$. In a perfectly competitive market, the equilibrium price would have been $p^* = 2$ and the quantity sold would have been $Q^* = 3$. The firm's revenue would have been 3*2 = 6 ($Q*p$), and its cost 3*2 = 6 ($Q*ATC$). Its profit would therefore have been zero. The consumer surplus (*CS*) would have been the sum of each customer's surplus, i.e. the difference between his or her valuation and how much he or she pays. For the first customer, the surplus is 4 - 2 = 2, for the second 3 - 2 = 1, and for the third 2 - 2 = 0. Therefore, we get a total consumer surplus of $CS = 3$. (Compare to Chapter 10.)

In a monopoly market of the type we analyzed in Chapter 11, the firm would have found the quantity at which $MC = MR$, i.e. $Q = 1.5$. If it can only sell whole units, it would have chosen to produce only one unit that it would have sold at a price of 4. The profit (= *PS*) would then be 1*(4 - 2) = 2, which is higher than in the perfectly competitive case. *CS* would be 4 - 4 = 0.

Suppose now instead that the monopolist knows the valuation of each consumer, and that the consumers cannot sell the goods to someone else if they have bought it. The monopolist can then use price discrimination of the first degree (also called *perfect price discrimination*): She charges a price from each customer that is equal to the maximum amount that customer is willing to pay. The first customer has to pay 4, the second 3, and the third 2. Since the monopolist has a marginal cost of production equal to 2, her surplus will be $PS = (4 - 2) + (3 - 2) + (2 - 2) = 3$. *CS* will be 0. Compared to a perfectly competitive market, the monopolist has won over all the *CS*.

Note that in this case, with first-degree price discrimination, the social surplus is as just large as in the case of perfect competition. The surplus has just been reallocated from the consumers to the producers. This means that this situation is actually efficient (see Section 18.2). It is another question whether it is fair.

12.2 Second Degree Price Discrimination

If the monopolist does not know the different valuations of different customers, she can instead use second-degree price discrimination. This amounts to offering different package solutions at different prices, and then the customers get to choose which package they prefer.

By choosing the composition of the packages in a clever way, she can get the customers to sort themselves into different groups. The goal of making them perform this type of self-sorting is to get the ones with high valuation to pay a high price and the ones with low valuation a lower price. An example of this type of price discrimination is quantity discounts.

12.3 Third Degree Price Discrimination

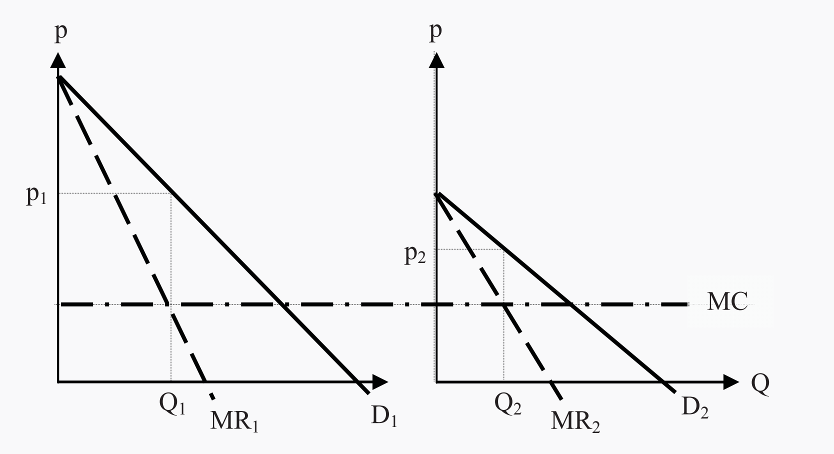
The third type of price discrimination amounts to dividing the market into two or more submarkets, where the valuations in the submarkets are different. Examples are different prices for children and grownups, or discounts for students and the unemployed. For this to work, it has to be possible to identify the consumer as actually belonging to a certain group.

The criterion for profit maximization under third degree price discrimination is, in principle, the same as before, but we have to separate demand and marginal revenue in the two (or more) submarkets. Mathematically, this can be written as

$$MC(Q_1 + Q_2) = MR_1(Q_1) = MR_2(Q_2)$$

In other words, the marginal cost of producing the total quantity has to be as large as the marginal revenue from the first submarket, and simultaneously as large as the marginal revenue from the second submarket. With constant *MC* and linear demand in both submarkets, this can be illustrated as in Figure 12.2. The firm chooses to produce the quantity $Q_1 + Q_2$, and then sell the quantity Q_1 at a price of p_1 in the first submarket and the quantity Q_2 at a price of p_2 in the second submarket.

Figure 12.2: Third Degree Price Discrimination



13 Game Theory

Before we go on to the other market forms, oligopoly and monopolistic competition, we will introduce a tool called *game theory*. Game theory is a much younger tool than most of the others we have discussed so far and has become a large field of research. Here, we will just present two different games. These will get to represent the two different groups of games: normal form games and extensive form games. We will later use these tools in the analysis of oligopolies.

13.1 The Basics of Game Theory

Game theory is used for analyzing how individual agents interact with each other. Depending on the structure, they may take into account how the other agents are choosing (or how they believe that they will choose), and then decide on their own strategy. This closely resembles the situation in many parlor games. Think, for instance, how the players act in chess: They only decide on their own moves, but they do so depending on how they believe that the opponent will respond. It is most often a bad strategy to hope that the opponent will not discover a trap. A better strategy is to assume that the opponent understands everything that one understands oneself, and then base one's strategy on that.

To turn this into a theory, we need to first define the basic building blocks. For a game to be a game in the relevant sense, we need players, rules, outcomes, and preferences over the outcomes. With chess, the players are “white” and “black,” the rules are the rules for chess, the usual outcomes are “white takes black’s king,” “black takes white’s king” or “draw.” The players’ preferences are usually that they prefer that they take the opponent’s king to a draw, and lastly that the opponent takes their king. In most games of game theory, we need to specify:

- *The players*. It could be individuals, firms, or countries. Often, there are only two or three players.
- *Actions*. All possible actions the different players can choose to do, for instance decide on quantity or price.
- *Information*. What each player knows at different stages of the game.
- *Strategies*. A strategy is a complete description of what a player will choose at each possible situation that could arise in a game. One can think of a strategy as a list. The list should be so exhaustive that another person could play instead of the player, and never actually have to decide anything by herself.
- *Payoffs*. The utility a player gets, given a certain outcome of the game.

13.2 The Prisoner's Dilemma

The game called The Prisoner's Dilemma is probably the most well known example of game theory. It is also an example of a **normal form game** (or **strategic form game**), which means that the players choose simultaneously. (Formally, a normal form game is a game that can be defined by only specifying the players, the strategies, and the payoffs.)

Normal / strategic form game: A game in which the players play simultaneously

One way to construct the game is the following. Two players, A and B, have been arrested (somewhere where the rule of law is somewhat substandard) and are kept in isolation. A prosecutor suggests A the following:

- If you confess and B does not, you will be set free as a sign of our gratitude. B will then get 10 years in prison.
- If the both of you confess, you each get 2 years in prison.
- If B confesses and you do not, you get 10 years in prison while B is set free.
- If none of you confesses, we will frame you for a petty crime and you will each have to pay a small fine.

At the same time, B gets the same suggestion. The two players cannot communicate with each other and therefore must consider a solution in solitude. Let us now identify the different elements that make this a game, in the game theoretical sense of the word.

- *The players*; Individuals A and B.
- *Actions*; For A: choose "Confess" or "Do not confess"; and similarly for B: choose "Confess" or "Do not confess.”

- *Information*; Both A and B know that the other has received the same offer, but they do not know how the other chooses.
- *Strategies*. Both A and B can only choose one of two different actions. Possible strategies for A are then "choose confess" or "choose not to confess", and similarly for B.
- *Payoffs*; Here we need to know the two players' preferences. For simplicity, we assume that they have the same preferences and that they are as follows: 10 years in prison (-10), 2 years in prison (-2), a small fine (-1), and freedom (+1).

Many normal form games can be represented with a so-called **payoff matrix**, where one player's strategies are displayed in the vertical direction and the other's strategies in the horizontal direction. Their respective payoffs that correspond to certain strategy pair are then indicated in the squares. If we do this for the present game, we get the payoff matrix in Figure 13.1. Note that player A's payoffs are to the left in the squares and player B's are to the right.

Payoff matrix: A matrix where the players' strategies are indicated to the left and at the top and the payoffs are indicated in the corresponding squares.

Figure 13.1: Payoff Matrix for the Prisoner's Dilemma

		Player B	
		"Confess"	"Do not confess"
Player A	"Confess"	-2, -2	+1, -10
	"Do not confess"	-10, +1	-1, -1

Let us first look at the game from the perspective of player A. She does not know how player B will choose, but she does know that player B will choose either "Confess" or "Do not confess." Say that player B would choose "Do not confess." Then, obviously, the best thing player A can do is to choose "Confess," since she will then get a utility of +1 (freedom) instead of -1 (a small fine). Now, say that player B chooses "Confess" instead. Then the best thing player A can do is still to choose "Confess," since she will then get a utility of -2 (2 years in prison) instead of -10 (10 years in prison). Consequently, player A has a strategy that is the best one, independently of what player B chooses. Such a strategy is called a **dominant strategy**.

Dominant strategy: A strategy that is never worse than any alternative, but is sometimes better.

Player B's problem is the same as player A's, and hence it is a dominant strategy for player B as well to choose "Confess." As a result, they both choose "Confess" and get two years in prison. This is so, even though it is possible for them both to get away with a small fine (if they both choose "Do not confess"). This is the dilemma. For both player A and B it is individually rational to confess, but acting that way they achieve an outcome worse than what is "collectively" possible. If they had been able to cooperate, they would both have been able to reach a higher utility level.

Games that have properties such as this one are called Prisoner's Dilemma-games. It could just as well be two countries deciding on whether to wage war on each other, two firms deciding on whether to start a price war or not, or two fishers deciding on whether to restrict their fishing or take the risk that the fish will go extinct. The players are kept from a rather good solution, because they choose their own individual best.

13.3 Nash Equilibrium

In the last section, we presented a solution of the Prisoner's Dilemma. With "solution", we here mean a prediction of how the players will play.

How does one generally solve a game? This is far from self evident, and in many games, there are several different reasonable solutions. The most popular concept for solving games is the **Nash equilibrium**. There are, however, several other ways in which to solve games, but most often, they are variations of a Nash equilibrium. Note also that, there can be more than one Nash equilibrium in a game.

A Nash equilibrium is:

- A set of strategies, one for each player.
- The strategies should be such that no player can improve her utility by unilaterally changing her own strategy.

13.3.1 Finding the Nash Equilibrium in a Game in Matrix Form

It is often easy to find the Nash equilibrium for a game in matrix form. Look at the game in Figure 13.1 again. We have four squares in the matrix. We can then find the Nash equilibrium by checking each square separately:

- {Do not confess, Do not confess}, i.e. the lower right square. Can any of the players improve her situation by unilaterally changing her own strategy? If, for instance, A changes to "Confess" she will get +1 instead of -1. (Similarly for B.) Consequently, she can improve her situation and this cannot be a Nash equilibrium.
- {Do not confess, Confess}, i.e. the lower left square. If A changes to "Confess", she will get -2 instead of -10. Consequently, this cannot be a Nash equilibrium either.
- {Confess, Do not confess}, i.e. the upper right square. If B changes to "Confess", she will get -2 instead of -10. Consequently, this cannot be a Nash equilibrium.
- {Confess, Confess}, i.e. the upper left square. If A would change to "Do not confess", she would reduce her utility from -2 to -10, and if B would change she would also reduce her utility from -2 to -10. None of the players can therefore improve on her situation by unilaterally changing her strategy, and this must be a Nash equilibrium.

The only Nash equilibrium in the Prisoner's Dilemma is that both players choose "Confess."

13.4 A Monopoly with No Barriers to Entry

We will now describe a game on so-called **extensive form**, where the question is whether a monopolist can uphold her monopoly if there are no barriers to entry. In a game on extensive form there is, in contrast to games on normal form, an order to the choices. One could say that we have added a time dimension.

Extensive form: A game where the players choose in an order.

There are two firms, The Incumbent (J) and the Entrant (E). J has, at the beginning, a monopoly in the market and E has to choose whether to enter the market or not. If she decides to enter it, J can choose to start a price war, i.e. lower the price to punish E, or to accept the competitor. The problem with a price war is that it also hurts J herself.

- The players; The Incumbent (J) and the Entrant (E).
- Actions; For E: choose "enter" or "not enter"; for J: choose "price war" or "accept."
- Information; E knows what the game structure looks like, but not how J will decide later on. J, on the contrary, knows how E has chosen when it is her time to choose. J consequently has more information than E.
- Strategies. For E there are two strategies:

  1. Choose "enter."
  2. Choose "not enter."

  For J, there are also two strategies:

  1. Choose "price war."
  2. Choose "accept."

- Payoffs. Here we need to know the players' preferences. Assume these are as in Figure 13.2.

This type of game is usually represented with a so-called **game tree**. The present game will look like in Figure 13.2.

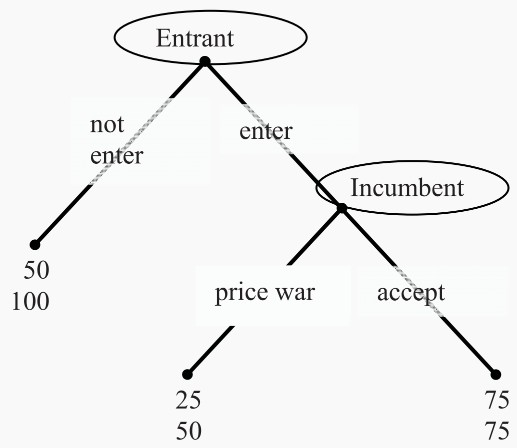
Game tree: A graphical illustration of a game on extensive form.

In the game tree, we have indicated where E and J decide, and what they can decide between at that point. At the far bottom, there are two rows of numbers. The number in the first row indicates the first player's (E's) payoff and the number in the second row the second player's (J's).

The game tree is read from top to bottom. It begins with E choosing between "not enter" and "enter." If the she chooses "not enter," the game ends and E gets 50 while J gets 100. If E, instead, chooses "enter," J gets to choose between "price war" and "accept." If she chooses "price war," the game ends and E gets 25 and J gets 50. Compared to the case when E chooses not to enter, both E and J get a lower payoff. If J, instead, chooses "accept," the game ends with E and J sharing the market and both getting a payoff of 75.

It is clear that J prefers that E does not enter the market (which gives J 100) to accepting (75), and both of these to a price war (50). E prefers to be accepted (75) to not entering (50), and both of these to a price war (25).

Figure 13.2: Game Tree



13.4.1 Finding the Nash Equilibrium for a Game Tree

To find the Nash Equilibrium for a game tree, we compare all different combinations of strategies. In the example in Section 13.4, E has two strategies and J has two. It is then possible to "translate" the game tree to a game on matrix form, as in Figure 13.3. Each strategy is translated into a row or a column. It now looks similar to the game from Section 13.2, but with other payoffs and strategies.

Figure 13.3: Payoff matrix for the Game Tree

		J	
		price war	accept
E	enter	25, 50	75, 75
	not enter	50, 100	50, 100

Note that in the case when E chooses "not enter", it does not matter what J chooses. Looking at the game tree in Figure 13.2, this is obvious since the game ends after such a choice and J never gets to choose. In the matrix, this translates into identical payoffs in all columns of the corresponding row, i.e. (50, 100).

To find the Nash equilibrium, we use the same method as in the last section and check each square separately. E chooses in the vertical direction and J in the horizontal. In Figure 13.3, we have inserted arrows from squares that have a better alternative to that alternative. Squares that have no arrows going *out* will then be Nash equilibria. In this case, there are two Nash equilibria:

- E chooses "enter" and J chooses "accept." If E unilaterally changes her strategy, she will diminish her payoff from 75 to 50, and if J does so, she will diminish her payoff from 75 to 50.
- E chooses "not enter" and J chooses "price war." If E changes her strategy, she will diminish her payoff from 50 to 25, and if J does so, she will get the same as before, i.e. 100. The latter is due to the fact that it does not matter what J chooses when E has chosen "not enter."

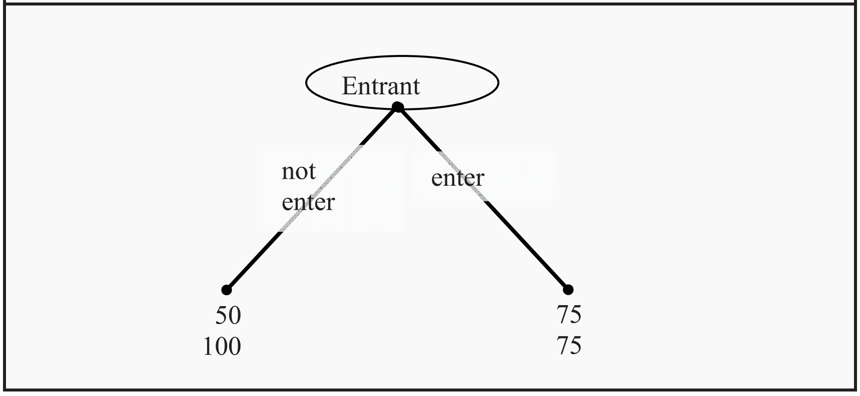
There is something odd with the latter Nash equilibrium. E chooses not to enter since J implicitly threatens with a price war. However, if E had established herself, J would have lost utility by actually starting a price war. According to the definition, this is a Nash equilibrium, but this objection leads us to introduce an alternative method of solving games on extensive form.

13.5 Backward Induction

For game trees, such as the one in Section 13.4, there is another, often used, solution method. The main idea is to start at the end of the tree, and then solve it backwards.

An example will make this clear. Consider the game tree in Figure 13.2 again. The last thing that happens in the game tree is that J chooses "price war" or "accept." If she chooses the first option, she gets 50 and if she chooses the second, she gets 75. Obviously, it cannot be optimal to choose the first. Consequently, given that E has chosen "enter," J will choose "accept." We can then reduce the game tree by omitting the alternative that J will not choose, and just keep the payoffs of the alternative that she does choose. The game tree will then look like in Figure 13.4.

Figure 13.4: Reduced Game Tree



Given that J will chose “accept,” the choice for E is simpler. If she chooses “not enter” she will get 50 and if she chooses “enter” she will get 75. Consequently, she chooses the latter. The solution, using backward induction, is then

- E: "enter"; J: "accept.”

Comparing this solution to the one in Section 13.4.1 (that had two different solutions, one of which is the same as the one here) this one seems more reasonable. Earlier, we found a Nash equilibrium in which E chose “not enter” and that included a (never realized) threat from J to start a price war. Using backward induction, the threat reveals itself as being empty, and the only solution is that E establishes herself and J chooses to accept. The solution one obtains by using backward induction is called **subgame perfect equilibrium**.

Subgame perfect equilibrium: A frequently used equilibrium concept for extensive form games.

14 Oligopoly

An **oligopoly** is a market in which there are only a few sellers. Most of the models in the literature only cover cases in which there are two sellers. Such markets are also called **duopolies**. As you will see, the analysis of oligopolies is quite complicated. Furthermore, there are several different models that all yield different results. This can be quite confusing. Take some time to see what the differences are in the assumptions and *why* they give different results. Which model to use, depends on what the situation is in a particular case. Different structures can have dramatically different effects on the market.

Oligopoly: A market with only a few sellers.

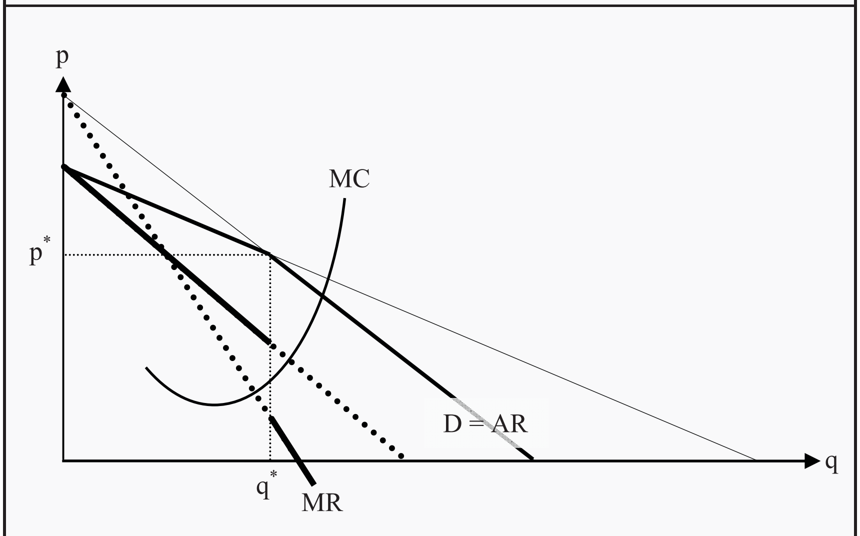
Duopoly: A market with only two sellers.

14.1 Kinked Demand Curve

Assume there are only a few firms in a market and that they all produce exactly the same good. Furthermore, assume that there is already a price has already been set. (For now, we will ignore the question from where this price has come.) If we were one of the firms, how would we reason regarding our own price setting?

What would happen if we would raise the price? Most of the customers would then buy from our competitors instead, to get the good at a lower price. The competitors would probably not lower their prices, as they would gain a larger market share instead. Consequently, we would sell *much* fewer goods. Conversely, what would happen if we lowered our price? If the competitors did not also lower their prices, we would gain a large part of their market shares. Since that would mean that they would reduce their profits, they would probably lower their prices as well.

Figure 14.1: Kinked Demand Curve



The slope of the demand curve that our firm faces is therefore different depending on whether we increase or decrease our price. This results in a so-called *kinked demand curve*, where the bend occurs at the existing price, p^*, and the corresponding quantity, q^* (see Figure 14.1).

The bend in the demand curve makes the construction of the *MR* curve more complicated. To find the *MR* curve, we extend the two parts of the demand curve, *D*, until they reach the two axes. (The extensions are the thin lines in the figure.) As the (real and imagined) demand curves are downward sloping, their corresponding *MR* curves will also be downward sloping, intercept the Y-axis at the same points, but have a slope which is twice as large. (Compare to the reasoning regarding the *MR* curve of a monopoly, Section 11.2.) Since *D* now has two parts, we do this for each part separately.

Since the demand curve is bent at the quantity q^*, the *MR* curve will also change at that quantity. To the left of q^*, we use the *MR* curve that is derived from that part of the demand curve that is valid to the left of that quantity. To the right of q^*, we instead use that *MR* curve that is derived from the part of the demand curve that is valid to the right of it. This causes the final *MR* curve to make a jump at q^*. In Figure 14.1, the final *MR* curve is indicated with thick full lines. The parts that are not used, since they correspond to the extensions of the demand curve, are indicated with thick dotted lines.

Just as before, a criterion for profit maximization is that the firm sets the quantity where $MR = MC$. In the models we have used this far, that criterion singled out exactly one point. However, since the *MR* curve now makes a jump at the quantity q^*, the *MC* curve can intersect the *MR* curve in that interval. At the prevailing price, it must do so by construction. This means that if the marginal cost, *MC*, only changes a little bit, for instance because a small tax is introduced on each unit sold, the firm might not change its produced quantity, and consequently not the price. As long as the *MC* curve still intersects the *MR* curve at the jump, the firm will produce the same quantity, q^*. However, the increase in *MC* will lead to a reduction in profit for the firm.

This is one way that the real-world phenomenon of **sticky prices** can be explained. According to the previous market-models (perfect competition and monopoly), prices should change immediately if quantities change. However, often we see that prices change more seldom; they seem to be stuck at a certain level for a while. If prices and quantities are set according to the kinked demand curve-model, however, this is exactly what we should expect.

Sticky prices: Prices are resistant to change.

14.1.1 How does the Price in the Kinked Demand Curve Arise?

In the analysis in last section, we ignored the question of how the price had arisen. One idea is that the sellers can have agreed on the price. If sellers can cooperate on the price setting, they will optimally agree to set a price that corresponds to the quantity a monopoly would have chosen, since the monopoly profit is the largest one can possible make in a market. Then they could split the monopoly profit between themselves.

However, that would amount to setting up a cartel, and that is against the law. Many people argue, however, that firms can have tacit agreements. There is no real cartel, but there is a sort of silent agreement that each seller should set a high price (see, however, Section 14.4). A frequently used example is the way gasoline distributors set their prices. Often, one firm announces that they will increase their price. Then the other firms follow immediately. Note however, that this type of behavior can also be against the law.

14.2 Cournot Duopoly

The Cournot model is a model of duopolies and is developed in line with the game theoretical approach we presented in last chapter. The Cournot model assumes that:

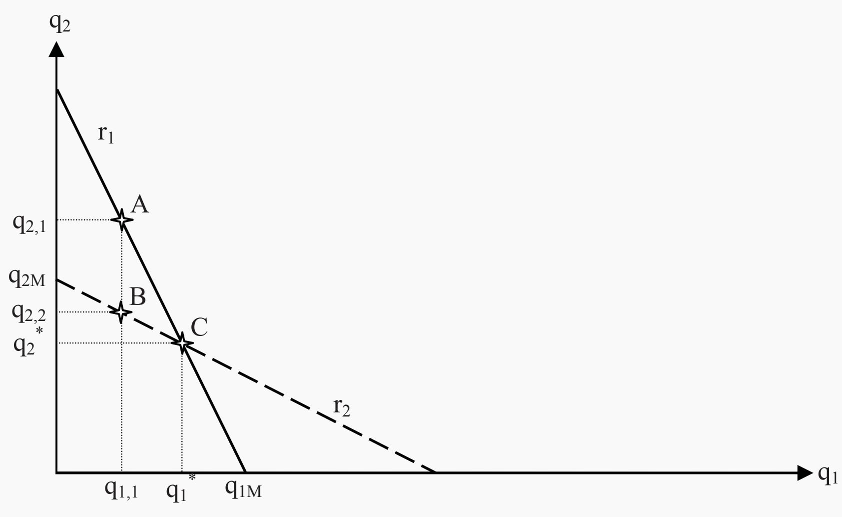
- We have two firms.
- They set quantities (and the price is then set by the market, given the quantity).
- They choose simultaneously, without knowing which quantity the other chooses.

How would these two firms reason? Both of them want to maximize their own profit. However, each firm's profit partly depends on the quantity set by the other firm, as total quantity determines the market price.

If a firm knows the quantity the other firm has chosen, then it is able to decide exactly which quantity that would maximize their own profit. There is an optimal response to each choice of the other firm. Let us use that observation, and determine that best response for each choice of quantity the other firm can possibly make. If we do that, we get a so-called **reaction function**. In Figure 14.2, r_1 is firm 1's reaction function and r_2 is firm 2's.

Reaction function: A function that describes the optimal response to someone else's choice.

Figure 14.2: The Cournot Model



To give an example of how to interpret the reaction function, suppose that firm 2 chooses to produce the quantity $q_{2,1}$. Which is firm 1's optimal response? Indicate $q_{2,1}$ on the Y-axis, go to line r_1 (point A) and read off the corresponding value on the X-axis: $q_{1,1}$ is firm 1's optimal response.

Note however that if firm 1 chooses the quantity $q_{1,1}$, then the quantity $q_{2,1}$ is not optimal for firm 2. Instead, the quantity $q_{2,2}$, at point B, is optimal for firm 2. However, then $q_{1,1}$ is not optimal for firm 1… and so on.

It is possible to show that the only point where both firms simultaneously respond optimally to the other's choice is point C, where the two reaction curves intersect each other. As no agent can achieve a better outcome by unilaterally changing her strategy, we have a Nash equilibrium (see Section 13.3). The conclusion of the Cournot model is then that, both firms will choose the Nash equilibrium quantities, q_1^* and q_2^*. Note that, if you continue to use the method of finding successive optimal responses as we did above, you will tend to get closer and closer to the Nash equilibrium in each round.

One should also note another thing in Figure 14.2. If firm 2 would produce nothing at all, firm 1 would be a monopolist in the market. The optimal quantity would then be the monopoly quantity. Similarly for firm 2. The reaction function of each firm must consequently hit the firm's own axis at the monopoly quantity. In the figure, these points are labeled q_{1M} and q_{2M}.

14.3 Stackelberg Duopoly

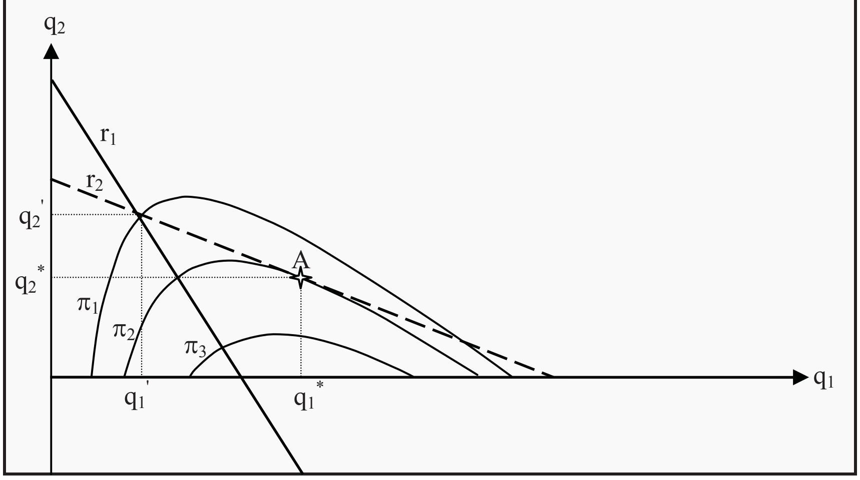
In the Cournot model, both firms made their decisions simultaneously and without knowing the other's decision. In the Stackelberg model, they decide one *after* the other. We call the one that chooses first, the *Leader* and the other one the *Follower*.

- We have two firms.
- They set quantities (and the price is set by the market).

- *Leader* first decides on her quantity, and then *Follower* decides on hers.

We will use the same reaction function as in the Cournot model, but the analysis will now be different since they do not choose simultaneously. *Leader*, who sets her quantity first, has an advantage. She knows that *Follower* will later set her quantity according to her reaction function. Therefore, *Leader* sets her quantity to maximize her own profit, given *Follower's* optimal response.

Figure 14.3: The Stackelberg Model



One way to illustrate this game is presented in Figure 14.3. We have drawn the reaction functions, r_1 and r_2, but we have also added a few curves indicating *Leader's* profit, π_1, π_2, and π_3; so-called **isoprofit curves**. Such curves show different combinations of q_1 and q_2 that give *Leader* the same profit. For instance, all combinations along π_1 give *Leader* a profit of π_1, etc. Note that *Leader's* profit increases *inwards*, the closer to the monopoly quantity (the point where r_1 intersects the X-axis) we get. The profit at π_2 is consequently higher than at π_1, and even higher at π_3.

Leader knows that *Follower* will choose her quantity along the reaction function r_2. *Leader* therefore finds an isoprofit curve that touches r_2 and that is as close to the monopoly quantity as possible. In the figure, the isoprofit curve π_2 touches r_2 in point A. *Leader* then chooses the quantity that corresponds to point A, i.e. q_1^*. As a response, *Follower* later chooses the quantity q_2^*.

Note that every other choice of quantity for *Leader*, higher or lower, must result in a lower profit for her. If she, for instance, would choose the quantity q_1' instead, *Follower's* reaction would be to choose q_2' and *Leader's* profit would be π_1, which is less than π_2.

14.4 Bertrand Duopoly

In the two preceding models, we have assumed that the firms set quantities. What happens if, instead, they set prices? The Bertrand model assumes that

- We have two firms.
- They set prices (and quantities are set by the market).
- They set prices simultaneously, without knowing which price the other one sets.

The previous models produced results that were very favorable for the firms but less so for the consumers. The Bertrand model, however, puts the two firms in a Prisoner's Dilemma-type of situation (see Section 13.2), and forces them to set $p = MC$, i.e. they set the same price as firms would do in a perfectly competitive market. This is, of course, unfavorable for the firms, but an improvement for consumers and society.

To see that the firms will set $p = MC$, suppose that we know that the other firm has set a high price. Which is then the best price we can set? Remember that we have homogenous (meaning identical) goods, so the consumers will not care from whom they buy it. Furthermore, they have perfect information about all prices. If we choose a price that is just below our competitor's, all customers will buy from us. This is a good situation for us, but far from optimal for the other firm. If they reason in the same way, they will want to set a price just below ours. Then we would lose all customers… and so forth.

No price above MC can consequently be an equilibrium. Regardless of which price the firm has set, the other will always want to undercut it and set a price just below its competitor. The only price that can be an equilibrium is then $p = MC$. At that price, none of the firms can lower their price since they would then make a loss. None of them would be able to make a profit by increasing the price either, since they would then lose all customers.

The surprising result is then that, since $p = MC$, we get the same outcome as in a perfectly competitive market, even though there are only two firms. If society is able to construct an oligopoly such that it becomes a Bertrand duopoly, there will be no loss of efficiency.

One way for the firms in a Bertrand market to increase profits anyway, is to try to differentiate their products. The customers will then not be indifferent between from whom they buy and the firms become two monopolists, however with goods that are very close substitutes. We will look at this type of situations in Chapter 15.

15 Monopolistic Competition

We ended last chapter by noting that a firm might be able to increase its profit by differentiating its products from those of its competitors. Most often, however, the products will still have many properties in common, which makes them close substitutes. Popular examples include Coca Cola and other cola- or soft drinks, and different brands of laundry detergent.

This behavior makes the firm a monopolist on their own product, for instance on Coca Cola, but with customers that have close substitutes to choose from, for instance Pepsi Cola. If the firm raises the price, some customers would move to the substitute, but not all of them. Similarly, if the firm would lower the price, they would attract some of the competitors' customers, but not all of them.

Note that, if the products were identical, we would have an oligopoly (see Chapter 14). If the firms, in addition, compete with prices, we would have a Bertrand situation (see Section 14.4) and none of the firms would make a profit.

15.1 Conditions for Monopolistic Competition

Criteria for monopolistic competition include

- There are several producers in the market
- The products are not identical, but they are close substitutes.
- There are no barriers to entry.

These conditions imply that each firm will face a downward sloping demand curve: If they increase the price, they will sell less and if they decrease it, they will sell more. However, the demand curve is very elastic (see Section 4.3.1) since there are close substitutes, so the customers will react quite strongly to price changes and quickly shift over to (or from) the competitors.

15.2 Market Equilibrium

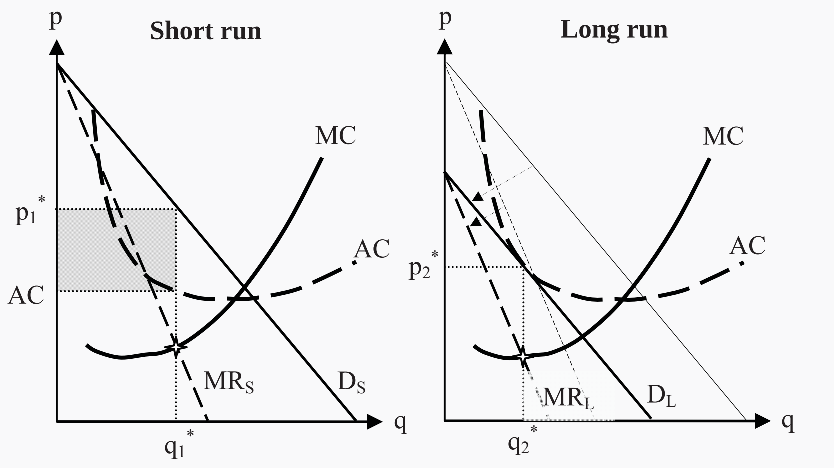
15.2.1 Short Run

In the short run, no new firms can establish themselves in the market (since the quantity of capital, by the definition of the short run, is fixed). To the left in Figure 15.1, D_S is the short-run demand curve an individual firm faces in a market with monopolistic competition, and MR_S is the corresponding marginal revenue. Similar to a monopoly, the MR curve is twice as steep as the demand curve. The firm, as always, maximizes its profit by choosing the quantity, q_1^*, that makes $MC = MR_S$. Since the average cost, AC, is below the price at that quantity, the firm makes a profit, $q_1^*(p_1^* - AC)$, corresponding to the grey rectangle in the figure.

15.2.2 Long Run

Since the firms make a short run profit and there are no barriers to entry, new firms will establish themselves in the market. Thereby, the demand curve that the individual firm faces changes so that at each price it is now possible to sell a smaller number of goods. This means that to the right in Figure 15.1, where we have the situation in the long run, the demand curve, D_L, and the marginal revenue, MR_L, have shifted inwards (see the arrows in the figure).

Figure 15.1: Equilibrium in the Short and Long Run for Monopolistic Competition



How far do they shift? They shift until there is no profit. Remember that, the firms choose the quantity that maximizes profit, i.e. the quantity that makes $MC = MR$. The demand curve, D_L, will consequently shift until the quantity where the firm maximizes its profit, q_2^*, is such that the price the firm can take for the good, p_2^*, is exactly equal to the average cost, AC. At that point, the profit is $q_2^*(p_2^* - AC) = 0$.

Note that the production is not efficient. Even in the long run we have that $p > MC$, which means that the cost of producing additional goods is lower than the consumers' valuations. If we compare to the results for perfect competition in the long run (see Section 9.5), we see that one difference is that long-run production in the case of monopolistic competition does not end up at the lowest point of the *AC* curve. This, in turn, means that there are unexploited economies of scale (compare to Section 8.3). Had we had fewer firms in the market, and thereby larger firms to satisfy the demand, they would have come closer to the lowest point on the *AC* curve. On the other hand, we would then have had fewer (close substitute) products between which to choose. It is not possible, without a more detailed analysis, to say what balance between these two – lower unit costs or more products to choose from – that is the best for the consumers.

16 Labor

To produce goods and services, a firm uses raw materials, labor, and capital. We will now look at the market for labor. The workers sell their labor, or alternatively the sell their leisure time, for a wage, and their supply depends on their valuations of leisure and wage, respectively.

From the firm's perspective, it buys labor as long as that gives a positive contribution to its profit. The firm's cost of labor is the wage, and its revenue of labor is the price at which they can sell the goods. The firm will consequently hire workers until the last produced unit of the good costs as much to produce as the firm is paid for it.

This means that the structure in the output market, i.e. the market where the firm sells its goods, will also affect what the firm will be willing to pay in wages, since it is in the output market that the price is set. We will study the cases when the output market is a perfectly competitive market and when it is a monopoly market. Furthermore, the structure of the labor market also affects the outcome. We study cases in which either the firm, or the workers, or both of them are in a monopoly position or in a perfectly competitive situation.

16.1 The Supply of Labor

We will assume that the workers prefer leisure to work and that they work for, and only for, the wage. There are 24 hours in a day, which sets an upper bound for how much labor a worker can sell. To analyze the supply of labor, it is useful to redefine the question: Instead of studying the supply of labor, we will study the demand for leisure. The supply of work will then be 24 minus the amount of leisure. Then, we can analyze the situation as in Figure 16.1.

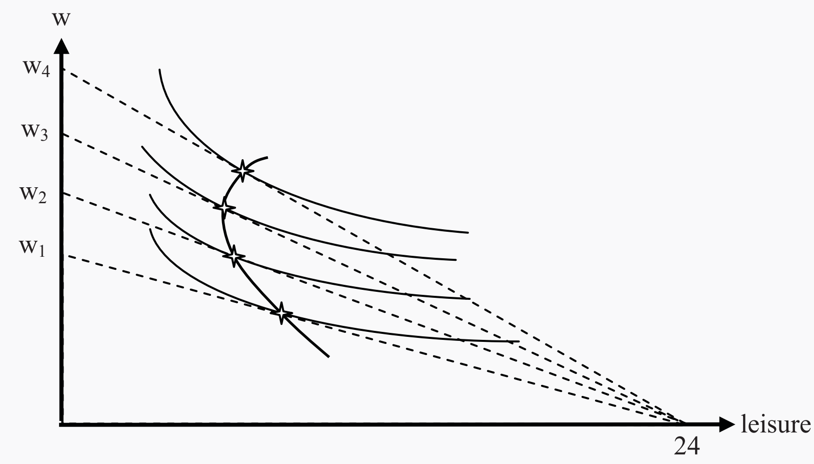
The situation here is a variation of the analyses we performed in Section 4.1.1 and in Chapter 5. As we are studying the consumption of two goods, leisure and wage (where, again, it is useful to think of money, the wage, as "all other goods than leisure"), increases and decreases in wage will have both substitution- and income effects.

To see the similarity with the analysis in Chapter 5, note that one can view an increase in wage as a decrease in the price of "all other goods." The budget line will then rotate around a fixed point at 24 (as a day has exactly 24 hours) on the X-axis and intersect the Y-axis at different points depending on the wage. Compare, for instance, to Figure 5.1. (However, note one thing: Here, the price of the *other* good changes (i.e. wage, not leisure), not the price of the one we are analyzing.)

In Figure 16.1, we have drawn four budget lines, corresponding to four different levels of income, *w* (for wage), and four indifference curves. The indifference curves, as always, indicate combinations of the two goods that the individual is indifferent between, and she strives to maximize her utility given the budget restriction. As before, the point of maximization occurs where an indifference curve just barely touches a budget line. We have indicated four such points of tangency in the figure, and then connected them to a curve. That curve corresponds to the individual's demand for leisure, and indirectly (if you take 24 minus her demand for leisure) to her supply of labor.

The odd thing about the supply curve for labor is that it slopes back again at high wages. Remember that the effect of a price change can be divided into a substitution effect and an income effect (see Chapter 5). At initially low wages, an increase in the wage often leads to an increase in the labor supplied. That is due to the substitution effect dominating over the income effect. The substitution effect makes the wage more attractive relative leisure, whereas the income effect makes the individual wealthier. The increase in wealth can lead to an increased consumption of both "other goods" (the wage) and of leisure.

Figure 16.1: The Individual's Supply of Labor



The higher the wage is, the more important the income effect will be, until finally it will start to dominate over the substitution effect. If a well-paid individual has her wage increased even further, she may choose to work less than she used to. This is what makes the supply curve for labor bend backwards for high wages.

16.2 The Marginal Revenue Product of Labor

In Chapter 7, we studied the firm's production, and we defined the production function as a function of labor, L, and capital, K, such that $q = f(L,K)$. In the long run, both L and K are variable, in the short run only L is variable.

If the firm buys one more unit (for instance, one more hour) of labor, how will that affect the profit? Remember that we defined the marginal product of labor, MP_L, in Section 7.2.1. In words, MP_L is how much more of the good that will be produced if we increase labor by a small amount (say, one more unit), given that everything else is held constant. When the firm decides whether to buy more labor, it first asks how large the *value* of the extra production is, i.e. how valuable is MP_L? We can express this value mathematically as

$$MRP_L = \frac{\Delta TR}{\Delta L} = \frac{\Delta TR}{\Delta q} \cdot \frac{\Delta q}{\Delta L} = MR \cdot MP_L$$

Here, MRP_L, is the **marginal revenue product of labor**. It corresponds to how much total revenue, TR, changes because of a small increase in L. In the third step above, we have divided and multiplied by Δq in order to show that $MRP_L = MR*MP_L$. In other words, if we hire one more unit of labor, the value added is how many additional units of the good is produced during that hour (MP_L) times at which price can we sell each additional unit (MR). That value added is called the marginal revenue product of labor.

Marginal revenue product of labor: If one more hour of work is done, what is the value of the units produced?

16.3 The Firm's Short-Run Demand for Labor

As we have mentioned, the firm's demand for labor depends on what the market for the firm's output looks like, as well as on the level of competition in the labor market.

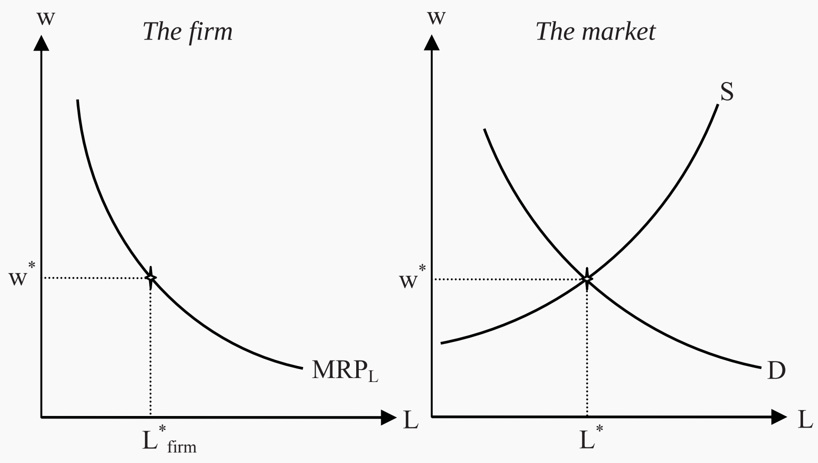
16.3.1 Perfect Competition in both the Input and Output Market

In the simplest case, we have perfect competition in both the input and in the output market. The firm cannot influence the price, p, on the good, which, in turn, makes the marginal revenue equal to the price (see Section 9.3.1): $MR = p$. The value of the marginal product of labor is consequently $MRP_L = p*MP_L$. Furthermore, the marginal cost of labor equals the wage, $MC_L = w$. The firm will then hire workers as long as $MRP_L > w$, i.e. as long as the revenue is higher than the cost of hiring, and the criterion for equilibrium is

$$MRP_L = w.$$

Note now that this means that the MRP_L curve will become the firm's demand curve for labor. Furthermore, remember that we have the law of diminishing marginal returns (see Section 7.2.2). $MRP_L = p*MP_L$ will therefore, eventually, start to diminish the more workers we hire (since MP_L will diminish while p is constant). We will then get a downward sloping demand curve for labor, as in the left part of Figure 16.2. (Compare also to Figure 7.1.)

Figure 16.2: The Firm's Demand for Labor and the Market Equilibrium

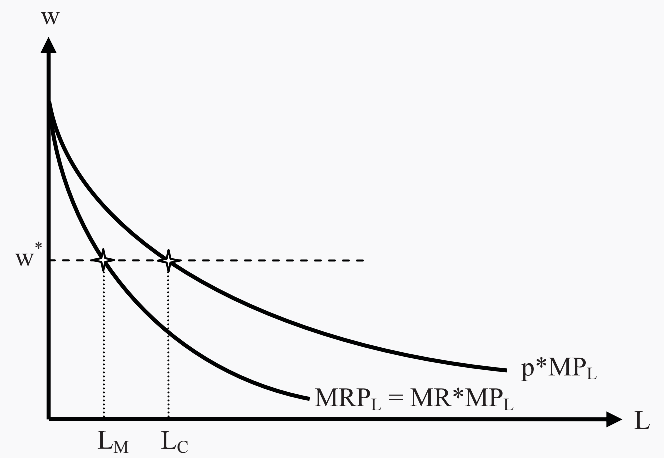


Since we have perfect competition in the labor market, both the firm and the workers take the wage, *w*, as given, and as we have perfect competition in the output market as well, the firm takes *p* as given. The market's demand for labor is the sum of all firms' demand curves, and the market's supply is the sum of all individuals' supply curves (to the right in Figure 16.2). The individual firm will then hire workers until $MRP_L = w$.

16.3.2 Monopoly in the Output Market

We continue to assume that there are many buyers and sellers of labor, but now we assume that the good is sold in a monopoly market. The firm maximizes profit in the same way as before, i.e. it hires workers until the cost, w, is as large as the marginal revenue product, MRP_L. However, since the good is now sold in a monopoly market, MR will not be equal to the price anymore. Instead, MR ($< p$) will fall with increased production (see Section 11.2). This, in turn, means that MRP_L ($= MR*MP_L$) will be steeper than in the case of perfect competition in the output market (since now both MRP_L and MR are downward-sloping curves). The monopolist produces a smaller quantity than a firm in a competitive market does, and therefore she will hire fewer workers.

Figure 16.3: The Demand for Labor when the Output Market is a Monopoly



In Figure 16.3, we have drawn the demand curves for labor, both for a monopolist and for a firm in a competitive market. The monopolist's demand curve is MRP_L ($= MR*MP_L$) and it will lie below the MRP_L ($= p*MP_L$) for a firm in a competitive market. The wage, *w*, is set in a competitive labor market and cannot be affected by either workers or firms. However, the firm hires fewer workers, L_M, than it would in a competitive market, L_C. A monopoly in the output market will consequently create inefficiencies in both the market for goods and in the labor market.

16.3.3 Monopsony in the Input Market

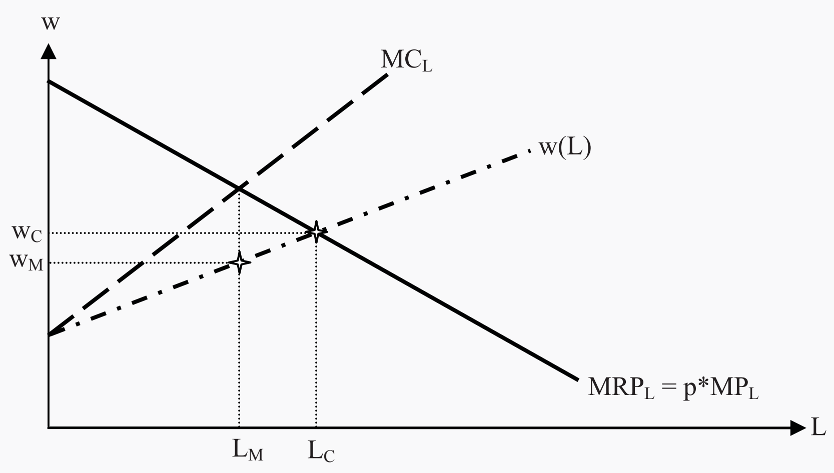
Monopsony: A market with only one seller.

Monopsony (mono = one; opsonia = buy) means that there is only one buyer. In the labor market, this means that there is only one buyer of labor. In countries where the government operates the health care system, it is, in effect, a monopsonist on, for instance, the market for nurses. The analysis of monopsonies parallels the one of monopolies.

Suppose, again, that there is perfect competition in the output market and that that there are many sellers of labor. However, there is only one buyer of labor. If the monopsonist increases the wage, she must do so for all workers, even the ones she has already hired. Thereby, her marginal cost of hiring one additional unit of labor is higher than the wage to the last worker. (Compare to the reasoning regarding *MR* for a monopolist in Section 11.2.)

Since the monopsonist is the only buyer in the market, she faces the whole supply of the market. (Compare to the monopolist, who faces the whole demand of the market.) Her marginal cost of labor, MC_L, will therefore have a steeper slope than the supply curve (see Figure 16.4).

Figure 16.4: Monopsony



As in the other cases, the monopsonist hires workers as long as she gets higher revenues from doing so than she has to pay in wages. Now, since *MC* is not equal to the wage any more, she does so as long as $MRP_L > MC_L$. In equilibrium

$$MRP_L = MC_L.$$

Note now that, in the previous sections, 16.3.1 and 16.3.2, we had that $MC_L = w$. Now, however, we have that $MC_L > w$. In Figure 16.4, the monopsonist hires L_M workers and pays a wage of w_M. Comparing to the case when we have competition in the labor market, we see that the wage in a monopsony, w_M, is lower and that the firm hires fewer workers, $L_M < L_C$. This parallels the results from the monopoly market, where the monopolist produced a smaller quantity than a perfectly competitive market did, and charged a higher price per unit.

16.3.4 Bilateral Monopoly

A **bilateral monopoly** is a situation in which there is only one buyer and one seller. Both parties will then have market power, and the outcome depends largely on negotiations, the business cycle, etc. This resembles the situation in some countries where there are centralized negotiations between unions representing the workers and other organizations representing the employers.

Bilateral monopoly: A market with only one seller and one buyer.

17 Capital

If the firm rents its capital, the problem of how much to rent is, more or less, the same as the problem of how much labor to use. It rents precisely so much that the rental rate becomes equal to the marginal revenue product of capital (compare to Chapter 16), i.e. until

$$MRP_K = r,$$

where r is the rental rate. In Chapter 8, we used the rental rate to price capital.

If instead the firm is to *invest* in capital, the problem is very different. The quantity of capital is not variable in the short run. Often, the firm decides about investments that are to remain in place during many years and that are expected to generate future profits. It is therefore necessary to compare flows from different points in time with each other: Is it ok to spend a large sum today to get access to future profits of certain expected size and that would come scattered over several years. How should one deal with the risk that the future profits will be lower than expected?

17.1 Present Value

What does time mean for the value of a certain sum of money? Suppose that we want to know how much 100 (of any currency) that we will get one year from now is worth today. We could then reason like this: If it were possible to put 98 in a bank account at an interest rate of 2%, then in a year that would be worth 98*1.02 = 100. Consequently, 100 in one year should be worth 98 today; the **present value** is 98. If we use that formula backwards, 100 in one year is worth 100/1.02 = 98 today, if the interest rate is 2 %. Similarly, 100 in two years is worth $100/(1.02)^2 = 96$ today, or more generally

Present value: The value today of future cash flows.

$$PV(100;R,n)=\frac{100}{(1+R)^n},$$

where *PV(*100,*R,n)* is the present value of 100, when the interest rate is *R* and the cash flow is in *n* years. If one knows which discount rate to use, it is easy to calculate the present value of future payments.

17.1.1 Bonds

We can use the present value formula to get the price of a bond. A bond is a security that gives the holder the right to a certain sum of money, the principal or face value, in a certain number of years. Often, the bond also pays a smaller amount each year, called the coupon. Some of the most important types of bonds are government bonds and corporate bonds.

Suppose we have a government bond that pays a coupon of 100 each year for ten years and then a face value of 1,000 in the tenth year. That means we will get ten payments. To calculate the present value of the future cash flows, we use the present value formula on each payment:

$$P_B=\frac{100}{1+R}+\frac{100}{(1+R)^2}+\frac{100}{(1+R)^3}+\ldots+\frac{100}{(1+R)^{10}}+\frac{1000}{(1+R)^{10}},$$

where P_B is the price of the bond. If the discount rate is 2% as before, the bond would be worth 1 719.

Note, however, that bonds usually are *not* priced this way. Instead, the price is decided on the market; either in an auction or is negotiated between buyers and sellers. Then, one uses the formula to back out which discount rate is compatible with the price, the coupons, and the face value. It is usually impossible to find a formula to that. Instead, one uses a numerical method (that is often preprogrammed into calculators or spreadsheet programs) to find the correct discount rate. Suppose that the ten year bond above does not cost 1 719, but is traded at a market price of 1,000. It is then possible to calculate that the discount rate is 10 %. In other words, if you use 0.10 instead of *R* in the formula above, the sum will be exactly 1,000. The discount rate one gets this way is called the **yield to maturity**, and can be viewed as an internal rate of return on the bond.

Yield to maturity: A sort of average return that one gets if one holds a bond until maturity.

There is also an odd type of bonds called perpetual bonds, which have no end date. They pay a coupon each year forever. This type of bond is surprisingly easy to price. It is possible to show that the value of perpetual bond that pays 100 each year is

$$P_{Perpetual} = \frac{100}{R}.$$

Conversely, if we know the price but not the yield: $R = 100/P$.

17.1.2 Stocks

Just as in the case of bonds, the price of stocks is usually set in the market. In principle though, one can use the present value formula to price them as well. As owner of a stock, one has the right to future **dividends**. With expectations of how large these will be, the present value of the future cash flows is:

Dividend: A (usually yearly or quarterly) payment to the stock-owners.

$$P_S = \frac{D_1}{1+R} + \frac{D_2}{(1+R)^2} + \frac{D_3}{(1+R)^3} + \dots$$

where P_S is the price of the stock and D_i is the expected dividend in year i. Unfortunately, this formula is impossible to use since it contains an infinite number of terms and D_i can vary between periods. However, if one is willing to assume that the dividends will grow at a constant growth rate, g, and that R is constant, it is possible to show that the formula can be simplified as (compare to the formula for a perpetual bond)

$$P_S = \frac{D_1}{R-g}.$$

Note, however, that these assumptions are quite unrealistic.

17.2 Correction for Risk

How do we get the correct discount rate? We have used R for all the discount rates above. However, interest rates for different securities differ substantially, and also change over time. The primary reason why, for instance, different bonds have different yields is that they have different levels of risk. A government bond is usually considered risk free whereas a corporate bond always has a risk that the issuer will not be able to fulfill her commitments. Interest rates for mortgages vary with the risk that the borrower will not be able to pay interest or installments. Discount rates for stocks are not quoted, but if calculated they differ wildly depending on the underlying risk. The discount rate for a risky asset will consequently be higher than for a risk free asset (such as a government bond). Often, the higher discount rate is called a risk-corrected discount rate.

17.2.1 Diversifiable and Nondiversifiable Risk

Regarding risk, we must differentiate between diversifiable and nondiversifiable risk. Suppose you invest all your savings in stocks in only one firm. You are then exposed to the risk that this stock will rise or fall in value in an unpredictable way. That risk can be quite big. You can decrease the risk by instead investing in a mutual fund. The part of the fund that is invested in a certain stock is still exposed to the risk in that stock. However, since a certain stock is only a small part of the fund, that risk is much smaller. Furthermore, other stocks might tend to move in the opposite way (they have negative **correlation** with each other), thereby offsetting some of the movements. That will reduce risk even more. You have now reduced the risk by **diversifying** your portfolio. All risk that you have discarded is diversifiable risk.

Correlation: A number between -1 and 1 that measures how two variables, such as the return on two stocks, tend to move in relation to each other.
Diversify: Invest in several different securities with different characteristics.

However, not all risk is diversifiable. To see that, imagine that you own all assets in the world. You would then have an enormous portfolio of stocks, bonds, currency, real estate, etc. You would still be exposed to some risk, though: The value of your portfolio would vary with the state of the world economy, for instance depending on the business cycle. The risk you are still exposed to is nondiversifiable (also called systematic risk or market risk). If you own a portfolio that has the same composition as the world economy, i.e. has the same percentage stocks etc., you are also fully diversified and have no diversifiable risk at all. If you have any other composition in your risky portfolio (and you probably do), you are exposed to both diversifiable and nondiversifiable risk.

To have diversifiable risk is unnecessary and most of those who have a large fraction of diversifiable risk are probably not aware of the problem. It is easy to get rid of that risk, either by diversifying or by investing in a well-diversified fund. One can therefore not expect any risk premium for holding diversifiable risk in ones portfolio. Remember that the risk premium (see Section 6.4) is the sum one has to pay a risk-averse agent to assume a certain risk. Since there is no reason to pay anyone to assume a risk that is possible to do away with, there is no risk premium on diversifiable risk. Only nondiversifiable risk carries a risk premium. The risk premium is the main reason why stock owners in the long run have larger returns than, for instance, people who save only in savings accounts.

17.3 CAPM: Pricing Assets

Suppose we have a fund that invests in the world market portfolio described above. That fund will carry no diversifiable risk. It will then generate a certain expected return, R_M (the market return). Part of that return is a compensation for risk. If one invests in the risk free rate, one gets a certain return of R_f. Therefore, the market risk premium must be the difference between the two: R_M - R_f. The risk premium is, consequently, the expected excess return one earns for assuming nondiversifiable risk.

We can use the market risk premium to price other securities. The most fundamental model for pricing financial assets is the *CAPM* (Capital Asset Pricing Model). It states that the risk premium on asset j is proportional to the risk premium on the whole market:

$$R_j - R_f = \beta_j \left(R_M - R_f\right).$$

Here, R_j is the expected return on asset j; β_j (beta) measures how asset j moves in relation to the market. If, for instance, asset j tends to rise 1 % when the market rises 1 %, $\beta_j = 1$, whereas if it only tends to rise by 0.5 %, $\beta_j = 0.5$. β_j is then a measure of the riskiness of asset j, where only nondiversifiable risk is measured. The more nondiversifiable risk a certain asset has, the more it will covary with the market and the higher its β will be. Conversely, an asset with no nondiversifiable risk will have $\beta = 0$, so that R_j - $R_f = 0$. In other words, the discount rate is then equal to the risk free rate.

17.4 Pricing Business Projects

Often, *CAPM* is used to determine whether an investment is profitable or not. The idea is to first estimate how large the future profits will be. Thereafter one estimates how large the discount rate should be. Finally, one calculates the present value of the investment and determines if it is worth its cost or not.

We assume that we already have estimates of the future profits. We must then determine the riskiness of the project. That is often done by comparing it to other similar projects that already exist. If the new project cannot be considered either more or less risky than other projects in the firm, one can use the firm's own internal rate of return as discount rate. However, suppose the firm is considering starting a new type of project. It is then possible to get an estimate of the riskiness by studying other firms with similar projects. The new project should have about the same (nondiversifiable) risk. Therefore, we use the other firms to get an estimate of β. Then we use that estimate in the *CAPM* formula and solve for what the discount rate, R_j, should be: $R_j = R_f + \beta(R_M - R_f)$.

To give an example: Suppose we have a project that we expect will make a loss of 10 during the first year and then generate a profit of 200 during each of three consecutive years. Then we can sell the remains for 100. The present value of the future cash flows is

$$PV = \frac{-10}{1+R} + \frac{200}{(1+R)^2} + \frac{200}{(1+R)^3} + \frac{200}{(1+R)^4} + \frac{100}{(1+R)^4}.$$

Furthermore, suppose that the market risk premium, R_M - R_f, is 6 %, that the risk free rate of return, R_f, is 2 % and that β for a firm with a similar project is 1.5. We then estimate the discount rate to 2 % + 1.5*6 % = 11 %. If we use that value in the formula, the present value is 497.

Should we invest or not? That depends on the price. If the cost of the project is less than 497, we invest; if it is more, we do not. Often, one includes the cost in the present value. This is then called the net present value (*NPV*). The investment criterion is then whether $NPV > 0$ or not. Suppose the investment costs 450. Then $NPV = -450 + 497 = 47$, and we should go on with the investment since it is profitable.

Sometimes, one performs a slightly different analysis. With an estimate of the future cash flows and the investment cost, one can solve for the *R* that makes *NPV* equal to zero. (One would then use the same type of numerical procedure as for calculating the yield a bond.) If the cost is still 450, $R = 14.6$ %. That rate is called the internal rate of return and can also be used as an investment criterion. If a project with the same risk (the same β) demands a return that is lower than 14.6 %, we invest, if it demands a higher return we do not.

18 General Equilibrium

When we have studied equilibria so far, it has always been so-called *partial equilibria*. (A partial equilibrium is one where we assume that "everything else is unchanged.") However, we have also seen that a change in one variable can lead to changes in many other variables, so the restriction that everything else is unchanged may not be very realistic. For example, a price change can affect the price of close substitutes and complementary goods. We will now study how interactions between two individuals in a very simple economy lead to a general equilibrium, i.e. a simultaneous equilibrium in all markets.

18.1 A "Robinson Crusoe" Economy

Consider an economy with only two agents on a desert island: Robinson and Friday. Those two are then the only consumers and the only producers. Let us assume that they also produce only two goods: Coconuts and fish. The question is then how much coconuts and fish to produce, and how to allocate them among themselves.

18.2 Efficiency

We have already mentioned efficiency several times in this book (for instance in Sections 11.4 and 12.1). Efficiency is about how much waste there is in an economy; less waste means more efficient. There are several related ways to define efficiency. An often-used measure of efficiency regarding allocations is Pareto efficiency. The definition of Pareto efficiency is:

- **Pareto improvement**. A change in allocation such that
  - No one is worse off; and
  - At least one, possibly several, is better off.
- **Pareto efficient** or **Pareto optimal**. An allocation such that
  - No Pareto improvements are possible.

Pareto improvement: A change in allocation that makes at least one individual better off, without making anyone worse off.

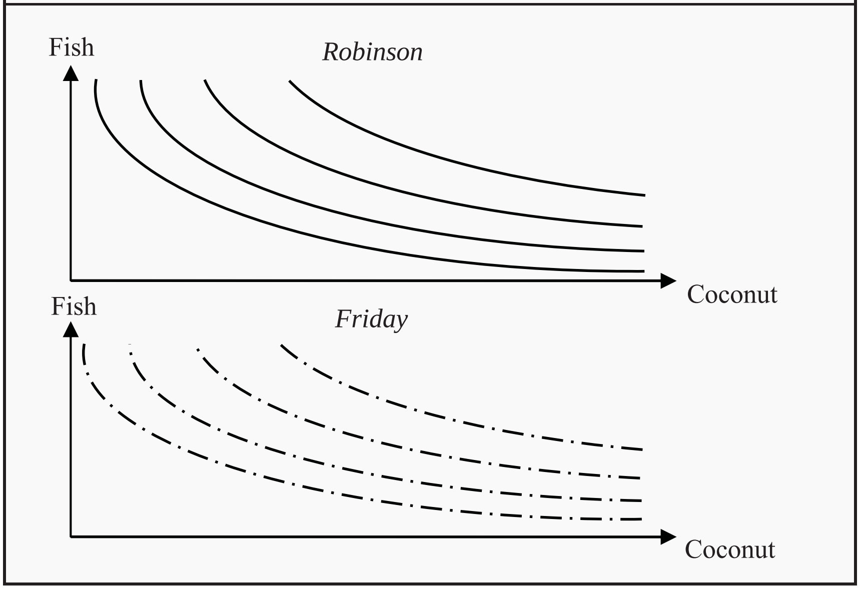
Pareto efficient / optimal: An allocation in which no Pareto improvements are possible.

18.3 The Edgeworth Box

In Chapter 3, we discussed the basics of consumer theory. We described, for instance, indifference curves and the budget line. If we consider an exchange economy in which the quantities of the goods are fixed, we have a **zero-sum game**. This means that one individual can only get quantities that the other ones do not get; there is, for instance, no growth in the economy. The name "zero-sum game" comes from the fact that, the sum of what some people gain and what others lose is always zero.

Zero-sum game: A game in which the sum of all gain and losses is zero, implying that if you gain something, someone else lost.

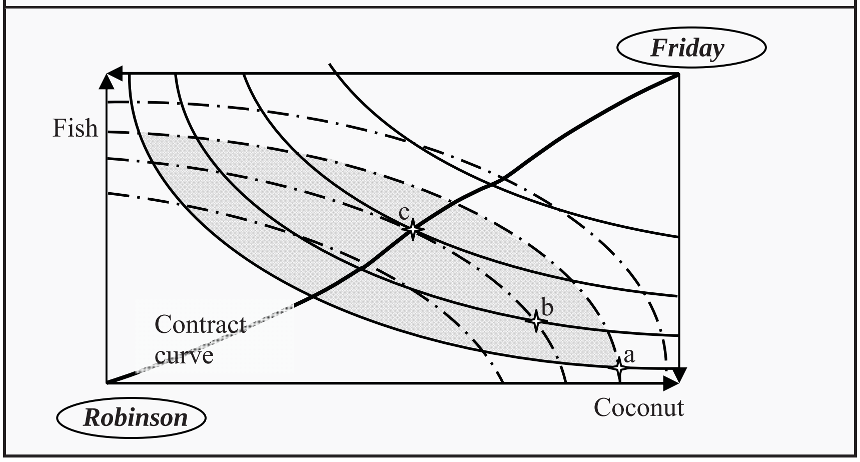
Figure 18.1: Two Indifference Maps



Suppose that Robinson and Friday have different preferences over coconuts and fish, and that these look like in Figure 18.1. Since the quantities of the two goods are fixed, we can combine these two indifference maps into one by taking one of them, for instance Fridays, turning it upside-down and putting it over Robinson's. We then get a picture such as the one in Figure 18.2. (The additional information in the figure will be explained below.) Such a diagram is called an **Edgeworth-box**.

Edgeworth-box: A diagram showing, for instance, the consumption possibilities of two agents in an economy with fixed total wealth.

Figure 18.2: An Edgeworth Box for Consumption



Note that the scales in the Edgeworth box are in opposite direction for the two agents. Upwards along the Y-axis, Robinson gets more fish while Friday gets less. To the right along the X-axis, Robinson gets more coconuts and Friday gets less. Also, be aware of which preference curves belong to which individual: The full lines belong to Robinson while the broken lines belong to Friday.

18.4 Efficient Consumption in an Exchange Economy

The clever thing with the construction in Figure 18.2 is that we can see directly which allocations of goods that are Pareto efficient.

- Consider, for instance, point a. Can it correspond to an efficient allocation? Compare it to point b. In b, both Robinson and Friday are better off (as both are on a higher indifference curve). Consequently, b is a Pareto improvement as compared to a, and then a cannot be an efficient allocation. Note that all points within the grey area are Pareto improvements as compared to a.
- Is then point b efficient? Compare b to c. In c, Robinson is better off while Friday is indifferent between b and c. Consequently, c is a Pareto improvement as compared to b, and then b cannot be efficient either.
- In point c, One of Robinson's indifference curves just barely touches one of Friday's indifference curves. That is the criterion for a Pareto efficient allocation of goods. Compared to c, every other allocation makes either Robinson or Friday (or both) worse off. c is consequently a Pareto efficient allocation.

Remember that the slope of an indifference curve is the same thing as the marginal rate of substitution; *MRS*. The criterion for an efficient allocation of goods can then be written

$$MRS_R = MRS_F,$$

where the subscripts refer to Robinson and Friday, respectively. In other words, for the allocation to be efficient, both agents are required to have the same marginal valuation of the goods.

Point c is, however, not the only Pareto efficient allocation in the diagram. We could repeat the procedure above for every possible indifference curve, and find all points of tangency. If we would do that, and then connect all points to a curve, we would get the so-called **contract curve**.

Contract curve: A curve indicating all Pareto efficient allocations in an Edgeworth-box.

If we assume that the initial allocation of coconuts and fish is as in point a and that we have free trade, then we would expect Robinson and Friday to start trading until they end up somewhere on the contract curve. Moreover, since only points in the grey area are Pareto improvements compared to a, we would expect them to end up on the part of the contract curve that lies within that area. Exactly where we they end up is, however, a question about negotiations between Robinson and Friday.

18.5 The Two Theorems of Welfare Economics

There are two important theorems regarding efficiency and competitive markets: the two welfare theorems:

- *1st theorem of welfare economics*: If all trade occurs in perfectly competitive markets, the allocation that arises in equilibrium is efficient.
- *2nd theorem of welfare economics*: Each point along the contract curve is a competitive equilibrium for some initial allocation of goods.

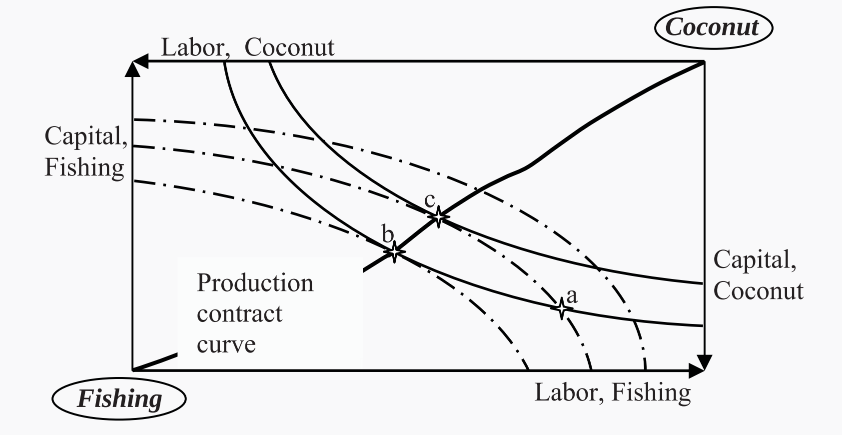
The first theorem is a variation of "the invisible hand" (see Section 9.7). It is enough to have perfect competition to get an efficient allocation. The second theorem states that there is no loss to efficiency from a reallocation. A competitive market will always find an efficient allocation.

18.6 Efficient Production

Regarding production, we can perform an analysis that is very similar to the one we did for consumption. Instead of two indifference maps, we put two isoquant maps (see Section 7.4) together. We imagine that Robinson and Friday have two firms, one that produces coconuts and one that produces fish. In the production, they use labor and capital, and their access to these input factors is fixed. They have a certain number of fishing tools, tools to pick coconuts with, and a maximum number of working hours. This allows us to construct an Edgeworth box for production (see Figure 18.3).

Let us start by assuming that Robinson and Friday have chosen point a. They consequently invest quite a large number of working hours, but not so much capital, on fishing and, vice versa, a small number of working hours but quite much capital on picking coconuts. Point a is not efficient. If they instead choose point b, they get, employing the same total number of working hours and the same total amount of capital, more coconuts and just as much fish as they did in a. If they would choose point c instead of a, they get more fish and just as many coconuts as they did before. Consequently, both point b and c constitute efficiency gains compared to point a.

Figure 18.3: Edgeworth Box for Production



What is "wrong" with point a? The isoquant for fishing (the full line) has a slope that is smaller than the isoquant for coconuts (the broken line). In Section 7.4, we defined the marginal rate of technical substitution, *MRTS*, as the slope of an isoquant. Remember what *MRTS* means: If we use one unit less of labor, how much more capital must we use in order to produce the same quantity of goods? The fact that the curves have different slopes implies that we can reduce the work in the fishing firm by a small amount and increase the capital in the same firm by a small amount to keep the quantity the same. This will free up labor that we can put into the coconut firm instead while we have to reduce the capital employed in that firm by a small amount. However, since the isoquant for coconuts is steeper than the one for fishing, the change means that they can now increase the production of coconuts.

In Figure 18.3, we see that the criterion for having an efficient production is that the isoquant for fishing just barely touches the isoquant for coconuts. In such a point, the two curves have the same slope and the criterion can be expressed as

$$MRTS_{good\ 1} = MRTS_{good\ 2}$$

If we, similarly to before, find all such efficient combinations of work and capital for the two goods and joint them into a curve, we get the **production contract curve**.

Production contract curve: A curve indicating all efficient factor combinations in an Edgeworth-box.

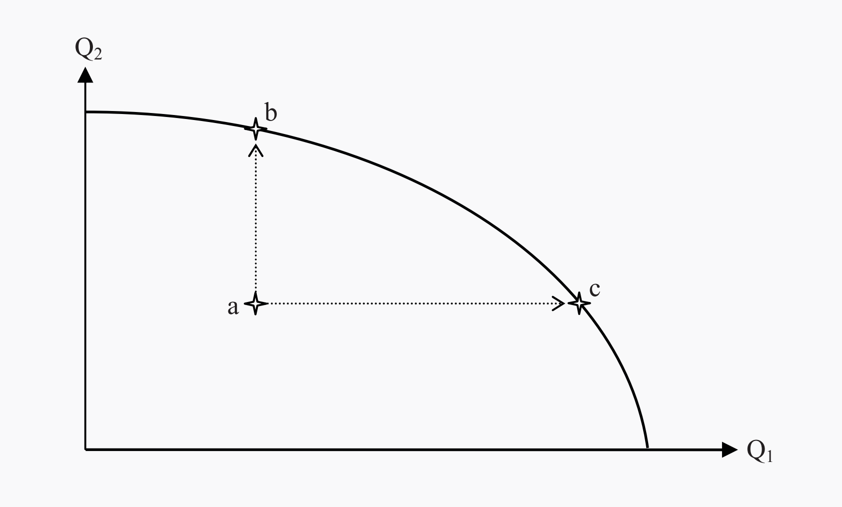
18.7 The Transformation Curve

In the last section, we derived the production contract curve. That curve is a collection of all efficient combinations of the two input factors. If we take those combinations, they also define how much we can maximally produce of one good, given a certain production of the other. Say, for instance, that we produce 50 fish. What is the maximum number of coconuts that we are then able to produce? For us to achieve the maximum, we have to produce at an efficient point, i.e. at some point on the production contract curve. At some point on this curve, we produce 50 fish. The maximum number of coconuts we are able to produce is then the number produced in that same point, say 100 coconuts.

If we take all points along the production contract curve and calculate which combinations of goods they each correspond to, and then use that information in a new graph, we can derive the so-called **transformation curve** (also called the **production-possibility frontier**). The point from the example above, 50 fish and 100 coconuts will then be one point on the transformation curve. On the curve, we have all efficient production possibilities and underneath it, we have all other possible, but inefficient, production possibilities.

Transformation curve / production-possibility frontier: A curve indicating all efficient combinations of quantities of goods produced.

Figure 18.4: The Transformation Curve



Consider point a under the transformation curve in Figure 18.4. What is the opportunity cost of producing one more unit of either good 1 or good 2? Since point a is not efficient, we do not have to give up anything to move to, for instance, point b. We just have to decrease the degree of waste in the economy. Consequently, the opportunity cost is zero! However, when we have reached point b, we cannot increase the production of good 2 more without reducing the production of good 1. If we want to move from point b to point c, we, instead, have to give up a certain quantity of good 2 to compensate for the increase in good 1. The quantity we have to give up is the opportunity cost. In other words, on the transformation curve the two goods have a price in terms of the other good, a relative price.

Note that the slope of the transformation curve is the same thing as what we in Section 3.1 defined as the marginal rate of transformation, *MRT*. To find *MRT*, we then used prices: $MRT = -p_1/p_2$. However, in the transformation curve there are no prices. Here, instead, we directly get the relative price of the goods. However, note that this is what we got in Section 3.1 as well. The example we used there was that one ice cream costs 10 units (of the appropriate currency) and a pizza 20 units. If we insert those prices into the formula, keeping all units, we get

$$MRT = -\frac{10\dfrac{units}{icecream}}{20\dfrac{units}{pizza}} = -\frac{10}{20}\frac{units}{icecream}\frac{pizza}{units} = -0{,}5\,\frac{pizza}{icecream}$$

MRT is consequently the relative price for one good, expressed in units of the other good. Note that this means that the slope of the budget line in Section 3.1 is directly related to which point on the transformation curve one has chosen.

18.8 Pareto Optimal Welfare

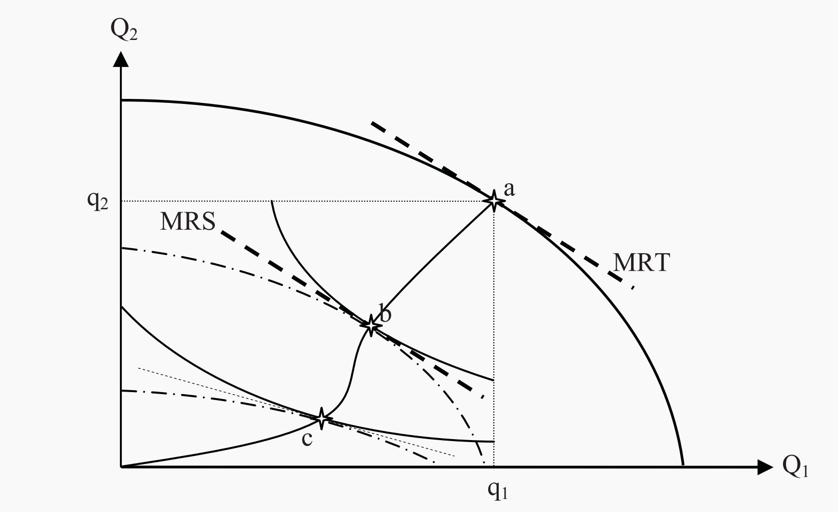
We will now put production and consumption together in one diagram. We start from the transformation curve, and assume that society has come up with an efficient mix of goods 1 and 2, i.e. of coconuts and fish in our example. The production will then lie on a point on the transformation curve, for instance point a in Figure 18.5. The marginal rate of transformation, *MRT*, is the slope in point a, and we produce the quantity q_1 of good 1 and q_2 of good 2.

Robinson and Friday now have to allocate the goods between themselves. We therefore add an Edgeworth box under the transformation curve, with one corner at point a and the opposite one at the origin. An efficient allocation then requires that their respective relative valuations of the goods are equal, i.e. that $MRS_R = MRS_F$ (where the subscripts refer to Robinson and Friday, respectively). In the figure, two such allocations are indicated: point b and point c, that both lie on the contract curve. (Also, compare to Figure 18.2.)

We now have efficiency in production (since the total production is on the transformation curve) and efficiency in consumption (since the allocation is on the contract curve). Is that enough for us to have general efficiency? No, it is not. It is also required that we produce what the consumers demand. There is one big difference between points b and c. In point b, the slope of the indifference curves, *MRS*, is the same as the slope of the transformation curve, *MRT*, in the point at which we have chosen to produce. That is not the case at point c. At point c, *MRS* is smaller in magnitude than *MRT*.

To see what the problem with that is, think of what *MRS* and *MRT* are. *MRS* is the price the consumer are *willing* to pay for one good in terms of the other, i.e. how many coconuts Robinson and Friday are willing to trade for one fish. Equilibrium in consumption demands that they have the same valuation. *MRT*, on the other hand, is the price the producers *have to* pay (given that the production is efficient) to produce one more unit of one good, again in terms of the other good. If the consumers are willing to pay more for one good than they have to, there are unexploited opportunities and the situation cannot constitute a general equilibrium. If we change the production such that we produce more of the good of which the consumers have a high valuation, then at least one consumer will be better off without anyone else being worse off.

Figure 18.5: Pareto Optimal Welfare



The criterion for an efficient **output mix** is then that

$$MRS = MRT$$

Output mix: A certain combination of quantities of goods produced.

18.8.1 A Definition of Pareto Optimal Welfare

We have now discussed three different types of efficiency:

- $MRS_R = MRS_F$; *Efficient consumption*. Robinson and Friday have the same marginal valuation of the goods. None of them can be made better off by a reallocation, without making the other one worse off.
- $MRTS_1 = MRTS_2$; *Efficient production*. The production of any of the goods cannot be increased without a reduction in the production of the other.
- $MRS = MRT$; *Efficient output mix*. It will cost as much to change from one good to the other as the relative valuation. No consumer can be made better off by another output mix without making the other worse off.

If all these three criteria are fulfilled, we talk of Pareto optimal welfare.

19 Externalities

In the remaining chapters, we will look at a few cases of **market failures**. A market failure is a situation in which the market fails to achieve an efficient allocation. A few such cases we have already seen. Both monopolies and oligopolies are, for instance, examples of market failures. In the following, we will briefly discuss externalities, public goods, and asymmetric information.

Market failure: A situation in which the market fails to achieve an efficient outcome.

Our consumption of goods does not occur in a social vacuum. Much of our consumption, perhaps all of it, indirectly affects other people. The most classical example is pollution. It does not have to be a big factory; it could be your neighbors having a barbecue party. You do not participate, but you still get a share of the smell. You might take your car to work; you pay for the fuel but not for the pollution or the congestion to which you expose others. Other examples include the use of penicillin (you are cured, but contribute to making bacteria penicillin resistant), vaccinations, and a well kept garden that your neighbors also enjoy looking at.

19.1 Definition

An **externality** is a situation in which the consumption or the production of goods has positive or negative effects on other people's utility where these effects are not reflected in the price.

Externality: A situation in which there are uncompensated costs (or benefits) not reflected in the price of a good.

It is common to distinguish between positive and negative externalities:

- Positive externalities. One person's consumption of a good also increases other people's utility without them having to pay for it.
- Negative externalities. One person's consumption of a good decreases other people's utility without them receiving any compensation.

Note that positive externalities are also a problem. Typically, we get too few goods with positive externalities and too many goods with negative externalities.

19.2 The Effect of a Negative Externality

Let us study the classical example of a negative externality: A firm produces a good, but in doing so they also pollute the environment. First, we need to define a few concepts:

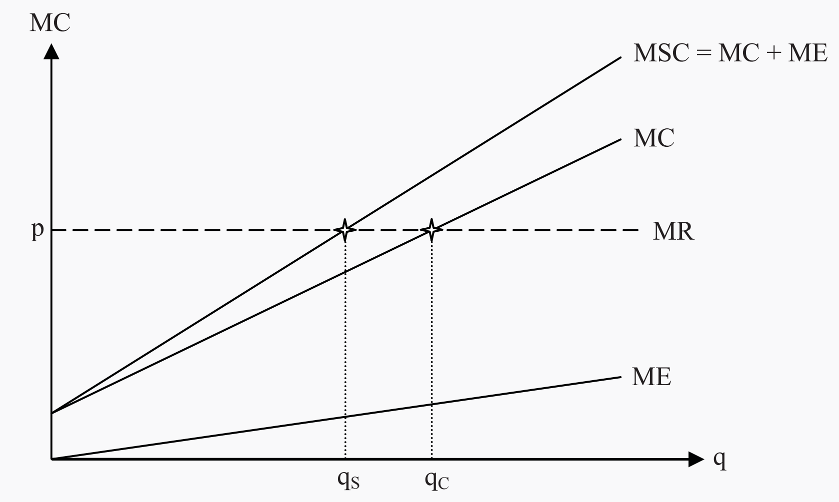
- *The marginal cost of the externality*, *ME*. The change in the cost of the marginal effect, when production is increased by one unit. This is similar to the concept of *MC*, but instead of concerning the cost for the firm, it concerns the (uncompensated) cost of the externality.
- *Social cost*. The sum of the cost of producing the good and the cost of the external effect.
- *Marginal social cost*, *MSC*. The sum of the firm's marginal cost and the marginal cost of the externality, i.e. $MC + ME$.

We can analyze this situation in a way that is similar to the one in Section 9.4 and 9.5. Look at Figure 19.1 (and compare, for instance, to Figure 9.2). The firm operates in a perfectly competitive market, so $MR = p$. To maximize its profit, the firm chooses to produce the quantity where $MC = MR$, i.e. the quantity q_C.

However, this firm also emits pollution. The pollution does not cost the firm anything, but there is a cost to society. The more the firm produces, the more it pollutes. In the figure, we have drawn the marginal cost of the external effect, *ME*, and the marginal social cost, $MSC = MC + ME$. We see that for society, the optimal quantity to produce is q_S.

The effect of the firm being able to ignore the cost of polluting is that it produces too much of the good. As an indirect effect of that, there will also be more pollution than at the optimum.

Figure 19.1: The Effect of a Negative Externality



19.3 Regulations of Markets with Externalities

One way to correct the situation in Section 19.2 is to put a tax on each unit of the product. If one knows the size of *ME* then the tax should be that same amount. Thereby, the marginal cost curve of the firm will coincide with *MSC*, and the firm will automatically correct its production to the optimal quantity, q_S. An obvious problem with this solution is that one rarely knows *ME*.

Other strategies to regulate the market include quantity regulations and the creation of transferable emissions permits.

20 Public Goods

20.1 Definition of Public and Private Goods

A **public good** is a good that fulfills both of the following two criteria:

Public good: A good that is both nonrival and nonexclusive.

- *Nonrival*. One individual's consumption of the good does not affect any other individual's consumption *of the same unit* of the good. Examples include lighthouses, television, parks, military defense, and streets with little traffic.
- *Nonexclusive*. It is not possible to exclude anyone from consuming the good. The examples above are usually nonexclusive.

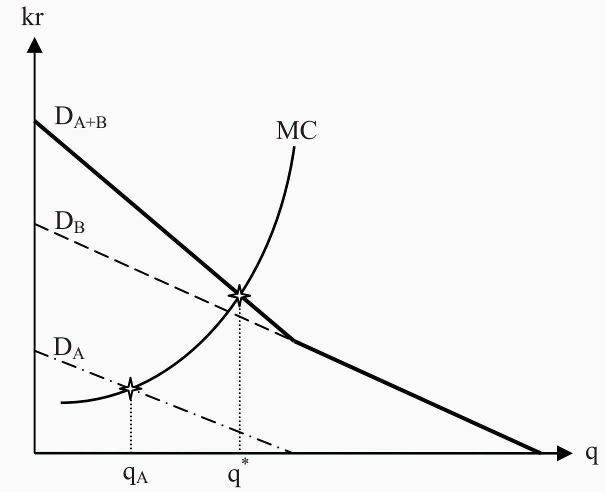
A private good is, instead, a good that does not fulfill any of the two criteria, i.e. one that is both rival and exclusive. Most goods are private goods.

20.2 The Aggregate Willingness to Pay

To find the market's demand curve for a public good we must know each individual's demand for it.

Suppose we have two individuals, A and B, and that they each have an individual demand curve regarding, say, a park, corresponding to D_A and D_B in Figure 19.1. When we, in Section 4.2, derived the market demand for a private good, we summed the individuals' demand curves horizontally. For public goods we, instead, have to sum them *vertically*.

Figure 19.1: Public Goods



To see why this is the case, think of what a public good is. Suppose we produce one unit of the good and that A value that unit to 10, whereas B values it to 15. Had it been a private good, only one of them could have consumed it. However, since a public good is nonrival, both A and B can consume it at the same time. Consequently, the aggregate willingness to pay for this unit is 10 + 15 = 25. Similarly for the second, third, and all following units, and we sum the demand curves in Figure 19.1 vertically. The market demand in this case, corresponds to the thick demand curve D_{A+B}.

We see that, the optimal quantity, q^*, is at the point where the marginal aggregate willingness to pay is equal to the marginal cost. As a comparison, we have also indicated what would happen if only one of the individuals had decided on the quantity. If A had done so, the quantity q_A would have been produced. As compared to the optimum, q^*, we would have seen a much smaller quantity.

20.3 Free Riding

In the last section, we derived the demand curve for the public good by summing the individuals' willingness to pay. The problem is that we usually do not know their willingness to pay. For private goods, this is not a problem, since it is optimal for the consumers to pay a price up to their willingness to pay. For instance, if the price of milk is 10 and an agent buys three liters of milk, her marginal willingness to pay for the last unit is at least 10 and her marginal willingness to pay for additional units is less than 10. We do not need to know it beforehand, as she will reveal it by her behavior.

However, for public goods it is not optimal for consumers to reveal their willingness to pay. If she will later have to pay an amount equal to the one she states, it is often individually better for her to understate her willingness to pay. If the good is still produced, she will make a sort of profit: She receives more utility than she has paid for. She is then said to be **free riding**.

Free riding: One lets other people pay by understating one's valuation of a public good.

21 Asymmetric Information

We have already mentioned that information is important in economics and, most often, we have assumed that the agents have perfect information. That is hardly a reasonable assumption. For instance, usually a seller knows more about a product than the buyer does, and a worker knows her skills better than the employer does. We will now look at some implications of the problem of **asymmetric information**. There are two important subcategories:

- **Adverse selection**. Depending on the fact that one side in a contractual agreement, the buyers or the sellers, have information that the other part does not have, only some buyers or sellers will want to enter into the contract. Only the ones that will profit the most from the contract will do so. Moreover, those are, typically, the ones the other part wants to avoid.
- **Moral hazard**. Sometimes, one's counterpart cannot check whether one fulfills one's obligations after having agreed on a contract. One may them be tempted to exploit the other's lack of knowledge.

Note the difference between the two types: Adverse selection is about what happens *before* the agreement has been made. Moral hazard is about what happens *after* it has been made.

Asymmetric information: A situation in which different agents have different amounts of information about a good.
Adverse selection: Depending on asymmetric information, different agent act in different ways before agreeing on a contract.

Moral hazard: Depending on asymmetric information, agents change their behavior after having agreed on a contract.

21.1 Adverse selection

We will give two classical examples of adverse selection: the market for insurance and the market for used cars. Note, however, that the concept is possible to apply on many types of goods and services.

21.1.1 Insurance

The price of insurance largely depends on the probability that the insurance firm will have to pay, for instance, on the probability that your bike is stolen. If there is a high probability, the price of insurance will also be high.

Different people differ in how well they keep after their belongings, and the risk that a careless person will get her bike stolen is much higher than that a carful person will get hers stolen. However, the insurance firm cannot see a difference between careless and careful people, and therefore charges them the same price corresponding to an average of the risks.

This, however, makes the insurance a good deal for the careless people, but it might make it too expensive for the careful people. They will probably not lose their bikes anyway. Then only the careless people remain; the ones that constitute a high risk for the insurer. When the insurer realizes that all people buying insurance are high-risk people, they will have to increase the price even more. The high-risk people will then have pushed the low-risk people out of the market, even though the latter might be fully willing to pay for insurance.

21.1.2 Used Cars

Suppose there are 100 used cars in a market, and that they are of two different levels of quality: Half of them are of high quality (H-cars) and half are of low quality (L-cars). The sellers want at least 50,000 for an L-car and at least

100,000 for an H-car, whereas the buyers are prepared to pay at most 60,000 for an L-car and 110,000 for an H-car. There are consequently possibilities for trades that are beneficial for both sides. If there had been two submarkets, one for L-cars and one for H-cars, people could have negotiated prices between 50,000 and 60,000 for L-cars and between 100,000 and 110,000 for H-cars.

However, if they are sold in the same market, the buyers cannot tell them apart. Neither can she ask the sellers, as all sellers would say that their car is an H-car. If the chance that she will get an L- or an H-car is 50% each, the buyer could think of this as a lottery. Suppose, for simplicity, that the buyer is risk neutral (so that she does not demand a risk premium for taking a risk). She would then be willing to pay the expected value of the car, i.e.

$$50\% * (60.000) + 50\% * (110.000) = 85.000$$

She will then maximally offer 85,000. However, at that price no seller is prepared to sell an H-car. Their lowest price for an H-car is 100,000. Consequently, they withdraw the H-cars from the market and only sell L-cars.

Then, however, the probability of getting an L-car is no longer 50%, but instead 100%. Since the buyer realizes this, she is prepared to pay a maximum of 60,000 for a used car, and the L-cars have pushed the H-cars out of the market.

This outcome is not efficient, since there are cars that the buyers have a higher valuation for than the sellers do, but that are not traded.

21.1.3 Signaling and How to Reduce Problems with Adverse Selection

There are several ways to reduce problems with adverse selection:

- *Legislation*. For instance, one could demand that sellers have to reveal the ingredients of (food) products. Thereby, buyers gain more information and we get less asymmetric information.
- *Demand more information*. Insurers often demand, for instance, a medical examination before selling insurance.
- A firm could acquire a *reputation for quality*. The cost of selling an L-car as if it was an H-car, i.e. lying about the product, would then be too high, since that would damage the reputation. Therefore, the customers know that all the seller's cars are H-cars.
- One could also offer a warranty for the cars. Since the probability that an L-car will break down is much higher than that an H-car will do that, a seller of L-cars cannot offer the guarantee. Thereby the sellers sort themselves into two groups, and for L-cars and one for H-cars.

The last two examples are variants of **signaling**. The idea with signaling is that the agents themselves signal to which group they belong. It is, of course, not enough that they *say* that they belong to a certain group. It must be a signal that the low-quality group cannot afford, so that truth telling is optimal.

Signaling: A behavior (sign) that makes it possible to distinguish an agent as being different from others.

21.2 Moral hazard

Moral hazard has to do with asymmetric information after an agreement has been made, for instance after a contract has been signed. We can continue the insurance example from above in the following way: Say that the careful person has managed to convince the insurance firm that she is, indeed, carful. Therefore, she constitutes a low-risk person, and she only has to pay a small premium to get the insurance.

Say that she bought insurance for her bike, and that this guarantees her a new bike if the one she has is stolen. Before she bought insurance, she would have lost the full value of the bike if it had been stolen; now she will only lose the time it takes to get a new one. Consequently, there is much less reason for her to go through the trouble of taking good care of her bike. Therefore, the risk that the bike is stolen increases and she might now constitute a risk for the insurer that is as big as the careless people are.

Because she has insurance, her risk behavior has changed to the insurer's disadvantage. She only had to pay a low price since she is careful, but after she got insurance, she is no longer careful and should have had to pay a high price. Since the insurer cannot check if her risk-behavior has changed, she can take advantage of the firm and offload a larger share of the risk on them than she has paid for.

21.2.1 How to Reduce Problems with Moral Hazard

The classical way to reduce problems with moral hazard in the insurance sector is to demand that the customer keeps a part of the risk. Usually, an insurer demands that the customer pays a certain amount herself, the so-called deductable. Thereby, the risk that she becomes overly careless is reduced.

22 Key Words

Chapter 1; Introduction
resources
scarce
opportunity / alternative cost
microeconomics
macroeconomics
agents
unemployment
inflation
goods
services
labor force
capital
market
models
positive economics
normative economics

Chapter 2; Supply, Demand, and Market Equilibrium
competition
demand
complementary goods
substitute goods
preferences
quantity
income
supply
factor prices
laws
regulations
equilibrium
regulated
secondary effects
minimum price
price floor
price ceiling

Chapter 3; Consumer Theory
budget
maximize
basket / bundle of goods
budget line
marginal rate of transformation
preference order
indifferent
complete
transitive
non-satiation
convexity
transaction costs
indifference curve
indifference map
marginal rate of substitution
corner solution
inner solution
utility function

Chapter 4; Demand
price-consumption curve
Engel-curve
income-consumption curve
elasticity
price elasticity
income elasticity
normal good
inferior good
necessary good
luxury good
cross-price elasticity

Chapter 5; Income and Substitution Effects
substitution effect
income effect
Giffen good

Chapter 6; Choice under Uncertainty
risk
expected value
fair lottery
expected utility
marginal utility
risk preferences
risk averse
certainty equivalent
risk premium
risk reduction

Chapter 7; Production
input
costs
revenue
production function
average product
marginal product
diminishing marginal returns
short run
long run
isoquant
marginal rate of technical substitution
returns to scale

Chapter 8; Costs
fixed costs
variable costs
total costs

average total cost
average variable cost
average fixed cost
marginal cost
isocost line
expansion path
economies of scale
diseconomies of scale

Chapter 9; Perfect Competition
product differentiation
barriers to entry
price taker
homogenous products
cartel
marginal revenue
abnormal / excess profit
normal profit
efficient

Chapter 10; Market Interventions and Welfare Effects
consumer surplus
producer surplus
welfare
deadweight loss

Chapter 11; Monopoly
monopoly

Chapter 12; Price Discrimination
price discrimination
perfect price discrimination

Chapter 13; Game Theory
game theory
player
actions
information
strategies
payoffs
prisoner's dilemma
normal form game
payoff matrix
dominant strategy
Nash equilibrium
extensive-form game
game tree
backward induction
subgame perfect equilibrium

Chapter 14; Oligopoly
oligopoly
duopoly
kinked demand curve
reaction function
isoprofit curves

Chapter 15; Monopolistic Competition
monopolistic competition

Chapter 16; Labor
output market
wage
marginal revenue product of labor
monopsony
bilateral monopoly

Chapter 17; Capital
present value
yield to maturity
bond
stock
risk correction
diversifiable risk
nondiversifiable risk
CAPM

Chapter 18; General Equilibrium
Pareto efficient
zero-sum game
Edgeworth box
contract curve
allocation
welfare theorem
production contract curve
transformation curve
production possibility curve

Chapter 19; Externalities
market failure
externality
external effect

Chapter 20; Public Goods
public good
non-rivalry goods
nonexclusive goods
private good
aggregate willingness to pay
free riding

Chapter 21; Asymmetric Information
asymmetric information
adverse selection
moral hazard
insurance
signaling