

The Science of Economics

By Fred E. Foldvary

1. The 3 sectors of an economy

Having analyzed consumption, we turn now to production. We already have seen how production uses three factors, land, labor and capital goods. We can now examine the second dimension of an economy, its three sectors: household, firms, and government. The third dimension was described in [Chapter 5](#), the categories of expenditure as consumption, investment, and waste.

Households consist of individuals, families, and other living-together arrangements, such as room mates and communes. Firms are the organizations which engage in the production of wealth. Government is the agency that has authority and power over the rules that firms and households are legally required to abide by in their consumption and production.

All wealth is owned by households. They either own firms personally or own shares in corporations. Households who are citizens are the ultimate owners of the land of a country and of its government's wealth. As owners, households rent their factors to firms, hiring themselves to firms as workers, loaning firms their capital goods, and renting land to firms. Firms may, of course, nominally (in name) own capital goods, but the ultimate owners are some households.

Households obtain an income from loaning the factors to firms, and they use this income to invest in capital goods (for firms or government), for their own consumption, or for waste imposed on them by thieves and government. Households are thus the sector that consumes the wealth produced by the firms.

Firms input factor resources and output consumer goods. Capital goods produced by firms and sold or rented to other firms in effect remain in the sector of firms.

Firms can be generally divided into four types:

- 1) families;
- 2) for-profit firms;
- 3) non-profit organizations;
- 4) government enterprises.

Non-profit organizations include churches and foundations. The motivation of **for-profit firms** is normally to maximize profits, but the ultimate goal is the maximization of utility, which can also involve achieving a large size of firm or getting prestige.

For-profit firms in turn have various possible forms of organization. The simplest is a **single proprietorship**, where an individual owns the firm. Next in complexity is a **partnership**, owned by several persons. A family is one possible partnership; it engages in household production, such as gardening. A marriage is also a partnership.

Most complex are **corporations**, firms whose owners have shares of stock and elect a board of directors to operate the firm. Typically, the board hires a president or manager. A corporation may take the form of a **cooperative**, in which a shareholder is also a member, and each member has one vote regardless of shares. Another form is a **non-stock corporation**, such as **condominium** housing, in which the owners are members and have various types of voting rights, but where ownership consists of having title to units, like apartments, which are tied into the membership of the firm.

In many countries, corporations have legally limited liability, which means that the ordinary shareholders may not be liable for debts of the company beyond the value of their shares. The board, though, is usually liable for the debt beyond their shares in the company. A disadvantage of corporations in the U.S. is that corporate income is taxed twice, once when the corporation earns it, and secondly when it distributes dividends to the shareholders, which is taxed again as personal income.

An advantage of corporations with shares of stock is that the shares trade on active exchange markets, so an investment in the corporation is highly "liquid," and also divisible. This enables a corporation to raise large amounts of money by issuing stock.

A problem with corporations is that the management may seek to promote its own well-being rather than that of the share holders. This is resolved by means of profit-sharing methods of payment as well as managers competing for their positions and the possibility of a takeover if a firm becomes too fat with management benefits. Laws preventing takeovers thus work to reduce corporate efficiency.

Besides stocks and units of ownership, companies can also issue debt in the form of bonds. These pay interest for a certain period of time, after which the bond "matures" and the company takes them back and pays back the principal.

The agents who organize the factors of production are called the **entrepreneurs**. Often they are owners of firms or executives with a great deal of control, but sometimes they can be sales persons or anyone that is able to organize production. Entrepreneurs are the drivers of a market economy; they actively seek out opportunities to bring resources to a more productive use, such as by developing new products or new methods, or bringing goods to markets they were not previously available in. If you have a vegetable garden in your back yard, you too are an entrepreneur, since you decide what to grow, how to grow it, and who gets the produce.

Government plays three different roles.

- First, some of its agencies act as firms. Governments run enterprises such as the post office, railroads, and street maintenance. In this role, government is a firm hiring factors

and producing wealth. So part of government is also in the category of firms. Like firms, governments also issue bonds.

- Secondly, government enacts rules that must be followed by the three sectors (including itself). These rules are of two types:
 - 1) rules creating markets, and
 - 2) rules that intervene or interfere in markets.

The distinction follows the universal ethic developed in Chapter 1. Rules which coerce households and firms, which impose costs or restrictions, i.e. any rule other than prohibiting and penalizing coercive harm to others, are interventions. Subsidies are also interventions.

Other rules create and maintain markets. For example, uniform and constant rules protecting property rights and enforcing contracts enable markets to operate efficiently. Laws setting up copyrights make a market in literature and art more efficient, preventing publishers from having to write a contract with each buyer regarding copying of the books. Laws setting up patents also simplify transaction costs, also preventing firms from having to write complex contracts with each user of the firm to avoid copying, and making the enforcement easier. Some people feel that copyrights and patents create monopolies and are thus inefficient, but this is a static view of these instruments. Patents and copyrights are general contracts between the sellers and buyers of new products, without which the new products might not be created or written.

- The third function of government is redistribution. Governments typically take much of the wealth from the first and second distributions (see Chapter 5) and forcefully redistributes it to others. Hence, much consumption is done by recipients of redistributed income who have not earned it by supplying factors to firms.

There is another agent in the economy, the thief. A thief also performs forced redistribution, taking wealth and income by force. Thieves in effect are underground governments, since their redistribution is not legally sanctioned; they impose their own independent rules and redistribution. Hence, their role can be subsumed under the redistributive and rule-making aspects of government, there being two categories of government, official, above-ground legal agents and underground, illegal agents called "criminals".

Thieves obtain some of the wealth of households by force and, as far as the rest of society is concerned, waste it. Some government expenditure is also typically wasted.

There are two circular flows in the economy among the sectors, for goods and money.

- Goods, produced by firms (including government firms), flow to households and to government. In turn, households supply factor resources to firms and government.
- The second flow, money, circulates in the opposite direction, since a consumer obtains goods in exchange for money. Firms pay money to households and to government for the use of factors. Households and government then use the money to pay the firms (including government) for goods, or lose some money to thieves or to government.

2. The production possibilities curve

When we disaggregate wealth into various products, the question arises as to how much of each is to be produced. We begin the analysis with a simple economy with only two products, bread and lettuce. The economy could produce only bread or only lettuce or some of each. Suppose that only bread is being produced, for a total of 100 loaves. Now we want to produce ten heads of lettuce. If the economy was fully employed in making the 100 loaves, some resources must be taken away from bread making to make the lettuce.

To analyze the economics of the trade-off between the two products, we again begin with the foundational principles. Proposition #1 states that some natural resources are scarce. So there is a finite amount of land and labor available in an economy, which is why only so much bread or lettuce can be produced, and not enough to satisfy everyone if they were free. Proposition #2 states that resources vary in quality, and proposition #4 tells us that different amounts of inputs will produce different amounts of outputs. Applying this, we see that some land is more productively used to grow wheat for bread, while other land is more productively used to grow lettuce; and likewise some labor has been trained to grow wheat and bake bread while other labor has been educating in growing lettuce. Then capital goods, of course, have been made specializing in one or the other.

When we use a bit less land, labor, and capital goods to produce bread and use it to make lettuce, how do we do this. Proposition #12 states that people economize. So we economize by giving up that wheat-growing land that is most suitable for growing lettuce, as well as labor, etc., that was trained to grow nice lettuce. Likewise, if we had been growing only lettuce, the most productive resources would be used to grow the first increment of wheat and to bake bread. The next most productive resources will be used to produce the next increments of the products.

If we then plot all combinations of bread and lettuce that can be produced, we get a curve. We have bread on one axis and lettuce on the other axis, and for each amount of bread, there is a certain maximum amount of lettuce that can be grown. This is the "production possibilities curve," PPC.

This curve demonstrates several principles.

- First is the **principle of efficiency**. Production is efficient if the total product is on the PPC. If the total output lies within the PPC, then one or more of the products can be increased without decreasing the other; hence, society is not producing efficiently. Efficient production means that the production of any product cannot be increased without reducing the amount of any other product. The existence of waste, of course, implies that society is inside the PPC relative to the desires of those obtaining the income of the first two distributions.
- A second principle is **opportunity cost**. To produce more bread, we must produce less lettuce; the opportunity cost of producing more bread is less lettuce.

- The third principle is the "**law of increasing cost.**" Economizing persons use the most productive resources first. Increased amounts of the good will require less productive resources, resources which may be used more productively in other uses.

So the law of increasing cost states that as we increase the production of one good, the opportunity cost of foregone production of other goods tends to increase. Therefore, the shape of the PPC is "bowed out," like a rainbow. Economists say it is "concave to the origin," but you can just think of it as bowed out unless you want to impress your friends. The curve is bowed because at first, giving up a little wheat gets you a lot of that first amount of lettuce, and vice versa.

The effect of better technology is to push out the whole PPC outwards, so that the same amounts of inputs yields a greater amount of output. The effect of accumulating more capital goods is also to push out the PPC, since land and labor become more productive. There is therefore a trade-off between consuming now and consuming tomorrow. We can consume less today and invest in new capital goods in order to consume more tomorrow.

3. The production function

Since a firm, in abstract, is an organization inputting resources and outputting products, we can describe it as a production function, or a product as a function of inputs. A function is a relationship between a dependent variable and some independent variables. The inputs are the independent variables, and output is dependent. The relationship can be concisely stated as $Q/T = f(N,L,K)$, where Q/T is output per time interval, N is the number and quality of workers, L is the amount and quality of land, and K represents capital goods. Note that these are all physical inputs and output; there is no financial capital such as money in the function.

The methods of production, the technology and rules (including government regulations, the role of luck, and the goals of the firm owners) are included in the functional variable f .

We can see then that if the amount of a resource such as N , labor, is varied, output will vary. The marginal product of labor is in fact the change in output Q/T caused by a change in labor, N .

In order to maximize profits, costs must be minimized, and this implies that the amount of each input will be determined by its marginal product, proportional to its cost. In the cost-minimizing combination of inputs, the marginal productivity of a dollar's worth of all inputs must be the same. If the marginal product of one input, divided by price, is lower than that of another, then costs can be reduced by switching, if possible, to the inputs with the higher relative productivity.

4. The theory of exchange

Some people have the idea that agriculture or manufacturing is "productive," but trading, buying goods in one place and selling in another, or exchanging one good for another among two persons, is not productive, but just moves things around. But Menger showed that this is not so.

Menger showed how in an exchange of goods, the goods have unequal rather than equal value. They may have an equal market value, but the subjective values must be different, otherwise the trade would not take place. Trade only takes place if person A has some good that is of less value to him than some good that B has.

Using Menger's (1871, p. 183) example, suppose A has horses and B has cows. Because of diminishing marginal utility, each extra horse has less and less value to A, and so with cows for B. Suppose the first cow or horse has a value of 50 to A and B, and that each extra one has a value of 10 less than the previous. Then if A has 5 horses, the fifth is only worth 10 to him. But it would be worth 50 to B. Same with cows. So they trade. A now has 4 horses and 1 cow. He lost 10 of value by giving up the horse, and gained 50 by getting the cow, for a net gain of 40. Likewise, B has a net gain of 40 from getting a horse.

As Menger states (p. 184), "each of the two traders obtained an economic gain from this first exchange equivalent to the gain that would accrue to him if his wealth had been increased by a good whose value to him is equal to 40... Trade is therefore no less productive than industrial or agricultural activity." Economic exchange contributes to consumers' utility and thus an increase in the subjective value of their wealth just as effectively as the physical increase of more goods. As Menger stated, "the end of economy is not the physical augmentation of goods but always the fullest satisfaction of human needs" (p. 190).

Both will continue to exchange as long as the marginal utility of the other's good is greater than that of the goods they have. The next horse or cow has a marginal utility of 20 to the owner and 40 to the other, so they exchange, each increasing utility by 20. After that, the marginal utilities are 30 for both, so they stop trading. Note that once the goods are of equal marginal value, trade comes to a halt. Trade went on because of unequal rather than equal subjective marginal values.

An important principle of exchange is that trade will continue until the economic gains are exhausted, until the marginal values of what one has is equal or greater than what others have.

5. Supply curves

We derived demand curves in Chapter 5, and now we will derive supply curves, the supply of goods offered by firms.

A market supply is the quantity of goods that a producer is willing to produce at particular prices. Like demand, the quantity is a function of (dependent on) price, but the quantity axis is horizontal and the price is on the vertical axis, by the convention set by Alfred Marshall.

Like demand schedules or curves, a supply can either be a flow, or quantity produced during some time interval, or a stock, a certain amount of goods at one particular moment in time.

The amount of goods competitive firms are willing to supply at various prices depends on the costs of production. Costs in the short run are either fixed or variable, fixed costs being those which cannot be changed during that time interval. In the long run, all costs can be changed;

indeed, the "long run" is defined for any particular firm as that time interval at which all costs are variable.

Average costs are simply total costs divided by the number of units, while marginal cost is the cost of producing one more unit (or tiny amount of product). A firm obtains a maximum profit when total revenues are greater than total costs and its marginal cost just equals its marginal revenue, since any extra unit would cost more than it gets in revenues. If the average variable cost and the long-term average cost are greater than the price, then the firm will shut down, unless the owners enjoy taking losses.

While demand curves slope down, there is no universal slope to supply curves. As noted in Chapter 3 on land, the supply curve for land as space is vertical, as is the supply curve for goods no longer in production, like rare coins and stamps, or old art.

But for goods in current production, supply curves usually slope up diagonally. This is due to the universal propositions about physical resources. Inputs are scarce, and some are more productively devoted to one use than another, relative to the values placed on them by consumers. So to increase the amount of bread, as we saw, resources must be drawn away from the production of lettuce, which may not be as productive for bread as the previous resources devoted to bread.

So the relative price of the new inputs is higher. Greater quantities are produced only if the price fetched in the market is higher, and the supply curve slopes up.

Therefore, short-run total marginal costs typically may decline at first with increasing production as the fixed cost is spread over more units, but then increase as the cost of using more inputs goes up and as the marginal product of variable factors declines, since some of the inputs such as land are fixed during that time.

However, over the long run, this tendency of input prices to increase can be offset by a change in f , the technology and method of production. At greater amounts of production, in some industries, more efficient ways of producing can be achieved. For example, it is more expensive to make a few cars than to mass-produce them in huge factories. There can be physical reasons for long-run economies of scale; the volume of facilities such as pipes and buildings increase at a greater rate than their surfaces, so the per-unit volume costs can go down with larger structures.

This is called "economies of scale": unit costs decrease with greater production. So an industry with such economies of scale will have a downward-sloping supply curve. Offsetting economies of scale are diseconomies or increasing costs of managing an ever larger firm, as management gets more and more complicated and bigger as a portion of costs, so at time amount of production, it is possible for these diseconomies to outweigh any continuing economies, and the supply curve would go up again.

It is also possible for the supply curve to be horizontal, if the costs of inputs are the same at all levels of production and there are no economies of scale. This is called a "constant cost" industry.

If the average cost of firm output first decreases and then increases, in a U-shaped curve, then the marginal cost curve must cross it at its lowest point, the marginal costs first pulling down and then pulling up the average.

Just as we distinguish a shift in a demand curve from a movement along a demand curve, we distinguish a shift in the supply curve (a change in the quantity supplied at all prices) from a movement along a supply curve (a change in quantity supplied as the price of a product changes). A decrease in government regulations or improvement in technology, for example, would shift a supply curve out, making it possible to produce more for any particular price.

6. Price equilibrium

Now that we have the two sides of the market, demand and supply, we can join the two curves in one graph. As the demand curve slopes down, it will in some cases intersect the supply curve. The point where the curves intersect determines the market price and quantity at that time. In practice, these curves are typically fuzzy rather than precise lines; there will be some range of prices for some product rather than one exact price everywhere, as consumers realize when they shop around.

It is quite possible that the curves will not intersect at all. You offer a poem on the market for only a dollar, but the demand curve starts at 25 cents for one poem and slopes down to 100 copies of the poem demanded if it is free. But your supply curve began at \$1 for the first poem; you refuse to sell it for any less, so the curves do not intersect. The quantity exchanged in that case is zero.

If the curves do intersect, then economists call the price and quantity an "equilibrium." If the time period is more than a moment (such as a week), then the equilibrium price is a range in which the trades have occurred. The equilibrium of a time duration consists of the equilibria which exist at each moment in time.

An equilibrium at a moment in time is a situation in which the gains from trade have become exhausted. If there were a shortage, gains from trade would be possible as sellers increased the price to buyers willing to pay more to get the goods. If there is a glut or surplus of goods, gains from trade can be made as sellers lower the price to get rid of the stuff. Either would be called a disequilibrium, since gains from trading can still be made. At equilibrium, trade halts, because gains from trade have become exhausted.

But wait a minute! If trade stops, then how can there be a market equilibrium price? Is this a paradox? The answer is that at each moment, markets are moving towards equilibrium, but never actually reach it (or if they do in limited situations, trade stops). People eat; they are in equilibrium and stop eating. Then they get hungry again, in disequilibrium, and go to market. They bid for food, while sellers make offers. The market price attains an equilibrium price as gains from trade are exhausted, but then hungry new buyers in disequilibrium keep coming afterwards, making new bids.

Hence, a market price constantly equilibrates or matches demands and supplies, and does not grind to a halt. This equilibrating clears the market - sellers are matched by buyers at the current price - but new buyers and sellers are always dynamically making bids and offers. Just because the price is stable does not mean equilibrium (in each moment) has been achieved. It just means that bids and offers are somewhat constant during some time interval. The price is always subject to change if there is some change in the amounts and flows of bids or offers. Of course if we consider a time range greater than a moment, then, looking back in time, we observe that often there is some narrow price range in which exchanges have taken place, and we can consider that to be a equilibrium during that duration.

7. Profit

The term "profit" is used in different ways by accountants and economists. To an accountant, profit is the difference between explicit or money revenues and money costs. An economist subtracts from this profit the implicit costs of a firm. Suppose a farmer owns his own land, which he could rent for \$10,000 per year. An accountant says his profit was \$50,000 that year. But the farmer could have rented his land out for \$10,000; by not doing so, he lost \$10,000 in potential revenue. This is an opportunity cost of using the land himself. Since the land rent is a cost regardless of to whom it is paid, an economist subtracts it as an implicit cost, not paid in money but a cost of using that factor nevertheless. So the economic profit is reduced to \$40,000. But wait: the farmer's own wages must be subtracted to. If he could have earned \$30,000 working for someone else, that too is an implicit cost, a wage. So we have \$10,000 left. But what about the capital goods? He could have hired them out for \$6,000 that year. Subtract this implicit yield on capital goods, and we are left with a \$4,000 economic profit. This is also called an entrepreneurial profit.

But since wealth is divided into a first distribution of wages, rent, and capital yields, where does this economic profit fit in? Since entrepreneurs are also workers, this entrepreneurial profit is really a type of wage.

But it is a special type of wage. Foundational proposition #15 states that the future is uncertain. Entrepreneurs are innovators, but they can't be sure whether they will earn a profit from trying or organize factors in what they expect to be better ways. Uncertainty cannot be insured against, unlike ordinary risk. Normal risks such as fires occur with some regularity in a larger-number environment, and insurance companies can measure how much loss there has been in an average year and provide insurance against it. But, as economist Frank Knight pointed out, uncertainty does not have probability distributions. New products are unique in time and circumstance, we cannot know what the probability of success is. So entrepreneurs and their fellow investors take a chance, and if they are right, their reward is entrepreneurial or economic profits. If they are wrong, they take losses.

Profits and losses are important signals in a market economy. Consistently high profits in an industry indicate that more resources can be devoted to this product. Losses indicate that too much production has taken place in that industry. Hence, taxes on profits skew these signals,

reducing the potential investment and entrepreneurship in an economy, reducing output, efficiency, and employment.

We see then, that profits induce firms to produce, innovate, and employ factors. How this is done by the economy as a whole will be the subject of the next chapter.